Zbornik 14. mednarodne multikonference

INFORMACIJSKA DRUŽBA – IS 2011

Zvezek A

Proceedings of the 14th International Multiconference

INFORMATION SOCIETY – IS 2011

Volume A

Uredili / Edited by

Marko Bohanec, Matjaž Gams, Dunja Mladenić, Marko Grobelnik,
Marjan Heričko, Urban Kordeš, Olga Markič, Jadran Lenarčič,
Leon Žlajpah, Andrej Gams, Vladimir A. Fomichov, Olga S. Fomichova,
Andrej Brodnik, Rok Sosič, Vladislav Rajkovič, Tanja Urbančič, Mojca Bernik

10. – 14. oktober 2011 / October 10 – 14, 2011
Ljubljana, Slovenia

Zbornik 14. mednarodne multikonference

INFORMACIJSKA DRUŽBA – IS 2011

Zvezek A

Proceedings of the 14th International Multiconference

INFORMATION SOCIETY – IS 2011

Volume A

Inteligentni sistemi
Izkopavanje znanja in podatkovna skladišča (SiKDD 2011)
Sodelovanje, programi in storitve v informacijski družbi
Kognitivne znanosti
Kognitonika
Robotika
Internet in Slovenija: 1985 - 1995
Vzgoja in izobraževanje v informacijski družbi

Intelligent Systems
Data Mining and Data Warehouses (SiKDD 2011)
Collaboration, Software and Services in Information Society
Cognitive Sciences
Cognitonics
Robotics
Internet and Slovenia: 1985 – 1995
Education in Information Society

Uredili / Edited by

Marko Bohanec, Matjaž Gams, Dunja Mladenić, Marko Grobelnik, Marjan Heričko, Urban Kordeš, Olga Markič, Jadran Lenarčič, Leon Žlajpah, Andrej Gams, Vladimir A. Fomichov, Olga S. Fomichova, Andrej Brodnik, Rok Sosič, Vladislav Rajkovič, Tanja Urbančič, Mojca Bernik

http://is.ijs.si

10.–14. oktober 2011 / October 10th–14th, 2011
Ljubljana, Slovenia

Založnik: Institut »Jožef Stefan«, Ljubljana
Tisk: Birografika BORI d.o.o.
Priprava zbornika: Mitja Lasič, Jana Krivec, Vesna Lasič
Oblikovanje naslovnice: Vesna Lasič, Miran Krivec
Tiskano iz predloga avtorjev
Naklada: 80

Ljubljana, oktober 2011

Konferenco IS 2011 sofinancirajo
Ministrstvo za visoko šolstvo, znanost in tehnologijo
Javna agencija za raziskovalno dejavnost RS (ARRS)
Institut »Jožef Stefan«

CIP - Kataložni zapis o publikaciji
Narodna in univerzitetna knjižnica, Ljubljana

659.2:316.42(082)
659.2:004(082)
004.8(082)

MEDNARODNA multikonferenca Informacijska družba (14 ; 2011 ; Ljubljana)
Zbornik 14. mednarodne multikonference Informacijska družba - IS 2011, 10.-14. oktober 2011 : zvezek A = Proceedings of the 14th International Multiconference Information Society - IS 2011, October 10th-14th, 2011, Ljubljana, Slovenia : volume A / uredili, edited by Marko Bohanec ... [et al.]. - Ljubljana : Institut Jožef Stefan, 2011. - (Informacijska družba, ISSN 1581-9973)

Vsebina na nasl. str.: Inteligentni sistemi = Intelligent systems ; Izkopavanje znanja in podatkovna skladišča (SiKDD 2011) = Data mining and data warehouses (SiKDD 2011) ; Sodelovanje, programi in storitve v informacijski družbi = Collaboration, software and services in information society ; Kognitivne znanosti = Cognitive sciences ; Kognitonika = Cognitonics ; Robotika = Robotics ; Internet in Slovenija: 1985-1995 = Internet and Slovenia: 1985-1995 ; Vzgoja in izobraževanje v informacijski družbi = Education in information society

ISBN 978-961-264-035-4
1. Informacijska družba 2. Information society 3. Bohanec, Marko, 1958-
258045696

PREDGOVOR MULTIKONFERENCI INFORMACIJSKA DRUŽBA 2011

V svojem štirinajstem letu je multikonferenca Informacijska družba (http://is.ijs.si) znova dokazala, da je ena vodilnih srednjeevropskih konferenc, ki združuje znanstvenike z različnih raziskovalnih področij, povezanih z informacijsko družbo. V letu 2011 smo v multikonferenco povezali devet odličnih neodvisnih konferenc. V Sloveniji in po svetu mrgoli konferenc, naša multikonferenca pa izstopa po širini in obsegu tem, ki jih obravnava, po akademski odprtosti in širini, ki spodbuja nove ideje, predvsem pa po tem, da ni tradicionalna konferenca, ampak se pogumno loteva vizionarskih tem, pogosto v interaktivni ali delavniški obliki.

Multikonferenca temelji na sinergiji interdisciplinarnih pristopov, ki obravnavajo različne vidike informacijske družbe ter poglabljajo razumevanje informacijskih, komunikacijskih in družbenih storitev v najširšem pomenu besede. Na multikonferenci predstavljamo, analiziramo in preverjamo nova odkritja in pripravljamo teren za njihovo praktično uporabo, saj je njen osnovni namen promocija raziskovalnih dosežkov in spodbujanje njihovega prenosa v prakso na različnih področjih informacijske družbe tako v Sloveniji kot tujini. Prepričani smo, da je svetla bodočnost človeštva povezana z razvojem in uporabo novih znanj, da bomo tako premagali vse trenutne izzive: podnebne, okoljske, pomanjkanja surovin, preobljudenost itd.

Na multikonferenci bo na vzporednih konferencah predstavljenih 150 referatov, vključevala pa bo tudi okrogle mize in razprave. Referati so objavljeni v zbornikih multikonference, izbrani prispevki pa bodo izšli tudi v posebnih številkah dveh znanstvenih revij, od katerih je ena Informatica, ki se ponaša s 35-letno tradicijo odlične znanstvene revije.

Multikonferenco Informacijska družba 2011 sestavljajo naslednje samostojne konference:

- Inteligentni sistemi
- Kognitivne znanosti
- Robotika
- Rudarjenje podatkov in podatkovna skladišča (SiKDD 2011)
- Sodelovanje, programska oprema in storitve v informacijski družbi
- Soočanje z demografskimi izzivi
- Vzgoja in izobraževanje v informacijski družbi
- Kognitronika
- Internet in Slovenija: 1985 – 1995.

Soorganizatorji in podporniki konference so različne raziskovalne institucije in združenja, med njimi tudi ACM Slovenija. Zahvaljujemo se tudi Agenciji za raziskovalno dejavnost RS ter Ministrstvu za visoko šolstvo, znanost in tehnologijo za sodelovanje in podporo. V imenu organizatorjev konference pa se želimo posebej zahvaliti udeležencem za njihove dragocene prispevke in priložnost, da z nami delijo svoje izkušnje o informacijski družbi. Zahvaljujemo se tudi recenzentom za njihovo pomoč pri recenziranju.

V letu 2011 sta se programski in organizacijski odbor odločila, da bosta podelila posebno priznanje Slovencu ali Slovenki za izjemen življenjski prispevek k razvoju in promociji informacijske družbe v našem okolju. Z večino glasov je letošnje priznanje pripadlo dr. Vladimirju Batagelju. Priznanje za dosežek leta je pripadlo Janezu Branku. V letu 2011 prvič podeljujemo nagrado »informacijska limona« in »informacijska jagoda« za najbolj (ne)uspešne poteze v zvezi z informacijsko družbo. Limono je dobila informatizacija zemljiške knjige, jagodo pa odprava avtomatov za zdravstvene kartice. Čestitke nagrajencem!

Niko Zimic, predsednik programskega odbora
Matjaž Gams, predsednik organizacijskega odbora

FOREWORD - INFORMATION SOCIETY 2011

In its 14th year, the Information Society Multiconference (http://is.ijs.si) continues as one of the leading conferences in Central Europe gathering scientific community with a wide range of research interests in information society. In 2011, we organized nine independent conferences forming the Multiconference. There are plenty of conferences in Slovenia and all over the world. The broad range of topics and the open academic environment fostering new ideas makes our event unique among similar conferences, promoting key visions in interactive, innovative ways.

The Multiconference flourishes the synergy of different interdisciplinary approaches dealing with the challenges of information society. The major driving forces of the Multiconference are search and demand for new knowledge related to information, communication, and computer services. We present, analyze, and verify new discoveries in order to prepare the ground for their enrichment and development in practice. The main objective of the Multiconference is presentation and promotion of research results, to encourage their practical application in new ICT products and information services in Slovenia and also broader region. We are confident that the new knowledge will overcame all the issues we face today: overpopulation, environment problems, diminishing natural resources and global heating.

The Multiconference is running in parallel sessions with 150 presentations of scientific papers. The papers are published in the conference proceedings, and in special issues of two journals. One of them is Informatica with its 35 years of tradition in excellent research publications.

The Information Society 2011 Multiconference consists of the following conferences:

- Intelligent Systems
- Cognitive Sciences
- Robotics
- Data Mining and Data Warehouses (SiKDD 2011)
- Collaboration, Software and Services in Information Society
- Demographic Challenges in Europe
- Education in Information Society
- Cognitronics
- Internet and Slovenia: 1985 – 1995.

The Multiconference is co-organized and supported by several major research institutions and societies, among them ACM Slovenia, i.e. the Slovenian chapter of the ACM. We would like to express our appreciation to the Slovenian Government for cooperation and support, in particular through the Ministry of Higher Education, Science and Technology and the Slovenian Research Agency.

In 2011, the Programme and Organizing Committees decided to award one Slovenian for his/her life-long outstanding contribution to development and promotion of information society in our country. With the majority of votes, this honor went to Dr. Vladimir Batagelj. In addition, a reward for current achievements was pronounced to Janez Brank for his support of the ACM computer competitions. In 2011, two additional awards were introduced: “information strawberry” and “information lemon”. The strawberry is pronounced to the end of medical cards permanent confirmations at the medical automata, and the lemon goes to the problems with the land register. Congratulations!

On behalf of the conference organizers we would like to thank all participants for their valuable contribution and their interest in this event, and particularly the reviewers for their thorough reviews.

Niko Zimic, Programme Committee Chair
Matjaž Gams, Organizing Committee Chair

KONFERENČNI ODBORI
CONFERENCE COMMITTEES

International Programme Committee

Vladimir Bajic, South Africa
Heiner Benking, Germany
Se Woo Cheon, Korea
Howie Firth, UK
Olga S. Fomichova, Russia
Vladimir A. Fomichov, Russia
Vesna Hljuz Dobric, Croatia
Alfred Inselberg, Izrael
Jay Liebowitz, USA
Huan Liu, Singapore
Henz Martin, Germany
Marcin Paprzycki, USA
Karl Pribram, USA
Claude Sammut, Australia
Jiri Wiedermann, Czech Republic
Xindong Wu, USA
Yiming Ye, USA
Ning Zhong, USA
Wray Buntine, Finland
Bezalel Gavish, USA
Gal A. Kaminka, Israel
Miklós Krész, Hungary
József Békési, Hungary

Organizing Committee

Matjaž Gams, chair
Mitja Luštrek, co-chair
Robert Blatnik
Vesna Koricki Špetič
Jana Krivec
Mitja Lasič

Programme Committee

Nikolaj Zimic, chair
Franc Solina, co-chair
Viljan Mahnič, co-chair
Cene Bavec, co-chair
Tomaž Kalin, co-chair
Jozsef Györkös, co-chair
Tadej Bajd
Jaroslav Berce
Mojca Bernik
Marko Bohanec
Ivan Bratko
Andrej Brodnik
Dušan Caf
Saša Divjak
Tomaž Erjavec
Bogdan Filipič
Andrej Gams
Matjaž Gams
Marko Grobelnik
Nikola Guid
Marjan Heričko
Borka Jerman Blažič Džonova
Gorazd Kandus
Urban Kordeš
Marjan Krisper
Andrej Kuščer
Jadran Lenarčič
Borut Likar
Janez Malačič
Olga Markič
Dunja Mladenič
Franc Novak
Marjan Pivka
Vladislav Rajkovič
Grega Repovš
Ivan Rozman
Niko Schlamberger
Stanko Strmčnik
Tomaž Šef
Jurij Šilc
Jurij Tasič
Denis Trček
Andrej Ule
Tanja Urbančič
Boštjan Vilfan
David B. Vodušek
Baldomir Zajc
Blaž Zupan
Boris Žemva
Leon Žlajpah

KAZALO / TABLE OF CONTENTS

Zbornik 14. mednarodne multikonference

INFORMACIJSKA DRUŽBA – IS 2011

Proceedings of the 14th International Multiconference

INFORMATION SOCIETY – IS 2011

Inteligentni sistemi

Intelligent Systems

Uredila / Edited by

Marko Bohanec, Matjaž Gams

http://is.ijs.si

12.–13. oktober 2011 / October 12th–13th, 2011
Ljubljana, Slovenia

PREDGOVOR

Konferenca *Inteligentni sistemi* je tudi letos, tako kot vsa pretekla leta od 1997 dalje, potekala v okviru multikonference *Informacijska družba*. Konferenca se ukvarja s pomembnim vidikom informacijske družbe: inteligentnimi sistemi in inteligentnimi storitvami. Ključna tema pa so programski sistemi v informacijski družbi oziroma konkretne tehnične rešitve v inteligentnih sistemih, možnosti njihove praktične uporabe, pa tudi trendi, perspektive, nujni ukrepi, prednosti in slabosti, priložnosti in nevarnosti, ki jih v informacijsko družbo prinašajo inteligentni sistemi.

Konferenca *Inteligentni sistemi* v letu 2011 ostaja mednarodna in vseslovenska; prispevki so tako v slovenskem kot angleškem jeziku. V osrednjem konferenčnem delu je predstavljenih pet prispevkov. Vabljeni referat Angela Montanarija s sodelavci govori o intervalni časovni logiki, pri drugih avtorjih pa letos prevladujejo teme o regresijskih modelih ter uporabi na področjih računalniško podprtega načrtovanja in zdravljenju raka. Vsi ti prispevki so bili recenzirani s strani vsaj po dveh anonimnih recenzentov. V ločeni sekciji pa so predstavljeni prispevki Delavnice E9. Večina prispevkov obravnava raziskovalne dosežke v Odseku za inteligentne sisteme Instituta Jožef Stefan. Hkrati s predstavitvijo poteka tudi aktivna analiza prispevkov vsakega predavatelja in diskusija o bodočih raziskavah.

Marko Bohanec in Matjaž Gams, predsednika konference

PREFACE

The conference *Intelligent Systems* remains a traditional part of the multiconference *Information Society* since its beginnings in 1997. The conference addresses an important aspect of information society: intelligent computer-based systems and the corresponding intelligent services. Specifically, it addresses technical aspects of intelligent systems, their practical applications, as well as trends, perspectives, advantage and disadvantages, opportunities and threats that are being brought by intelligent systems into the information society.

The conference *Intelligent Systems 2011* remains a national and international event and presents papers written in both English and Slovenian languages. The international part of this Proceedings includes five papers. The invited paper by Angelo Montanari et al. presents interval temporal logics, and the other papers address the topics related to regression models and applications in computer-aided design and cancer treatment. All these papers have been reviewed by at least two anonymous reviewers. In a separate section, papers from the E9 workshop are presented. Most of them present research at the Department of Intelligent Systems at the Jožef Stefan Institute, Ljubljana, Slovenia. Each presentation consists of the classical paper report, and further includes analysis of the researcher's achievements and future research of each presenter in the workshop manner.

Marko Bohanec and Matjaž Gams, Conference Chairs

PROGRAMSKI ODBOR / PROGRAMME COMMITTEE

Marko Bohanec, *predsednik*
Institut Jožef Stefan

Matjaž Gams, *predsednik*
Institut Jožef Stefan

Tomaž Banovec
Statistični urad Republike Slovenije

Cene Bavec
Univerza na Primorskem, Fakulteta za Management Koper; IBM

Jaro Berce
Univerza v Ljubljani, Fakulteta za družbene vede

Marko Bonač
ARNES

Ivan Bratko
Univerza v Ljubljani, Fakulteta za računalništvo in informatiko, IJS

Dušan Caf
Telekom Slovenije

Bojan Cestnik
Temida

Aleš Dobnikar
Center Vlade RS za informatiko

Bogdan Filipič
Institut Jožef Stefan

Nikola Guid
Univerza v Mariboru, Fakulteta za elektrotehniko računalništvo in informatiko

Borka Jerman Blažič
Institut Jožef Stefan

Tomaž Kalin
DANTE

Marjan Krisper
Univerza v Ljubljani, Fakulteta za računalništvo in informatiko

Marjan Mernik
Univerza v Mariboru, Fakulteta za elektrotehniko računalništvo in informatiko

Zbornik 14. mednarodne multikonference IS 2011

INTELIGENTNI SISTEMI

Proceedings of the 14th International Multiconference IS 2011

INTELLIGENT SYSTEMS

Vabljeno predavanje

Invited lecture

A SHORT WALK THROUGH INTERVAL TEMPORAL LOGICS

Angelo Montanari
Dipartimento di Matematica e Informatica, Università di Udine
Via delle Scienze 206, I-33100 Udine, Italy
e-mail: angelo.montanari@uniud.it

Dario Della Monica
Dipartimento di Informatica e Applicazioni, Università di Salerno
Via Ponte don Melillo, I-84084 Fisciano (SA), Italy
e-mail: ddellamonica@unisa.it

Valentin Goranko
Department of Informatics and Mathematical
Modelling, Technical University of Denmark
Bld. 321, Richard Petersens Plads, DK-2800 Kgs. Lyngby, Denmark
e-mail: vfgo@imm.dtu.dk

Guido Sciavicco
Departamento de Ingenieria de la Información y
las comunicaciones, Universidad de Murcia
Campus de Espinardo, E-30100 Espinardo (Murcia), Spain
e-mail: guido@um.es

ABSTRACT

Interval temporal logics are a family of modal logics for reasoning about relational interval structures over linear orders. The set of all possible binary relations between such intervals is known as the set of Allen's interval relations. A distinct modal operator can be associated with each of them. Formulae of interval temporal logics are evaluated at time intervals rather than time points. This results in a substantially higher expressiveness and computational complexity of interval temporal logics as compared to point-based ones. This extended abstract is a short walk through interval temporal logics aimed at illustrating the main achievements of research in the field.

1 INTRODUCTION

Temporal reasoning plays a main role in many areas of computer science and artificial intelligence. In most cases, time points are assumed to be the basic ontological temporal entities. However, time points are not suitable to properly reason about real-world events with an intrinsic duration. Point-based and interval-based temporal ontologies are clearly closely related and technically reducible to each other: on the one hand, time intervals can be determined by pairs of time instants (their endpoints); on the other hand, a time instant can be viewed as a degenerated 'point interval', whose left and right endpoints coincide. However, these reductions do not resolve the main semantic issue arising when developing logical formalisms for temporal reasoning: should formulae in the given logical language be interpreted as referring to points or to intervals? There are at least three possible answers to such a question, that lead to three different classes of logical formalisms, namely, point-based temporal logics, interval-based temporal logics, and mixed, two-sorted temporal logics, where points and intervals are considered as separate sorts on a par and formulae for both sorts are built.

One of the first interval-based logical formalisms is Moszkowski's Propositional Interval Temporal Logic (PITL), which has been successfully applied to the specification and

verification of hardware components [21]. An extension of PITL with metric features, called Duration Calculus (DC), has been developed and extensively studied since the early '90s [24]. DC features a notion of duration of an event over a time interval in order to reason about specification and design of time-critical systems. DC is definitely one of the most popular and applicable interval-based logical formalism; however, its semantics is built on a point-based temporal ontology, and thus we will not analyze it here (we refer the interested reader to [23]).

The first temporal reasoning formalism proposed in the literature, which is based on a purely interval-based temporal ontology, is Allen's interval algebra (IA) [1]. IA allows one to deal with the family of binary relations arising between two intervals in a given linear order, later called Allen's interval relations. Besides Allen's relations, the natural and important operation of chopping an interval into two subintervals, which gives rise to the ternary interval relation `chop', was proposed and studied by Moszkowski in [21]. A systematic logical study of purely interval-based temporal reasoning started with Halpern and Shoham's seminal work on the multi-modal logic HS featuring one modality for each Allen's relation [15]. Concurrently with [15], Venema devised the even more expressive interval logic CDT involving binary modal operators associated with the ternary relation Chop (C) and its two residual relations D and T [22].

2 INTERVAL TEMPORAL LOGICS: MAIN ISSUES AND RESULTS

Halpern and Shoham's work initiated a stream of active research on the family of HS fragments devoted to the analysis of their expressiveness and to the study of the decidability/undecidability and complexity of their validity/satisfiability problem. Expressiveness and decidability issues for interval temporal logics are the main themes of this extended abstract (see the survey paper [14] for details).

While decidability is a common feature of point-based temporal logics for computer science, it turned out that undecidability dominates among interval-based logics. The first such undecidability results were obtained for PITL by Moszkowski already in [21]. Furthermore, general undecidability results for HS are given in [15], where Halpern and Shoham proved that validity of HS formulae in any class of interval models on linear orders satisfying very weak conditions, including the classes of all, discrete, and dense linear models, is undecidable. Moreover, they proved that the validities of HS in any of the standard numerical orderings of the natural numbers, integers, and reals (all of them are Dedekind complete) are not even recursively axiomatizable. Since then, a number of undecidability results for extremely simple fragments of HS have been obtained by sharpening the original proof techniques [2,6,16,17].

For a long time, these strong undecidability results have discouraged the search for practical applications and further theoretical investigation on purely interval-based temporal logics. Meanwhile, several semantic modifications or restrictions, that basically reduce the interval-based semantics to a point-based one, have been proposed to remedy the problem and to obtain decidable systems. As an example, in [21] Moszkowski showed that PITL decidability can be recovered by constraining atomic propositions to be point-wise and defining truth over an interval as truth over its initial point (the so-called locality principle). This bleak picture started lightening up in the last few years when various non-trivial decidable fragments of HS have been identified [5,8,10,11,12,18,19,20]. (Un)decidability of HS fragments depends on two main factors: (i) the set of interval modalities, and (ii) the class of interval models (the linear order) over which the fragment is interpreted. Gradually, it became evident that the trade-off between expressiveness and computational affordability in the family of HS fragments is rather subtle and sometimes unpredictable, with the border between decidability and undecidability cutting right across the core of that family.

The study and classification of decidable and undecidable fragments of HS have also invoked a systematic comparative analysis of their expressiveness. On the one hand, that line of research has led to several correspondence results between fragments of HS and meaningful fragments of first-order logic; on the other hand, it motivated the classification of the family of HS fragments with respect to expressiveness. By a systematic use of bisimulations between interval models, we have been able to identify a complete set of inter-definability equations among HS modalities over the class of all linear orders, thus obtaining a complete classification of HS fragments with respect to expressiveness [13]. Using that result, we have been able to establish that there are exactly 1347 expressively different such fragments out of the 4096 subsets of HS modal operators.

Finally, the strive for obtaining even more expressive, yet decidable interval logics has naturally led to the study of quantitative extensions of HS fragments with metric constraints on the length of intervals. Metric interval logics are based on interval structures over linear orders endowed with distance between points, and thus with a natural notion of interval length, and with language extended with arithmetic constraints on interval length. Decidability and expressive completeness issues for metric interval logics have been addressed in [3,7] and further elaborated in [4,9] with respect to the class of finite linear orders, the natural numbers, and the integers.

3 THE ROADS AHEAD

Despite the very substantial progress over the past ten years in the area of interval temporal logics, the field is still very rich with interesting challenges and unexplored paths. Here we outline our present view of the main immediate and long-term challenges in the field.

The main items in the current research agenda are the following ones: (i) extending the expressiveness classification result for the family of HS fragments to other important classes of linear orders, e.g., the classes of finite, discrete, and dense linear orders; (ii) obtaining a complete classification of the family of HS fragments with respect to decidability/undecidability of their satisfiability problem, first on the class of all interval models over linear orders, and then on meaningful subclasses of it. More than 90% of HS fragments have already been classified with respect to various classes of linear orders (for a summary of the current state of the classification, see the web page: https://itl.dimi.uniud.it/content/logic-hs), but the remaining cases are expected to be among the most difficult to settle; (iii) extending the study of metric extensions of interval logics to other HS fragments and over other metrizable linear orders, notably that of rational numbers.

The long-term research lines in the field include the following ones: (i) quest for automaton-based techniques for proving decidability of interval temporal logics; (ii) development of methods and algorithms for model-checking in finite and finitely presentable infinite interval structures, such as ultimately periodic ones; (iii) last but not least, identification and development of major applications of interval temporal logics. Besides the classical field of system specification, verification, and synthesis, we expect interval temporal logics to be successfully applied in the areas of planning and plan validation, to represent and to reason about temporally extended goals and constraints, temporal databases, to deal with temporal aggregation, and natural language processing, to model features like progressive tenses.

Acknowledgments

This extended abstract is a shortened version of a paper by the same authors that will appear in the Bulletin of the European Association for Theoretical Computer Science.

References

1. J. F. Allen. Maintaining knowledge about temporal intervals. *Communications of the ACM 26(11)*, pp. 832--843, 1983.

2. D. Bresolin, D. Della Monica, V. Goranko, A. Montanari, and G. Sciavicco. Decidable and Undecidable Fragments of Halpern and Shoham's Interval Temporal Logic: Towards a Complete Cassification. *Proc. of the 15th Int. Conference on Logic for Programming, Artificial Intelligence, and Reasoning*, volume 5330 of LNCS, pp. 590--604. Springer 2008.

3. D. Bresolin, D. Della Monica, V. Goranko, A. Montanari, and G. Sciavicco. Metric Propositional Neighborhood Logics: Expressiveness, Decidability, and Undecidability. *Proc. of the 19th European Conference on Artificial Intelligence*, pp. 695--700, Lisbon, Portugal, 2010.

4. D. Bresolin, D. Della Monica, V. Goranko, A. Montanari, and G. Sciavicco. Metric propositional neighborhood interval logics on natural numbers. *Software and Systems Modeling*, accepted for publication, January 2011 (doi: 10.1007/s10270-011-0195-y, online since February 2011).

5. D. Bresolin, V. Goranko, A. Montanari, and P. Sala. Tableaux for logics of subinterval structures over dense orderings. *Journal of Logic and Computation 20*, pp. 133--166, 2010.

6. D. Bresolin, V. Goranko, A. Montanari, and G. Sciavicco. Propositional interval neighborhood logics: Expressiveness, decidability, and undecidable extensions. *Annals of Pure and Applied Logic 161(3)*, pp. 289--304, 2009.

7. D. Bresolin, V. Goranko, A. Montanari, and G. Sciavicco. Right propositional neighborhood logic over natural numbers with integer constraints for interval lengths. *Proc. of the 7th IEEE Conference on Software Engineering and Formal Methods*, pp. 240--249, Hanoi, Vietnam, 2009.

8. D. Bresolin, A. Montanari, and P. Sala. An optimal tableau-based decision algorithm for Propositional Neighborhood Logic. *Proc. of the 24th Int. Symposium on Theoretical Aspects of Computer Science*, volume 4393 of LNCS, pp. 549--560. Springer, 2007.

9. D. Bresolin, A. Montanari, P. Sala, and G. Sciavicco. An Optimal Decision Procedure for MPNL over the Integers. *Proc. of the 2nd Int. Symposium on Games, Automata, Logics and Formal Verification*, pp. 192--206, Minori (SA), Italy, 2011.

10. D. Bresolin, A. Montanari, P. Sala, and G. Sciavicco. Optimal tableau systems for propositional neighborhood logic over all, dense, and discrete linear orders. *Proc. of TABLEAUX-11*, volume 6793 of LNAI, pp. 73--87. Springer, 2011.

11. D. Bresolin, A. Montanari, P. Sala, and G. Sciavicco.What's decidable about Halpern and Shoham's interval logic? The maximal fragment ABBL. *Proc. of the 26th Symposium on Logic in Computer Science*, pp. 387—396. IEEE Computer Society, 2011.

12. D. Bresolin, A. Montanari, and G. Sciavicco. An optimal decision procedure for Right Propositional Neighborhood Logic. *Journal of Automated Reasoning 38(1-*

3), pp. 173--199, 2007.

13. D. Della Monica, V. Goranko, A. Montanari, and G. Sciavicco. Expressiveness of the interval logics of Allen's Rrelations on the class of all linear orders: complete classification. *Proc. of the 22nd Int. Joint Conference on Artificial Intelligence*, pp. 845—850, Barcelona, Spain, 2011.

14. V. Goranko, A. Montanari, and G. Sciavicco. A road map of interval temporal logics and duration calculi. *Journal of Applied Non-Classical Logics 14(1-2)*, pp. 9--54, 2004.

15. J. Halpern and Y. Shoham. A propositional modal logic of time intervals. *Journal of the ACM 38(4)*, pp. 935--962, 1991.

16. K. Lodaya. Sharpening the undecidability of interval temporal logic. *Proc. of 6th Asian Computing Science Conference*, volume 1961 of {LNCS, pp. 290--298. Springer, 2000.

17. J. Marcinkowski and J. Michaliszyn. The Ultimate Undecidability Result for the Halpern-Shoham Logic. *Proc. of the 26th Symposium on Logic in Computer Science*, pp. 377—386. IEEE Computer Society 2011.

18. A. Montanari, G. Puppis, and P. Sala. A decidable spatial logic with cone-shaped cardinal directions. *Proc. of the 18th EACSL Annual Conference on Computer Science Logic*, volume 5771 of LNCS, pp. 394—408. Springer 2009.

19. A. Montanari, G. Puppis, and P. Sala. Maximal decidable fragments of Halpern and Shoham's modal logic of Intervals. *Proc. of the 37th Int. Colloquium on Automata, Languages, and Programming - Part II* , volume 6199 of LNCS, pp. 345—356. Springer 2010.

20. A. Montanari, G. Puppis, P. Sala, and G. Sciavicco. Decidability of the interval temporal logic ABB over the natural numbers. *Proc. of the 27th Int. Symposium on Theoretical Aspects of Computer Science*, pp. 597--608, 2010.

21. B. Moszkowski. Reasoning about digital circuits. *Tech. Rep. STAN-CS-83-970*, Dept. of Computer Science. Stanford University, Stanford, CA, 1983.

22. Y. Venema. A modal logic for chopping intervals. *Journal of Logic and Computation 1(4)*, pp. 453--476, 1991.

23. C. Zhou and M. R. Hansen. Duration Calculus: A Formal Approach to Real-Time Systems. *EATCS: Monographs in Theoretical Computer Science*. Springer, 2004.

24. C. Zhou, C. A. R. Hoare, and A. P. Ravn. A calculus of durations. *Information Processing Letters*, 40(5), pp. 269--276, 1991.

Zbornik 14. mednarodne multikonference IS 2011

INTELIGENTNI SISTEMI

Proceedings of the 14th International Multiconference IS 2011

INTELLIGENT SYSTEMS

Mednarodni del konference

International part of the conference

Feature extraction and selection of breast cancer on digital mammogram imaging by using support vector machines

Azir Aliu[1], Ana Madevska-Bogdanova[2], Margita Kon-Popovska[2]
[1]Faculty of Contemporary Sciences and Technologies
South East European University
Inlindenska bb, 1200 Tetovo, Macedonia
Tel: +389 44 356 000; fax: +389 44 356 001
e-mail: azir.aliu@seeu.edu.mk
[2]Faculty of Computer Science and Engineering
Ss. Cyril and Methodius University
Rugjer Boskovic 16, 1000 Skopje, Macedonia
Tel: +389 2 3249 999; fax: +389 2 316 2078
e-mail: ana.madevska.bogdanova@finki.ukim.mk;
margita.kon.popovska@finki.ukim.mk

ABSTRACT

Digital mammography is one of the most promising methods to diagnose breast cancer, which is the most prevalent woman cancer. Constant improvements in the breast cancer detection and diagnosis is lowering the rate of mortality, but still, there is a need for further progress, particularly for better computer-aided detection and diagnosis (CAD). We investigate two types of objects, microcalcifications and masses that appear in mammography image. Their accurate detection is a key concern in computer aided detection scheme. Masses are categorized by their shape and density, while microcalcifications appear as small, bright objects. Microcalcifications represent calcium deposits and are considered as indicative of breast cancer. Among the machine learning methods such as SVM, LDA, NDA, PCA and ANN, SVM were able to achieve better classification accuracy.

In this paper, we proposed a novel method for feature extraction and selection, which we named intelligent algorithm for feature extraction and selection (IAFES). The experimental results has shown that IAFES method can extract small feature set from 6,3% to 12% of the total number of features and achieve 89.7% accuracy (94.6% for microcalcifications and 83% for masses). Average error recognition rates are less than 5% and depend on input mammogram images (if they are with high density, error recognition rate is bigger).

Keyword: *Digital mammography image, feature selection, support vector machine*

1 INTRODUCTION

Brest cancer is one of the most prevalent forms of cancer-cause of women mortality. According WHO (World Health Organization) and published statistics, 16% of all cancer deaths are due to breast cancer. Primary prevention seems impossible since the cause of this disease still remains unknown. However, efficient diagnosis of breast cancer in its early stages can reduce mortality and full recovery from disease. Mammography may detect cancer one and a half to four years before a cancer becomes clinically evident [1]. Digital mammography is the most popular method worldwide. It is one of these methods that use low-dose of x-ray. Mortality has been declining during the last few years due to the mammogram double reading protocol by two different human readers [2]. Recently, there have been many efforts to implement computer-aided detection and diagnosis systems, which can assist radiologists in reading mammograms and providing support to their diagnosis. These systems are used as second decision criteria by radiologists. CAD systems used computer technologies to detect pathological area in mammograms (calcifications, masses, etc.) and diagnose abnormal lesions that cannot be recognized by clinical breast palpation or even eyes inspection of mammograms. Therefore, CAD can play a key role in the early detection of breast cancer, obtaining a successful treatment of the disease and reduce the mortality rate.

Mass and calcification are two most significant early signs of breast cancer in mammograms. In the past years, CAD systems with artificial intelligence and machine learning techniques applications have attracted the attention of both researchers and radiologists. Recent advances in research community have concentrated on improving the performance of these CAD systems. Improvements in computer-aided detection and diagnosis systems can be obtained by high efficient, high accurate, lesion detection algorithms that can make distinction of malignant cancers from benign ones. Effectiveness of clinical applications of CAD methods is of great importance for radiologists. To

achieve better performance of CAD systems, many different approaches have been applied. From the field of machine learning techniques, both classic and modern machine learning approaches such as Bayesian networks [3, 4], artificial neural networks [5, 6] and support vector machines [7, 8, 9] have been applied. However, there is still a need for further improvements of computer-aided detection and diagnosis systems, particularly for better accuracy and performance of CAD systems, not being as high as required for practical usage. This problem can be partially solved by using a better feature selection and classification model, which will reduce number of input features excluding those with low detection accuracy.

This work proposes and evaluates a new feature selection method using SVMs. Our method is based on ensemble of two Support Vector Machines that represent subsystems to detect microcalcifications and masses with a set of combined images. They are automatically extracted for each image grid. Feature extraction is very important part of CAD systems and classification accuracy. One of the main aims of the proposed research is to obtain a balance between set of positive and negative features, and select the best features in order to obtain the highest classification rate for microcalcifications and masses detection. For this purpose we developed a system presented in figure 2. The different stages of the feature selection process and the classification task are performed by several sequentially classifiers.

2 SUPPORT VECTOR MACHINES (SVM)

Support vector machines (SVM) are a proven success method in many areas and promising machine learning technique proposed by Vapnik and his group [10]. SVM is one of the most modern classification methods based on VC dimensional theory and statistical learning theory. The goal of SVM is to define an optimal hyperplane that classifies correctly all the training vectors. Hyperplane is characterized by its direction (ω) and its exact position in space (ω_0). The goal is to search for the direction that gives maximum possible margin of hyperplane from nearest sample. Moreover, in this work we will deal with problems that are not linearly separable and for which the design of linear classifier, even in an optimal way, does not lead to satisfactory performance. For nonlinear case there exist a mapping from the input feature space into a k dimensional space, where the classes can satisfactorily be separated by a hypperplane and the only difference is that the involved vectors will be the k-dimensional mappings of the original input feature vectors. We can use any kernel function meeting Mercer's Theorem for SVM, typical kernels used in pattern recognition applications are polynomials, Gaussian radial basis function (RBF) and sparse kernels, which are used and in this research.

3 MATERIAL AND METHODS

Intelligent computer-aided detection (ICAD) architecture [12] algorithm operates in a sequence of three stages. First stage is preprocessing one, used for noise reduction, extraction of artifacts and pectoral muscle regions, which are not of interest for additional processing. In the second stage, each region is characterized by a set of features and selected features for classification. The third part is segmentation and classification stage, where the suspected lesions are marked and separated from the rest of the regions by their contour identification. Segmentation and classification stage is automatic. It can significantly reduce human effort on segmentation of the region of interest. The process of feature extraction, which is part of the second stage of ICAD system, is used to reduce the vast number of features and selects the most representative ones the mass and the microcalcification lesion and their surrounding tissue and background tissue. This way one generates the final training set. Many feature selection algorithms have been proposed. Generally, they can be classified into three categories: filter, wrapper and hybrid filter-wrapper. However, the filter, wrapper and hybrid approaches ignore the unbalances and heterogeneity of high-dimensional multi-modal features and they can only select the features of the same type [11].

This paper focuses largely on the design of the feature selection procedure, classification and normalization for the second stage of ICAD architecture (IAFES algorithm) that has capability to eliminate both irrelevant and redundant features and select the most representative feature components. As a result, the dimensionality of the required training images and the training time complexity is significantly reduced.

4 AUTOMATIC FEATURE SELECTION AND CLASSIFICATION

After the first preprocessing stage, when artifacts and pectoral muscle are removed from the mammogram, the task of ICAD is segmentation of specific image regions that have suspect lesions or abnormalities. The mammograms, both CC and MLO views are partitioned into a set of regular image grids with different sizes and levels of hierarchy, starting from the highest level $2048\,x\,2048$ to 4x4 pixel resolution. Features are automatically extracted from each region as a feature vector x. The proposed algorithm is shown in Fig.1. It includes the following steps for extraction and selection of features for the classification task.

- For each extracted image grid [11] in the mammogram, apply feature extraction methods to get the combined feature vector x.
- Use the trained classifiers, SVM-C for microcalcification detection and SVM-M for mass detection to make decision whether $x\ belongs$ to microcalcification class (1) or not (0) for SVM-C, and mass class (1) or not (0) for SVM-M.
- Use the trained SVM3 and SVM4 where the input feature vector for SVM3 is T1 and for SVM4 is T2. T1 and T2 are set of feature vectors, which

are classified as negative (0) from the previous classifier SVM-C and SVM-M respectively.

- Use the trained $SVM3_1$ and $SVM4_1$ where the input feature vector for $SVM3_1$ is T_n and T_m is for $SVM4_1$. $T_n = T1 - T3$ and $T_m = T2 - T4$ are set of feature vectors, which are classified as negative (0) from the predeceasing classifiers SVM3 and SVM4 respectively.
- Use the trained $SVM3_2$ and $SVM4_2$ where the input feature vector for $SVM3_2$ is T_{cc} and T_{mm} for $SVM4_2$. $T_{cc} = T_n - T3_1$ and $T_{mm} = T_m - T4_1$ are set of feature vectors, which are classified as negative (0) from the predeceasing classifiers $SVM3_1$ and $SVM4_1$.respectively

The task of the SVM-C and SVM-M classifiers subsystem is to decide input vector at each location whether is microcalcification/mass pattern (z=1) or not (z=0). Before eliminate both irrelevant and redundant features and select the most representative feature components, IAFES first use the feature extraction algorithm to extract microcalcifications (MC) and masses for each image grids. For each microcalcifications and masses in a training mammogram set, a regular image grids with different sizes of $k\,x\,k$ pixels where $k = 4, 8, \ldots, 256$ for microcalcification and $k = 64,\ 128, \ldots 2048$ for masses is extracted and for each of them the n dimension feature vector will be calculated for each image block, where n is the number of features ($n = 127$ for MC and $n = 165$ for masses). After that the feature vector x_i was normalized as real number in the range of [0,1]. The normalization is accomplished by the following equation.

$$z_i = \frac{z_i - z_{min}}{z_{max} - z_{min}}$$

where, z_{max}, z_{min} are maximum/minimum value of feature. Then x_i is treated as an input pattern ($z_i = 1$ for positive sample and $z_i = 0$ for negative sample). In order to find optimal parameters of (C, γ) for SVM-C and SVM-M classifiers with RBF kernel, they are trained by using 10-fold cross-validation with different parametric value of (C, γ). It is worth noting that in mammography imaging the set of positive features and negative features is unbalanced and heterogeneous. For this reason in order to minimize the error rate that is defined by the equation

$$SG = \frac{1}{2}\left(\frac{\#FP}{\#P} + \frac{\#FN}{\#N}\right)$$

where, #FP-Number of false positive, #FN-number of false negative, #P-total number of positive samples, #N- total number of negative samples. The other part of the IAFES system consists from three other classifiers (SVM3, $SVM3_1$, $SVM3_2$, for MC selection and SVM4, $SVM4_1$, $SVM4_2$ for mass selection). For each classifier evaluation was made through 10-fold cross-validation and to exploit inset feature correlations, principal component analysis (PCA) is performed on each feature subset (T1, T_n, T_{cc}, T3, $T3_1$, $T3_2$, T2, T_m, T_{mm}, T4, $T4_1$, $T4_2$) to select the most representative feature components. All classifiers in the subsystem were trained with more specific training images. By taking advantage of interconnection of classifiers and feature hierarchy (i.e., the number of features is small), system performs feature selection in four levels and learns the accurate image classifiers by using a small number of training images. At the end the IAFES system incorporated all results in order to make new training set which is smaller and the selected features are with the high detection accuracy.

5 IMPLEMENTATION

The IAFES system was implemented using MatLab, and its Image processing, SVMstruct and Symbolic math toolboxes. It was executed on an SunFire X4600 M2 x 64 Server with 4x146GB 10K RPM 2.5" SAS HDD, 8xX4600 M2 quad core processor board with 1x834HE (1.9 GHz Quad Core, ACP 55W)/8 DIMM, 16 x 4GB memory kit DDR2- 667, Solaris 10 5/08 Operating System x64/x86.

6 RESULTS

Table 1. *Experimental results of Health, Mudigonda, IAFES MC and IAFES Mass. Experimental data set is obtained from The Mammographic Image Analysis Society (MIAS) [13]*

Method	Sensitivity	#FN per Image	# Images	AUC
Heath [14]	65%	1.75	246	0.712
Mudigonda [13]	85%	2.45	56	0.901
IAFES MC [12]	94.4%	0.4	430	0.917
IAFES Mass [12]	83%	0.89	385	0.802

From figure 1, it can be shown that the IAFES system has a higher detection accuracy rate compared to other systems with same extracted features for masses and microcalcifications.

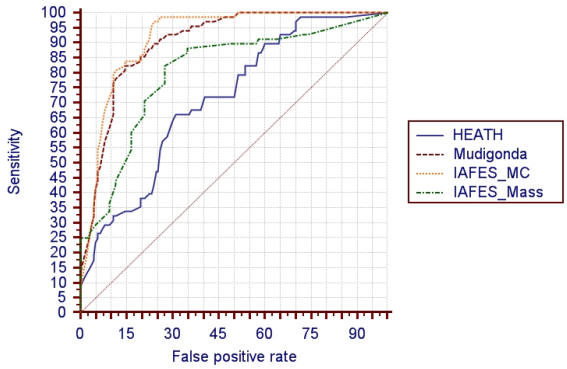


Figure 1. *ROC Curves of Microcalcification and mass detection using Health [14], Mudigonda [13], IAFES MC [12] and IAFES mass systems [12], where AUC (0.712, 0.901, 0.917, 0802).*

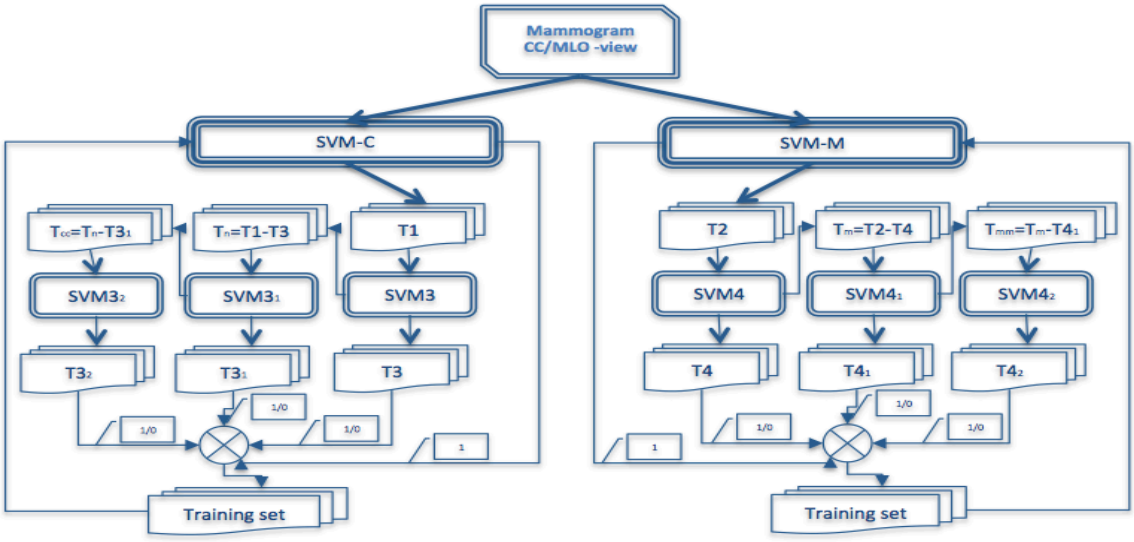


Figure 2: *Intelligent algorithm for feature extraction and selection (IAFES)*

7 CONCLUSIONS

We developed a new complete system for feature extraction and selection of breast cancer on digital mammogram imaging. First, decries processing time (balanced training sets between positive and negative features and training sets with small number of features) and increase sensitivity of system (features with high detection rate). Second advantage of our proposed classification method is that our system has two SVM main image classifiers SVM-C that is classifier for microcalcifications and SVM-M that is classifier for masses. This makes our system to have a common approach. The experiment showed that the system is an appropriate tool for extraction and selection features with high detection accuracy.

ACKNOWLEDGEMENT

This research was supported by the Computer Science - Research Center at South East European University

References

[1] Tabár, L., et al.: Beyond randomized controlled trials: organized mammographic screening substantially reduces breast carcinoma mortality. Cancer 91:1724–1731, 2000

[2] P. Autier, M. Boniol, A. Gavin, and L. J Vatten: Breast cancer mortality in neighboring European countries with different levels of screening but similar access to treatment: trend analysis of WHO mortality database. *BMJ 2011; 343:d4411 (Published 28 July 2011)*

[3] T. Xiaojian, X. Xiangmin, Z. Weiqiang, M. Yunfeng A Novel Semantic-based CAD System for Mammography. *Computer Science and Information Technology, 2008. ICCSIT '08.* International Conference. 97 – 101

[4] E. A. Fischer, J. Y. Lo, M. K. Markey:Bayesian Networks of BI-RADS Descriptors for Breast Lesion Classification. *Proc of the 26th IEEE EMBS San Francisco*, CA, USA September 1-5, 2004 2:3031-3034

[5] F. Keivanfard, M. Teshnehlab, M. A. Schoorehdeli: Feature selection and classification of breast cancer on dynamic magnetic resonance imaging by using Artificial neural network. *Proc of the 17th ICBME* 2010 1-4.

[6] R.R.Janghel, A. Shukla, R. Tiwari, R. Kala: Breast Cancer Diagnosis using Artificial Neural Network Models. *Information Sciences and Interaction Sciences (ICIS), 2010 3rd International Conference*

[7] S. W. Purnami, S.P. Rahayu, A. Embong: Feature selection and classification of breast cancer diagnosis based on support vector machines. *Information Technology, 2008. International Symposium.1-6*

[8] Wang Yi, Wan Fuyong: Breast Cancer Diagnosis via Supp ort Vector Machines. *Proc. of the 25th Chinese Control Conference 2006*, Harbin, Heilongjiang

[9] S.I. Niwas, P. Palanisamy, W. J. Zhang, N. Ashidi M. Isa, R. Chibbar: Log-Gabor wavelets based carcinoma classification using least square support vector machines. *IST, 2011 IEEE International Conference* 219-223.

[10] V. Vapnik: The nature of statistical learning theory. *Springer*, *Heidelberg* 1995.

[11] G.H. John, R. Kohavi, K. Pfleger: Irrelevant features and the subset selection problem, *ICML* 1994

[12] A. Aliu: Advance techniques for identifying digital mammography images. *PhD thesis*, 2011 University Ss Kiril and Methodius, Skopje

[13] N. R. Mudigonda R. M. Rangayyan, J. E. Leo Desautels, Detection of Breast Masses in Mammograms by Density Slicing and Texture Flow-field Analysis, *IEEE Transactions on Medical Imaging* (2001), pp. 1215—1227

[14] M. D. Heath and K. W. Bowyer, Mass detection by relative image intensity (2000)

A METHOD FOR CALIBRATING REGRESSION MODELS

Rok Kreslin, Igor Kononenko
University of Ljubljana, Faculty of Computer and Information Science
Tržaška 25, 1000 Ljubljana, Slovenia
e-mail: rok.kreslin@gmail.com; igor.kononenko@fri.uni-lj.si

ABSTRACT

We propose a method for calibrating regression models. All previous methods only focused on calibrating probabilities in classification. Our approach is based on isotonic regression and tries to improve the predicted regression value for individual instances in dataset. We evaluate the proposed method with five different regression algorithms including M5P, LR, KNN, SVM and MP. Each of listed algorithms is evaluated on a number of real and artificially constructed datasets. The proposed method improves poor predicting models but sometimes decreases the performance of good models.

1 INTRODUCTION

Calibrating probabilities in classification has been a subject of many papers in late years. Two main techniques have been provided. The first is Platt Scaling [1]; the second one is Isotonic regression [2]. This method is more general and can correct any type of isotonic distortion. We are unaware of any previous attempts of applying calibration techniques in regression. In favor of correcting any isotonic type of distortion we choose Isotonic regression. Instead of trying to fit the predicted probabilities to true probabilities we fit predicted values to true values. The difference between both approaches is shown in Figure 1.

In Section 2 we present principles of calibration and our proposed method. We describe how to transform Isotonic regression from classification to regression domain. Pseudocode of the algorithm for calibrating regression models is also given. In Section 3 we evaluate the proposed method with different regression algorithms on many different datasets. Datasets include real-world problems from UCI repository [3] and artificially constructed datasets. We present the results of evaluation and point out (dis)advantages of our method. In Section 4 we summarize the contribution of this paper and suggest directions for future improvements.

2 CALIBRATING METHOD

We begin with the assumption that we have a regression model which assigns a regression value to each instance in the dataset. We can compare that value with true value from the dataset and determine the performance of our regression model. Ideally we would always get result with zero error. Our task is to minimize the error. One way of doing it is to run original model on the dataset and then modify results via

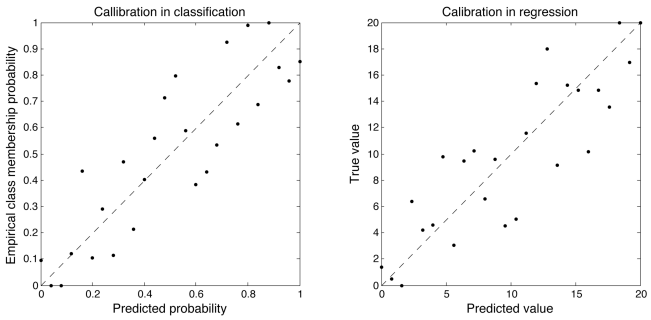


Fig. 1. **Left:** Calibration in classification: on x-axis we have predicted probabilities and on y-axis true class membership probabilities. The goal is to align both probabilities on dashed line. **Right**: Calibration in regression: on x-axis we have predicted values and on y-axis true values for individual instances. We also notice that in regression the target values do not lie on [0,1] interval as it is the case in classification. Goal remains the same - to align both values on dashed line.

calibration. By doing so we must be careful not to overfit the model with calibration. The goal of Isotonic regression is to build an isotonicaly increasing function which minimizes the squared error between predicted and true values. The reason the isotonic regression is suitable for calibration lies in an assumption of correctly ordered results given by regression models. Isotonic regression therefore corrects regression models to conform them to the above assumption. One commonly used algorithm for learning isotonic regression is PAV (pair adjacent violators) [4]. We modified it for our needs and used it for calibration of regression models. The pseudocode of the algorithm is given in Table 1.

TABLE 1: MODIFIED PAV ALGORITHM

Modified PAV algorithm for building calibrator for regression models	
1.	Sort instances according to predicted values in ascending order.
2.	**foreach** instance $inst_i$
3.	**while** $inst_i$.true < $inst_{i-1}$.true
4.	compute average
5.	s = s ∪ {$inst_i$, $inst_{i-1}$}
6.	$inst_i$ = $inst_{i-1}$
7.	Replace true values of instances in s with average value

The first step of the algorithm is to sort instances according to their predicted values in an ascending order. Then we iterate over instances and check if the true value of the previous instance is greater than the true value of the

current instance. If these two instances violate the previously introduced order constraint (hence the name pair adjacent violators) we replace the true values of these instances with their average value so that we regain the order of instances. But as we do that we have to check that the new average value is bigger than the true value of the instance we checked two steps ago. If it is not we compute the average of all three instances and assign it to them. The algorithm continues while there are still some adjacent violators.

The algorithm averages true values of instances in the area where our regression model does not perform well, where there are originally many instances violating the order constraint. In areas where the performance of our regression model is good, algorithm just copies true values.

Given a new dataset we first split it in three subsets: the training set, calibration set and the test set. By doing so we minimize possibilities for overfiting either in learning or in the calibrating phase. Next we train our regression model with the training set. The proposed method for calibration can calibrate any type of regression model. After we trained our model we run it on the calibration set. We put output of this as an input to the modified PAV algorithm presented in Table 1. This ends the learning phase. As a result we get a regression model and a calibrator.

Next we take the test set and put it as an input to previously built regression model. The model outputs predictions which we need to calibrate with the calibrator. The algorithm for calibrating (after we have build the calibrator) new instances is presented in Table 2.

TABLE 2: CALIBRATING NEW INSTANCES

Algorithm for calibrating new instances	
1.	**for each** instance $inst_i$
2.	**while** calibrator.$inst_k$.predicted < $inst_i$.predicted
3.	k++
4.	$inst_i$.predicted = calibrator.$inst_k$.true

This algorithm iterates over predicted values in the test set and finds first instance in calibrator that has the predicted value greater than the predicted value of the current instance in iteration. Then it replaces the current instance predicted value with true value of the instance in calibrator. Figure 2 graphically represents the process of calibration.

3 EXPERIMENTAL EVALUATION AND RESULTS

We evaluated the proposed method with five different regression models: LR (linear regression), KNN (with k = 4), M5P (regression tree), SVM (support vector machine) and MP (multilayer perceptron). The evaluation was done using Weka software package [5]. All of the models were built with default parameters. We did not focus on good performance of regression models on individual domains but only on the performance gain obtained by calibrating original regression models.

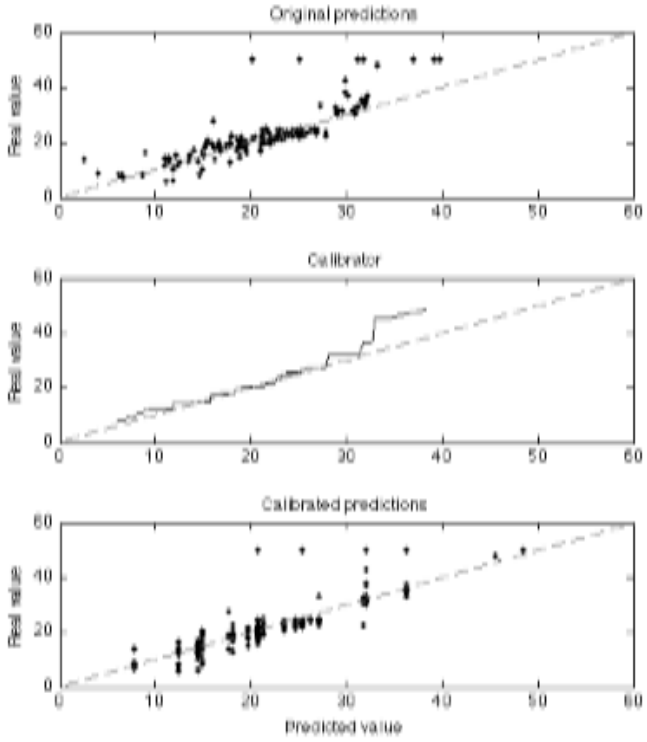


Fig. 2. **Top:** original predictions of the regression model. **Middle:** calibrator learned on the calibration set. We can see that predicted values are highly averaged on the areas where our regression model performs poorly. **Bottom:** calibrated predictions.

We used two packs of datasets. The first pack contains 20 real world datasets from UCI repository [3]. The smallest dataset contains 40 instances and the largest contains 8192 instances. The second pack contains 10 artificially constructed datasets used in [6]. Datasets are constructed in a way that some of the regression models perform well on a specific dataset and others do not.

Before evaluating we binarized nominal attributes and replaced missing values with the average values of attributes. Each model was evaluated on every domain using 10 times repeated random subsampling validation. Each time we randomly shuffled dataset and split in two parts. The first half of dataset was used as a training set. The second half of dataset we split once again and used the first half for calibration and the second half for testing. The performance measure was RMSE [7]:

$$RMSE = \frac{n \cdot MSE}{\sum_i \left(f_i - \overline{f}\right)^2} \qquad (1)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(f_i - g_i\right)^2$$

where f_i is true value, $\overline{f}$ the average of real values and g_i the predicted value. Acceptable values for RMSE lie on interval $0 \le RMSE \le 1$ because values bigger than 1 can be trivially outperformed by using the average. However in our evaluation we sometimes get values bigger than 1 but

this is not of concern because we are only interested in the performance gain after calibration.

Table 3 presents RMSE gains after calibrating real-world datasets. With LR, KNN, M5P and SVM we get worse results than if we use no calibration. But with MP we greatly improve RMSE score as we can see many peeks at MP values in Figure 3. The first set of peaks we get with **bolts** dataset. Calibration decreases accuracy for all of the regression methods. The dataset only contains 40 instances. So we get 20 instances for learning, 10 for calibrating and 10 for testing. This is a very small number of instances and as we can not induct good model from such a low number of instances we can not make good calibrator either.

TABLE 3: PERFORMANCE GAIN AFTER CALIBRATION

RMSE gain after calibration on real world datasets

Dataset	Method				
	LR	KNN	M5P	SVM	MP
abalone	0.00	0.00	0.00	0.01	0.11
auto_price	-0.07	0.02	-0.08	-0.02	0.09
autoMpg	0.00	-0.01	-0.01	-0.01	0.05
bank8FM	0.02	0.07	0.00	0.02	0.02
bodyfat	-0.03	-0.02	-0.03	-0.03	-0.03
bolts	-0.52	-0.46	-0.41	-0.14	-0.04
cloud	-0.21	0.02	-0.20	-0.28	-0.24
cpu	-0.24	-0.20	-0.32	-0.69	-0.40
cpu_small	0.19	0.00	0.00	0.28	0.01
delta_ailerons	-0.02	-0.01	-0.01	-0.01	0.04
echoMonths	-0.14	0.13	-0.16	-0.04	1.07
fishcatch	-0.01	-0.07	-0.04	0.00	-0.03
housing	0.03	-0.03	-0.01	0.05	0.06
lowbwt	-0.18	-0.01	-0.23	0.00	0.70
pollution	0.94	-0.38	-0.41	-0.16	1.64
pwLinear	-0.07	-0.05	-0.03	-0.07	0.02
sensory	-0.01	-0.01	-0.02	0.00	1.83
stock	0.01	0.00	0.00	0.01	0.01
strike	-0.03	0.14	0.12	-0.07	1.19
triazines	0.21	-0.08	-0.02	0.17	1.08
AVG	**-0.01**	**-0.05**	**-0.09**	**-0.05**	**0.36**

The next set of peaks appears with **cloud** dataset. The only method that is not affected by calibration is KNN. The reason for decreased performance of methods after calibration is shown in Fig. 4. We can see two outliers after calibration; two right-most points. This dataset clearly shows a flaw in our algorithm for calibrating new instances (Table 2). Points whose original predicted values lies near 3 (Fig. 5, top) get calibrated to next bigger value in calibrator and that is 6 (Fig. 5, bottom). By doing this we introduce large error which is clearly shown as decreased RMSE.

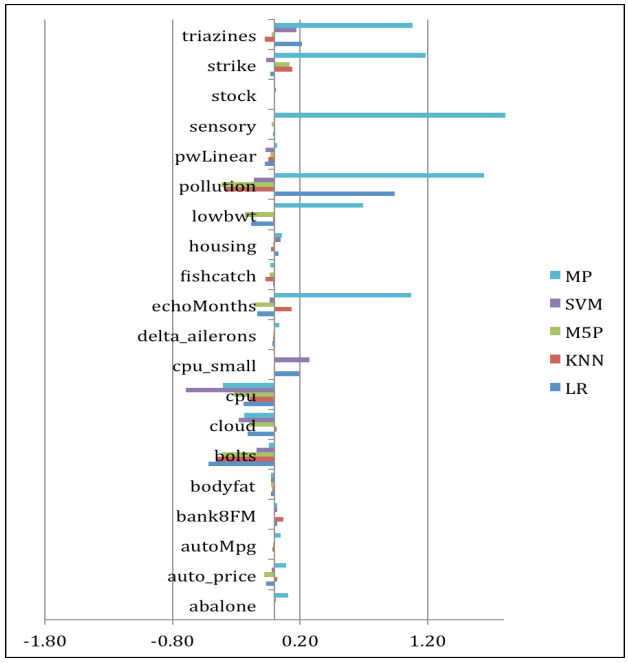


Fig. 3. Histogram of performance gain after calibration on real-world datasets. The reason for improved performance of MP is clearly visible as many light blue peaks.

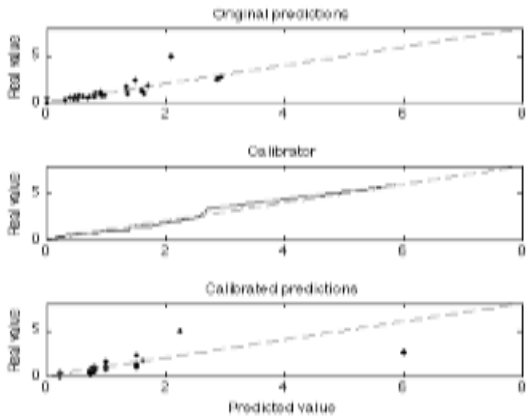


Fig. 4. **Top**: original predictions of SVM model on cloud dataset. **Middle**: calibrator. **Bottom:** calibrated predictions.

The next set of peaks appears with **cpu** dataset where all models are affected. The reason is exactly the same as with bolts dataset. The effect is visible once again in **pollution** dataset. Better way of calibrating new instances would be to find two instances in calibrator: one with just lower predicted value and one with just greater predicted value, fit linear model truth their real values and transform the real value of the given instance with this linear model. However, the effect does not seem to show if datasets are large enough even with the current calibrating algorithm. All of the above mentioned datasets contain less than 300 instances.The set of domains with peaks in MP models include **echoMonths**, **lowbwt**, **pollution**, **sensory**, **strike** and **triazines**. The reason lies in poor initial predictions of

MP that are later averaged by the calibrator. The effect is shown in Figure 5. Even though we gain performance with calibration, RMSE values remain above 0.9 so models are not very usable even after calibration.

Table 4 presents RMSE gains after calibrating artificially constructed datasets. With LR and M5P we get worse results than with no calibration, the performance of KNN stays the same on average, and performances of SVM and MP do get better after calibration.

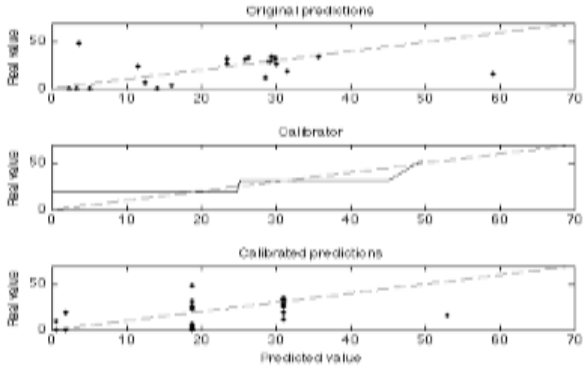


Fig. 5. **Top**: original predictions of MP model on echoMonths dataset. **Middle**: calibrator, highly averaged. **Bottom:** calibrated predictions.

TABLE 4: PERFORMANCE GAIN AFTER CALIBRATION

RMSE gain after calibration o artifical datasets

Dataset	Method				
	LR	KNN	M5P	SVM	MP
dDisjunct	0.22	-0.12	-0.03	0.59	0.07
dDisjunctBin	0.59	0.00	0.00	1.04	0.00
dLinear	0.00	0.00	0.00	0.00	0.00
dLocLinear	0.11	0.00	0.00	0.03	0.00
dNonLinPoly	-3.32	0.01	0.01	0.01	0.06
dNonLinSinCos	0.12	0.00	0.01	0.11	0.05
dRandom	-1.70	0.20	-1.70	0.00	0.18
dRedundant	0.00	0.00	0.00	0.00	0.00
dXor	-0.40	-0.05	-0.04	0.47	0.12
dXorBin	-0.57	0.00	0.07	0.95	0.18
AVG	**-0.50**	**0.00**	**-0.17**	**0.32**	**0.07**

The first dataset is **dDisjunct**. Initially M5P is the most succesful method for modelling this dataset. The error after calibration remains almost the same. KNN is second best method for this dataset. Calibration slightly affects it in a negative way and increases RMSE error for 0.12. Three remaining methods gain from calibration because it averages out predictions in areas where models perform poorly. Performances on **dDisjuntBin** are the same except the fact that MP gives no error even without calibration.

dLinear dataset presents a linear problem which all methods solve with no error before calibration. And calibration does not negatively affect this. **dLocLinear** dataset is a piecewise linear dataset. MP, M5P and KNN solved the dataset without error with or without calibration. LR and SVM produced low error before calibration. After calibration errors of these two methods became almost zero.

The next dataset is **dNonLinPoly**. Best methods for solving this dataset before calibration are KNN, M5P and MP. Calibration does not affect them. SVM performs badly with or without calibration. But in case of LR the calibration greatly decreases performance of the model. The reason behind it is the same as in real-world datasets – flaw in calibration method that produces outliers. In **dNonLinSinCos** all methods gain from calibration. The gain is proportional to the error before calibration. **dRandom** dataset is completely random so none of the methods perform well. Calibration slightly decreases RMSE error with all methods except M5P and LR. The reason for this is once again flaw in our calibration method and outliers produced by it. **dRedundant** dataset presents linear problem where one of attributes is redundant (has the same values as another attribute). All methods solve the problem without error and calibration has no effect.

The last two datasets are **dXor** and **dXorBin**. This datasets have only two resulting values, 0 and 1. The performance of different regression algorithms is similar on both of them. SVM performs worst and calibration greatly improves the results. Because datasets present XOR problem which is highly non-linear, LR performs badly.

4 CONCLUSION

The proposed method deals with calibration of regression models. It is simple to implement and can be used with any regression method. Results show that it can greatly improve performance of otherwise poorly performing models and slightly improves it for models with good performance without calibration. The only requirement is that the original dataset is large enough. It has a flaw that it produces outliers which greatly enlarge overall error after calibration. In the future we will investigate this flaw of the algorithm even further and try to overcome it. One way worth investigating is already presented in Section 3.

References

[1] J. Platt, Probalistic outputs for support vector machines and comparison to regularized likelihood methods*, Advances in large margin classifiers,* 1999, pp. 61-74

[2] B. Zadrozny, C. Elkan, Obtaining calibrated probability estimates from decision trees and naïve bayesian classifiers*, Proc. ICML*, 2001, pp. 609-616

[3] A. Frank, A. Asuncion, *UCI Machine Learning Repository* [http://archive.ics.uci.edu/ml], CA: University of California, School of Information and Computer Science, 2011

[4] M. Ayer, H. Brunk, G. Ewing, W. Reid, E. Silverman, An empirical distribution function for sampling with imcomplete information*, Annals of mathematical statistics,* 1955, pp. 641-647

[5] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten, The WEKA data mining software: An update, *SIGKDD Explorations 11(1)*, 2009, pp. 1–3

[6] E. Štrumbelj, I. Kononenko, A General Method for Visualizing and Explaining Black-Box Regression Models, *Lect. notes comput. sci.*, part 2, 2011, pp. 21-30.

[7] I. Kononenko, M. Kukar, *Machine Learning and Data Mining: Introduction to Principles and Algorithms*, Horwood publ., 2007

LOCAL NEIGHBOURHOOD BASED ESTIMATION OF PREDICTION INTERVALS FOR ARBITRARY REGRESSION MODELS

Darko Pevec, Erik Štrumbelj, Igor Kononenko
University of Ljubljana
Faculty of Computer and Information Science
Tržaška 25, 1000 Ljubljana, Slovenia
e-mail: darko.pevec@fri.uni-lj.si

ABSTRACT

We consider the general case of empirically constructing input dependent prediction intervals for regression models. The common approach with neural networks is to estimate the model's error through bootstrap and to use the residuals to estimate the data error (or noise) by maximum likelihood estimation. Both error estimates are then combined to form prediction intervals and the two steps can be done in a number of ways. We compare them with procedures that form prediction intervals directly from individual samples' local neighbourhoods. Because bootstrap approaches aim to produce valid intervals, they generally outperform the simpler but much more time efficient methods in terms of consistency.

1 INTRODUCTION

We focus on the general regression case where there is an unknown underlying function $f(\mathbf{x})$ that generated the available observations $(\mathbf{x}_i, y_i)$, which are corrupted with zero mean noise: $y_i = f(\mathbf{x}_i) + \varepsilon(\mathbf{x}_i)$. If we assume the noise has a Gaussian distribution with zero mean and constant variance throughout the attribute space, then the prediction intervals are straightforward to calculate with analytical methods from elementary statistics for constructing confidence and prediction intervals.

Noise in data and non-uniform distribution of examples represent a challenge for learning algorithms, leading to different prediction accuracies in different parts of the problem space. This component is called the data noise variance, σ_d^2. Apart from distribution of learning examples there are also other causes that influence the accuracy of prediction models: their generalization ability, bias, resistance to noise, avoidance of over-fitting, etc. These factors form the component called the model uncertainty variance, σ_m^2. The two components are assumed to be independent of each other and their sum is the total prediction variance: $\sigma^2 = \sigma_m^2 + \sigma_d^2$ [7].

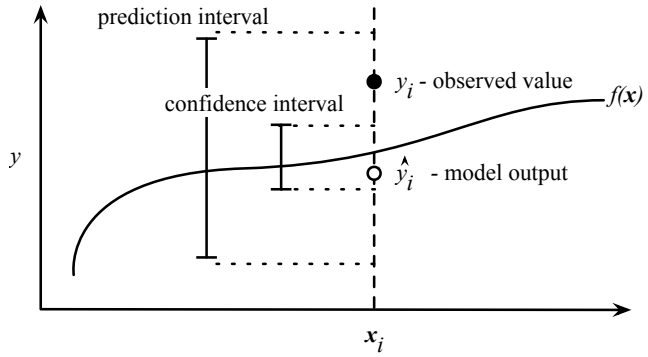


Figure 1: *Illustration of the confidence and prediction interval for sample* $\mathbf{x}_i$.

Figure 1 illustrates the relationship between the prediction interval (PI), the confidence interval and the three relevant values. Confidence intervals are concerned with the accuracy of the model's estimate $\hat{y}_i = \hat{y}(\mathbf{x}_i)$ of the true but unknown function $f(\mathbf{x}_i)$. Therefore confidence intervals strive to capture the distribution of the quantity $f(\mathbf{x}_i) - \hat{y}_i$. However in applications, it is more informative to quantify the accuracy of the model's output with respect to the realized observations. Prediction intervals predict the distribution of individual future points and are concerned with the quantity $y_i - \hat{y}_i$. If we expand the first term,

$$y_i - \hat{y}_i = f(\mathbf{x}_i) + \varepsilon(\mathbf{x}_i) - \hat{y}_i = \left[f(\mathbf{x}_i) - \hat{y}_i\right] + \varepsilon(\mathbf{x}_i),$$

we see that the corresponding confidence interval is necessarily enclosed within the prediction interval. In real world applications, PI is more practical than the confidence interval because the former is concerned with the accuracy with which it is possible to predict the observed value itself, but not just the accuracy of the estimate of the true regression value [4].

It is common to approach the estimation of the model's uncertainty variance σ_m^2 through bootstrap and use the residuals to estimate the data noise variance σ_d^2 by means of maximum likelihood estimation [2,7]. Both estimates are then combined to form PI. Bootstrap based methods are good estimators but their major drawback is computational complexity. Having a more efficient method for constructing PI would be of great benefit in real-world applications. In this paper we compare these procedures with procedures that empirically construct the PIs in local neighbourhoods. The latter can be applied in cases where there is no access to the learning algorithm.

In the following section we first overview the related work and the most promising approaches for model-independent modelling of PIs found in the literature. In Section 3 we adopt three existing approaches, present our variation of one of the former and propose a local neighbourhood based method (two variants) that use the residuals in the local neighbourhood of an individual sample to formulate its estimates. Methodology behind our empirical evaluation is presented in Section 4 and the results in Section 5. Finally, we conclude in Section 6.

2 RELATED WORK

In the machine learning community, Nix and Weigend [3] were among the first to address the problem of estimating the mean and variance of the target probability distribution of the target, when it is not assumed the noise is uniform across the input space. The method extends neural networks with an additional output neuron that computes $s^2(\mathbf{x})$, the localized estimate of the variance $\sigma^2(\mathbf{x})$ of the target around the true underlying function $f(\mathbf{x})$. This provides a measure of uncertainty of the usual network output for each input. Assumptions that there is a sufficiently large data set, that there is no risk of over-fitting and that the neural network finds the correct regression, are usually too strict in practical applications with limited data. Tibshirani [5] has performed a comparison of three statistical methods for forming confidence intervals: the delta method based on the Hessian, bootstrap estimators, and the sandwich estimator. Tests performed with neural networks show that the bootstrap methods provide the most accurate estimates of the standard errors of the predicted values.

Later, the uncertainty of prediction errors was decomposed into two independent components, namely the model uncertainty and inherent noise of the data. While confidence intervals deal with the deviations of the models predictions and the mean of the target, PIs need to account for the variation of the data. Heskes [2] proposed a method for computing PIs in two stages. Confidence intervals are formed for the ensemble of neural networks from bootstrap replicates of the original data set. The variance of the target distribution is estimated with the samples left out of the bootstrap sample. Residuals of these samples are then fed to a separate neural network with an exponential activation function for the output neuron to ensure positive variance.

Zapranis and Livanis [7] have combined the works of Heskes and Tibshirani, and compared the methods with analytical approaches. Their results indicate clear superiority of the combination of the bootstrap and maximum likelihood approaches in constructing PIs.

A different approach to constructing PIs is presented by Shreshta and Solomatine [4]. They partition the input space into fuzzy clusters and form PIs for each cluster based on the empirical distribution of the errors associated with all instances within the cluster. Prediction intervals for new samples are formed according to the grade of their membership in identified clusters, without the calculation of confidence intervals.

These approaches inspired us to test their general use. In the following section we generalize different methods of constructing PIs so they can be applied to arbitrary regression models.

3 ESTIMATION OF PREDICTION INTERVALS

The most straightforward approach is from [7]. Given a fixed model with available learning algorithm and training set, the reinterpretations of the method go as follows. To estimate $\sigma_m^2(\mathbf{x})$, bagging is done with the training data, using the model at hand. Confidence intervals are formed by assuming the normal distribution and calculating $\sigma_m^2(\mathbf{x})$, the variance of the bagged predictions. Then a radial basis function network (RBFN) is trained on the residuals of the bagging predictions on the test data set and used to provide the estimate $\sigma_d^2(\mathbf{x})$. Assuming normal distribution, $\sigma^2(\mathbf{x}) = \sigma_m^2(\mathbf{x})+\sigma_d^2(\mathbf{x})$, so the PI is $\hat{y}_{bag}(\mathbf{x}) \pm z_{\alpha/2}\, \sigma^2(\mathbf{x})$. We propose a variation of this method, where residuals from a separate validation set are used to estimate $\sigma_d^2(\mathbf{x})$. These two methods are noted in tables as BagMLa and BagMLb. Generalizing the method from [2], $\sigma_m^2(\mathbf{x})$ is again obtained with calculating the variance of the bagged model. The estimation of $\sigma_d^2(\mathbf{x})$ is done by using the out-of-sample bagged residuals to train the RBFN. Assuming normal distributions, the BagMLc PI is $\hat{y}(\mathbf{x}) \pm z_{\alpha/2}\, \sigma^2(\mathbf{x})$. Keen readers would notice that PIs from methods BagMLa and BagMLb are centred at $\hat{y}_{bag}(\mathbf{x})$ but BagMLc is centred at $\hat{y}(\mathbf{x})$. The idea is that $\hat{y}_{bag}(\mathbf{x})$ provides a more stable estimate of the true function $f(\mathbf{x})$ than $\hat{y}(\mathbf{x})$ does. Authors of [7] have actually gone further in lowering the variance of the estimate, by using an ensemble of ensembles to form $\hat{y}'_{bag}(\mathbf{x})$.

The second family of methods relies on local neighbourhoods (LN) for estimation of σ^2. Adopting the idea from [4], we implemented k-means clustering on the training data. The number of clusters is defined with the common heuristic $k = \sqrt{n/2}$, where n is the size of the training set. LNcl PIs are constructed for each cluster directly from the empiric distribution of residuals, by taking the appropriate percentiles (i.e. the 2.5 and 97.5 percentiles for 95% PIs). For an unseen example, the nearest cluster defines the PI.

Finally, we go a step further in the simplification of this idea. Instead of using clustering, we use the nearest neighbours' algorithm to constructs PIs in the following way. First, signed residuals of the training set are obtained. Enforcing zero mean of the residuals produces bias correction. For the distribution of the nearest neighbours residuals, we assume the normal distribution and calculate the standard deviation, which gives us $\sigma^2(\boldsymbol{x})$. The number of used neighbours is relative to the size of the data set. We tested two variants with different sizes of neighbourhoods. With method LN5, the size of the neighbourhood is 5% of the total population. The second variant LN100 is computationally even simpler, as it

covers the whole (100%) population and is therefore equivalent to analytic methods that assume constant variance.

4 EXPERIMENTS

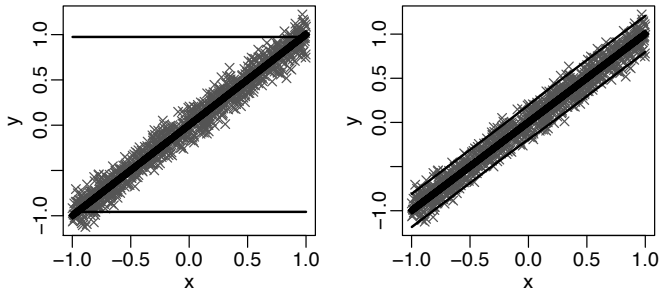
The experiments were done in Java with the use of Weka [1], from which we used seven regression models with default model parameters: Linear Regression, M5P - regression trees, Multi Layer Perceptron with 5 hidden neurons, RBFN, IBk - 11-nearest neighbours, regression SVM, Bagging with 20 regression trees.

For the test datasets, we created a compilation of 56 artificial and real-world data sets, which is available from the author on request. Seven artificial datasets comprise of one-dimensional sets of 3000 instances of the normal Gaussian noise; linear function with added 10% noise from uniform and standard normal distributions; three-region piece-wise constant function, which regions have either 1%, 2% and 5% or 10%, 20% and 50% added standard normal noise; a quadratic function with added standard normal noise multiplied by the square of x; sinusoidal function with mixed Gaussian noise. Five multi-dimensional artificial datasets from [6] have five features, 1000 instances and comprise of linear, polynomial and trigonometric functions. We included the artificial data set used in [7]. The 33 real-world data sets are compiled from Weka's numeric, regression and UCI collections.

When evaluating the PIs, we are interested in two quantities. The first quantity is Prediction Interval Coverage Probability (PICP). It is defined as the percentage of test instances for which the target value is contained within the PIs. By definition, the PIs should on average enclose the realized observations in $(1-\alpha)$% of all cases (typically 95%); hence PICP is a quantization of validity.

The second quantity of interest is the average interval length or the Mean Prediction Interval (MPI). As we want to quantify the optimality of different approaches, we introduce the measure Relative Mean Prediction Interval (RMPI), for which we need a baseline or default for comparison. We construct default PIs directly from the distribution of all target values in the test data set by taking the appropriate quantiles (i.e. 2.5 and 97.5 percentiles for 95% PIs). MPI_{def} is the length of the PI from the default method. RMPI is then calculated by taking the method's MPI and normalizing it with MPI_{def}. Figure 2 shows the difference between constant and data-dependent PIs. How the default PIs look on the artificial data set generated from a linear function with Gaussian noise of constant variance is shown in Fig. 2(a).

Testing is performed in the following way. Every data set is first randomly split into three equally sized sets: training set, validation set and test set. For training of models, only the training set is available, but for construction of the PIs, the validation set can also be used. The test set is used only for evaluation. On this set, default PIs are constructed, then the PICP and RMPI are calculated for all methods. The runtimes are also measured so we can compare the methods' efficiency.



(a) PICP=0.950 (b) PICP=0.941

Figure 2: *Difference between constant and data-dependent PIs for the linear regression model. Figures show predictions intervals for the linear function with added 10% Gaussian noise constructed with (a) default method and (b) the Heskes approach (BagMLc). The observed values are grey points, the models predictions are black.*

5 RESULTS

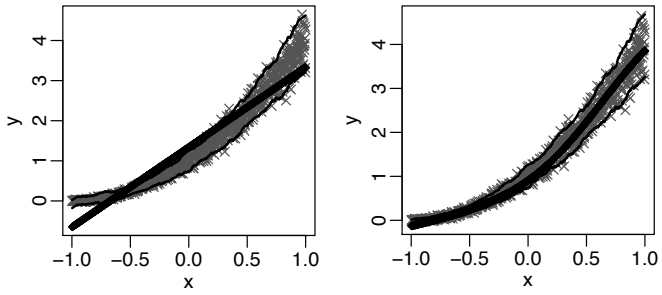
The evaluation on artificial data sets of the six different methods for constructing PIs is summed up in Table 1. PICP columns present the 2.5 percentile, the mean and 97.5 percentile of the PICP distribution. The runtime of construction is averaged over datasets and models. As we are constructing 95% PIs, the target PICP is 0.95 (the last row). We see that the local neighbourhood methods LN5 and LN100 on average produce intervals with correct coverage on the artificial sets, but the clustering method (LNcl) underestimates the PIs. The reason for this is over-fitting to the training data. On the real-world domains, as presented in Table 2, all local methods underestimate the mean PICP. We see from Fig. 3 that neighbourhood based methods actually produce intervals that are independent of the model, so if the model follows the data, the PIs are credible. In some real-world datasets, these methods fail and this is shown in their very low $PICP_{2.5}$.

method	$PICP_{2.5}$	$PICP_{50}$	$PICP_{97.5}$	RMPI	runtime [ms]
BagMLa	0.766	0.950	1.000	0.618	9489
BagMLb	0.745	0.954	1.000	0.632	9485
BagMLc	0.726	0.951	1.000	0.638	9245
LNcl	0.866	0.927	0.945	0.372	212
LN5	0.893	0.951	0.994	0.388	225
LN100	0.902	0.951	1.000	0.584	91
default	0.950	0.950	0.950	1.000	52

Table 1: *Experimental evaluation of methods for constructing PIs on artificial data sets.*

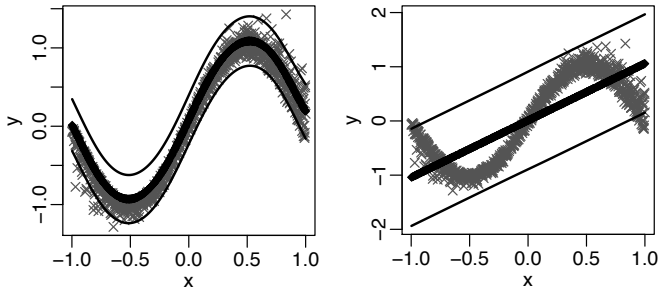
method	$PICP_{2.5}$	$PICP_{50}$	$PICP_{97.5}$	RMPI	runtime [ms]
BagMLa	0.701	0.931	1.000	0.750	12502
BagMLb	0.858	0.957	1.000	0.918	12466
BagMLc	0.848	0.951	1.000	0.858	12720
LNcl	0.080	0.762	0.945	0.498	501
LN5	0.022	0.665	0.989	0.460	381
LN100	0.012	0.831	1.000	0.627	96
Default	0.950	0.956	1.000	1.000	84

Table 2: *Experimental evaluation of methods for constructing PIs on real-world data sets.*



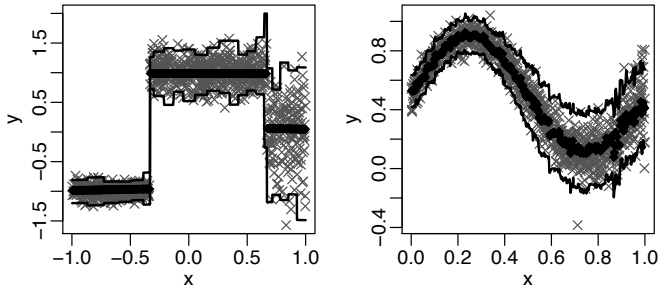
(a) PICP=0.958 (b) PICP=0.953

Figure 3: *Predictions intervals for the quadratic function with input-dependent noise, constructed with the LN5 method. The used models are* (a) *linear regression and* (b) *multi-layer perceptron.*



(a) PICP=0.976 (b) PICP=0.979

Figure 4: Valid *predictions intervals constructed for the sinusoid function with mixed noise with (a) BagMLa and MLP (b) BagMLb and linear regression.*



(a) PICP=0.939 (b) PICP=0.951

Figure 5: *Examples of credible PIs constructed with (a) LNcl and regression trees on the piece-wise constant function, (b) BagMLc and bagging of regression trees on the Zapranis and Livanis artificial test set.*

On the contrary, Figure 4 shows how the BagML methods strive to produce valid PIs. As is evident from both tables, local neighbourhood methods on average produce narrower intervals (lower RMPI) than BagML methods, and are almost five times more efficient.

Figure 5 shows two cases when the methods produce credible PIs. In Fig. 5(a), the LNcl method fits the model and data well, though its PIs are a bit underestimated (PICP is 93.2%). In Fig. 5(b) we show the Heskes approach on the Zapranis and Livanis artificial test set. Among the BagML approaches, the Heskes variant produces the most consistent intervals.

6 CONCLUSIONS

We reinterpreted existing approaches that use bootstrap and maximum likelihood estimation for use with arbitrary regression models and added a variation that makes use of a validation set. We presented simpler methods that use local neighbourhoods to form PIs. A method with clustering was adopted and we proposed an even simpler method that uses signed residuals. The comparison of the two families of methods shows that the former approaches generally outperform the simpler ones in terms of consistency. Local neighbourhood methods have proven to be much more time efficient with promising results on the artificial datasets, but the results on real-world datasets have shown that these methods can badly underestimate PIs due to their ignorance of the model's errors. We have also shown that BagML methods fit the models and therefore produce valid PIs, so when the model does not fit the data well, the RMPI of these methods is considerably larger than that of the local neighbourhood based methods. In further work we will continue our search of better methods for constructing PIs. Our focus will be on finding more efficient ways of estimating the models' uncertainty and in ways to improve local neighbourhood based methods. An adequate graphic representation of PIs for multi-dimensional data is of our interest as well.

Acknowledgments. This work was supported by a grant from the Slovenian Research Agency (P2-0209).

References

[1] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten. The WEKA data mining software: an update. SIGKDD Explorations 11(1), 10–18. 2009.

[2] T.Heskes. Practical confidence and prediction intervals. Advances in Neural Information Processing Systems 9. pp. 176–182. MIT press. 1997.

[3] D.A. Nix, A.S. Weigend. Estimating the mean and variance of the target probability distribution. IJCNN. 1994.

[4] D.L. Shrestha, D.P. Solomatine. 2006 special issue: Machine learning approaches for estimation of prediction interval for the model output. Neural Netw. 19, 225–235. 2006.

[5] R. Tibshirani. A comparison of some error estimates for neural network models. Neural Comput. 8, 152–163. 1996.

[6] E. Štrumbelj, I. Kononenko. A general method for visualizing and explaining black-box regression models. Adaptive and Natural Computing Algorithms. LNCS, vol. 6594, pp. 21–30. 2011.

[7] A. Zapranis, E. Livanis. Prediction intervals for neural network models. Proceedings of the 9th WSEAS International Conference on Computers. pp. 76:1–76:7. 2005.

SISTEM ZA PODPORO KONSTRUIRANJU KOMBINIRANEGA PREOBLIKOVALNEGA ORODJA

David Potočnik[1], Boštjan Pesan[1], Miran Ulbin[2]
[1]EMO-orodjarna d.o.o., [2]Fakulteta za strojništvo v Mariboru
Bežigrajska 10, 3000 Celje, Slovenija
Tel: +386 3 4282100; fax: +386 3 4282101
e-mail: david.potocnik@emo-orodjarna.si

POVZETEK

Konstruiranje orodij za preoblikovanje pločevine je kompleksen postopek, ki temelji na opredelitvi in preračunu številnih preoblikovalnih parametrov. Tovrstno opravilo predstavlja veliko težavo predvsem neizkušenim konstrukterjem, ki pogosto nimajo znanja, kako kakovostno izvesti ta postopek. Dejstvo, da jim večina sodobnih CAD-sistemov ne omogoča podpore v tej smeri, še dodatno razširi omenjeno problematiko. Da bi naredili korak k razrešitvi te pomanjkljivosti CAD-sistemov, smo s pomočjo programskega paketa CATIA V5 razvili sistem namenjen podpori konstruiranju kombiniranega preoblikovalnega orodja za izdelavo votlega lončka s prirobnico. Njegovo delovanje temelji na podlagi integriranega konstrukterskega znanja vzpostavljenega s pomočjo uporabniških parametrov in programskimi relacijami (povezavami) med njimi. Uporaba sistema omogoča enostavno opredelitev konstrukcijskega problema, avtomatski preračun parametrov, s katerim uporabnik pridobi vse potrebne informacije za konstruiranje orodja ter avtomatsko izgradnjo CAD-objektov kvadratne pločevine, krožne platine ter končnega izdelka. Ker ti postopki tradicionalno potekajo ročno in brez računalniške podpore, naš sistem bistveno prispeva k skrajšanju časa konstruiranja, dviguje kakovost konstrukcije ter omogoča izobraževanje neizkušenih konstrukterjev.

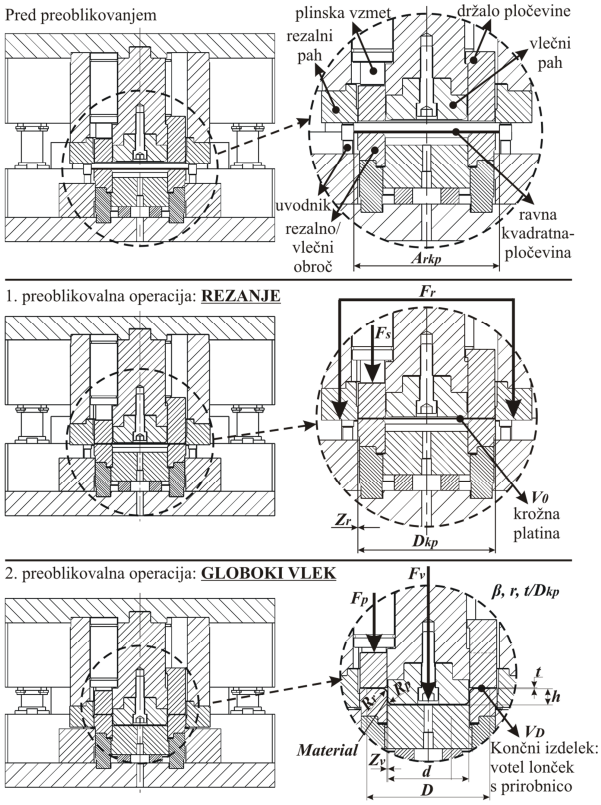
1 UVOD

Visoko konkurenčno okolje na področju orodjarstva danes od podjetij zahteva hitrejše, kakovostnejše in ekonomsko upravičene konstrukcijske rešitve, pri čemer zavzema uporaba računalniško podprtega konstruiranja ali krajše CAD (angl.: Computer Aided Design = CAD) eno izmed pomembnejših mest za doseganje teh kriterijev. Navkljub prednostim, ki jih ponujajo sodobni 3D-CAD programski paketi, pa ti samostojno ne zagotavljajo potrebnega in zadostnega znanja namenjenega konstrukterski podpori. Te pomanjkljivosti lahko odpravimo z uporabo 3D-CAD sistemov, ki omogočajo parametrično modeliranje in dovoljujejo programiranje konstrukcijskega procesa. Parametrično modeliranje omogoča uporabniku nadzor, spreminjanje in upravljanje grafičnih objektov z uporabo matematičnih spremenljivk-parametrov. Opis CAD-modela z uporabo parametrov dovoljuje uporabniku enostavno prilagoditev ali posodobitev atributov objekta preko spremembe vrednosti parametra. Nekateri parametrični modelirniki, kot je CATIA V5, vključujejo tudi dodatna orodja, s katerimi lahko preko uporabe raznovrstnih parametrov na različne načine zajemamo in vgrajujemo inženirsko znanje (glej [1] in [2]). Številni raziskovalci v svojih delih [3, 4, 5] učinkovito izkoriščajo omenjene prednosti parametričnih modelirnikov za razvoj avtomatiziranih, na znanju temelječih in inteligentnih sistemov.

Prva faza konstruiranja vsakega preoblikovalnega orodja zajema skrben pregled 3D-CAD modela pločevinskega izdelka, ki običajno ponuja omejen niz informacij. Zadovoljivo so predstavljeni le geometrijska oblika in končne dimenzije, določitev in preračun preostalih pomembnih preoblikovalnih parametrov pa sta prepuščena konstrukterju. Pri tem pa predvsem neizkušeni konstrukterji brez zadostnega znanja pogosto naletijo na velik problem, saj jim sodobni CAD-sistemi ne omogočajo kakovostne podpore v tej smeri. Da bi naredili korak k rešitvi tega problema, smo s pomočjo programskega paketa CATIA V5 razvili sistem namenjen podpori konstruiranju kombiniranega preoblikovalnega orodja za izdelavo votlega lončka s prirobnico. Skozi prispevek opisujemo izgradnjo sistema, njegovo delovanje ter podajamo njegovo uporabno vrednost v smislu podpore neizkušenim konstrukterjem.

2 OPIS TER RAZVOJ SISTEMA

Zamišljen končni izdelek je izdelan iz ravne kvadratne pločevine in z uporabo kombiniranega preoblikovalnega orodja, ki omogoča dve različni preoblikovalni operaciji. Prva tehnološka operacija predstavlja rezanje, kjer iz kvadratne pločevine pridobimo krožno platino. Ta se v naslednji fazi preoblikuje s postopkom globokega vleka, s katerim tako pridobimo končno obliko izdelka. Na sliki 1 so s pomočjo 2-D konstrukcije orodja jasno prikazani preoblikovalni parametri, ki bodo predstavljali osnovo za izgradnjo in delovanje sistema, preoblikovalne operacije za pridobitev izdelka ter pomembne komponente orodja. V nadaljevanju podajamo postopek opredelitve odvisnih in neodvisnih preoblikovalnih parametrov, njihov pomen pri konstruiranju orodja ter postopek njihove parametrične vgradnje v bazo znanja znotraj 3D-CAD sistema CATIA V5.



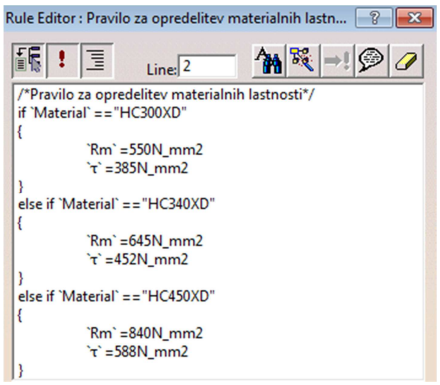
Slika 1: *Preoblikovalni parametri, tehnološki operaciji ter sestavni deli orodja*

2.1 Neodvisni preoblikovalni parametri

Vse, s čimer konstrukter običajno razpolaga na začetku konstrukcijskega postopka, je CAD-model z geometrijo in dimenzijami končnega pločevinskega izdelka. Te dimenzije zato predstavljajo neodvisne geometrijske preoblikovalne parametre, kot so premer prirobnice *D*, premer lončka/paha *d*, debelina pločevine *t* in višina votlega lončka *h*, katere smo z istoimenskimi dolžinskimi parametri opredelili znotraj geometrijske skice končnega izdelka.

Poleg neodvisnih geometrijskih parametrov obstaja še eden neodvisen parameter, ki ima velik vpliv na preoblikovalni postopek in pogosto ni zadovoljivo predstavljen znotraj CAD-modela izdelka. To je material pločevine, ki smo ga v sistem vključili z opisnim parametrom poimenovanim »*Material*«. Njegove vrednosti predstavljajo tri različna dvofazna jekla za hladno preoblikovanje: XC300XD, XC340XD in XC450XD. Za opredelitev njihove natezne R_m in strižne trdnosti τ smo v nadaljevanju ustvarili dva večvrednostna parametra ter s pomočjo produkcijskega pravila povezali njune vrednosti z ustreznimi vrednostmi parametra »*Material*« (slika 2).

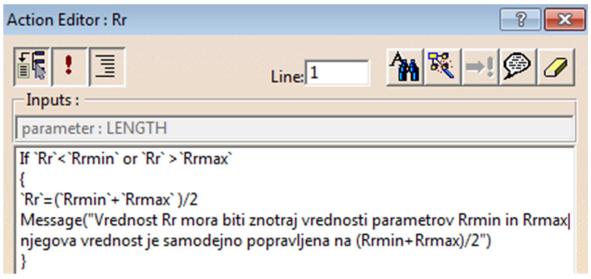
Vsi zgoraj navedeni neodvisni parametri predstavljajo osnovo za opredelitev različnih konstrukterskih problemov, ki zahtevajo opredelitev in preračun odvisnih parametrov podanih v nadaljevanju. V prispevku navedeni parametri so bili nameščeni znotraj parametričnih nizov, katerih poimenovanja uporabniku podajajo jasen odgovor, na katero konstruktersko lastnost se določen parameter nanaša.



Slika 2: *Pravilo za opredelitev materialnih lastnosti*

2.2 Odvisni geometrijski preoblikovalni parametri

Odvisni geometrijski parametri predstavljajo dimenzije preostale geometrije, ki ne morejo biti prosto določene. Opredelitev odvisnih geometrijskih parametrov smo pričeli z operacijo globokega vleka. Pomembnost pravilne izbire velikosti parametrov, ki predstavljata radij vlečnega paha R_p ter radij vlečnega obroča R_r nazorno podajata [6] in [7]. Da bi bili zmožni zapisati ta priporočila znotraj sistema, smo ustvarili štiri dodatne dolžinske parametre R_{pmax}, R_{pmin}, R_{rmax} in R_{rmin}, ki glede na priporočila podana v [8] ločeno predstavljajo zgornje in spodnje dovoljene vrednosti parametrov R_p in R_r, ter ustvarili reakciji (slika 3), ki uporabnika opozarjata na dovoljene vrednosti teh dveh parametrov. V primeru prekoračitve dovoljenih vrednosti sistem samodejno popravi vrednosti parametrov na srednjo vrednost med spodnjo in zgornjo dovoljeno vrednostjo posameznega parametra.



Slika 3: *Reakcija, ki omogoča preverjanje in samodejni popravek vrednosti parametra Rr*

Zapis vseh v nadaljevanju predstavljenih enačb za opredelitev vrednosti odvisnih parametrov je potekal s pomočjo urejevalnika enačb integriranem znotraj sistema CATIA V5.

Vlečna zračnost Z_v je razdalja med stenami vlečnega paha in vlečnega obroča ter predstavlja zelo pomemben parameter operacije globokega vleka, ki smo ga opredelili skladno z enačbo 1. Vpliv nepravilne določitve vrednosti tega parametra lahko povzroči neugodne pojave pri postopku globokega vleka [8].

$$Z_v = t + 0.07 \cdot \sqrt{10 \cdot t} \quad (1)$$

Da lahko konstrukter zagotovi zahtevane dimenzije in geometrijo vlečenega izdelka, mora izračunati premer krožne platine D_{kp}. Ob predpostavki, da se debelina in prostornina izdelka po postopku globokega vleka ne bosta spremenili ($t = konst. \text{ in } V_0 = V_D$), smo dolžinski parameter D_{kp} izrazili z uporabo enačbe 2 ter ga kasneje uporabili pri izgradnji CAD-gradnika krožne platine.

$$D_{kp} = \sqrt{\frac{4 \cdot V_D}{\pi \cdot t}} \quad (2)$$

Postopek rezanja zahteva opredelitev vrednosti parametra rezalne zračnosti Z_r. Optimalna vrednost tega parametra, katero smo opredelili s pomočjo enačbe 3, zagotavlja kakovost rezalnega roba in daljšo življenjsko dobo rezalnih elementov.

$$Z_r = 0.075 \cdot t \quad (3)$$

Ker je krožna platina izdelana s postopkom rezanja iz ravne kvadratne pločevine, mora konstrukter predvideti nekaj dodatnega materiala za rezanje. Tega smo, skladno z empirično enačbo 4, vključili znotraj dolžinskega parametra A_{rkp}, ki predstavlja stranico ravne kvadratne pločevine.

$$A_{rkp} = D_{kp} + 4 \cdot t \quad (4)$$

2.3 Odvisni procesni preoblikovalni parametri

Pri operaciji reza je najprej potrebno izvesti preračun parametra rezalne sile F_r, ki je znotraj sistema izveden po enačbi 5. Vrednost tega parametra predstavlja enega izmed faktorjev za izbiro stiskalnice z ustrezno kapaciteto.

$$F_r = t \cdot \pi \cdot D_{kp} \cdot \boldsymbol{\tau} \quad (5)$$

Med postopkom rezanja morajo plinske vzmeti delovati na držalo pločevine s silo, ki je enaka ali večja od snemalne sile F_s. Slednja je po enačbi 6 izražena kot odstotek F_r [9]. Izračun snemalne sile predstavlja osnovo za izbiro tipa in števila plinskih vzmeti v orodju.

$$F_s = 0.08 \cdot F_r \quad (6)$$

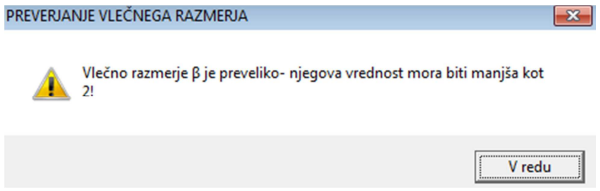
Stabilnost postopka globokega vleka lahko nadzorujemo s preprostimi merili, ki jih lahko določimo z uporabo neodvisnih geometrijskih parametrov. Ta merila predstavljajo vlečno razmerje β, redukcijo r in razmerje debelina-premer t/D_{kp}. Da bi lahko dosegli stabilen postopek globokega vleka in se izognili gubanju, morajo biti vrednosti teh parametrov znotraj določenih mej.

$$\beta = \frac{D_{kp}}{d} < 2 \quad (7)$$

$$r = \frac{D_{kp} - d}{D_{kp}} < 0.5 \quad (8)$$

$$\frac{t}{D_{kp}} > 0.01 \quad (9)$$

Z vzpostavitvijo treh preverjanj smo na podlagi omejitvenih enačb 7, 8 in 9 dosegli, da sistem v primeru neustrezne izbire vrednosti posameznega parametra uporabnika opozori na napako in ga usmeri k pravilni izbiri njegove vrednosti (slika 4).



Slika 4: *Pojavno okno, ki uporabnika opozarja o preveliki vrednosti parametra β*

Ko vlečni pah potuje navzdol, krožna platina teče pod držalom pločevine preko radijusa v odprtino orodja. Da bi preprečili tok pločevine in njeno gubanje, moramo z zadostno silo delovati na krožno platino in zato je potrebno izračunati silo pridržavanja F_p, ki jo sistem preračuna po enačbi 10. V njej nastopa A_h kot parameter, ki predstavlja velikost površine, na katero deluje držalo pločevine. Poleg zadostne snemalne sile morajo plinske vzmeti prav tako zagotoviti potrebno silo pridržavanja.

$$F_p = 2.75 \cdot A_p \quad (10)$$

Na koncu je potrebno izračunati še vrednost sile vlečenja F_d, katere rezultat v primerjavi s silo rezanja F_r poda odgovor o izbiri stiskalnice z ustrezno kapaciteto.

$$F_d = \pi \cdot t \cdot d \cdot R_m \cdot (\beta - 0.7) \quad (11)$$

2.4 Vzpostavitev relacij za preprečevanje geometrijske porušitve CAD-objekta končnega izdelka

Z zapisom treh reakcij smo preprečili porušitev CAD-objekta, ki predstavlja končni izdelek. Z uporabo prve smo omejili vrednost parametera D tako, da se ta vedno nahaja znotraj dovoljenih vrednosti. V primeru kršenja te omejitve sistem skladno z enačbo 12 samodejno popravi vrednost parametra.

$$D \geq d + 2 \cdot t + 2 \cdot R_\mathrm{r} + 1\mathrm{mm} \quad (12)$$

Na enak način je preko omejitev predstavljenimi z enačbama 13 in 14 z drugo reakcijo vodena vrednost parametra d.

$$d \geq 2 \cdot R_\mathrm{p} \quad (13)$$

$$d \leq D - 2 \cdot t - 2 \cdot R_\mathrm{r} - 1\mathrm{mm} \quad (14)$$

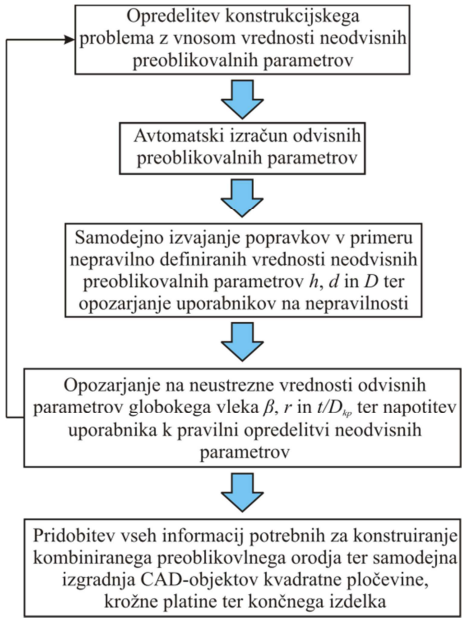
Tretja reakcija na podlagi omejitvene enačbe 15 preprečuje porušitev CAD-objekta zaradi nepravilne izbire vrednosti parametra h.

$$h \geq R_\mathrm{r} + R_\mathrm{p} + 2 \cdot t \quad (15)$$

3 DELOVANJE SISTEMA

Sistem omogoča enostavno opredelitev konstrukcijskega problema skozi določitev vrednosti neodvisnih preoblikovalnih parametrov vzpostavljenimi neposredno v drevesnem meniju programa CATIA V5. Znanje o opredelitvi in avtomatskemu izračunu vseh preostalih odvisnih preoblikovalnih parametrov, vezanih na konstruiranje predlaganega orodja je zbrano in vključeno v

sistem na podlagi študija knjig o konstruiranju preoblikovalnih orodij, priporočil strokovnjakov ter internih standardov podjetja. Pravilne povezave med različnimi parametri so programsko vzpostavljene, način izračuna določenega parametra znotraj sistema pa je opredeljen s pomočjo relacij glede na preostale parametre v dokumentu. Pri tem preverjanja in reakcije predstavljajo inteligentne algoritme za avtomatsko prepoznavo, opozarjanje in odpravljanje napak glede na omejitve vzpostavljene s parametri, ki morajo biti zadoščene, da bi se lahko izognili neželjenim napakam pri konstrukciji orodja.



Slika 5: *Delovanje sistema*

Logiko delovanja sistema lahko prikažemo na osnovi opredelitve vrednosti parametrov *d* in »*Material*«. Vnos vrednosti parametra *d* povzroči avtomatski preračun odvisnih parametrov D_{kp}, A_{rkp}, β, r, t/D_{kp}, F_r, F_s, F_p in F_v. Sistem nadzoruje vrednost parametra *d* z reakcijami, ki temeljita na enačbah 13 in 14, s čimer zagotavlja njegovo pravilno določitev v povezavi s parametri R_p, D, t in R_r. S pomočjo treh preverjanj istočasno zagotavlja pravilno opredelitev njegove vrednosti glede na preoblikovalne omejitve postavljene s parametri β, r ali t/D_{kp}. Z vnosom vrednosti parametra *d* je izvedenena tudi geometrijska izgradnja CAD-objektov kvadratne pločevine, krožne platine ter končnega izdelka. S pomočjo orodja »Knowledge Inspector« je razvidno, da sprememba vrednosti parametra »*Material*« povzroči avtomatski preračun parametrov F_b, F_s in F_d, kjer sta parametra R_m in τ, ki predstavljata materialne lastnosti vodena s pravilom prikazanem na sliki 2. Vpliv vrednosti posameznih parametrov na proces konstruiranja orodja je razložen v prejšnjih poglavjih.

Eksperimentalni rezulati kažejo, da naš sistem pri konstruiranju desetih različnih variant preoblikovalnega orodja za izdelavo votlega lončka s prirobnico skrajša čas konstruiranja za skoraj uro.

4 ZAKLJUČEK

Razvili smo sistem namenjen podpori konstruiranju kombiniranega preoblikovalnega orodja za izdelavo votlega lončka s prirobnico. Sistem deluje na podlagi konstrukterskega znanja, ki je z uporabo uporabniških parametrov in relacijami med njimi integrirano znotraj programskega paketa CATIA V5. Njegova uporaba omogoča enostavno opredelitev konstrukcijskega problema, avtomatski preračun parametrov, s katerim uporabnik pridobi vse potrebne informacije za konstruiranje orodja ter avtomatsko izgradnjo CAD-objektov kvadratne pločevine, krožne platine ter končnega izdelka. Ker ti postopki tradicionalno potekajo ročno in brez računalniške podpore, naš sistem bistveno prispeva k skrajšanju časa konstruiranja, dviguje kakovost konstrukcije ter omogoča izobraževanje neizkušenih konstrukterjev.

Viri

[1] D. Potočnik, B. Pesan, M. Ulbin. Parametrično-asociativni pristop h konstruiranju preoblikovalnega orodja v programskem paketu Catia V5. *Vir znanja in izkušenj za stroko/Industrijski forum IRT*. PROFIDTP. Škofljica. 2010.

[2] D. Potočnik, B. Pesan, M. Ulbin. Konstruiranje z vgrajenim znanjem v programskem paketu CATIA V5. *Sistemi na ključ-priložnosti dodane vrednosti*. GZS Združenje kovinske industrije, Odbor za orodjarstvo in strojegradnjo. Ljubljana. 2010.

[3] BT. Lin, SH. Hsu. Automated Design system for drawing die. *Expert Systems with Applications*, vol. 34, no. 3, pp.1586-1598. 2008.

[4] BT. Lin, CK. Chan, JC. Wang. A knowledge-based parametric design system for drawing dies. *The International Journal of Advanced Manufacturing Technology*, vol. 36, no. 7-8., pp. 671-680. 2008.

[5] ZX. Jia, HL. Li, XC. Zhang, JQ. Li, BJ. Chen. Computer-aided structural design of punches and dies for progressive die based in functional component. *The International Journal of Advanced Manufacturing Technology*, vol. 54, no. 9-12, pp. 837-852. 2011.

[6] A. Hendrick. Key design principles for successful deep drawing. *Stamping Journal*, March-April. 1999.

[7] V. Boljanovic. *Sheet metal forming processes and die design*. Industrial, New York. 2004.

[8] J. Suchy. *Handbook of Die design*. The McGraw-Hill Companies, USA. 1998.

[9] SME. *Die design handbook*. Third edition. SME. USA. 1990.

»Operacijo delno financira Evropska unija, in sicer iz Evropskega socialnega sklada.«

Zbornik 14. mednarodne multikonference IS 2011

INTELIGENTNI SISTEMI

Proceedings of the 14th International Multiconference IS 2011

INTELLIGENT SYSTEMS

Slovenski del konference

Slovenian part of the conference

Vpliv PESQ MOS na uspešnost samodejnega razpoznavanja govorcev

Robert Blatnik, Tomaž Šef
Odsek za inteligentne sisteme, Institut "Jožef Stefan"
Jamova cesta 39, 1000 Ljubljana, Slovenija
robert.blatnik@ijs.si, tomaz.sef@ijs.si

POVZETEK

V članku predstavljamo meritve vpliva kakovosti prenosa govora po metodi PESQ na uspešnost samodejnega razpoznavanja govorcev (SRG) in primerjavo rezultatov za brezžično lokalno omrežje (VoWLAN), stacionarno telefonijo PSTN in mobilno telefonijo GSM. Primerjavo rezultatov uspešnosti SRG v odvisnosti od izmerjene kakovosti govora smo predstavili v obliki grafov DET. Rezultati kažejo na neposredno povezavo med slabšanjem pogojev pri prenosu govornih posnetkov preko telefonskega omrežja in slabšanjem uspešnosti razpoznavanja.

1 UVOD

Eden od najbolj razširjenih načinov uporabe sistemov za samodejno razpoznavanje govorcev (SRG) je verifikacija neznanega govornega vzorca na osnovi primerjave z znanim govornim vzorcem, pri čemer je vsaj eden od govornih vzorcev prenesen preko telefonije, oba govorna vzorca pa sta na nek način zajeta in digitalizirana [1]. Znano je, da imajo različne tehnologije procesiranja in prenosa govora različen vpliv na uspešnost razpoznavanja sistemov za SRG [2]. Zmanjševanje uspešnosti, oziroma povečevanje stopnje napake pri razpoznavanju v sistemih za SRG zaradi vpliva kanala, kot so vplivi tehnologij procesiranja govora v literaturi pogosto poimenovani, je bilo v zadnjih letih večkrat predstavljeno v okviru tekmovanj uspešnosti sistemov za SRG, ki jih redno vsako leto organizira ameriški NIST [3]. Prav tako je vprašanje vpliva kanala eden večjih izzivov pri razvoju in raziskavah metod za SRG. Kljub temu, po vedenju avtorjev, do sedaj še ni bilo poglobljene raziskave, ki bi poiskala korelacije med stopnjo napake sistema za SRG in izmerjeno kakovostjo prenosa govora.

Vprašanje, ki si ga zastavljamo je naslednje: *Ali je mogoče na osnovi meritev kakovosti govora predvidevati uspešnost razpoznavanja sistema za SRG.* Za odgovor na to vprašanje, je potrebno najprej zagotoviti možnost objektivnih meritev kakovosti govora, ki se neizbežno poslabša z vsakokratnim prenosom preko različnih tehnologij prenosa. V nadaljevanju je potrebno popačene posnetke govora uporabiti pri razpoznavanju s sistemom za SRG in primerjati stopnjo napake razpoznavanja popačenih posnetkov s stopnjo razpoznavanja pri nepopačenih posnetkih oziroma posnetkih, ki so popačeni preko drugih tehnologij prenosa oziroma pri drugih pogojih prenosa, kot je na primer spreminjanje obremenitev omrežja in samih pogojev delovanja omrežja.

V raziskavi, ki jo obravnavamo v tem delu, smo kakovost govora izmerili po standardizirani metodi PESQ [4]. Zgradili smo okolje za popačenje in merjenje kakovosti govora v stacionarni telefoniji PSTN, mobilni telefoniji GSM in VoIP v brezžičnem lokalnem omrežju (VoWLAN) [5]. Rezultate izmerjene kakovosti govora po metodi PESQ, smo primerjali z rezultati uspešnosti razpoznavanja sistema za SRG. Na ta način smo lahko opazovali obnašanje sistema za SRG v odvisnosti od spreminjanja izbrane telefonije in v primeru VoWLAN, s spreminjanjem pogojev delovanja lokalnega brezžičnega omrežja WLAN pri različnih močeh RF signala in različnih obremenitvah, ki smo jih povzročali s simulacijo govornega prometa v obliki generiranja več sočasnih RTP tokov.

2 MERJENJE KAKOVOSTI GOVORA PO METODI PESQ IN USPEŠNOST SISTEMOV ZA SRG

Uspešnost sistemov za SRG merimo na osnovi preizkušanja, v kakšni meri sistem pravilno sprejema osebe z istovetnim govornim vzorcem in pravilno zavrača osebe z neistovetnim govornim vzorcem. V primeru, ko so govorni vzorci slabše kakovosti, bodisi zaradi popačenja v telefoniji, izgube informacije pri kodiranju, slabih pogojev snemanja in drugih vzrokov, ki izvirajo iz kanala, to nedvomno vpliva na uspešnost SRG. Zato ni naključje, da vpliv kanala predstavlja enega od ključnih izzivov pri razvoju sistemov za SRG.

Različne tehnologije prenosa različno vplivajo na kakovost govornega signala, ki nosi informacijo o identiteti osebe, ki jo razpoznavamo v sistemu za SRG. Iz tega razloga je zmanjševanje vpliva kanala na sisteme za SRG ena ključnih usmeritev raziskav in razvoja na tem področju. Pri tem je pomembno, da znamo te razlike tudi objektivno izmeriti.

Kakovost prenosa govora v telefoniji je mogoče oceniti subjektivno, torej s poslušanjem, ali objektivno, to je s strojno analizo govornega signala [6]. V zadnjih letih se zaradi številnih prednosti uporabljajo predvsem objektivne metode za ocenjevanje kakovosti govora. V tem delu smo izvajali meritve kakovosti govora po standardizirani metodi PESQ. [4].

3 IZVEDBA MERITEV PESQ IN USPEŠNOSTI SISTEMA ZA SRG Z REZULTATI

Izvedba meritev vpliva kakovosti govora na sistem za SRG je potekala v petih korakih: izbor in priprava testnih posnetkov, popačenje testnih posnetkov v telefonskem omrežju, meritve kakovosti govora, meritve uspešnosti sistema za SRG in analiza pridobljenih rezultatov. V prvem koraku smo iz zbirke govornih posnetkov izbrali posnetke 381 angleško govorečih žensk in jih pripravili za izvajanje meritev. V drugem koraku smo testne govorne posnetke popačili tako, da smo najprej vzpostavili in preverili komunikacijsko povezavo pri danih pogojih v izbranem telefonskem omrežju in pognali predvajanje govornega posnetka na izvorni strani telefonske povezave, pri čemer smo hkrati zajemali isti govorni posnetek na drugi strani komunikacijske povezave. Na ta način smo v stereo datoteko WAV zapisali oba časovno sinhronizirana posnetka; na prvi kanal je bil posnet originalni govorni signal, na drugi kanal pa je bil posnet popačeni govorni signal. V tretjem koraku smo izmerili kakovost govora s komercialno dostopnim programom za analizo po metodi PESQ [7]. V četrtem koraku smo izbrali popačene testne posnetke iz drugega koraka in jih uporabili za meritve uspešnosti razpoznavanja govorcev s testnim sistemom SRG. Na koncu, v petem koraku, smo analizirali uspešnost razpoznavanja in predstavili rezultate v obliki grafov DET, kar nam je omogočilo opazovanje obnašanja sistema pri različni kakovosti govora. V nadaljevanju predstavljamo izvedbo posameznih korakov meritev.

3.1 Izbor in priprava testnih posnetkov

Testne posnetke smo najprej izbrali iz zbirke NIST 2008 glede na spol govorca, tip kanala, tip govora in jezik. Izbrali smo posnetke angleško govorečih žensk, ki se pogovarjajo preko telefona. Testne posnetke je bilo potrebno pretvoriti iz formata sphere v format VAW z ustreznim kodiranjem in frekvenco vzorčenja (PCM, 16 bit, 8000 Hz), ki je bila primerna za razpoznavanje v sistemu SRG. Za ženske govorce smo se odločili zato, ker se sistemi SRG praviloma obnašajo nekoliko slabše na ženskih populacijah govorcev [2]. Za angleščino smo se odločili preprosto zato, ker je bilo v zbirki na voljo največ angleških posnetkov, kar nam je omogočilo izvajanje testov na večjem številu različnih govorcev. Na vsakem posnetku govori samo po ena oseba. Drugega sogovornika se na posnetku praviloma ne sliši, saj je bil vsak od obeh govorcev v telefonskem pogovoru posnet na ločen kanal. Obstaja nekaj izjem, kjer se zaradi medkanalskih prisluhov delno sliši sogovornika, vendar bistveno tišje od govora osebe, ki smo jo razpoznavali. Tišine, ki je sestavni del vseh posnetkov, nismo izločevali. Vsi posnetki so dolgi približno 5 minut, skupaj s tišino.

Za testiranje in učenje smo uporabili skupino posnetkov 381 žensk. Za izgradnjo modela ozadja pa smo izbrali telefonske pogovore 269 žensk, ki niso zastopane v skupini za testiranje in učenje. Skupina posnetkov za testiranje vsebuje po en posneti telefonski pogovor za vsako osebo iz testne množice. Skupina posnetkov za učenje pa vsebuje različno število posnetkov; osebe govorijo na dveh do nekaj deset posnetkih, ki smo jih uporabili za učenje.

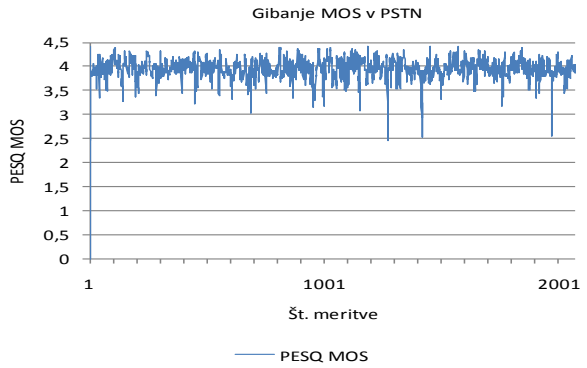
3.2 Popačenje testnih posnetkov v telefonskem omrežju

Popačili smo le posnetke, ki so bili namenjeni za testiranje oziroma za razpoznavanje v sistemu za SRG. Posnetkov, ki so bili izbrani za učenje modelov govorcev in izdelavo modela ozadja nismo popačili. Najprej smo izbrali telefonsko omrežje in vzpostavili komunikacijsko povezavo, nato smo izbrano skupino posnetkov predvajali na izvorni strani telefonske povezave s hkratnim zajemom popačenih testnih govornih posnetkov na drugi strani komunikacijske povezave v stereo datoteke WAV. Popačenja testnih govornih posnetkov smo izvajali v telefoniji PSTN, mobilni telefoniji GSM in VoWLAN. Za vzpostavitev govorne telefonske povezave v telefoniji PSTN in GSM ni bilo potrebno nastavljati konfiguracije omrežja, saj smo uporabljali omrežji komercialnega ponudnika. Za vzpostavitev povezave v VoWLAN pa smo morali ustrezno nastaviti testno brezžično lokalno omrežje in vzpostaviti VoIP povezavo. Medtem, ko pri izvajanju testiranja v PSTN, GSM in VoIP nismo imeli možnosti spreminjanja pogojev delovanja omrežja, smo pri VoWLAN lahko spreminjali pogoje delovanja omrežja WLAN na dva načina. Prvi način je bil spreminjanje obsega govornega prometa v obliki generiranja več sočasnih RTP tokov, drugi način pa je bil poslabševanje pogojev delovanja s slabljenjem RF signala. RF signal smo slabili s spreminjanjem nastavitev moči signala na brezžični točki dostopa in s spreminjanjem fizične oddaljenosti WLAN klienta od točke dostopa. Na WLAN klientu smo tako izmerili tri različne stopnje moči RF signala. Pri postavitvi WLAN klienta v neposredno bližino točke dostopa smo z aplikacijo NetStumbler izmerili moč signala okoli -30 dBm, pri postavitvi WLAN klienta v isti sobi na oddaljenosti 6 metrov od točke dostopa brez ovir med točko dostopa in klientom smo izmerili moč signala okoli -40 dBm, pri postavitvi klienta v sosednjo sobo na oddaljenosti 5 metrov, pri čemer je bila vmes stena, smo izmerili -50 dBm, pri oddaljenosti približno 40 metrov, pri čemer je bilo vmes nekaj sten, pa je moč signala padla na -70 do -80 dBm. V zadnjem primeru je klient deloval na meji sprejema in pogosto padel iz omrežja. Prav tako so se občutno povečale napake v prometu in zakasnitve ter trepetanje. Spreminjanje obremenitev omrežja z RTP prometom smo izvajali pri 5, 10, 15 in 20 sočasnih RTP tokovih. Na osnovi pridobljenih popačenih posnetkov smo v nadaljevanju lahko izvedli PESQ analizo kakovosti prenosa govora v izbrani telefoniji in v primeru VoWLAN tudi pri izbranih obremenitvah omrežja in različnih močeh RF signala.

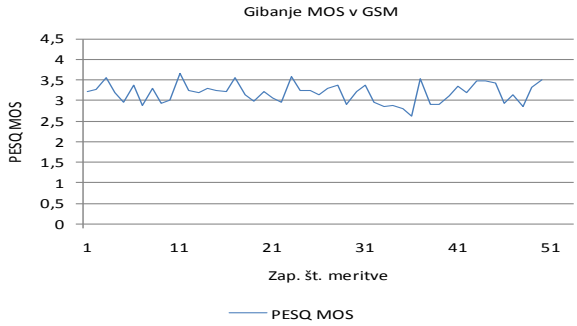
3.3 Meritve kakovosti govora

Meritve kakovosti prenosa govora smo izvajali s PESQ analizo stereo WAV datotek, ki smo jih pridobili s popačenjem govornih posnetkov v izbrani telefoniji. Pri izvedbi meritev smo bili omejeni z dolžino posnetkov, saj običajni postopki merjenja kakovosti govora s PESQ predvidevajo analizo testnih posnetkov v dolžini 10 do 20

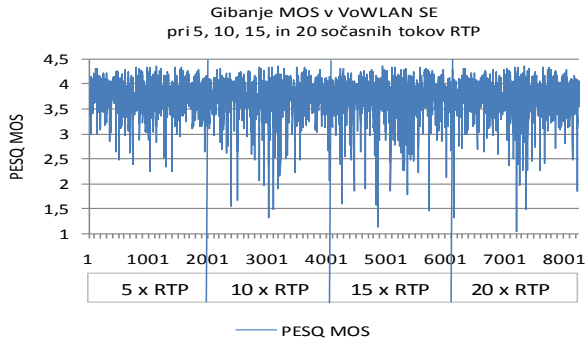
sekund, naši testni posnetki pa so bili dolgi 5 minut. Izvajanje meritev z datotekami, ki so daljše od 1 minute lahko privede do prekomernega povprečenja rezultata PESQ MOS.



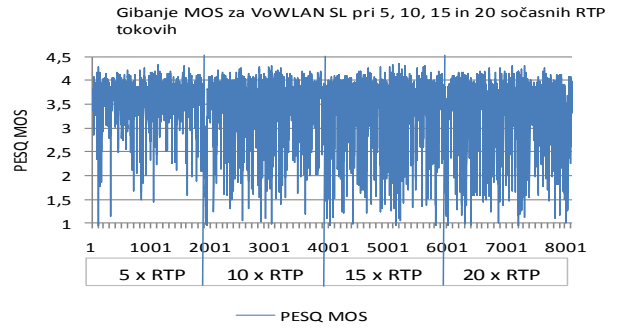
Slika 1: *Gibanje PESQ MOS v telefoniji PSTN*



Slika 2: *Gibanje PESQ MOS v telefoniji GSM*



Slika 3: *Gibanje PESQ MOS in zakasnitev v VoWLAN z obremenitvami omrežja z od 5 do 20 sočasnimi RTP tokovi*



Slika 4: *Gibanje PESQ MOS in zakasnitev v VoWLAN pri nizkem RF signalu z 5 do 20 sočasnimi RTP tokovi*

Okolje Opera ima omejitev procesiranja daljših datotek od 1 minute. Zaradi teh omejitev smo vse stereo WAV datoteke razrezali na 1 minutne sekvence in tako dobili iz vsake testne datoteke pet enominutnih datotek. Tako pripravljene posnetke smo lahko analizirali z orodjem Opera v celotni dolžini trajanja, pri čemer smo lahko opazovali nihanje MOS tudi znotraj posameznih testnih posnetkov.

Na grafih so predstavljeni rezultati meritve PESQ za telefonijo PSTN (slika 1), GSM (slika 2), VoWLAN SE (slika 3) in VoWLAN SL (slika 4).

3.4 Meritve uspešnosti sistema za SRG

Meritev uspešnosti sistema za SRG smo izvajali v več korakih: izbor posnetkov, generiranje modela ozadja, učenje sistema, testiranje in analiza rezultatov. Najprej smo izbrali dve skupini posnetkov. Prva skupina posnetkov je bila namenjena za generiranjem modela ozadja, druga skupina pa je bila razdeljena na podskupino za učenje in podskupino za testiranje sistema. Podskupino za testiranje smo popačili v različnih pogojih prenosa in na ta način dobili skupine posnetkov za PSTN, GSM, VoWLAN SE in VoWLAN SL. Podskupina za učenje sistema je ostala nepopačena. Pri posnetkih v testni skupini smo imeli vedno na voljo podatek, ali gre za avtentičen ali neavtentičen posnetek. Na osnovi tega podatka smo lahko uredili rezultate verifikacije na dve skupini, na skupino rezultatov za avtentične posnetke in na skupino rezultatov za neavtentične posnetke. Tako smo lahko analizirali razmerje FAR in FRR v grafih DET (sliki 5 in 6).

5 ANALIZA VPLIVA PESQ MOS NA USPEŠNOST RAZPOZNAVANJA

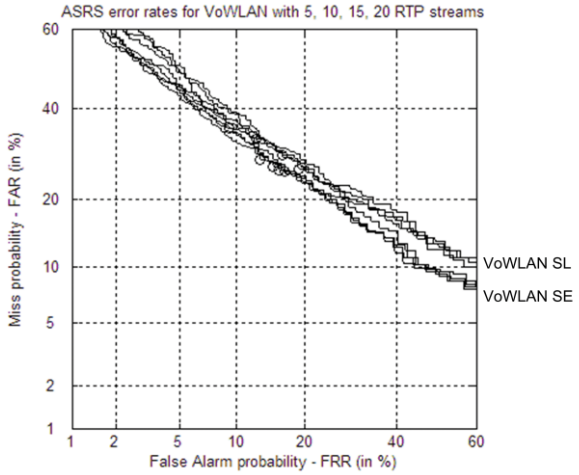
5.1 Rezultati meritev kakovosti govora

V rezultatih kakovosti govora za **PSTN in GSM** opazimo nihanje rezultata PESQ MOS, ki je dokaj konstantno skozi vse zaporedne meritve. Pri GSM je v primerjavi s PSTN opazna manjša gostota krivulje, kar je posledica manjšega števila meritev pri GSM. Povprečen rezultat PESQ MOS za PSTN je okoli vrednosti 3.9, za GSM pa okoli vrednosti 3.2, kar je pričakovano in primerljivo s sorodnimi meritvami.

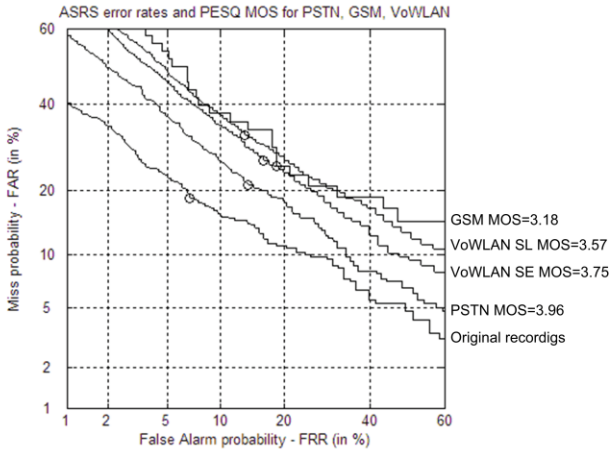
Pri rezultatih kakovosti govora v VoWLAN SE je lepo vidno nihanje izmerjenega MOS, ki je bistveno bolj izrazito, kot v primeru PSTN in GSM. Prav tako je vidno povečevanje nihanja MOS, ki se povečuje s povečevanjem števila tokov RTP. Prav tako lahko opazimo poslabšanje povprečnega rezultata PESQ MOS in povečevanje zakasnitev v primerjavi s PSTN, kar je pričakovano.

Pri VoWLAN SL je povprečni rezultat PESQ MOS za nizek RF signal pričakovano slabši od meritev z odličnim RF signalom. V primerjavi z VoWLAN SE je lepo vidno občutno povečano nihanje rezultata PESQ MOS.

Za meritve kakovosti govora v VoWLAN lahko sklenemo, da ima slabljenje signala RF največji vpliv na kakovost govora v brezžičnem omrežju WLAN, nekoliko slabši vpliv pa je možno zaznati pri povečevanju obremenitev omrežja z govornim prometom do 20 sočasnih RTP tokov.



Slika 5: *Rezultati uspešnosti sistema za SRG v VoWLAN pri SL in SE*



Slika 6: *Prikaz vpliva uspešnosti sistema za SRG od PESQ MOS za telefonije PSTN, GSM in VoWLAN*

5.2 Vpliv PESQ MOS na uspešnost sistema za SRG

Sistem za SRG se pričakovano najbolje obnaša pri nepopačenih posnetkih, kjer se EER giblje okoli 15%. Pričakovano poslabšanje EER zaradi poslabšanja kakovosti govora v testnih telefonijah se postopoma stopnjuje od PSTN s povprečnim MOS 3.96 in z EER pod 20%, sledi VoWLAN SE s povprečnim MOS 3.75 in z EER nekoliko na 20% in še za nekaj odstotkov slabše v primeru VoWLAN SL s še nekoliko nižjim povprečnim MOS 3.57. Najslabši rezultat EER doseže GSM s povprečnim MOS 3.18 in EER, ki se giblje nekoliko slabše od VoWLAN SL.

Podobno, kot za kakovost govora, lahko ugotovimo tudi za uspešnost sistema za SRG, da je občutno večje spremembe v uspešnosti razpoznavanja moč zaznati pri spreminjanju moči RF signala, kot pa pri povečevanju prometa RTP do 20 sočasnih RTP povezav.

Lahko torej potrdimo, da je uspešnost sistema za SRG manjša, kadar se izvaja razpoznavanje govorcev nad posnetki, ki imajo zaradi pogojev prenosa v različnih telefonijah slabšo kakovost govora. Zanimiva pa je ugotovitev, da ima slabljenje signala RF v omrežju WLAN večji vpliv na kakovost govora v brezžičnem omrežju, kot povečevanje obremenitev s prometom RTP do 20 sočasnih RTP povezav. To je lepo vidno na sliki 5, kjer se celotna skupina SL jasno razlikuje od skupine SE. Kakovost se je znotraj ene skupine meritev za SE oziroma SL v tem primeru zmanjšala minimalno, prav tako pa tudi spreminjanje EER.

6 ZAKLJUČEK

Rezultati meritev kakovosti govora v izbranih telefonijah so skladni s primerljivimi meritvami. Rezultati meritev uspešnosti sistema za SRG pričakovano kažejo na konsistentno poslabševanje stopnje napak s poslabševanjem pogojev pri prenosu govornih posnetkov preko telefonskega omrežja. Ugotovitve nam potrjujejo tezo, da z meritvami kakovosti govora s PESQ MOS lahko napovemo uspešnost sistema za SRG pri določenih pogojih delovanja poljubnega telefonskega omrežja. To ugotovite potrjujemo z meritvami, ki kažejo konsistentno povečevanje napake z zmanjševanjem kakovosti, izražene s PESQ MOS pri različnih pogojih prenosa govora.

Literatura

[1] Rosenberg, A. E., Bimbot, F., Parthasarathy, S. Overview of speaker recognition. V: Benesty, J.; Sondhi, M. M.; Huang, Y. (Eds.) Springer Handbook of Speech Processing. 725 – 741 (Springer-Verlag, Berlin Heidelberg, 2008).

[2] Vesničer, B. Postopki normalizacije v sistemih za samodejno razpoznavanje govorcev: doktorska disertacija. Ljubljana, 2010.

[3] Doddington, G. R., Przybocki, M. A., Martin, A. F., Reynolds, D. A., The NIST speaker recognition evaluation - overview methodology, systems, results, perspective. *Speech Commun.* 31, 2-3 (June 2000), 225-254, 2000.

[4] Rix, A. W., Beerends, J. G., Hollier, M.P., Hekstra, A. P. Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs. IEEE ICASSP 2001. Salt Lake City, UT, USA. 2001.

[5] Blatnik, R., Kandus, G., Šef, T. Influence of the perceptual speech quality on the performance of the text-independent speaker recognition system. Int. j. circuits syst. signal process., 2011, vol. 5, no. 4, str. 346-353.

[6] Grancharov, V., Kleijn, W. B. Speech quality assessment. V: Benesty J., Sondhi, M. M., Huang, Y. (ur.). Springer Handbook of Speech Processing. 83 – 96. Springer-Verlag, Berlin Heidelberg, 2008.

[7] Technical Specification for the Opera Software Suite V3.0. OPTICOM GmbH, Erlangen, Germany. Dostopno na spletu: http://www.opticom.de/. 2001.

OCENA PORABE ENERGIJE FIZIČNE AKTIVNOSTI S POSPEŠKOMEROM

Božidara Cvetković, Mitja Luštrek
Department of Intelligent Systems
Jozef Stefan Institute
Jamova 39, 1000 Ljubljana, Slovenia
e-mail: {boza.cvetkovic, mitja.lustrek}@ijs.si

POVZETEK

V prispevku je predstavljena raziskava in preliminarni rezultati iskanja razumnega kompromisa med računsko kompleksnostjo in številom senzorjev ter točnostjo algoritma za oceno porabe energije pri fizični aktivnosti. Rezultati so pokazali, da lahko le z enim pospeškomerom in s kombinacijo enostavnih linearnih regresijskih modelov zmanjšamo povprečno absolutno napako na 1.06 MET in z dvema pospeškomeroma na 0.96 MET glede na splošni regresijski model.

1 UVOD

Vredotenje porabe energije je pomembno tako v športni, medicinski kot tudi v domeni ambientalne inteligence. Redna in predvsem ustrezna telesna aktivnost je ključnega pomena za zdravje in dobro počutje ljudi vseh starosti [1]. Poznavanje vrednosti porabljene energije je v pomoč tako definiranju ustrezne vadbe, kot tudi pri detekciji nepravilnosti v delovanju telesa.

Za oceno porabe energije obstaja več metod a) direktna kalorimetrija, b) indirektna kalorimetrija, c) dvojno označena voda (*ang.* DLW Doubly Labeled Water), d) pisanje dnevnika, e) uporaba pedometrov, f) uporaba srčnega utripa in g) uporaba pospeškomerov.

Metode a), b) in c) so znane po visoki točnosti, vendar za svoje delovanje potrebujeta drago opremo. Uporaba teh metod je omejena ali na določene aktivnosti (direktna kalorimetrija), na poseben ter nadzorovan prostor (indirektna kalorimetrija) ali na zmožnost merjenja povprečne porabe energije od enega do treh tednov aktivnosti (dvojno označena voda). Te metode so zelo okorne in neprimerne za vsakodnevno uporabo.

Ostale metode ne potrebujejo drage opreme, imajo pa nižjo točnost zaradi sebi specifičnih problemov.

Metoda pisanja dnevnika je časovno potratna in nenatančna, vendar z njo pridobimo kontekstualne, kvalitativne in kvantitativne informacije.

Metode z uporabo senzorjev so za uporabnika enostavnejše. Pedometri so pripeti na gležnje in s tem je ocena porabe energije omejena le na aktivnost spodnjih udov. Težava metode srčnega utripa je vpliv različnih faktorjev na frekvenco utripa (psihično stanje, temperatura, nikotin, fizična pripravljenost, starost, spol…). Ocena porabe energije iz signalov pospeškomera je omejena le na tiste dela telesa, na katerega so pospeškomeri pripeti. Točnost ocene je nižja tudi zato, ker ne morejo meriti porabe energije delovanja mišic.

Podatek o vrednosti porabljene energije se najpogosteje izrazi kot metabolični ekvivalent aktivnosti (MET). MET je enota ki označuje količino porabljenega kisika med telesno dejavnostjo. En MET je enak energiji porabljeni pri sedenju.

Cilj naše raziskave je razviti natančen vendar računsko nezahteven algoritem za ocenjevanje porabe energije. Uporabili smo pospeškomere, optimizirali njihovo število, razvili enostavne atribute in poiskali linearno odvisnost s porabo energije.

Nadaljevanje prispevka je organizirano po sledečem zaporedju: v poglavju 2 smo predstavili sorodna dela, pridobivanje podatkov in uporabljene atribute smo predstavili v poglavju 3, eksperimente ter rezultate v poglavju 4 in zaključili s poglavjem 5.

2 PREGLED SORODNIH DEL

S problemom ocenjevanja porabe energije se raziskovalci ukvarjajo že vrsto let in temu primerno obstaja širok spekter pristopov. Te lahko razdelimo na dve večji skupini in sicer na pristope a) ki uporabljajo model človeka in b) ki temeljijo na regresijskih algoritmih.

Metode, ki uporabljajo model človeka temeljijo na hitrosti in poziciji. Ti se določijo z integracijo pospeška, kar pa se izkaže za zelo zahtevno in nenatančno nalogo saj se z integriranjem signala napake hitro seštevajo. Ko se ocenita hitrost in pozicija se uporabijo kinematični modeli za oceno porabe energije [2]. Ni znano, da bi kakšna rešitev z uporabo kinematičnega modela delovala bolje od najboljše rešitve z regresijskimi algoritmi [3].

Regresijski pristopi poiščejo linearno ali nelinearno povezavo med porabo energije in signali, ki jih vrača pospeškomer. Enostavne rešitve uporabljajo en pospeškomer pritrjen na bok ali zapestje uporabnika in linearen regresijski model [4]. V preteklosti se je izkazalo, da je za višjo točnost ocene potrebno uporabiti več regresijskih modelov [5] in kompleksnejše atribute poračunane iz signala pospeškomera Bouten et al. [6] in Crouter et al. [3].

Algoritem Crouter et al. [3], je sestavljen iz dveh korakov: a) klasifikacija aktivnosti v eno od treh kategorij: sedeča

aktivnost, potujoča aktivnost, aktivnost življenjskega stila in b) glede na rezultat klasifikacije se uporabi regresijski model za oceno porabe energije. Pri klasificirani sedeči aktivnosti je ocena porabljene energije vedno enaka 1 MET, potujoče aktivnosti, hoja in tek se ocenita z modelom linearne regresije, ostale aktivnosti življenjskega sloga pa se ocenijo z eksponentnim regresijskim modelom. Težava te metode je, da popolnoma izključuje nekatere vsakdanje aktivnosti (kolesarjenje) in ima slabo ocena porabe energije za aktivnosti, kjer se uporablja zgornji del telesa. Ta težava izhaja iz lokacije pospeškomera, ker je pripet na bok.
Uporaba srčnega utripa lahko pomaga bolj točno oceniti porabo energije pri zahtevnih aktivnostih, vendar je sam odvisen od veliko dejavnikov in s tem lahko pride do večje napake, kot pa če bi uporabili le pospeškomere. Berge et al. [7] je razvil algoritem v obliki drevesa, ki uporablja izolirane podatke o srčnem utripu in pospeškomerih. Prva vejitev se izvede na podlagi intenzitete premikanja, naslednja na podlagi srčnega utripa. Listi drevesa vsebujejo regresijske enačbe, ki ocenijo porabo energije.
Naša rešitev uporablja tako kinematične modele kot kompleksne atribute za iskanje linearne odvisnosti s porabljeno energijo. Analizirali smo tudi kako srčni utrip vpliva na oceno.

3 MERITVE IN METODA

Izziv ki smo si ga postavili zajema tako optimizacijo števila uporabljenih senzorjev, kot željo po visoki točnosti ocene večjega števila aktivnosti. Rešitev temelji na linearnem regresijskem modelu. Uspešnost rešitve je odvisna od kvalitete in količine podatkov v procesu učenja. Iz zajetih podatkov smo izluščili čim bolj informativne atribute in zgradili regresijski model.
V nadaljevanju bomo predstavili zajete podatke in izbrane atribute.

3.1 PODATKI

Za zajem podatkov smo sestavili seznam 23 scenarijev. Scenariji se med seboj razlikujejo tako po aktivnostih , kot tudi po lokaciji. Delimo jih lahko na tri večje skupine: a) vsakdanje življenje: pisarniško delo, čiščenje, pomivanje tal, spanje, igranje, pometanje, hoja, hoja po stopnicah, b) različno intenzivne vadbe: kolesarjenje, tek, trebušnjaki, skoki in c) zunanje aktivnosti: kolesarjenje, vožnja z avtomobilom, vožnja z dvigalom.
Vsak scenarij je sestavljen iz množice krajših aktivnosti:

VREDNOST MET	INTENIVNOST AKTIVNOSTI
< 3	Normalna
3 < > 6	Srednja
> 6	Zahtevna

Tabela 1. Delitev intenzivnosti aktivnosti glede na vrednost MET.

stanje, hoja, tek, sedenje, ležanje, na vseh štirih, klečanje, kolesarjenje in tranzicija. Vse aktivnosti so ovrednotene z številom MET. Vrednosti so privzete iz svetovno uveljavljenega seznama aktivnosti [8]. Tabela 1 prikazuje delitev aktivnosti, glede na intenzivnost. Zajeti podatki vsebujejo aktivnosti od enega do 11 MET-ov.

3.2 ATRIBUTI

Uspešnost regresijskega modela je odvisna tudi od informativnosti izbranih atributov.
Uporabljeni atributi so izpeljani iz pospeška v treh smereh, za vsak uporabljen pospeškomer. Pomemben atribut, ki ga bomo uporabili je najdaljša aktivnost v trenutnem časovnem oknu. Privzeto je, da je prepoznana aktivnost pravilna.
Regresijski modeli v katerih bomo uporabili srčni utrip vsebujejo poleg atributov pospeškomerov tudi povprečni srčni utrip in temperaturo kože v izbranem časovnem oknu. Računanje atributov se izvaja v premikajočem oknu desetih sekund.
Ostali atributi so: povprečna dolžina vektorja, ploščina pod grafom absolutnega pospeška v treh smereh, vsota ploščin vseh smeri, kvadrat ploščine pod grafom absolutnega pospeška v vseh treh smereh, vsota kvadratov ploščin vseh smeri, ploščina pod grafom absolutnega pospeška z odšteto gravitacijo v vseh smereh, število sprememb smeri pospeška v vsaki smeri, vsota vrednosti sprememb smeri pospeška v vsaki smeri, sprememba hitrosti pospeška brez gravitacije v vseh smereh, integral spremembe kinetične energije v vsaki smeri posebej, vsota sprememb kinetične energije vseh smeri.
Izbrani atributi imajo konstantno ali linearno časovno zahtevnost $O(n)$. Kjer se atribut računa posebej v vseh treh smereh je časovna zahtevnost enaka $3*O(n)$.

4 EKSPERIMENT IN REZULTATI

Zajeli smo posnetke petih ljudi (4 moški, 1 ženska) ob izvajanju prej omenjenih 23 scenarijev. Opremljeni so bili s štirimi Shimmer pospeškomeri [9], na prsih, zapestju, stegnu in gležnju ter z oprsnico Zephyr [10], ki meri srčni utrip in temperaturo kože.
Želeli smo ugotoviti katera kombinacija senzorjev vpliva na boljšo točnost ocene energije, zato smo zgradili 15 senzorskih kombinacij, kjer smo spreminjali število in lokacijo senzorjev. Vsem nastalim kombinacijam smo v izolirani množici dodali še srčni utrip in temperaturo kože, kar je skupaj 30 kombinacij. Regresijski model smo iskali s petimi algoritmi. Rezultati točnosti modelov kažejo, da uporaba več senzorjev ne doprinese k točnosti ocene, zato smo se omejili na uporabo enega senzorja ali kombinacijo dveh. Tabela 2 vsebuje rezultate regresijskih algoritmov za podatke enega senzorja. Ovrednoteni so z relativno absolutno napako, ki smo jo dobili s prečnim preverjanjem ene osebe na ostalih štirih.
Pospeškomer na prsih in algoritem REPTree vrneta najboljši rezultat. Težava tega algoritma je, omjenost napovedi na vrednosti, ki so bile predstavljene algoritmu z učnimi podatki. V realnem življenju bodo uporabniki izvajali aktivnosti, ki jih v učni množici ni, zato smo se odločili da za učni algoritem enega senzorja izberemo

SVR, kadar imamo le podatke pospeškomera in algoritem M5P, kadar imamo na voljo še srčni utrip. Omenjena algoritma zgradita model, ki lahko vrne vrednosti, ki jih algoritem še ni videl.
Privzeli smo da je senzor na prsih vedno prisoten in analizirali kako vpliva dodaten senzor na oceno. Rezultati so predstavljeni v tabeli 3. Ovrednoteni so z relativno absolutno napako, ki smo jo dobili s prečnim preverjanjem ene osebe na ostalih štirih.

	Položaj	LR	SVR	MLP	REPTree	M5P
Srčni utrip	Prsi	44.6	24.7	38.2	**18.7**	24.2
	Stegno	46.1	26.4	33.8	19.0	41.6
	Gleženj	49.7	29.7	46.8	**18.2**	30.6
	Zapestje	46.4	26.4	35.4	21.9	38.3
Brez srčnega utripa	Prsi	44.1	26.6	29.5	**18.1**	28.3
	Stegno	46.0	41.8	29.8	19.7	26.5
	Gleženj	43.7	39.4	29.2	22.8	26.2
	Zapestje	45.0	41.0	25.7	24.6	30.1

Tabela 2. Rezultati regresijskih algoritmov (relativna absolutna napaka) z enim senzorjem in s podatkom o srčnem utripu.

Algoritem REPTree se je znova izkazal za najboljšega, vendar iz omenjenih razlogov je ta neprimeren. Odločili smo se, da za učenje privzamemo algoritem SVR saj je njegova prednost linearni regresijski model, ki ni nagnjen k »prenaučenosti« na učne podatke (*ang.* overfitting).

	Položaj	LR	SVR	MLP	REPTree	M5P
Srčni utrip	Prsi + stegno	43.9	23.3	26.7	19.1	25.1
	prsi + gleženj	48.5	21.0	24.4	16.4	19.1
	prsi + zapestje	44.3	23.3	24.2	24.5	35.2
Brez srčnega utripa	Prsi + stegno	42.6	36.5	27.9	20.2	35.0
	prsi + gleženj	41.2	21.6	23.1	19.2	21.2
	prsi + zapestje	40.6	21.1	41.6	21.4	26.7

Tabela 3. Rezultati regresijskih algoritmov kombinacije dveh senzorjev z in brez srčnega utripa, kjer je privzeto da je senzor na prsih vedno prisoten.

Izbrano imamo lokacijo in število senzorjev, ki jih bomo uporabili pri meritvah točnosti regresijskega modela ocene. Pri uporabi enega senzorja bomo uporabili senzor na prsih, pri dveh pa senzor na prsih in gležnju.
Točnost regresijskega modela smo merili s povprečno absolutno napako (*ang.* Mean Absolute Error) pri prečnem preverjanju ene osebe na ostalih štirih.
Tabela 4 vsebuje rezultate splošnih modelov. *Splošni regresijski model za pospeškomer* vsebuje atribute pospeškomera. *Splošni regresijski model pospeškomer + srčni utrip* pa poleg atributov pospeškomera vsebuje še srčni utrip in temperaturo kože.

	Prsi (MET)	Prsi + Gleženj (MET)
Splošni regresijski model za pospeškomer	1.7	1.7
Splošni regresijski model za pospeškomer + srčni utrip	1.6	2.2

Tabela 4. Povprečna absolutna napaka prečnega preverjanja splošnih modelov.

Dva senzorja in srčni utrip vrneta višjo napako. Pri analizi vzroka za napako smo ugotovili da srčni utrip zniža točnost ocene pri normalnih in srednje intenzivnih aktivnostih in izboljša točnost pri zahtevnejših aktivnostih. Testirali smo kakšna je napaka, če uporabimo regresijski model s srčnim utripom le na zahtevnejših aktivnostih. Rezultati so pokazali zmanjšanje napake.

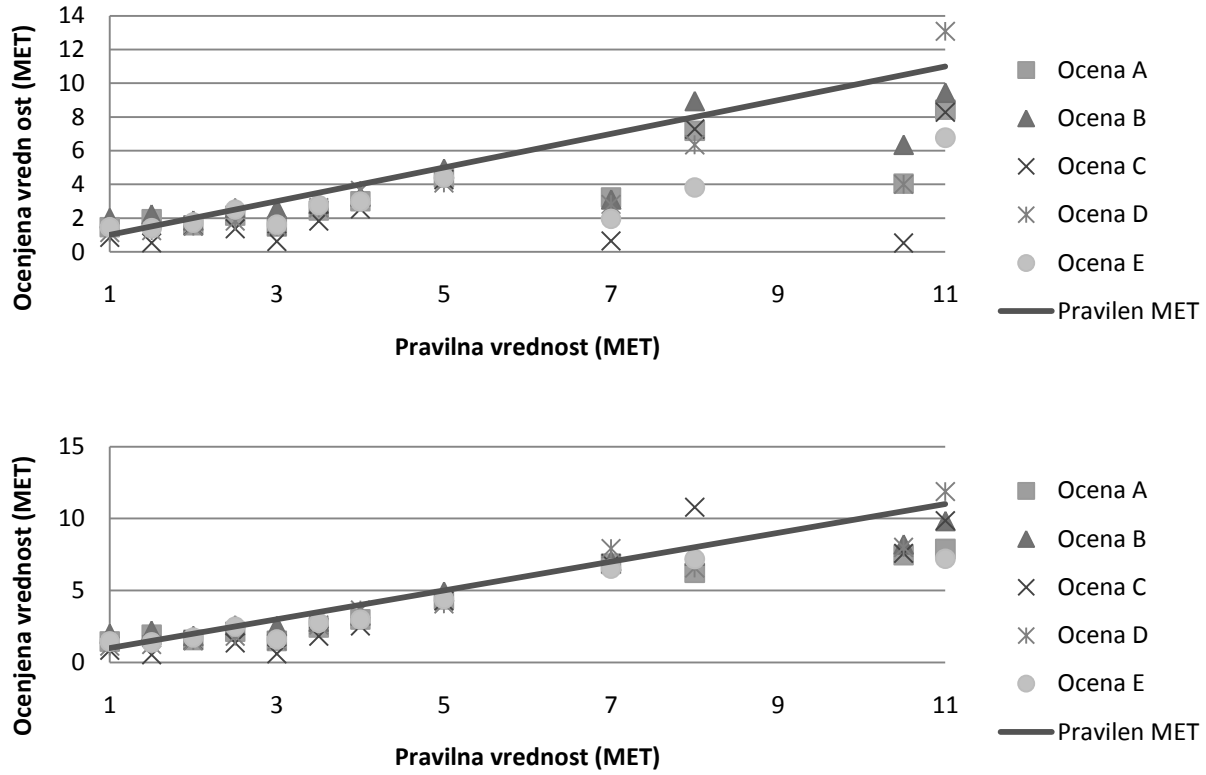
	Prsi (MET)	Prsi + Gleženj (MET)
Splošni regresijski model s pospeškomerom + model po razredu (s srčnim utripom)	1.07	0.989
Splošni regresijski model s pospeškomerom + model po razredu (brez srčnega utripa)	1.06	0.96

Tabela 5. Povprečna absolutna napaka prečnega preverjanja. Kombinacija splošnega regresijskega modela in modela po razredih.

Iz rezultatov je razvidno, da je napaka višja pri aktivnosti, kot sta tek in kolesarjenje. V naslednjem koraku smo uporabili poseben razredni regresijski model za te aktivnosti. Regresijski model smo naučili na izoliranih aktivnostih s srčnim utripom in brez, ter zmerili novo povprečno absolutno napako. Rezultati v tabeli 5 kažejo, da se napaka zniža na 0.96 MET v primeru dveh senzorjev in 1.06 pri enem ter da bistvene razlike med uporabo modela s srčnim utripom ali brez ni.
Slika 1 prikazuje zvišanje točnosti pri uporabi razrednih regresijskih modelov.

6 ZAKLJUČEK

V prispevku smo predstavili analizo lokacije in števila pospeškomerov za oceno porabe energije pri fizičnih aktivnostih. Ugotovili smo da če privzamemo da je senzor na prsih vedno na voljo, je kombinacija s senzorjem na gležnju najbolj točna.
Rezultati kažejo, da uporaba srčnega utripa bistveno ne izboljša točnosti ocene, to smo dosegli s kombinacijo splošnega in razrednih regresijskih modelov za kolesarjenje in tek. S to kombinacijo smo zmanjšali povprečno absolutno napako z 1.7 MET-a na 0.96 MET-a.



Slika 1. Graf pravih vrednosti MET v primerjavi z vrednostjo MET ocene. Zgornji graf je rezultat splošnega regresijskega modela brez srčenga utripa. Spodnji graf je rezultat splošnega regresijskega modela brez srčnega utripa, kjer so aktivnosti kolesarjenje in tek ocenjeni s posebnimi razrednimi modeli.

Zaradi linearnosti modelov in linearne časovne zahtevnosti atributov lahko zaključimo, da lahko ta metoda deluje v realnem času na napravah s slabšo procesno enoto, kot je telefon. Nadaljnje delo bo vsebovalo testiranje predstavljenega regresijskega modela na daljših posnetkih, ki bodo zajemali več dni vsakdanjih aktivnosti povprečne osebe. Metodo bomo primerjali z metodo Coruter et al. [3].

Literatura

[1] Warburton DE, Nicol CW, Bredin SS. Health benefits of physical activity: the evidence. *Canadian Medical Association Journal.*174:801–9, 2006

[2] Yongwon J., M.W. Jung, J. Kang, H. C. Kim. An Wearable Energy Expenditure Analysis System based on the 15-channel Whole-body Segment Acceleration Measurement, *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the* , pp.3834-3836, 2005

[3] Crouter, S. E., Clowers, K. G. & Bassett, D. R. A novel method for using accelerometer data to predict energy expenditure. *Journal of applied physiology.* vol. 100, p.p.1324-1331, 2006.

[4] Nichols J.F., Morgan C.G., Chabot L.E., Sallis J.F., Calfas K.J. Assessment of physical activity with the Computer Science and Applications, Inc., accelerometer: laboratory versus field validation. *Research Quarterly of Excercise and Sport.* vol. 71, 1, p.p. 36-43, 2000.

[5] Heil, D. P. Predicting activity energy expenditure using the actical activity monitor. *Research quarterly for exercise and sport.* vol. 77, p.p. 64-80, 2006.

[6] Bouten, C. V., Westerterp, K. R., Verduin, M. & Janssen, J. D. Assessment of energy expenditure for physical activity using a triaxial accelerometer. *Medicine and science in sports and exercise.* vol 26, p.p. 1516-1523, 1994.

[7] Brage S, Brage N, Franks P.W., Ekelund U, Wong M.Y., Andersen L.B., Froberg K, Wareham N.J. Branched equation modeling of simultaneous accelerometry and heart rate monitoring improves estimate of directly measured physical activity energy expenditure. Journal of Applied Physiology, vol. 96, p.p.343-351, 2004.

[8] Ainsworth B.E., Haskell W.L., Herrmann S.D., Meckes N., Bassett D.R. Jr., Tudor-Locke C., Greer J.L., Vezina J., Whitt-Glover M.C., Leon A.S. 2011 Compendium of Physical Activities: a second update of codes and MET values. *Medicine and science in sports and exercise.* vol.43, p.p. 1575-81, 2011

[9] Shimmer, http://www.shimmer-research.com/

[10] Zephyr, http://www.zephyr-technology.com/

A TWO-LEVEL APPROACH FOR DISCOVERING DRIVING STRATEGIES ACCORDING TO CONFLICTING OBJECTIVES

Erik Dovgan [1], *Matija Javorski* [2], *Matjaž Gams* [1], *Bogdan Filipič* [1]
[1] Department of Intelligent Systems, Jožef Stefan Institute, Jamova cesta 39, 1000 Ljubljana, Slovenia
[2] Faculty of Mechanical Engineering, University of Ljubljana, Aškerčeva cesta 6, 1000 Ljubljana, Slovenia
Tel: +386 1 4773393; fax: +386 1 4773131
e-mail: erik.dovgan@ijs.si

ABSTRACT

This paper presents a two-level multiobjective optimization approach for discovering driving strategies. The algorithm at the lower level is a deterministic multiobjective algorithm that searches for driving strategies with respect to two conflicting objectives: traveling time and fuel consumption. The algorithm at the upper level is an evolutionary algorithm that searches for the best input data for the lower-level algorithm. This paper describes the improvements of the lower-level algorithm that we presented in the past. Moreover, it presents the upper-level algorithm and the preliminary tests of the two-level approach. Finally, ideas for future work are given.

1 INTRODUCTION

The awareness of the need to protect our environment is increasing. This trend suggests that each person reduces the environment pollution. When taking into account vehicles, e.g., cars and trucks, the pollution reduction is proportional to the reduced fuel consumption. This has an interesting side effect: the traveling cost is reduced. However, there is always a price to pay. In this case, the traveling time increases. It is clear that it is not optimal to heavily reduce the fuel consumption because the traveling time becomes unacceptably long. Therefore, both objectives have to be taken into account simultaneously when constructing a driving strategy.

There is a lot of research done in the field of discovering driving strategies. The most commonly used approach uses the weighted sum of several objectives that has to be minimized. Another approach minimizes only one objective, e.g., fuel consumption, where the other objectives are used as constraints. Such approaches enable to use single objective algorithms, such as predictive control or dynamic programming. These algorithms find only one solution. When a weighted sum is used, the solution significantly depends on the used weights. However, it is usually not clear what weights are optimal. In order to find a set of good strategies without defining the weights, a multiobjective optimization approach has to be used. It finds a set of nondominated strategies where no strategy is better than any other strategy with respect to both objectives. Afterwards, the user selects the most preferable strategy without limiting the time in advance and without defining the weights. Such an approach is specially suitable for users that frequently travel on the same route, e.g., transportation companies. It enables to search for strategies for one route only once. Afterwards, the user selects a different strategy from the set of found strategies when the requirements change.

This paper presents a two-level approach that searches for driving strategies with respect to conflicting objectives. The lower-level algorithm is a deterministic multiobjective algorithm based on breadth-first search [1] that includes mechanisms from the Nondominated Sorting Genetic Algorithm NSGA-II [2]. It searches for a set of nondominated strategies, where a strategy is presented as a set of actions that are applied to the vehicle in various states. It requires a predefined state space discretization, i.e., hypercube discretization, and other parameters, which are defined with the upper-level algorithm. The upper-level algorithm is an evolutionary algorithm that searches for good hypercube discretizations and other parameters. The final set of nondominated strategies consists of the strategies found at the best discretizations.

2 MULTIOBJECTIVE DISCOVERY OF DRIVING STRATEGIES

In [3] we described an initial implementation of the lower-level algorithm, i.e., the deterministic multiobjective algorithm that searches for driving strategies. This algorithm has been further improved. The algorithm and its improvements are presented in Subsection 2.1. This algorithm has been included into the two-level approach at the lower level. The upper-level algorithm searches for hypercube discretizations which are required as input for the lower-level algorithm. The upper-level algorithm and the integration of the two algorithms are presented in Subsection 2.2.

2.1 The lower-level algorithm for discovering driving strategies

This subsection presents the mutiobjective algorithm for discovering driving strategies with respect to two objectives: traveling time t and fuel consumption c. The inputs for the algorithm are the hypercube discretization of the vehicle and route state space, the number of predicted steps N_p, and the minimum ω_{min} and maximum fuel consumption weight ω_{max} from the vector of consumption weights Ω. The discretization is defined with the vectors of interval limits for the dimensions of the state space as presented in Table 1.

Table 1: *The discretization definition*

Discretized space dimension	*Vector of interval limits*
Vehicle velocity and velocity limit	D_v
Route length of current segment	D_s
Segment inclination	D_α
Throttle percentage	D_ε

The discretization is used to discretize a seven-dimensional vehicle and route state space. The obtained hypercubes store fuel consumption weights ω. The weight is used to determine the optimal action that is applied to the vehicle if its state and the current route state is inside the hypercube. The hypercube dimensions are presented in Table 2. For further information about hypercube discretizations see [3].

Table 2: *The hypercube dimensions*

Hypercube dimension	*State type*	*Variable*
Velocity	Vehicle state	v_V
Inclination of the current segment	Route state	α_{CS}
Inclination of the next segment	Route state	α_{NS}
Velocity limit of the current segment	Route state	$v_{CS,limit}$
Velocity limit of the next segment	Route state	$v_{NS,limit}$
Route length to the end of the current segment	Route state	s_{SE}
Route length from the start of the current segment	Route state	s_{SS}

The lower-level algorithm for discovering driving strategies is outlined in Algorithm 1. For more detailed explanation see [3]. The main improvements with respect to the previously presented algorithm are the following. First, predictive search is added to the algorithm. Consequently, the best actions are selected based on the prediction of the vehicle driving for more than one step when an action is applied. Second, a weighted-sum approach is used to select the best action at each step from the set of all possible actions A. Each action consists of throttle percentage ε and gear g. The throttle percentages are discretized with the D_ε. The hypercubes store fuel consumption weights that are used to select the appropriate actions applied to the vehicle.

Algorithm 1: *Lower-level algorithm*

$S_{pop} = \{S_{init}\}$ $\{S_{init}$ *is empty strategy*$\}$
$S_{final} = \{\}$
repeat
  $S_{pop,nextStep} = \{\}$
  for all $S \in S_{pop}$ *do*
    $S_{pop,temp} = \{\}$
    if $\{$*currently observed hypercube of strategy S stores weight*$\}$ then
      $S_{pop,temp} = \{S\}$
    else
      for each $\omega \in \Omega, \omega_{min} \le \omega \le \omega_{max}$ *do*
        $S_{clone} = S.clone()$
        $S_{clone}.add(\omega)$
        $S_{pop,temp}.add(S_{clone})$
      end for
    end if
    for all $S_{temp} \in S_{pop,temp}$ *do*
      for each $a \in A, a = \{\varepsilon, g\}$
        $S_{temp}.drivingPrediction(N_p)$
      end for
      $a = S_{temp}.selectBestAction(S_{temp}.\omega)$
        $\{$*select* a_i *where* $\min_{i=1\ldots|A|}(c_{a_i}\omega + t_{a_i}(1-\omega))\}$
      $S_{temp}.simulateOneStep(a)$
      if $\{S_{temp}$ *simulated driving on whole route and feasible*$\}$ then
        $S_{final}.add(S_{temp})$
      else if $\{S_{temp}$ *feasible*$\}$ *then*
        $S_{pop,nextStep}.add\{S_{temp}\}$
      end if
    end for
  end for
  $reduceNumberOfStrategies(S_{pop,nextStep})$
    $\{$*apply Fast Nondominated Sort and Crowding Distance* [2]$\}$
  $S_{pop} = S_{pop,nextStep}$
until $S_{pop} = \{\}$
$S_{final} = returnNondominatedStrategies(S_{final})$
  $\{$*apply Fast Nondominated Sort* [2]$\}$
return S_{final}

2.2 The upper-level algorithm for discovering good discretizations

This subsection presents the upper-level evolutionary algorithm that searches for good discretizations and other

parameters required by the lower-level algorithm. Its pseudocode is presented in Algorithm 2. The details of the algorithm are presented in the following subsections.

Algorithm 2: *Upper-level algorithm*

```
randomly initialize a population of solutions
for each generation
        for each solution in the population
                select two solutions
                crossover
                mutation
                evaluate solutions and add to population
        end for
end for
```

2.2.1 Representation of solutions

The solution is encoded as a vector of Boolean values and three integer values forming the chromosome. The vector of Boolean values defines the discretization as follows. The finest discretization has to be given in advance as $\{D_{v,finest}, D_{s,finest}, D_{\propto,finest}, D_{\varepsilon,finest}\}$. The discretization used by the solution is a subset of the finest discretization. This is defined with the vector of Boolean values D, $|D| = |D_{v,finest}| + |D_{s,finest}| + |D_{\alpha,finest}| + |D_{\varepsilon,finest}|$. If the Boolean value is true, the corresponding interval bound in the discretized hypercube dimension is used, resulting in more fine discretization. Otherwise, the bound is not used, resulting in more coarse discretization. The obtained discretization is given as follows:

$D_v = \{v; v = D_{v,finest}[i],\ where\ D[i] = true\ and\ i = 1\ to\ |D_{v,finest}|\}$

$D_s = \{s; s = D_{s,finest}[i],\ where\ D[i] = true\ and\ i = |D_{v,finest}| + 1\ to\ |D_{v,finest}| + |D_{s,finest}|\}$

$D_\alpha = \{\alpha; \alpha = D_{\alpha,finest}[i],\ where\ D[i] = true\ and\ i = |D_{v,finest}| + |D_{s,finest}| + 1\ to\ |D_{v,finest}| + |D_{s,finest}| + |D_{\alpha,finest}|\}$

$D_\varepsilon = \{\varepsilon; \varepsilon = D_{\varepsilon,finest}[i],\ where\ D[i] = true\ and\ i = |D_{v,finest}| + |D_{s,finest}| + |D_{\alpha,finest}| + 1\ to\ |D_{v,finest}| + |D_{s,finest}| + |D_{\alpha,finest}| + |D_{\varepsilon,finest}|\}$

The three integer values represent the following parameters:

- the number of predicted steps N_p, between 1 and maximum value $N_{P,max}$.
- the index of the minimum fuel consumption weight $i_{\omega_{min}}$; $\omega_{min} = \Omega[i_{\omega_{min}}]$,
- the index of the maximum fuel consumption weight $i_{\omega_{max}}$; $\omega_{max} = \Omega[i_{\omega_{max}}]$.

The vector of possible consumption weights Ω is given in advance. These values are sorted in ascending order. The solution uses only the fuel consumption weights between ω_{min} and ω_{max}.

2.2.2 Genetic operators

The algorithm uses multi-point crossover and mutation operators. Crossover is performed by selecting crossover points and exchanging data between each second pair of consecutive crossover points. Mutation of Boolean values is carried out by negating them, while the integer values are mutated by replacement with a random value between minimum and maximum bounds [4].

2.2.3 Algorithm evaluation

The goal of the algorithm is to find good discretizations. Therefore, a population of discretizations is the result. Consequently, when the evaluation is performed, the whole population has to be considered.

The population consists of a set of solutions, i.e., $P = \{s_1, s_2, \dots\}$. When a solution is created, the driving strategies are found by taking into account the parameters $D_v, D_s, D_\alpha, D_\varepsilon, N_p, \omega_{min}$ and ω_{max}. Therefore, the solution s_1 contains a set of driving strategies $DS_{s_1} = \{DS_{s_1,1}, DS_{s_1,2}, \dots\}$. When the population is evaluated, the strategies from all solutions are applied, i.e., $DS_{population} = \{DS_{s_1,1}, DS_{s_1,2}, \dots, DS_{s_2,1}, DS_{s_2,2}, \dots\}$.

Finally, the hypervolume [5] is calculated.

When a child is obtained, the driving strategies of all solutions in the population with the exception of the parent are united with the driving strategies of the child. Next, the new hypervolume is calculated. Finally, if the new hypervolume is higher than the original hypervolume, the parent is removed from the population and the child is added to the population.

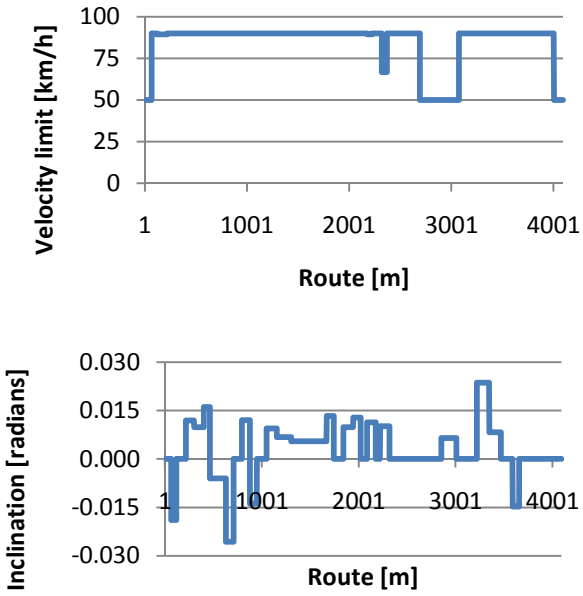


Figure 1: *Inclinations and radii of the testing route*

3 TESTING AND RESULTS

The presented approach has been tested on one route defined as shown in Figure 1. The parameters of the evolutionary algorithm are given in Table 3. Three tests

have been done. The first and the third tests have been done using the finest discretization given in Table 4. The second test has been carried out using the updated finest discretization as presented in Table 5. The values of other parameters in the chromosome are given in Table 6. To sum up, the three tests can be described as follows. The second test has more fine route discretization with respect to the first test. The third test has a higher maximum number of predicted steps compared to the first test.
The results are shown in Figure 2. The results show that better strategies are obtained if more prediction steps are used. On the other hand, the quality of the strategies does not change if more fine hypercube discretization is used.

Table 3: *Parameters of evolutionary algorithm*

Parameter	*Value*
Number of solutions	20
Number of generations	400
Crossover probability	0.9
Number of crossover points	3
Mutation probability	0.1

Table 4: *The finest discretization values*

Variable	*Unit*	*Min*	*Max*	*Step*
$D_{v,finest}$	km/h	0	130	5
$D_{s,finest}$	m	20	50	30
$D_{\propto,finest}$	$radians$	-0.16	0.16	0.02
$D_{\varepsilon,finest}$		-1	1	0.05

Table 5: *The updated finest discretization values*

Variable	*Unit*	*Min*	*Max*	*Step*
$D_{s,finest}$	m	10	30	10
		50	70	20
		70	100	30

Table 6: *Other chromosome values*

Variable	*Test number*	*Min*	*Max*	*Step*
$N_{P,max}$	1, 2	1	10	1
$N_{P,max}$	3	1	80	1
ω	1, 2, 3	0.00	0.95	0.05
		0.99	1.00	0.0003

4 CONCLUSION

We presented a two-level approach that searches for driving strategies with respect to two conflicting objectives: traveling time and fuel consumption. The lower-level algorithm is a deterministic multiobjective algorithm for discovering driving strategies based on breadth-first search that includes mechanisms from the Nondominated Sorting Genetic Algorithm NSGA-II. The upper-level algorithm is an evolutionary algorithm that searches for good vehicle and route state space discretizations and other parameters that are input to the lower-level algorithm. The final set of nondominated driving strategies consists of the strategies found at the best discretizations. The approach has been preliminary tested and the results already confirm the effectiveness of the approach.
The future work will include additional testing and comparison with the existing weighted-sum approach. Moreover, additional routes will be tested with the presented approach and weighted-sum approach, and the practical potential of the results will be analyzed.

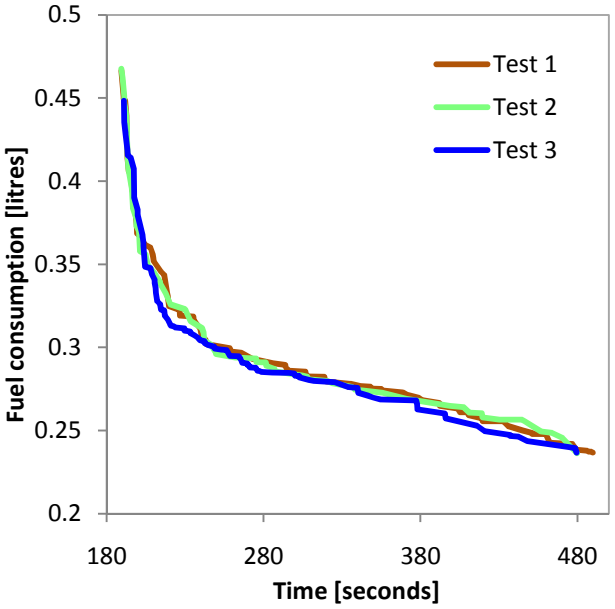


Figure 2: *Results obtained with the three tests*

References

[1] S. J. Russell, P. Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall. New Jersey. 2010.

[2] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*. Vol. 6. No. 2. pp. 182-197. 2002.

[3] E. Dovgan, M. Gams, B. Filipič. A multiobjective optimization algorithm for discovering driving strategies. *13th Annual Genetic and Evolutionary Computation Conference GECCO 2011*. Dublin. Ireland. 2011.

[4] L. Davis. *Handbook of Genetic Algorithms*. Van Nostrand Reinhold. New York. 1991.

[5] E. Zitzler, L. Thiele. Multiobjective evolutionary algorithms: A comparative case study and the strength Pareto approach. *IEEE Transactions on Evolutionary Computation*. Vol. 3. No. 4. pp. 257-271. 1999.

RAZVOJ SISTEMA ZA RAČUNALNIŠKO PODPRTO EVIDENTIRANJE IN SESTAVLJANJE FRAGMENTOV STENSKIH POSLIKAV

Bogdan Filipič, *Miha Mlakar*, *Erik Dovgan*, *Tea Tušar*
Odsek za inteligentne sisteme, Institut "Jožef Stefan", Jamova cesta 39, 1000 Ljubljana
Tel: (01) 477 33 52; faks: (01) 477 31 31
E-pošta: bogdan.filipic@ijs.si

POVZETEK

Restavriranje stenskih poslikav iz fragmentov z arheoloških najdišč je zahtevno zaradi množice fragmentov različnih velikosti in oblik, poškodb fragmentov in manjkajočih delov poslikav. V prispevku predstavljamo projekt razvoja računalniške podpore za evidentiranje in podporo pri sestavljanju fragmentov stenskih poslikav, ki bo olajšala in pohitrila restavratorsko delo, ter prikazujemo zmožnosti programske opreme, ki smo jo razvili v ta namen.

1. UVOD

Ena od dejavnosti Ateljeja za stensko slikarstvo Restavratorskega centra, ki deluje v okviru Zavoda za varstvo kulturne dediščine Slovenije [1], je restavriranje stenskih poslikav, pridobljenih z arheoloških najdišč v obliki fragmentov. Atelje sprejema, evidentira in hrani fragmente ter z njihovim sestavljanjem restavrira stenske poslikave. Vsi postopki obravnave fragmentov so do sedaj potekali ročno. Ker je fragmentov z določenega najdišča ponavadi veliko in so lahko na različne načine poškodovani, mnogi pa tudi manjkajo, je restavriranje stenskih poslikav iz fragmentov zahtevno. Da bi ga olajšali in pohitrili, smo razvili sistem za računalniško podprto evidentiranje in sestavljanje fragmentov stenskih poslikav Pedius [2, 3].

Projekt razvoja računalniške podpore je obsegal določitev ustreznega načina označevanja fragmentov in zajemanja njihovih digitalnih posnetkov, specifikacijo baze podatkov za evidentiranje fragmentov, razvoj programske opreme za evidentiranje fragmentov in pomoč pri njihovem sestavljanju ter njeno uvedbo v prakso.

V nadaljevanju predstavljamo arheološko nahajališče iz rimske Celeie, bogato s fragmenti stenskih poslikav, katerih načrtovano restavriranje je zaradi zahtevnosti spodbudilo razvoj računalniške podpore, ter opisujemo postopke označevanja fragmentov, zajema digitalnih slik fragmentov in računalniško podprtega sestavljanja fragmentov v poslikave.

2. ARHEOLOŠKO NAJDIŠČE IZ RIMSKE CELEIE

Območje Celja se ponaša z bogatimi ostanki rimskega mesta Celeie, ki potrjujejo navedbe kronistov in zgodovinarjev o takratni urbanizaciji in visoki ravni bivalne kulture. Razvoj današnjega mesta je bil ves čas tesno povezan s številnimi odkritji in arheološkimi najdbami. Prvi podatki o tem so znani iz druge polovice 15. stoletja, arheološke raziskave so se pričele sredi 19. stoletja, posebej intenzivne pa so bile v drugi polovici 20. stoletja.

Med pomembnejše najdbe v osrednjem delu mesta spadajo ostanki mozaikov iz rimskih zgradb. Eno takšnih najdišč je Turška mačka, ki je dobilo ime po starem gostišču na Gledališki ulici, pod katerim so ob prenavljanju odkrili ostanke rimskih hiš iz prvega stoletja (slika 1). Po ostankih sklepamo, da so bile hiše pretežno pritlične, nekaj pa je bilo tudi nadstropnih.

Slika 1: *Arheološko najdišče Turška mačka iz rimske Celeie*

Izkopavanja na najdišču Turška mačka so potekala leta 1978. V podolgovatem prostoru velikosti 13 m × 4 m je bil na tleh mozaik, stene pa so bile bogato okrašene s freskami. Freske so bile raznih barv in so vsebovale veliko vzorcev in motivov. Deli fresk, ki so jih po izkopu prenesli v Restavratorski center, so zelo različnih velikosti in oblik (slika 2). Zbranih fragmentov je preko 9000. Večinoma so dobro ohranjeni, tako da bi z uspešnim sestavljanjem lahko dobili lepe freske z rimskimi motivi.

Slika 2: *Primeri fragmentov antične freske*

V preteklosti so že začeli s sestavljanjem fragmentov s tega najdišča v freske, vendar se je delo brez računalniške podpore zaradi velikega števila fragmentov izkazalo za prezahtevno in je zastalo.

3. OZNAČEVANJE FRAGMENTOV POSLIKAV

Za potrebe evidentiranja, shranjevanja in restavriranja je potrebno fragmente poslikav primerno označiti. Ob tem vsakemu fragmentu dodelimo enolično oznako ali kodo. Takšno označevanje kasneje ob ustrezni računalniški podpori omogoča učinkovito shranjevanje in dostop do podatkov o fragmentih, iskanje uskladiščenih fragmentov, njihove primerjave v postopku sestavljanja poslikav ipd. Za označevanje fragmentov imamo na voljo različne tehnologije: papirne oznake, črtno kodo, identifikacijo z radijskimi valovi in zapisovanje oznak na fragmente s posebnimi snovmi.

Označevanje z oznakami, napisanimi na papir, ki se jih priloži k fragmentu ali pritrdi nanj, je najpreprostejši in najcenejši način označevanja fragmentov. V ta namen lahko uporabljamo običajen ali dodatno zaščiten papir. Na fragmente ga lahko pritrjujemo s sponkami ali z lepljenjem. Glavna prednost takega označevanja je nizka cena, največja pomanjkljivost pa časovna zahtevnost evidentiranja in obdelave fragmentov ter možnost napak zaradi ročnih postopkov. Ta način zato uporabljamo le v primerih manjšega števila fragmentov oziroma v kombinaciji z drugimi načini označevanja. Tako so na primer v projektu Pylos pri restavriranju fresk z arheoloških najdišč na Peloponezu fragmente označili s pisanjem na zadnjo stran s posebno snovjo, zaboje, v katere so shranili tako označene fragmente, pa z zapisi na papirju [4].

Označevanje s črtnimi kodami obsega tiskanje črtnih kod na papirne etikete in njihovo pritrjevanje na fragmente. Etikete so različnih velikosti in kakovosti. V primerjavi z označevanjem s papirnimi oznakami je ta način dražji zaradi potrebne strojne opreme, to je tiskalnika in čitalnika črtnih kod, njegova prednost pa je, da omogoča avtomatizacijo evidentiranja fragmentov. Nameščanje etiket s črtnimi kodami na fragmente, običajno na njihovo zadnjo stran, zahteva pazljivost oziroma dodatno delo. Neravna ometna podlaga namreč lahko povzroči težave pri odčitavanju črtnih kod, zato je treba omet po potrebi zbrusiti, da zagotovimo ravno podlago. To težavo lahko omilimo z uporabo etiket manjših dimenzij. Po drugi strani pa so večje etikete s kodo enostavnejše za odčitavanje, saj pri tem ni potrebna tolikšna natančnost pri usmerjanju laserskega žarka čitalnika na kodo.

Pri identifikaciji z radijskimi valovi (angl. Radio Frequency Identification, RFID [5]) na predmet namestimo majhno elektronsko vezje, imenovano RFID oddajnik, ki omogoča identifikacijo oziroma sledenje predmeta. Najpogosteje imajo RFID oddajniki obliko nekoliko debelejše etikete, lahko pa so vgrajeni tudi v kartice, žetone itd. Glede na napajanje se delijo na pasivne, ki nimajo svojega napajanja, in aktivne, ki vsebujejo tudi baterijo za napajanje. Signale RFID oddajnika sprejema RFID čitalnik, kar omogoča identifikacijo predmeta oziroma RFID oddajnika. Z brezžičnim čitalnikom lahko hitro in učinkovito ugotovimo stanje inventarja, saj se je s čitalnikom potrebno predmetom z RFID oddajniki le približati in ga ne točno usmeriti vanje, kot je to potrebno pri branju črtnih kod. Takšno označevanje tudi zmanjša možnosti poškodb, saj fragmentov ni potrebno veliko premikati. Identifikacija z radijskimi valovi tako nudi številne prednosti v nadaljnjih postopkih obdelave fragmentov, vendar je zaradi cene oddajnikov ter potrebne strojne in programske podpore to najdražji način.

Eden od načinov označevanja fragmentov je pisanje na fragmente s posebnimi snovmi. Če oznake fragmentov zapisujemo na njihovo sprednjo stran (lice), mora biti uporabljena barva taka, da jo je moč enostavno odstraniti brez poškodovanja fragmenta. Ob pisanju na zadnjo stran pa je možno oznake pisati neposredno na omet ali pa najprej na omet nanesti pomožno plast, ki služi kot podlaga za pisanje.

Po analizi možnih načinov označevanja fragmentov smo izbrali označevanje s črtnimi kodami, ki glede na veliko število fragmentov omogoča ustrezno stopnjo avtomatizacije evidentiranja ob sprejemljivih stroških. Izbrana velikost etiket s črtnimi kodami je 13 mm × 6 mm.

4. ZAJEM DIGITALNIH SLIK FRAGMENTOV

Osnovni namen označevanja fragmentov je poenostavitev računalniškega evidentiranja in sestavljanja fragmentov v originalne freske. Temelj tovrstnih restavratorskih projektov je baza podatkov o fragmentih, med katerimi so najpomembnejše slike fragmentov. Digitalni zapis slik fragmentov omogoča programsko iskanje in primerjanje fragmentov pri zlaganju v fresko. Šele, ko na ta način ugotovimo, katere fragmente bomo združili, to opravimo z dejanskimi fragmenti. Tako podprto restavriranje poslikav je lažje, učinkovitejše, manjša pa je tudi možnost dodatnih poškodb fragmentov.

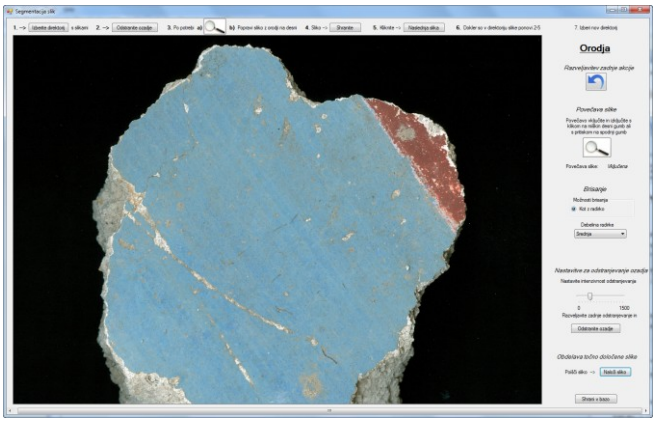
Metode za zajem digitalnih slik oziroma digitalizacijo fragmentov se delijo v dve skupini glede na to, kaj zajemamo: 2D slike lic fragmentov ali 3D posnetke fragmentov. Od tega je odvisna tudi potrebna strojna oprema. Za 2D slikanje fragmentov običajno uporabljamo optične čitalnike in digitalne fotoaparate, za zajem 3D posnetkov pa kombinacije optičnih čitalnikov in drugih namenskih laserskih čitalnikov.

Za rekonstrukcijo freske so sicer najpomembnejša lica fragmentov, a slikovna informacija o njih včasih ne zadostuje. To zlasti velja za dele fresk brez izrazitih vzorcev, ko so edine informacije, ki lahko pomagajo pri sestavljanju fragmentov, oblika robov ter sestava in barva ometa. Če imamo na voljo le 2D slike, si pri sestavljanju pomagamo z obliko robov fragmentov, kar pa pogosto ni zadostno, saj so robovi fragmentov pogosto okrušeni in se zato ne prilegajo dobro. V takih primerih si lahko pomagamo s 3D modeli fragmentov, ki poleg informacije o robu lica fragmenta vsebujejo tudi informacije o obliki pripadajočega ometa, njegovi debelini in grobosti materiala. Ta pristop so uporabili npr. pri restavriranju fresk z najdišč na Santorinu [6].

Po analizi prednosti in slabosti obeh načinov zajemanja slik fragmentov smo izbrali zajemanje 2D slik lic fragmentov. Ta način je cenejši, zajem slik je občutno hitrejši in tudi razvoj in uporaba programske opreme sta enostavnejša.

Zajemanje slik lic fragmentov poteka na dva načina. Pri veliki večini fragmentov sliko lica fragmenta pridobimo z optičnim čitalnikom, fragmente, ki so za optični čitalnik preveliki, pa slikamo z digitalnim fotoaparatom. Na osnovi podanega merila in ločljivosti sliki programsko spremenimo velikost tako, da dobi lastnosti, kot jih imajo slike, dobljene z optičnim čitalnikom.

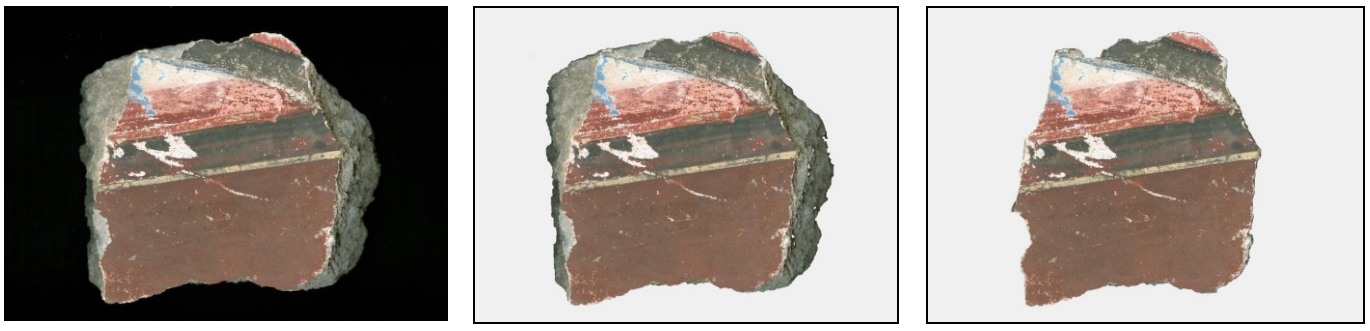
S programskim modulom za evidentiranje fragmentov stenskih poslikav (slika 3) dobljene 2D slike fragmentov nato skupaj s pripadajočimi lastnostmi zapišemo v podatkovno bazo. Slike fragmentov je potrebno pred shranjevanjem v bazo računalniško obdelati, kot prikazuje slika 4. Slika fragmenta, ki jo dobimo z optičnim čitalnikom ali digitalnim fotoapatatom, ima črno ozadje in poleg lica fragmenta je običajno viden tudi omet. Razvita programska oprema omogoča samodejno odstranitev ozadja, uporabnik pa z dodatnimi orodji na sliki odstrani tudi vidne sledi ometa. Rezultat te obdelave je slika, na kateri je le lice fragmenta. Program nato izračuna površino fragmenta ter vsebovanost in deleže barv na tem fragmentu. Uporabnik poleg tega izbere še grafično in črtno predlogo, katerima pripada fragment. Na osnovi teh podatkov je nato v fazi sestavljanja možno iskanje podobnih fragmentov.



Slika 3: *Programski modul za evidentiranje fragmentov stenskih poslikav: glavno okno za obdelavo digitalnih slik fragmentov*

5 SESTAVLJANJE FRAGMENTOV

Osrednji modul razvite programske opreme podpira sestavljanje fragmentov v originalno stensko poslikavo. Po vnosu slik in ostalih podatkov o fragmentih v podatkovno bazo se uporabnik lahko loti interaktivnega sestavljanja fragmentov. Modul omogoča izvedbo tega postopka na računalniškem zaslonu, ko pa je uporabnik s sestavljeno stensko poslikavo ali njenim delom zadovoljen, fragmente na enak način sestavi tudi fizično. Z računalniško podporo se sestavljanje zelo pohitri, manj je tudi rokovanja s fragmenti in s tem manj možnosti njihovih poškodb.



a) b) c)

Slika 4: *Računalniška obdelava digitalne slike fragmenta: a) izvirna slika z optičnega čitalnika, b) odstranitev ozadja, c) odstranitev delov slike, ki predstavljajo omet*

Pred uvedbo računalniške podpore so restavratorji fragmente sestavljali v poslikave v za to namenjenem peskovniku. V njem je moč fragmente zložiti tako, da njihova lica ležijo vodoravno in na isti višini. Tako pripravljene fragmente je moč primerjati, prilegati in s poskušanjem postopoma sestaviti v prvotno poslikavo. Z razvito programsko opremo smo na računalnik prenesli koncept peskovnika in ga nadgradili z novimi možnostmi, ki so mogoče samo na računalniku. Osnovni potek dela obsega iskanje fragmentov v podatkovni bazi, njihovo dodajanje na delovno površino in manipuliranje na njej.

Učinkovito iskanje podobnih fragmentov je zlasti pomembno, kadar imamo opraviti z velikim številom fragmentov. Podobni fragmenti ponavadi na stenski poslikavi sodijo skupaj. Da bi jih v veliki množici laže našli, programski modul za podporo sestavljanja omogoča iskanje po različnih kriterijih in njihovih kombinacijah. Fragmente lahko iščemo glede na površino, ki jo zavzemajo, barvo, delež določene barve ter grafične in črtne predloge, ki jim fragmenti pripadajo.

Fragmente, ki so dodani na delovno površino, je možno programsko premikati, jih obračati, spreminjati pogled na delovno površino in s tem vplivati na podrobnosti, ki jih vidimo, fragmente je mogoče tudi povezovati v skupine itd. Ko za dva ali več fragmentov ugotovimo, da sodijo skupaj, jih lahko programsko združimo in tako tvorimo nov fragment. Ta fragment dobi svojo unikatno identifikacijsko oznako in zanj se določijo tudi ostale lastnosti, enako kot za izvorne fragmente. Tako virtualno združujemo fragmente, cilj združevanja pa je postopno sestavljanje stenske poslikave v celoto. Primer delno sestavljene poslikave s tem postopkom prikazuje slika 5.

Slika 5: *Z razvito programsko opremo delno sestavljena stenska poslikava*

Pomembna prednost računalniškega sestavljanja fragmentov pred sestavljanjem v peskovniku pa je tudi možnost sestavljanja alternativnih postavitev fragmentov. Pogosto se namreč zgodi, da restavrator ni povsem prepričan, kako najbolje sestaviti določene fragmente. Razlogi za to so lahko poškodovanost robov fragmentov, odsotnost vsakršnega vzorca ali drugih sprememb na licu fragmenta, kar bi lahko pripomoglo k ugotavljanju sosednosti fragmentov, manjkanje določenih fragmentov ipd. Programsko je mogoče pripraviti več različnih postavitev fragmentov ter jih nato med seboj primerjati. Možno je tudi shranjevanje postavitev za nadaljnje delo, kopiranje izbrane postavitve fragmentov v drugo postavitev ter selektivno združevanje postavitev. Če restavrator v nekem trenutku ni prepričan v postavitev, lahko poskuša z več alternativnimi postavitvami, dokler ne ugotovi, katera je najustreznejša.

6 ZAKLJUČEK

Razvoj in praktična uporaba računalniške podpore za evidentiranje in sestavljanje fragmentov stenskih poslikav sta v času vsesplošne informatizacije temeljnega pomena za nadaljnji razvoj restavratorske stroke. V okviru teh prizadevanj smo za Restavratorski center v Ljubljani razvili računalniški sistem Pedius z bazo podatkov za evidentiranje in sestavljanje fragmentov stenskih poslikav. Spodbuda za razvoj sistema je bil zahteven projekt restavriranja antičnih fresk iz rimske Celeie na osnovi preko 9000 najdenih fragmentov.

Rrazvoj računalniške podpore je obsegal določitev ustreznega načina označevanja fragmentov in zajemanja njihovih digitalnih posnetkov, specifikacijo baze podatkov za evidentiranje fragmentov, razvoj programske opreme za evidentiranje in pomoč pri sestavljanju fragmentov ter njeno predajo v uporabo. Začetni rezultati na omenjenem zahtevnem restavratorskem projektu so spodbudni, v nadaljnjem razvoju sistema pa nameravamo povečati stopnjo avtomatizacije sestavljanja fragmentov.

LITERATURA

[1] ZVKDS Restavratorski center, Predstavitev Ateljeja za stensko slikarstvo. Dostopno na http://www.rescen.si/index.php?id=42, pridobljeno 14. 9. 2011.

[2] M. Mlakar, E. Dovgan, B. Filipič. Raziskava, razvoj in vzpostavitev baze podatkov za evidentiranje fragmentov stenskih poslikav. Projekt PR-02598, vmesno poročilo. Institut "Jožef Stefan", Ljubljana, 2010.

[3] M. Mlakar, E. Dovgan, T. Tušar, B. Filipič. Pedius, program za evidentiranje in podporo zlaganja fragmentov stenskih poslikav. Navodila za uporabo. Institut "Jožef Stefan", Ljubljana, 2011.

[4] The Pylos regional archeological project. Dostopno na: http://classics.uc.edu/prap/, pridobljeno 18. 6. 2010.

[5] R. Weinstein. RFID: A technical overview and its application to the enterprise. *IT Professional*, 7 (3), pp. 27–33, 2005.

[6] B. J. Brown, C. Toler-Franklin, D. Nehab, M. Burns, D. Dobkin, A. Vlachopoulos, C. Doumas, S. Rusinkiewicz, T. Weyrich. A system for high-volume acquisition and matching of fresco fragments: Reassembling Theran wall paintings. *ACM Transactions on Graphics*, 27 (3), 2008.

ACCELEROMETER DATA PREPARATAION FOR ACTIVITY RECOGNITION

Hristijan Gjoreski, Matjaž Gams
Department of Intelligent Systems
Jozef Stefan Institute
Jamova 39, 1000 Ljubljana, Slovenia
Tel: +386 1 4773419; fax: +386 1 4251038
e-mail: {hristijan.gjoreski, matjaz.gams}@ijs.si

ABSTRACT

This paper presents an approach to activity recognition using wearable accelerometers. The focus of this paper is on the process of data preparation which is of great importance for the activity recognition. The data preparation starts with the low and high pass filters which are applied on the raw data. Then, 25 attributes are computed for each accelerometer. Machine learning algorithm – Random Forest is used for evaluation. Results achieved with one, two and three accelerometers are presented. The results showed that with appropriate data preparation techniques small number (i.e. one or two) of accelerometers is sufficient for achieving acceptable performance (F-measure above 93%).

1 INTRODUCTION

There are a lot of projects and useful applications aiming at nursing older people. Most of them belong to the area of ambient intelligence and are trying to make an everyday life easier, simpler and safer for elderly. The research presented in this paper is a part of the Confidence project [1], which aims to create a remote care system to detect health problems of the elderly by monitoring their posture and activities. In this paper we present our work on the activity recognition using wearable 3-axis accelerometers.

After analyzing the related studies in terms of data preparation techniques, we noticed that usually researchers do not analyze or quickly go through the data from the accelerometers. Usually they are more focused on the algorithms. On contrary, in this research we were more interested in understanding the accelerometer's data and implementation of data preparation techniques.

In activity recognition process we analyzed seven target activities*: standing, sitting, lying, sitting on the ground, on all fours, going down,* and *standing up*. We also investigated the performance with different numbers of accelerometers (1 to 3). This way we showed the improvement of the system as the number of accelerometers is increased. The final system should be as non-intrusive as possible (fewer wearable accelerometers), but still accurate enough to detect each activity.

The classification algorithm used for the research is Random Forest [2]. This was the algorithm yielding the best results after analyzing several classification algorithms. For the experiments 11 young people were recorded performing the same scenario 5 times each. The leave-one-person-out (i.e. cross validation with 11 folds) technique was used for evaluation.

2 RELATED WORK

Activity recognition is an exciting area for the development of robust techniques, as applications in this field typically require to deal with high-dimensional, multimodal streams of data that are characterised by a large variability (e.g. due to changes in the user's behaviour or as a result of noise). However, unlike other applications, there is a lack of established benchmarking problems and datasets. Typically, each research group tests and reports the performance of their algorithms on their own datasets using experimental setups specially conceived for that specific purpose. For this reason, it is difficult to compare the performance of different methods.

Most of the researchers investigate the activity recognition problem **using machine learning techniques.**

Of particular interest are the results presented in [3]. A mobile phone was used as an accelerometer. The target activities differ from ours. Only three out of eight were common. The process of attribute computation consists only of the computation of statistical attributes (mean, variance, etc.). When they used the same person's data for training and testing the achieved accuracy was 90%, but when they used a different person's data for testing the accuracy drops to 65%. In our research we use data from different people for training and testing.

A similar evaluation approach to ours is used in [4]. The leave-one-person-out cross-validation evaluation technique is performed. They used the data from six people data for training and one person's data for testing. They used a Neuro-Fuzzy classifier and one accelerometer fixed on the wrist. In the process of attribute computation standard statistical attributes were computed, but also an analysis in the frequency domain was performed. Unfortunately, we

had only two out of eight activities in common. The reported overall accuracy is 93%.

In [5] the authors collected data from three users using two accelerometers to recognise five activities, i.e. walking, sitting, standing, running, and lying down. This paper claimed that data from a thigh accelerometer was insufficient for classifying activities such as sitting, lying down, walking, and running, and thus multiple accelerometers were necessary (a claim that is also proved with our research). They achieved an accuracy of 62% for the ankle accelerometer, 83% for the thigh and 95% for both accelerometers.

3 ACCELEROMETER

A **3-axis accelerometer** is a sensor that returns a real-valued estimate of the acceleration along the axes *x*, *y* and *z*. It measures the acceleration and output the projections of the acceleration vector represented in a 3D coordinate system. In Figure 1 an accelerometer with its coordinate system is presented. Because of the Earth's gravity, all objects experience a gravitational pull towards the Earth's centre. When the accelerometer is at rest, the only force that is affecting the sensor is the Earth's gravity. The acceleration unit of the pull is referred to as *g* or *g*-force. Consequently all objects are subject to 1 *g* acceleration. Figure 2 shows the accelerometer with its coordinate system and the *g*-force that is influencing it. This information about the *g*-force is of great interest to us. Using the gravity component we can find out the orientation of the sensor (e.g. vertical, horizontal), which enables us to distinguish between different activities (e.g. standing, lying).

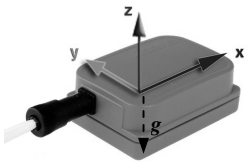


Figure 1: *Sensor-specific 3D coordinate systems.*

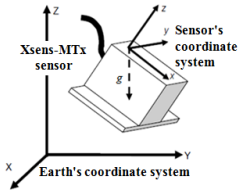


Figure 2: *Earth and sensor specific 3D coordinate systems.*

The measured acceleration vector is directed upwards (positive value for the *z* axis), even though the gravitational force pulls downwards. That means that only when the accelerometer is in free fall it will measure a value of zero. Even though its speed is increasing, it is in an inertial frame of reference, in which it is weightless. When the accelerometer is at rest, it will measure 1 *g* upwards. The accelerometer is not measuring gravity, but the force of the surface on the body that counteracts gravity.

4 DATA PREPARATION

4.1 Filters

Because the data is received from sensors (which usually means noisy data), additional filtering techniques were applied. There are several techniques that perform signal filtering. In our research we applied a **low-pass** and **high-pass** filter [6].

A simple **low-pass filter** for data in time domain is a smoothing function. In other words, the filtered signal is smoother and less dependent on short changes.

For the activity recognition task one of the most important features is the orientation (inclination) of the accelerometer. Therefore we had to use techniques that will filter out the portion of the acceleration data caused by gravity from the portion of the data that is caused by motion of the accelerometer. To do this, we used a low-pass filter to reduce the influence of sudden changes on the accelerometer data. The resulting filtered values then reflect the more constant effects of gravity. Actually with the low-pass filter we isolated the gravity component from the acceleration data.

It is also possible to filter a series such that the low-frequency variations are reduced and the high-frequency variations are unaffected. This type of filter is called a **high-pass filter**. This is especially important in the acceleration data, in which this filter allows us to eliminate the gravity component and take into consideration only the isolated sudden changes in acceleration.

Algorithm 1 shows the version of the low-pass and high-pass filters that we use in our research. It uses a low-value filtering factor to generate a value that uses 20% of the unfiltered acceleration data and 80% of the previously filtered value. This factor was chosen empirically.

Algorithm 1: *Low-Pass and High-Pass Filter.*

```
alpha = 0.8
Filter  BEGIN
        Low_X = alpha × Prev_X + (1 − alpha) × Curr_X
        Low_Y = alpha × Prev_Y + (1 − alpha) × Curr_Y
        Low_Z = alpha × Prev_Z + (1 − alpha) × Curr_Z
        High_X = Curr_X − Low_X
        High_Y = Curr_Y − Low_Y
        High_Z = Curr_Z − Low_Z
Filter END
```

As shown with the sample code above, the low and high passed acceleration values are stored in *Low* and *High* 3D vectors respectively. The previous acceleration values are stored in the *Prev* 3D vector and the current values are stored in the *Curr* 3D vector.

4.2 Attribute Computation

This subsection describes the process of computing the attributes. These attributes are later combined to create the final attribute vector which is used in the machine learning – classification stage. All the attributes are computed by using the technique of **overlapping sliding windows**.

Sliding window is a common approach to solving the problem activity recognition. Usually the algorithms do not try to recognise each data sample that is received from the sensors, but are trying to recognise some pattern in the data that is over some time interval (window).

In time series analysis, a sliding window is a technique that combines set of data samples in one window. A window

size is the time interval for which the data is collected. If the widows have some data samples as intersection, then this technique is named overlapping sliding windows.
Because the final sampling frequency of our accelerometers was 6 Hz, we chose a window size of six, which is one-second time interval. We decided for one-second time interval because in our target activities there are transitional activities (standing up and going down) that usually last from one to four seconds.

Length of the Acceleration Vector

The first computed attribute is the length of the acceleration vector. It is a simple but very useful attribute, which is also used further in the process of the computation of new attributes. It is not used as separate attribute in the final attribute vector because of the sliding window technique.

Statistical Attributes

The first sets of attributes that are used in the final attribute vector are the statistical attributes. They are computed for the low-passed filtered values of each of the axes and for the length of the acceleration vector. Three statistical features are computed: *Mean Value, Root Mean Square* and *Standard Deviation*.

Accelerometer Movement Detection

When a person's body is static, the accelerometer responds only to the gravity, producing a constant 1 *g* total acceleration. During motion the accelerometer produces a changing acceleration signal and the fiercer the motion, the greater the change in the signal. Using these changes in the acceleration vector, an attribute is computed for the detection of the accelerometer movement: *Acceleration Vector Changes (AVC)*. The AVC value of this attribute increases as the accelerometer is in motion (walking, going down, standing up, etc.). This attribute takes into consideration the data from the current window (six data samples). It sums up the last six differences of lengths of the acceleration vector and divides the sum by the time interval (one second) of the data. The AVC is computed as follows:

$$AVC = \frac{\sum_{i=1}^{n} | length_i - length_{i-1} |}{T_n - T_0} \quad (1)$$

T_0 is the time stamp for the first data sample in the window, and T_n is the time stamp of the last data sample. With this attribute the movement of the person can be detected: it distinguishes static from dynamic activities. For this attribute the raw value for the length of the acceleration vector is used instead of the low-passed value. The reason for this is that we are more interested in the small changes in the acceleration signal and the low-pass filter smoothes these changes.

Max-Min Value

An additional attribute that is computed is the difference between the maximum and the minimum value of the acceleration vector in the current data window. The difference between these two values is bigger in transitional activities (e.g. going down, standing up).

Accelerometer Inclination Angles

With accelerometer inclination angles we conclude the attribute computation subsection. The most important characteristic for activity recognition is the inclination (i.e. orientation, tilt) of the accelerometers. Accelerometer's data is used to measure the static angle of inclination. The inclination angles are calculated as the angles between the actual acceleration (e.g. the Earth's gravity for static activities) and each of the axes (Figure 3).

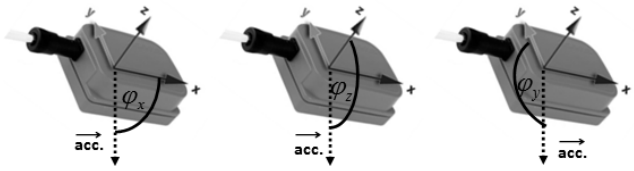


Figure 3: *Accelerometer inclination angles.*

For instance, the angle φ_x between the acceleration vector and the *x* axis is computed as follows:

$$\varphi_x = \arccos\left(\frac{a_x}{\sqrt{a_x^2 + a_y^2 + a_z^2}}\right), \quad (2)$$

where the values a_x, a_y and a_z represent the actual acceleration vector.
It should be noted that for the computation of these angles, low-passed filtered data is used, because it has fewer changes and the angle has fewer variations. Without the low pass filter the angles were sensitive to each small change of the accelerometer. These angles improve the classification of activities that have different accelerometer angle inclinations. For instance, when the chest accelerometer is in vertical position the user is probably standing or sitting. The horizontal position of the accelerometer indicates that the person is lying or is on all fours. The sitting on the ground activity is user dependent and in most of the cases is in between these two groups. Similarly the accelerometer placed on the thigh can distinguish between standing and sitting, but has problems distinguishing between sitting and lying. Different body placements of the accelerometers can give different information about the target activities.

5 ACTIVITY RECOGNITION

After the step of data preprocessing and computation of all additional attributes, the final attribute vector is created. This attribute vector is passed to the classification model which tries to recognise (classify) the appropriate activity of the user. This process is shown in Figure 4. Because the main focus of this research is the data preparation process, the next steps are briefly described.
Machine learning approach was used for the activity recognition. In this study, the machine learning task is to learn a model that will be able to classify the target activities (e.g. standing, sitting, etc.) of the person wearing accelerometers. The classification step was performed using

the application program interface of the software toolkit WEKA (Waikato Environment for Knowledge Analysis) [7]. Random Forest [2] was the algorithm that yielded the best results. It is an ensemble method for decision trees.

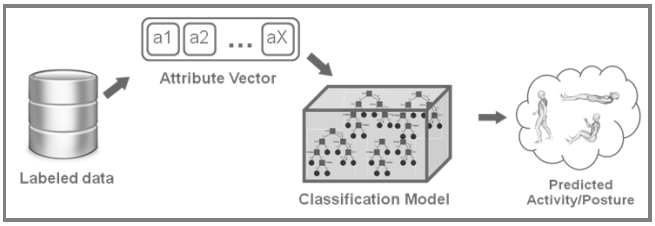


Figure 4*: Activity Recognition Flow Chart*

5.1 Classification Results

The experimental dataset was recorded by 11 people (7 males and 4 females). Special test scenario was created, which included all target activities. The test scenario was performed 5 times by each person. The events in the scenario were recorded in a single recording. The tests were conducted in an experimental laboratory. The total number of instances without the overlapping sliding window technique was **316 314**. After implementing the sliding window technique, the final amount of instances was reduced to **105 438.**

The leave-one-person-out cross-validation technique was used for evaluation. Thus, each fold was represented by the data of one person. This means the model was trained on the data recorded for ten people and tested on the remaining person's data. This procedure was repeated for each person data (11 times) and the average performance was measured.

The F-measure evaluation metric was used as the most suitable for our research. The results are shown in Figure 5. The F-measure for each activity is presented. The body placements of the accelerometers were chosen to be: the chest, the thigh and the ankle.

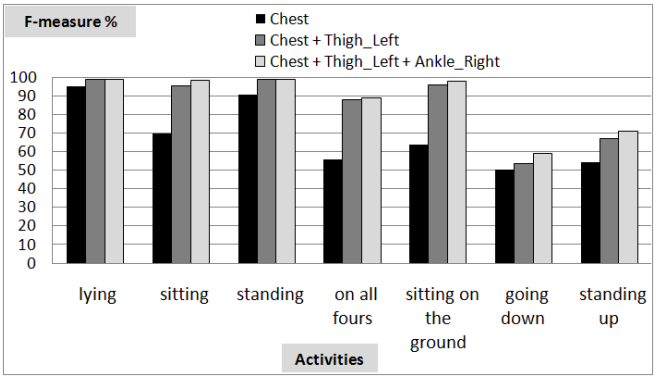


Figure 5: *Results achieved for each activity.*

The overall F-measure was 93% when the system was using one accelerometer; 96.4% and 98% with two and three accelerometers respectively. The results show that there is obvious improvement in the system when the number of accelerometers increases from one to two. When the third accelerometer is added to the system, the improvements are minimal (by 1.5 percentage point overall).

6 CONCLUSION

Data preparation is challenging task in machine learning. It is especially important step in activity recognition systems that rely on sensor data. In our research we used wearable sensors (i.e. accelerometers); therefore, the data is noisy and requires additional preprocessing techniques.

The first step towards the final solution was the step of understanding the raw data. After this, filtering techniques were applied.

Second step was the attribute computation step. Finding the appropriate attributes that will describe the user's behaviour was of particular interest. The behaviour needs to be represented by simple and general attributes, so that the algorithm using these attributes will also be general and work well on behaviours different from those in our scenario.

The final step in our research was the evaluation using machine learning algorithms. The results showed that with appropriate data preparation techniques small number of accelerometers is sufficient for achieving overall F-measure above 93%.

ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Community's Framework Programme FP7/2007–2013 under grant agreement No. 214986.

References

[1] The Confidence Project. http://www.confidence-eu.org (accessed August 2011).

[2] Breiman, L. Random forests. *Machine Learning*, **45**, 5–32 (2001).

[3] Ravi, N.; Dandekar, N.; Mysore, P.; Littman, M. L. Activity Recognition from Accelerometer Data. *Neural Networks*, (2009).

[4] Yang, J. Y.; Chen, Y. P.; Lee, G. Y.; Liou, S. N.; Wang, J. S. Activity Recognition Using One Triaxial Accelerometer: A Neuro-fuzzy Classifier with Feature Reduction. *Ifip International Federation For Information Processing*, 395-400 (2007).

[5] Krishnan, N.; Colbry, D.; Juillard, C.; Panchanathan, S. Real time human activity recognition using tri-Axial accelerometers. *Sensors, Signals and Information Processing Workshop* (2008).

[6] Meko, D. M. Course: Applied Time Series Analysis. Laboratory of Tree-Ring Research, University of Arizona. http://www.ltrr.arizona.edu/~dmeko/geos585a.html (accessed August 2011).

[7] Witten, I.; Frank, E. *Data Mining: Practical machine learning tools and techniques*. (Morgan Kaufmann. 2nd Edition, 2005).

DETEKCIJA SUMLJIVEGA OBNAŠANJA Z UPORABO FUNKCIJE KORISTNOSTI

Boštjan Kaluža
Odsek za inteligentne sisteme, Institut "Jožef Stefan"
Jamova cesta 39, 1000 Ljubljana, Slovenija
bostjan.kaluza@ijs.si

POVZETEK

V mnogih domenah, kjer je detekcija sumljivega obnašanja posameznikov ključnega pomena, ni na voljo posameznih signifikantnih dogodkov oz. incidentov, ki bi to omogočali. Namesto tega smo soočeni s kombinacijo dogodkov, ki vodijo do zaključka ali se posameznik obnaša sumljivo ali ne. Trenutni algoritmi za detekcijo takšnega obnašanja ne upoštevajo dejstva, da se evalvacija zaznanih dogodkov spremeni glede na obnašanje posameznika v preteklosti. V tem prispevku predstavimo razširitev pristopa za detekcijo planov z uporabo koristnosti: koristnost modeliramo kot potencial, sestavljen iz več funkcij, ki so odvisne tako od prehodov med različnimi dogodki, kot tudi od obnašanja posameznika v preteklosti. Predlagani pristop eksperimentalno preizkusimo na simulirani domeni, kjer je cilj detektirati sumljivo obnašanje potnika na letališču. Rezultati kažejo, da predlagan pristop dosega boljše rezultate od obstoječih metod.

1 UVOD

Dosedanje delo na področju detekcije sumljivega obnašanja se je osredotočalo predvsem na detekcijo posameznik signifikantnih dogodkov oz. incidentov, ki so razkrili namene sumljivega potnika. Primeri vključujejo detekcijo pretepov [5], kraj in zapuščenih predmetov [4], nenavadnih oblik trajektorij [7] itd. V tem prispevku se osredotočimo na razred aplikacij, kjer posamezen dogodek ne zadostuje za oblikovanje dokončne odločitve ali se posameznik obnaša sumljivo ali ne. Namesto tega se posvetimo zaporedju *prožilnih dogodkov*, ki razložijo posameznikovo obnašanje. Pri tem se soočamo z naslednjimi izzivi: (1) noben posamezen dogodek (oz. incident) ni dovolj za odločitev; (2) natančen načrt sumljivega posameznika ni znan; (3) prožilni dogodki vključujejo interakcije med več osebami, kar oteži njihovo zaznavo; in (4) prepričanje, da se posameznik obnaša sumljivo, narašča s številom nenavadnih prožilnih dogodkov ne-linearno. Ti problemi onemogočajo neposredno uporabo obstoječih metod za detekcijo sumljivih planov, saj plani niso znani v naprej. Enostavno štetje nenavadnih prožilnih dogodkov prav tako spodleti, saj ne omogoča ne-linearne akumulacije.

Detekcije sumljivega obnašanja se lotimo v dveh korakih. V prvem koraku predstavimo metodo za detekcijo prožilnih dogodkov, ki temelji na prepletenih skritih modelih Markova [3]. S to metodo je mogoče modelirati interakcije med dvema posameznikoma. V drugem koraku predstavimo razširitev metode za detekcijo planov z uporabo koristnosti [2]. Koristnost definiramo kot množico funkcij, odvisnih od prehodov med dogodki in zaporedjem že opaženih dogodkov. Predlagano metodo preizkusimo na simuliranem letališkem terminalu, kjer je cilj najti potnike, ki se obnašajo sumljivo. Rezultati kažejo, da predlagana detekcija prekaša sorodne metode.

2 OPIS PROBLEMA

Posameznika modeliramo kot agenta v večagentnem okolju in predpostavimo, da lahko natančno beležimo njegove akcije. Obnašanje posameznika opišemo z naključnim procesom H, ki generira prožilne dogodke. Proces H je sestavljen iz dveh naključnih podprocesov N (običajni) in S (sumljivi), ki modelirata običajnega oz. sumljivega posameznika.

Definicija 1. *Akcija je n-terka atributov $\langle f_1, ..., f_n \rangle_t$ ob času t, ki določajo stanje agenta.*

Definicija 2. *Sled akcije a je po času urejeno zaporedje akcij $\mathbf{a}^{(t)} = (a_1, a_2, ..., a_t)$.*

Definicija 3. *Prožilni dogodek x je podzaporedje sledi akcij, ki označuje del sledi. Opisan je z verjetnostima $n(x) = Pr\{x|y = 0\}$ in $s(x) = Pr\{x|y = 1\}$, kjer naključna spremenljivka $y = 0$ označuje, da je bil prožilni dogodek generiran s procesom N (oz. $y = 1$, če je x generiral S).*

Definicija 4. *Sled dogodkov je po času urejeno zaporedje prožilnih dogodkov $\mathbf{x}^{(t)} = (x_1, x_2, ..., x_t)$.*

Glede na apriorno verjetnost sumljivega dogodka λ, lahko verjetnost dogodka x_t opišemo kot

$$Pr\{H(t) = x_t\} = \lambda Pr\{H(t) = x_t|y_t = 1\} + \\ +(1 - \lambda)Pr\{H(t) = x_t|y_t = 0\}. \quad (1)$$

Cilj detekcije sumljivega obnašanja posameznikov je identifikacija tistih sledi dogodkov $\mathbf{x}^{(t)} = (x_1, x_2, ..., x_t)$, za katere je verjetnost

$$Pr\{y_t = 1 | H(t) = x_t, t = 1, ..., k\} > \tau, \qquad (2)$$

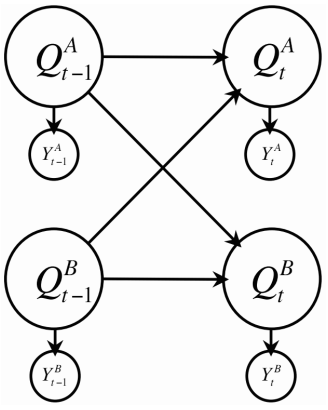
nad neko izbrano mejo τ ali relativno velika v primerjavi z verjetnostmi ostalih sledi.

3 DETEKCIJA DOGODKOV

Prožilni dogodek je lahko kakršnokoli podzaporedje sledi akcij, ki lahko kot *dokaz* podpre hipotezo bodisi o sumljivem bodisi o običajnem obnašanju. V našem primeru se osredotočimo na interakcije med posamezniki, natančneje med potniki in varnostnim osebjem. Ideja temelji na predpostavki, da se sumljivi potniki zaradi živčnosti v bližini varnostnega osebja obnašajo drugače kot običajni potniki.

Proces detekcije naprej izbere vse interakcije med potniki in varnostnim osebjem v določenem radiju (med potnikom in varnostnikom), kar navrže množico parov sledi akcij. Posamezen par tako predstavlja en prožilni dogodek x_i za vpletenega posameznika.

Za določanje verjetnosti $n(x)$ in $s(x)$ uporabimo prepletene skrite modele Markov (*Coupled Hidden Markov Models* [3], CHMM). Takšen model je sestavljen iz dveh verig skritega modela Markova, kjer skrita stanje iz ene verige neposredno vplivajo na skrita stanja v drugi verigi, kot je prikazano na sliki 1. Trenutno stanje Q_t^A agenta A je odvisno tako od prejšnjega stanja Q_{t-1}^A, kot tudi od prejšnjega stanja Q_{t-1}^B agenta B (podobno je stanje Q_t^B odvisno od Q_{t-1}^B in Q_{t-1}^A). Vsako skrito stanje Q_i neposredno vpliva na stanje opaženih vhodov Y_i.



Slika 1: Prepleteni skriti model Markova je sestavljen iz dveh medsobojno povezanih verig.

Za ocenjevanje verjetnosti zgradimo dva prepletena skrita modela Markova: model $\hat{N}_I$, ki modelira interakcije med običajnimi potniki in varnostnim osebjem, ter model $\hat{S}_I$, ki modelira interakcije varnostnega osebja s sumljivimi potniki. Ob novem prožilnem dogodku x izračunamo posteriorno verjetnost, da je bil dogodek generiran s prvim modelom $n(x) = Pr\{\hat{N}_I(x)\}$ oz. drugim $s(x) = Pr\{\hat{S}_I(x)\}$.

4 DETEKCIJA OBNAŠANJA

4.1 Razpoznavanje planov z uporabo koristnosti

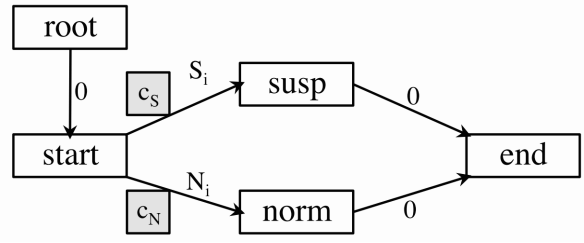
Za detekcijo obnašanja oz. odločitev ali sled dogodkov ustreza sumljivemu ali običajnemu posamezniku smo uporabili algoritem za detekcijo planov z uporabo koristnosti (*Utility-based Plan Recognizer*, Zilberbrand and Kaminka, 2007 [2]), ki je na kratko opisan v nadaljevanju. Podroben opis je mogoče najti v [1].

Algoritem sestavlja knjižnica planov in postopek, ki glede na opažene vhodne simbole poišče ujemajoče se plane v knjižnici. Korak v planu q je zaporedje akcij, ki vzdržujejo ali dosegajo nek cilj. Korak je lahko nedeljiv ali deljiv, torej sestavljen iz podkorakov. Plan je zaporedje korakov $\langle q_1, q_2, ..., q_n \rangle$ povezanih preko zaporednih povezav, ki opisujejo zaporedje izvajanja tega plana in pripadajočih podkorakov. Vsak korak v planu je opisan s tremi verjetnostmi prehodov: (a) verjetnost zaporednega prehoda, (b) verjetnost prekinitve plana in (c) verjetnost dekompozicije koraka (t.j. verjetnost prehoda iz koraka v podkorake). Vsak nedeljiv korak je opisan še z verjetnostno distribucijo nad vhodnimi simboli.

Algoritem predstavi tudi koristnost posameznih prehodov med koraki. V primeru, ko iščemo sumljivo obnašanje lahko koristnost obravnavamo tudi kot ceno prehoda. Prehodi z višjo ceno so tako manj ugodni. Vsak korak vsebuje tri koristnosti glede na vrsto povezave: (a) koristnost zaporednega prehoda $u_s(q_{i-1}, q_i)$, (b) koristnost prekinitve plana $u_e(q_{i-1}, q_{end})$ in (c) koristnost dekompozicije koraka $u_d(q_{i-1}^d, q_i^{d+1})$.

Vhod v algoritem je zaporedje simbolov o, ki v našem primeru ustrezajo kar sledi dogodkov $\mathbf{x}$. Zaporedje dogodkov se primerja s knjižnico planov, izbere plane, ki so konsistentni z zaporedjem o, in sortira plane glede na pričakovano koristnost.

Sledili smo pristopu za detekcijo nevarnih voznikov [1] in pripravili enostavno knjižnico planov, kot je prikazano na sliki 2. Posameznik lahko izvede sumljiv dogodek z verjetnostjo $s(x_t)$ in ceno $c_s < 0$ ali običajen dogodek z verjetnostjo $n(x_t)$ in ceno $c_n > 0$. Vse ostale cene se enake 0. Akomulirano pričakovano ceno (koristnost)



Slika 2: Knjižnica planov vsebuje dva koraka.

sledi dogodkov lahko zapišemo kot:

$$U_{UPR}(\mathbf{x}^{(k)}) = \sum_{t=1}^{k} u(x_t), \quad (3)$$

$$u(x_t) = \begin{cases} c_n n(x_t); & \text{if } |c_n n(x_t)| > |c_s s(x_t)| \\ c_s s(x_t); & \text{if } |c_n n(x_t)| \leq |c_s s(x_t)| \end{cases}.$$

Če akumulirana cena preseže neko izbrano mejno vrednost, potem sled dogodkov označimo kot sumljivo.

4.2 Koristnost kot funkcija

Predstavljeni pristop definira koristnost kot konstanto, kar vnaša določene omejitve. Tako ni npr. mogoče spremeniti koristnosti nekega prehoda glede na obnašanja posameznika v preteklosti. V tem prispevku definiramo koristnost kot funkcijo na naslednji način.

Definicija 5. *Funkcija koristnosti U, ki je odvisna od trenutnega koraka v planu q_t, prejšnjega koraka v planu q_{t-1} in celotnega vhodnega zaporedja $\mathbf{x}^{(t)}$, je funkcija, ki preslika*

$$U : \langle q_t, q_{t+1}, \mathbf{x} \rangle^k \to \mathbb{R}.$$

Funkcijo koristnosti lahko zapišemo kot

$$U(q_t, q_{t+1}, \mathbf{x}^{(t)}) = \sum_{j=1}^{k} \lambda_j u_j(q_t, q_{t+1}, \mathbf{x}^{(t)}),$$

kje vsaka funkcija u_j lahko predstavlja koristnost bodisi zaporednega prehoda bodisi prekinitve plana bodisi dekompozicije koraka. Parametri λ_j so konstante, ki jih določi uporabnik.

Ta definicija omogoča vpeljavo množice dodatnih funkcij, ki opisujejo ne le koristnost prehodov med koraki plana, temveč tudi koristnost dodatnih lastnosti vhodnega zaporedja $\mathbf{x}$.

Definirajmo število sumljivih dogodkov v sledi dogodkov $\mathbf{x}^{(k)}$ kot

$$\eta_s(\mathbf{x}^{(k)}) = \sum_{t=1}^{k} \Delta(x_t), \quad (4)$$

$$\Delta(x_t) = \begin{cases} 1; & \text{če } s'(x_t) \geq \tau' \\ 0; & \text{sicer} \end{cases}, \quad (5)$$

$$s'(x_t) = \frac{\lambda \cdot s(x_t)}{\lambda \cdot s(x_t) + (1-\lambda) \cdot n(x_t)}. \quad (6)$$

λ je apriorna verjetnost, da je prožilni dogodek x_t sumljiv (λ lahko ocenimo iz podatkov ali privzamemo $\lambda = 0.5$). Podobno definirajmo $\eta_n(\mathbf{x}^{(k)}) = k - \eta_s(\mathbf{x}^{(k)})$ kot število običajnih dogodkov v izbrani sledi $\mathbf{x}^{(k)}$.

Predpostavimo, da smo opazili vhodno sled dogodkov $\mathbf{x}^{(k)}$, ki vsebuje le sumljive prožilne dogodke ($\forall t, t = 1, ..., k : \Delta(x_t) = 1$). Naše prepričanje, da celotna sled pripada sumljivemu posamezniku, narašča z vsakim dogodkom x_t, npr., če je naše prepričanje ob prvem dogodku w, potem ob drugem sumljivem dogodku naraste za več kot $2w$, kar pomeni, da se evalvacija drugega dogodka spremeni glede na zgodovino obnašanja posameznika v preteklosti. To lahko modeliramo z eksponentno funkcijo. Po drugi strani, če opazimo zaporedne dogodke $\Delta(x_t) = 1$, naše prepričanje, da sled pripada običajnemu posamezniku, upada z vsakim novim dogodkom (saj so običajni dogodki sami po sebi dokaj pogosti, zato vsak dodaten dokaz, da je obnašanje običajno, nima enakega doprinosa), kar lahko opišemo z eksponentnim upadanjem. To opažanje lahko zapišemo s funkcijo koristnosti, ki je rekurzivno definirana nad vhodno sledjo dogodkov na naslednji način:

$$\begin{aligned} u_e(x_t, \mathbf{x}^{(t-1)}) &= a_t \cdot (f_e(\mathbf{x}^{(t-1)}) + b_t), \\ u_e(\mathbf{x}^{(0)}) &= 0, \\ b_t &= \beta \cdot \eta_s(\mathbf{x}^{(t)})^{\alpha(\tilde{s}(x_t) - \tilde{\tau})}, \\ a_t &= e^{-(\delta + \eta_n^*(\mathbf{x}^{(t)}))/(\gamma \cdot \eta_s(\mathbf{x}^{(t)}))}. \end{aligned} \quad (7)$$

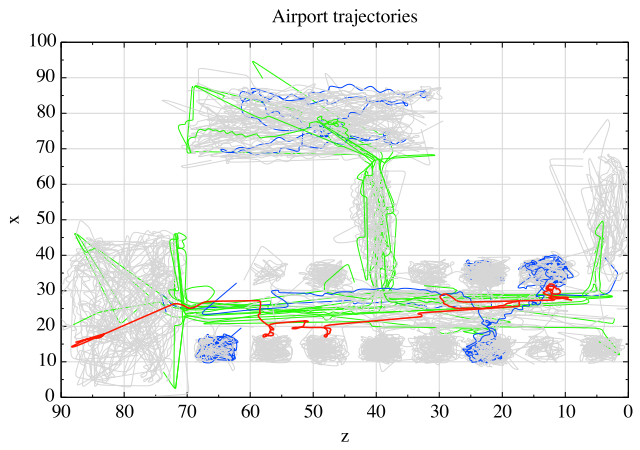
Člen b_t opisuje zgoraj omenjeno opažanje z eksponentno funkcijo, ki uporablja število sumljivih dogodkov v sledi $\eta_s(\mathbf{x}^{(k)})$ kot eksponent. Člen a_t opisuje *mehanizem pozabljanja* oz. eksponentno upadanje koristnosti s številom običajnih dogodkov v sledi. Konstante α, β, γ in δ se določi iz podatkov.

5 EKSPERIMENTI

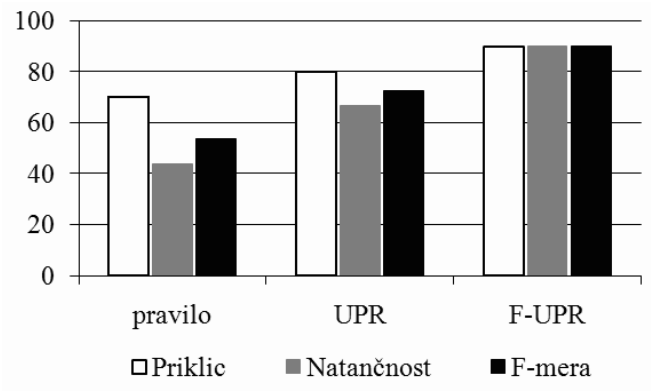
Predlagani pristop predstavimo na simulaciji letaliških potnikov na terminalu. V sodelovanju z domenskimi eksperti iz področja varnosti na letališču smo definirali scenarij, kjer sumljivi potnik skuša doseči točko B iz točke A, pri čemer se skuša izogniti stiku z varnostnim osebjem. Mogoče se zdi, da se bo potnik, ki nima dobrih namenov, skušal obnašati karseda normalno. To mogoče drži za posebej izurjene posameznike, a povprečen posameznik, izpostavljen visoki stopnji stresa, kaže znake strahu, napetosti, živčnosti itd[1]. Potemtakem je varno privzeti, da sumljivi potnik skuša minimizirati število kontaktov z varnostnim osebjem.

Za izvedbo eksperimenta smo uporabili sistem ESCAPES [6], simulator za evakuacijo potnikov na letališču, ki modelira številne tipe potnikov (družine, osebje, potniki, nakupovalci, ...). Simulacijo smo pognali z določeno mapo letališkega terminala, varnostnim osebjem, običajnimi potniki in enim sumljivim potnikom. Kot izhod smo dobili 2-D sledi akcij vseh agentov. Primer je prikazan na sliki 3, kjer so sledi varnostnega osebja prikazane z zeleno, sledi sumljivega potnika z rdečo in sledi običajnih potnikov z modro in sivo barvo. Simulator smo zagnali s 100 agenti, od tega je 10 agentov

[1] Eden izmed pristopov za detekcijo sumljivega obnašanja preko daljših časovnih intervalov temelji na opazovanju letaliških potnikov s specializiranim varnostnim osebjem (*behavior detection officers*), ki patruljira po letališču. Njihov cilj je odkriti potnike, ki kažejo znake živčnosti in napetosti. Ameriška agencija za promet (US Transportation Security Administration, www.tsa.gov) je izurila preko 5000 takšnih policistov, ki so nameščeni na preko 160 letališč širom ZDA.



Slika 3: Sledi vseh agentov na letališkem terminalu.



Slika 4: Rezultati evalvacije prikazujejo priklic, natančnost in F-mero za izbrane tri metode.

imelo vlogo varnostnega osebja in en agent vlogo sumljivega potnika. Pognali smo 20 simulacij, vsaka je vsebovala sled akcij z dolžino 1500-3000 akcij. Skupno število sledi akcij je bilo 2000, število interakcij med potniki in varnostnim osebjem pa 4316.

Za evalvacijo smo uporabili *natančnost*, *priklic* in *F-mero*. Natančnost pr (*precision*) je definirana kot kvocient števila resnično pozitivnih primerov (število vseh sumljivih potnikov pravilno označenih kot sumljivi) in števila vseh primerov označenih kot pozitivni (število vseh potnikov, ki jih označimo za sumljive). Idealna natančnost ima vrednost $pr = 1$, kar pomeni, da so vsi sumljivi potniki pravilno označeni kot sumljivi. Vrednost $1 - pr$ tako predstavlja odstotek *lažnih alarmov*. Priklic re (*recall*) je definiran kot kvocient števila vseh resnično pozitivnih primerov in števila vseh sumljivih potnikov. Idealni rezultat $re = 1$ pomeni, da so bili odkriti vsi sumljivi potniki. F-mera je harmonična sredina natančnosti in priklica.

Primerjali smo tri metode: (1) enostavno pravilo, ki za dano sled dogodkov $\mathbf{x}^{(t)}$ in izbrani parameter k pravi če $\exists k : \eta_s(\mathbf{x}^{(t)}) \geq k$, potem označi sled $\mathbf{x}^{(t)}$ kot sumljivo; (2) osnovno verzijo algoritma za razpoznavanje planov z uporabo koristnosti (UPR); in (3) razširjeno verzijo algoritma UPR, ki uporablja funkcijo koristnosti (F-UPR). Rezultate, prikazane na sliki 4, smo pridobili z 10-kratnim prečnim preverjanjem. Enostavno pravilo pričakovano dosega nižji priklic z visoko stopnjo lažnih alarmov, saj predstavlja navadno štetje sumljivih dogodkov v sledi. Algoritem UPR dosega precej boljši priklic in natančnost, saj dogodkom dodeli različne uteži glede na tip. Kot najboljša se izkaže predlagana metoda F-UPR, saj uteži dinamično prilagaja glede na obnašanje posameznika v preteklosti.

6 ZAKLJUČEK

V prispevku smo obravnavali problem detekcije nenavadnega obnašanja iz zaporedja dogodkov, ki sami po sebi niso dovolj, da bi lahko obnašanje nekega posameznika označili kot sumljivo. Predstavili smo metodo, ki v prvem koraku detektira prožilne dogodke in jim dodeli verjetnost, da so sumljivi oz. normalni. V drugem koraku predlagamo razširitev metode UPR, ki oceni zaporedje prožilnih dogodkov glede na obnašanje posameznika v preteklosti.

Metodo smo demonstrirali na simulaciji letališkega terminala, kjer je bil cilj zaznati sumljivega potnika, ki se skuša izmakniti varnostnemu osebju pri premiku iz enega dela terminala v drugi. Predlagana metoda je dosegla boljše rezultate tako od pravila, ki šteje število nenavadnih prožilnih dogodkov, kot tudi od osnovne verzije algoritma UPR.

Literatura

[1] Dorit Avrahami-Zilberbrand. *Efficient Hybrid Algorithms for Plan Recognition and Detection of Suspicious and Anomalous Behavior*. Doktorska disertacija, Bar-Ilan University, 2009.

[2] Dorit Avrahami-Zilberbrand in Gal A. Kaminka. Incorporating observer biases in keyhole plan recognition (efficiently!). V zborniku *Proceedings of AAAI-07*, 2007.

[3] Matthew Brand, Nuria Oliver in Alex Pentland. Coupled Hidden markov models for complex action recognition. V zborniku *CVPR'97*, str. 994 – 999, 1997.

[4] Somboon Hongeng in Ramakant Nevatia. Large-scale event detection using semi-hidden Markov models. V zborniku *IEEE International Conference on Computer Vision*, str. 1455–1462, August 2003.

[5] M. Naylor in C. I. Attwood. Advisor: Annotated digital video for intelligent surveillance and optimised retrieval, 2003. Final report.

[6] Jason Tsai, Gal Kaminka, Shira Epstein, Avishay Zilka, Inbal Rika, Xuezhi Wang, Andrew Ogden, Matthew Brown, Natalie Fridman, Matthew Taylor, Emma Bowring, Stacy Marsella, Milind Tambe in Ankur Sheel. ESCAPES - Evacuation Simulation with Children, Authorities, Parents, Emotions, and Social comparison. V zborniku *AAMAS-2011*, 2011.

[7] Namrata Vaswani, Amit Roy Chowdhury in Rama Chellappa. Shape Activity": A Continuous State HMM for Moving/Deforming Shapes with Application to Abnormal Activity Detection. V zborniku *IEEE Transactions on Image Processing*, str. 1603 – 1616, 2005.

SAMODEJNO RAZVRŠČANJE VPRAŠANJ GLEDE NA VRSTO ODGOVORA

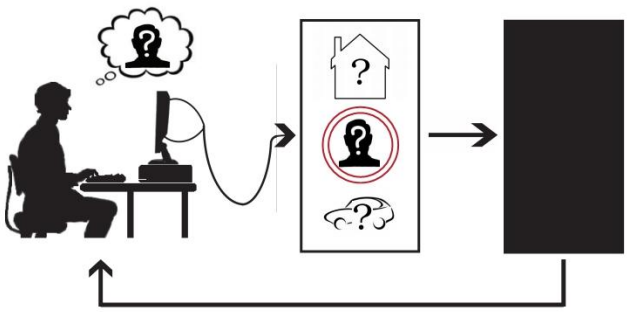
Tomaž Kompara, Domen Marinčič, Matjaž Gams
Odsek za Inteligentne sisteme
Institu Jožef Stefan
Jamova cesta 39, 1000 Ljubljana, Slovenija
Tel: +386 1 4773419; fax: +386 1 4251038
e-naslovi: {tomaz.kompara, domen.marincic, matjaz.gams}@ijs.si

POVZETEK

V zadnjem času se vse več sistemov za odgovarjanje na vprašanja (OV) uporablja tudi v komercialne namene, kjer se zahteva visok odstotek pravilno odgovorjenih vprašanj. V ta namen je razvitih več tipov sistemov za OV, ki imajo določene prednosti, kot tudi slabosti in so prilagojeni za uporabo v različnih domenah.
Če želimo pravilno odgovoriti na vprašanje, moramo ugotoviti kakšno vrsto odgovora želi uporabnik. V ta namen smo razvili samodejni razvrščevalnik vprašanj, ki določi vrsto pričakovanega odgovora. Razvrščevalnik deluje v domeni računalništva za slovenščino. Z njim lahko ločimo vprašanja, kar omogoča kakovostnejše odgovarjanje, saj lahko posebej obravnavamo vsako skupino vprašanj in podamo zelo natančen odgovor.
Razvrščevalnik smo naučili na učni množici s 1000 primeri, najboljša dosežena točnost razvrščevanja je 87,6%.

1 UVOD

Pri iskanju informacij v spletu običajno uporabljamo spletne iskalnike. Uspešnost iskanja je v veliki meri odvisna od primerne izbire ključnih besed, kar od uporabnika zahteva določeno mero spretnosti. Alternativa so sistemi za odgovarjanje na vprašanja (sistemi OV), kjer uporabnik vprašanje zastavi v naravnem jeziku. Ti sistemi postajajo vedno pomembnejši del informacijskih tehnologij, saj uporabnikom omogočajo pridobivanje informacij na naraven način brez ročnega pregledovanja množice spletnih strani [1, stran 561-563].
Sistemi OV so običajno sestavljeni iz treh delov [1, stran 567-568]. V prvem delu sistem sprejme vprašanje in tega analizira ter obdela (ang. Question Processing). Tako dobimo računalniku razumljivo obliko vprašanja. Nato s pridobljenimi podatki preiščemo bazo (ang. Document Processing), oziroma le del baze, odvisno od algoritma in iskanega odgovora. V zadnjem modulu pa pridobimo in sestavimo odgovor (ang. Answer Extraction and formulation) tako, da je primeren in razumljiv za uporabnika.
V tem članku smo se osredotočili na analizo vprašanj in sicer na določitev vrste pričakovanega odgovora. S tem podatkom si v naslednjih korakih pomagamo pri iskanju, kar vpliva tako na hitrost kot tudi na kvaliteto odgovora.
V ta namen je bilo razvitih že veliko razvrščevalnikov (ang. Classifiers) [2, 3], vendar ti niso prilagojeni slovenskemu jeziku. V tem članku je opisan razvoj razvrščevalnika vprašanj po vrsti odgovora, ki je posebej prilagojen za slovenski jezik in deluje za domeno računalništva.



Slika 1: *Razvrščevalnik vprašanj po vrsti odgovora*

2 NAČRTOVANJE RAZVRŠČEVALNIKA

V tem poglavju bomo opisali izbiro razredov, značilk in učnega algoritma. Pri načrtovanju smo upoštevali lastnosti in zakonitosti slovenskega jezika ter tako razvrščevalnik prilagodili slovenskemu jeziku.

2.1 Množica vprašanj

Za analizo in učenje ter testiranje smo potrebovali množico vprašanj. Te smo pridobili z avtomatskim iskanjem povedi, ki se zaključijo z vprašajem. Dolga in kompleksna vprašanja smo odstranili s pogojem, da je največja dolžina vprašanj 10 besed.
Iskanje po referenčnem slovenskem korpusu FidaPLUS [4] ni dalo primernih vprašanj za učenje. Primer takih vprašanj: ”Kaj pa hrana?”, ”Premalo prostora?”, ”Preveč časa v službi?”, itd.
Nato smo poskušali z iskanjem po bolj tehnično usmerjenemu besedilu in sicer po korpusu iFpX, ki vsebuje besedila, ki se nanašajo na računalništvo. Pridobili smo

približno 10.000 vprašanj, od katerih smo jih ročno izbrali 500.
Ker z uporabljenima korpusoma nismo zbrali dovolj vzorcev za izpeljavo učenja, smo iskanje nadaljevali na svetovnem spletu. Osredotočili smo se predvsem na vprašanja, ki jih ljudje pogosto zastavljajo. S pomočjo spletnega pajka smo preiskali preko 400 slovenskih spletnih strani, ki jih dobimo kot rezultat v spletnem iskalniku Google [7] na iskalni niz 'pogosta vprašanja'. Na ta način nam je uspelo zbrati še približno 4.000 različnih vprašanj, ki so bila primerna za učenje razvrščevalnika. Večina teh vprašanj se je nanašala na temo računalništva.

2.2 Razredi

Ker je razvrščevalnik namenjen sistemom za OV, morajo biti tudi razredi temu prilagojeni. Ker se teme odgovorov v različnih jezikih ne razlikujejo, smo si pri izbiri razredov pomagali s tujo literature [2]. Avtorja članka sta odgovore razdelila v 6 glavnih razredov in skupno 50 podrazredov.
Pri izbiri teme smo se zgledovali po glavnih šestih razredih, ki sta jih avtorja uporabljala za razvrščanje, pri čemer smo en razred izločili in dodali še dva nova. S tem smo zajeli vse vrste odgovorov in razrede prilagodili ciljem našega razvrščevalnika. Končni razredi so navedeni in opisani v tabeli 2.1.

Tabela 2.1: *Razredi in njihovi opisi*

Razred	**Opis**
Opis	Defnicije, opisi, vzroki, načini, itd.
Entiteta	Živali, pojmi, predmeti, rastline, itd.
Količina	Velikosti, količine, denar, čas, datum, itd.
Kraj	Kraji, države, mesta, gore, itd.
Oseba	Ljudje, skupine ljudi.
Da oz. ne	Vprašanja, na katera lahko odgovorimo z da oziroma ne.
Ostalo	Vsa vprašanja, ki ne spadajo v nobenega izmed zgornjih razredov.

2.3 Značilke

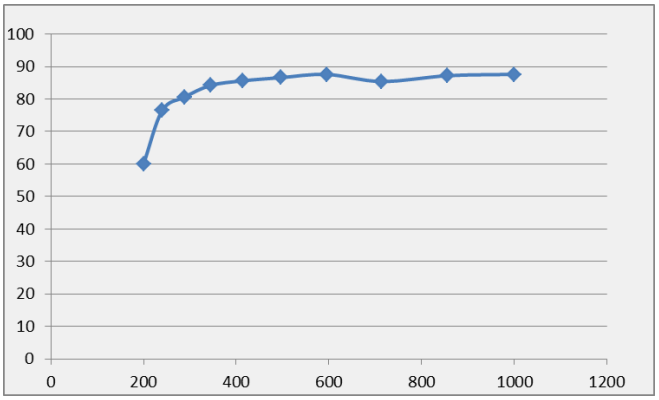
Da bi pridobili čim več podatkov o vprašanju, tega najprej slogovno označimo in lematiziramo. Za ta korak smo uporabili slogovni označevalnik in lematizrator PosTaggerTag, ki je prosto dostopen na spletu [6].
Pri izpeljavi značilk smo uporabili leme, besedne vrste in sklone. Ker je za pomen vprašanja pomemben tudi zaporedje besed, smo v model značilk vključili tudi bigrame (dve zaporedni značilki) ter dodali značilko, ki označuje število besed v vprašanju. Na koncu smo izbrali najprimernejše značilke in tako pridobili nekaj več kot 2.000 značilk.
V modelu značilk smo posebej obravnavali prvi unigram in bigram v vprašanju, saj ta nosi največ informacije o temi odgovora, kar smo tudi pokazali z preizkusom vpliva izbire značilk na točnost razvrščanja.

3 REZULTATI

Po končanem postopku učenja smo ocenili točnost razvrščanja s pomočjo neodvisne testne množice. V nadaljevanju je predstavljen vpliv velikosti učne množice, izbira značilk in algoritma na razvrščanje.

3.1 Točnost glede na število učnih primerov

S prvim poskusom smo ugotovili, kako na točnost razvrščanja vpliva število učnih vzorcev.



Slika 3.1: *Točnost razvrščevanja (navpična os v %) glede na število učnih primerov (vodoravna os)*

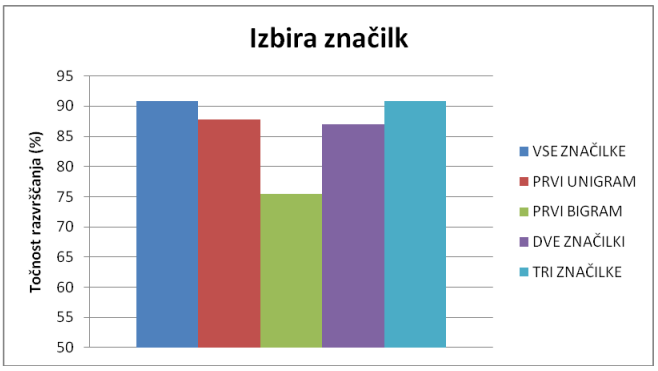
Začetna množica je vsebovala 200 učnih primerov, končna, največja množica 1000 primerov. Opazimo lahko, da od velikost množice približno 500 primerov naprej ni velike pridobitve točnosti.

3.2 Izbira značilk

Primerjali smo, kako izbira značilk vpliva na točnost razvrščanja. Z vsakim naborom značilk smo naučili razvrščevalnik z celotno učno množico. Pri vseh preizkusih smo uporabili enak algoritem učenja in sicer AdaBoostM1 in J48. Za primerjavo smo izbrali pet naborov značilk:

- Vse značilke
- Prvi bigram
- Prvi unigram
- Dve značilki: prvi unigram, prvi bigram
- Tri značilke: prvi unigram, prvi bigram, dolžina

Rezultati so prikazani na sliki 3.2. Razvrščevalnik, ki je za učenje uporabljal le prvi unigram, dosega zelo dobre rezultate, kar dokazuje, da ta nosi največ informacij o vrsti odgovora.

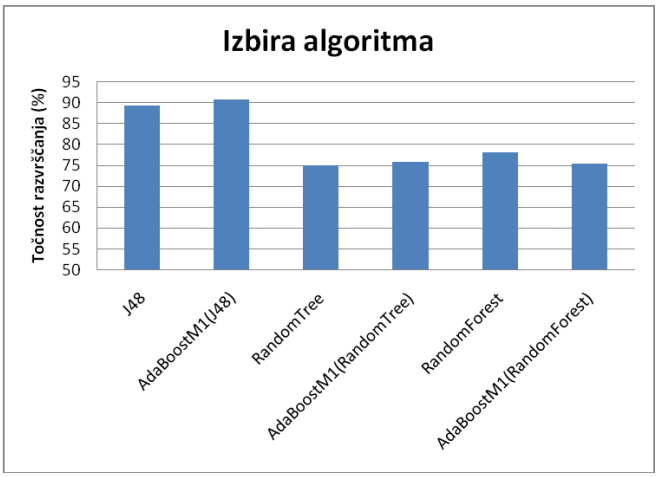


Slika 3.2: *Primerjava točnosti razvrščanja glede na izbiro značilk.*

Najslabše rezultate dobimo z ravrščevalnikom, ki je bil naučen le na podlagi prvega bigrama. Razlog za to je število učnih vzorcev, saj pri učenju zajamemo le manjši del začetnih bigramov. Opazimo lahko tudi da z uporabo treh značilk dosežemo enake rezultate kot z uporabo vseh.

3.3 Primerjava učnih algoritmov

V tem odstavku bomo pokazali, kako na točnost razvrščanja, vpliva izbira algoritma. Zaradi sistemske zahtevnosti, smo se odločili uporabiti nabor treh značilk (prvi unigrami, prvi bigram in dolžina). Pri vseh primerih smo za učenje uporabili celotno učno množico. Rezultati so prikazani na sliki 3.3.



Slika 3.3: *Primerjava učnih algoritmov*

Najboljše rezultate smo dobili prav z uporabljenim algoritmom AdaBoostM1 in J48. Zelo dobro se je odrezal tudi sam algoritem J48, medtem ko so bili ostali veliko slabši. V dveh od treh meritev z uporabo algoritma AdaBoost se je izkazalo, da z njim izboljšamo točnost razvrščanja, v primerjavi z osnovnim algoritmom.

4 ZAKLJUČEK

Razvili smo razvrščevalnik vprašanj, glede na vrsto pričakovanega odgovora, namenjen uporabi v sistemih OV.

Izmerili smo kako na točnost razvrščanja vplivajo število učnih primerov in izbira značilk in izbira učnega algoritma. Primerjali smo kako na točnost razvrščanja vpliva izbira učnega algoritma. Najboljše rezultate smo dobili z uporabo algoritma AdaBoostM1 skupaj z osnovnim razvrščevalnikom J48.

V nadaljnjem delu je potrebno preveriti nove značilke in algoritme, s katerimi bi bilo mogoče še izboljšati dobljene rezultate. Potrebno je preizkusiti, kako na točnost razvrščanja vplivajo napake lematizatorja in slogovnega označevalnika.

Viri

[1] Ruslan Mitkov. The Oxford handbook of computational linguistics. Oxford handbook. Oxford University Press, 2003.

[2] Xin Li, Dan Roth. *Learning question classifiers. Proceedings of the 19th international conference on Computational linguistics - Volume 1*, COLING'02, str. 1–7, Stroudsburg, PA, USA, 2002. Association for Computational Linguistics.

[3] Zhiheng Huang, Marcus Thint, Asli Celikyilmaz. Investigation of question classifier in question answering. *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2 - Volume 2*, EMNLP '09, str. 543–550, Stroudsburg, PA, USA, 2009. Association for Computational Linguistics.

[4], Špela Arhar Holdt, Vojko Gorjanc, Simon Krek. FidaPLUS corpus of Slovenian : the new generation of the Slovenian reference corpus: its design and tools. *V: DAVIES, Matthew (ur.). Proceedings of the Corpus Linguistics Conference, CL2007,* University of Birmingham, UK, 27-30 Julij 2007.

[5] Špela Vintar, Tomaž Erjavec. iKorpus in luščenje izrazja za Islovar. *Zbornik Šeste konference Jezikovne tehnologije, 16. do 17. oktober 2008, Ljubljana, Slovenija : zbornik 11. mednarodne multikonference Informacijska družba - IS 2008.*

[6] *PosTagerTag*, Avgust 2011. Dostopno na: http://oznacevalnik.slovenscina.eu/Vsebine/Sl/ProgramskaOprema/Oblikoslovni.aspx.

[7] *Google*, Avgust 2011. Dostopno na: http://www.google.si/.

GAIT RECOGNITION WITH INERTIAL SENSORS

Simon Kozina
Department of Intelligent Systems
Jozef Stefan Institute
Jamova 39, 1000 Ljubljana, Slovenia
Tel: +386 1 4773195; fax: +386 1 4251038
e-mail: simon.kozina@ijs.si

ABSTRACT

Due to the rapid aging of the European population, an effort needs to be made to ensure that the elderly can live longer independently with minimal support of the working-age population. In ambient intelligence there are many new approaches under development to address this problem. One of the possible solutions are intelligent systems which are constantly monitoring the health of the elderly people and inform doctors and relatives, if some major changes occur to their health condition. In such systems, where there is more than one user, the information about a current user is very important. Using their walking signature, we can distinguish between users to find out which user is currently using the system. Four machine learning algorithms were compared and different placements of the sensors were considered. At each sensors placement, different inertial sensors combinations were evaluated. The highest classification accuracy of 100% is achieved by Support Vector Machine classifier using accelerometer sensors placed on the right thigh.

1 INTRODUCTION

The European population is aging due to the increase in life expectancy and decrease in birth rate. The percentage of population aged over 65 years is anticipated to rise from 17.9% in 2007 to 53.5% in 2060 [1]. As a consequence, the number of the elderly will exceed the society's capacity for taking care of them. Thus an effort needs to be made to ensure that the elderly can live longer independently with minimal support of the working-age population. This is currently one of the main research problems in the area of ambient intelligence. Several intelligent systems have been designed to address this problem: Confidence [2], Chiron [3], etc.

Information about current user/users of the system is important, since the performance of such systems is improved, when the system is adapted to a specific user [4]. Our goal is to maximize the performance of distinguishing between users with inertial sensors when trying to minimize the number of sensor locations and inertial sensors.

Gait recognition field is divided into two subfields on the basis of the sensors that are used to identify a walking signature of a user. First subfield uses motion capture from cameras to detect human gait and has been researched extensively. Hayfron-Acquah et al. [5] were recognized gait of four people with accuracy of 100% and of six people with accuracy of 97%. Lam et al. [6] tried to recognize gait of a large number of people, 115. To create a learning model, they have used 2128 walking sequences and achieved accuracy of more than 80%. Second subfield of gait recognition uses inertial sensors, such as accelerometers, gyroscopes and magnetometers. It has become an important area of research today. First researches on this field were conducted by Ailisto et al. [7] and Mäntyjävri et al. [8]. Both used one accelerometer attached to a person's hip. Testing their methods on gait signature of 36 persons, their classification accuracy was 85%.

The paper is structured as follows. Section 2 explains data acquisition. Section 3 lists attributes extracted from the input data that are fed into the machine learning algorithms. Section 4 presents the experiments. Finally, Section 5 concludes the paper and presents some ideas for future work.

2 DATA ACQUISITION

The goal of our research was to distinguish among users on the basis of their gait signature. Equipment used in our experiments consisted of ten wearable Xsens-MTx sensors [9] and X-bus Mater unit. The Xsens-MTx is a complete MEMS inertial measurement unit (IMU) with integrated 3-axis accelerometer, 3-axis gyroscope and 3-axis magnetometer. An accelerometer measures acceleration, a gyroscope orientation and a magnetometer measures the strength of a magnetic field. Xsens hardware is capable of acquiring the data with a frequency of 50 Hz.

The data for our experiment consisted of recordings of five individuals. An individual recorded two scenarios consisted of walking with two different speeds: normal walking (5 km/h) and quick walking (6.5 km/h). Each scenario was recorded for four minutes.

In our experiments, we used ten Xsens sensors, placed at: (0) left ankle, (1) right ankle, (2) left knee, (3) right knee,

(4) waist, (5) left wrist, (6) left upper arm, (7) chest, (8) right upper arm and (9) right wrist, as shown on Figure 1.

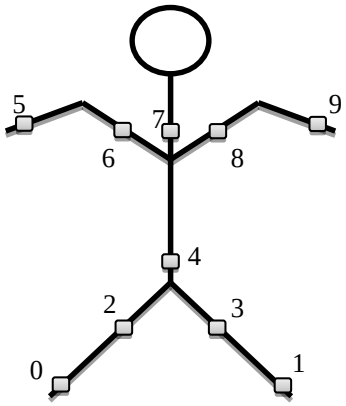


Figure 1:Xsens sensor placement.

3 FEATURE EXTRACTION

Features were extracted from raw sensor data using sliding time window. Data in one time window are used to compute one feature vector. Three settings for the length of the sliding window were used: 1 second, 2 second and 4 second time window.

As stated above, data from ten sensors was gathered. As we want to find the minimal number of sensors to achieve the highest accuracy, experiments were only conducted using one sensor and a combination of one or more inertial sensors at a time. The features described below were derived for one sensor and one type of inertial data. Therefore, if we are using more than one sensor or more than one type of inertial data, we have to compute several feature vectors and then concatenate them. One feature vector includes the following attributes:

- The average value along the x, y and z axes and the average length of the data vector ($\bar{x}, \bar{y}, \bar{z}, \bar{v}$).
- The variance of the values along x, y and z axes and the variance of the length of the data vector ($\delta_x^2, \delta_y^2, \delta_z^2, \delta_v^2$).
- The maximum and the minimum values along the x, y and z axes and the maximum and the minimum length of the data vector ($M_x, M_y, M_z, M_v, m_x, m_y, m_z, m_v$).
- The difference between the maximum and the minimum value along the x, y and z axes and the difference between the maximum and the minimum length of the data vector ($diff_x, diff_y, diff_z, diff_v$).
- The angle of change in the acceleration between the maximum and the minimum value along the x, y and z axes and the length of the data vector ($spd_x, spd_y, spd_z, spd_v$):

$$spd_x = \tan^{-1} \frac{M_x - m_x}{t(M_x) - t(m_x)}$$

$t(M_x)$ and $t(m_x)$ are the times when the maximum and the minimum acceleration M_x and m_x were measured. Figure 2 shows how the angle of change is computed.

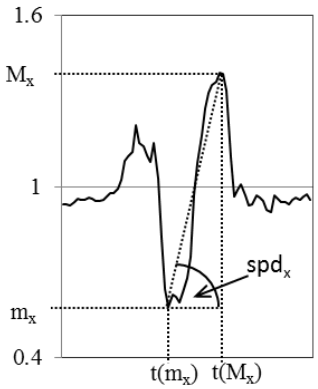


Figure 2: The angle of change in acceleration between the maximum and the minimum values along the x axis.

- The orientation of the inertial sensor along the x, y and z axes ($\varphi_x, \varphi_y, \varphi_z$):

$$\varphi_x = \cos^{-1}\left(\frac{\bar{x}}{\sqrt{\bar{x}^2 + \bar{y}^2 + \bar{z}^2}} \right)$$

- The index of dispersion along the x, y and z axes and the index of dispersion of the length of the data vector (d_x, d_y, d_z, d_v):

$$d_z = \frac{\delta_x^2}{\bar{x}}$$

- The sum of absolute differences between the consecutive lengths of the data vector (*s*):

$$s = \frac{\sum_{i=1}^{N-1} | v_{i+1} - v_i |}{t_N - t_1}$$

t_1 and t_N are the starting and the ending time of the window, N is the number of examples in the time window and v_i is the length of the data vector at data point *i* in the time window.

- The detection of movement of the inertial sensor:

$$moving = \begin{cases} 0; & d_v < \tau \\ 1; & otherwise \end{cases}$$

τ is set experimentally for each of the inertial sensors.

- The frequency of movement along the x, y and z axes. The frequency is computed as the index of the maximum element of Fast Fourier Transformation (fft_x, fft_y, fft_z).
- The energy of movement (W) along x, y and z axes is the squared sum of all elements of FFT of the corresponding axis (W_x, W_y, W_z):
  - $$W_x = \frac{\sum_{i=1}^{N} fft_{x_i}^2}{N}$$
- The correlation between all pairs of axes of the accelerometer (r_{xy}, r_{xz}, r_{yz}), computed as the Pearson product-moment correlation coefficient:

$$r_{xy} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}}$$

To sum it up, 42 features were computed for every sensor and for every inertial data.

4 RESULTS

In our experiments we have tested four machine learning algorithms to distinguish between persons: Random Forest algorithm, Support Vector Machine (SVM) algorithm, Naive Bayes algorithm and Decision Trees algorithm, all implemented in Weka machine learning suite [10]. For each machine learning algorithm, we have tried three settings for time window length: 1 second, 2 seconds and 4 seconds time window. Together, 12 settings for every sensor position.

Learning and testing sets are, as stated above, recordings from five individuals. Evaluation was done with 10-fold cross validation and the results for combinations at every position are presented in Table 1. In Table 1, only the data from accelerometers were used in evaluation.

First, we evaluated which time window setting is optimal. Comparing the results in Average II column, we concluded, that for every machine learning algorithm average value over all the sensor placements is highest at 4 second time window. Secondly, we determined that the best machine learning algorithm for gait recognition is SVM. Comparing the average values for four second time window in Average II column, it is clear that the SVM is the most successful. Finally, the best placement for sensor was evaluated. Taking into consideration the highest average value in Average I row, and one of the best results, if we take into account two of the previous findings, four second time window and SVM as machine learning algorithm, the sensor placement is at sensor number 3 – right knee.

Some other patterns can be observed from Table 1. If we focus only on the row, where machine learning algorithm is SVM and time window is set to four seconds, we concluded, that sensor placements on the waist and below (ankles and knees) are superior to the placements of sensor on upper body (chest, wrists and upper arms). This effect is also seen at other settings and confirmed by the results in Average I row.

		Sensor placement										
	Time window	0	1	2	3	4	5	6	7	8	9	Average II
Random Forest	1 s	93.51	96.38	97.25	98.25	98.63	92.13	94.26	92.63	93.26	91.14	94.74
	2 s	94.01	98.75	99	99.25	99.5	94.26	95.01	93.02	94.01	94.01	96.08
	4 s	96.02	98.01	97.01	98.01	100	95.02	96.02	96.52	95.52	96.52	96.87
SVM	1 s	96.38	97.88	99	98.75	98.5	91.89	95.26	92.26	92.63	89.51	95.21
	2 s	99.25	100	99.75	99.5	99.75	96.51	98	96.51	96.51	92.77	97.86
	4 s	100	100	100	**100**	100	96.52	97.51	97.01	97.01	95.02	98.31
Naive Bayes	1 s	83.9	88.26	91.26	94.01	89.76	73.41	84.02	77.28	87.27	74.16	84.33
	2 s	89.03	93.52	96.26	94.51	92.02	76.56	86.78	82.79	89.53	78.8	87.98
	4 s	93.53	97.51	94.53	97.51	94.03	81.09	88.56	82.09	94.03	83.08	90.6
Decision Tree	1 s	89.76	93.88	93.88	95.26	94.51	89.39	89.39	91.26	88.14	87.14	91.26
	2 s	88.78	97.51	95.76	96.51	95.01	90.02	89.78	90.52	91.27	88.03	92.32
	4 s	94.03	100	93.03	95.02	93.53	89.05	89.05	96.02	92.54	88.06	93.03
Average I		93.18	96.81	96.4	97.22	96.27	88.82	91.97	90.66	92.64	88.19	

Table 1: Classification accuracy for combinations of every sensor position, all machine learning algorithms and time window lengths are listed. Only accelerometers were used to compute these results.

All the findings stated above and the results in Table 1 are evaluated only from the accelerometer data. To compare all the combinations of inertial sensors, we have focused on only one combination from Table 1. We choose sensor number 3 – right knee and Naive Bayes classificator with 1 second time window. The reason why the best classificator and optimal time window was not selected is simple: we want to observe, if some other combination of inertial sensors could produce better results than only accelerometers. We could not observe this, if we were using the setting where classification accuracy is already 100%.

In Figure 3 different combinations of inertial sensors are shown. Initial value, which we want to improve, is with accelerometers only, 94.01%. The best improvement is when using accelerometer and gyroscope data, for 3.12 percentage points. When using only a gyroscope or a magnetometer data, classification accuracy drops compared to experiment with only accelerometer data. A magnetometer is apparently not the optimal inertial sensor for gait recognition. The results with magnetometer data included in a testing in a learning set are not as good as results without magnetometer.

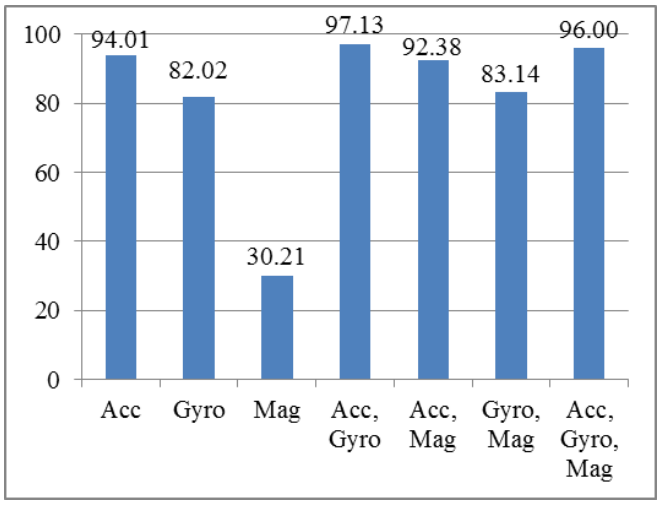


Figure 3: Classification accuracy with different combinations of inertial sensors, tested with Naive Bayes classificator and 1 second time window.

5 CONCLUSION

Gait recognition with inertial sensors is becoming an important part of ambient intelligence applications, since their performance increases, if there is information about current user.

We have presented a survey of different sensor placement, machine learning algorithms and time window settings to find an optimal sensor placement for gait recognition. As it turns out, the best sensor placement is on the waist or on knees or ankles. Support Vector Machine and 4 second time window should be used to achieve the best classification accuracy. If the inertial sensor, which we are using, also includes gyroscope data, it should be used, as it improves classification accuracy of gait recognition.

There are several directions for further work. First we should record more people to increase our sample. All the previous tests should be repeated in order to confirm that the current selected settings are still the best. Furthermore, gyroscope data should be included into the evaluation to confirm, that performance increases. Finally, combinations with more than one sensor should be used if the performance with only one sensor is not optimal.

References

[1] Eurostat. http://epp.eurostat.ec.europa.eu, 2011.

[2] Confidence. http://www.confidence-eu.org, 2011.

[3] Chiron. http://www.chiron-project.eu, 2011.

[4] B. Cvetkovič, M. Luštrek, B. Kaluža, M. Gams. Semi-supervised Learning for Adaptation of Human Activity Recognition Classifier to the User. *In Proc. Space, Time and Ambient Intelligence*, pp. 24-29. Barcelona. 2011.

[5] J. B. Hayfron-Acquah, M. S. Nixon and J. N. Carter. Automatic gait recognition by symmetry analysis. Pattern Recognition Letters, vol. 24, pp. 272–277, 2003.

[6] T. H. W. Lam and R. S. T. Lee, A new representation for human gait recognition: Motion silhouettes image (MSI*), in International Conference on Biometrics*, pp. 612–618, 2006

[7] H. J. Ailisto, M. Lindholm, J. Mäntyjärvi, E. Vildjiounaite, and S.-M. Mäkelä, Identifying people from gait pattern with accelerometers. *In Proceedings of SPIE*, 2005.

[8] J. Mäntyjärvi, M. Lindholm, E. Vildjiounaite, S.-M. Mäkelä, and H. J. Ailisto, Identifying users of portable devices from gait pattern with accelerometers. *In IEEE International Conference on Acoustics, Speech, and Signal Processing*, March 2005.

[9] Xsens. http://www.xsens.com/en/general/mtx, 2011.

[10] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H.Witten. The weka data mining software: An update. *SIGKDD Explorations*, 11, 2009.

STROJNO UČENJE EPITOPOV IZ PEPTIDNIH MIKROMREŽ

Mitja Luštrek
Odsek za inteligentne sisteme
Institut Jožef Stefan
Jamova cesta 39, 1000 Ljubljana, Slovenija
Telefon: +386 1 4773380; telefaks: +386 1 4773131
E-pošta: mitja.lustrek@ijs.si

POVZETEK

Epitopi so delci beljakovin, ki jih prepozna imunski sistem. Peptidne mreže omogočajo določanje velikega števila epitopov z laboratorijskimi poizkusi. Podatke, pridobljene s peptidnimi mrežami, smo uporabili za učenje klasifikatorja za prepoznavanje epitopov. Klasifikator je bil sestavljen iz ansambla osmih osnovnih klasifikatorjev, od katerih je vsak uporabljal svoj nabor atributov, združeval pa jih je meta-klasifikator. Z njim smo dosegli klasifikacijsko točnost 83,7 do 85,9 %, kar je za 0,9 do 2 odstotni točki bolje od metode, ki je sodeč po literaturi trenutno najboljša.

1 UVOD

Antigen je molekula, ki jo imunski sistem prepozna kot škodljivo, epitop pa je del antigena, ki je zaslužen za prepoznavo. Poznavanje epitopov je koristno iz vsej dveh razlogov. Mogoče jih je uporabljati v cepivih, saj izzovejo imunski odziv, na da bi bilo cepljeni osebi treba vbrizgati celoten antigen, ki utegne biti škodljiv. Poleg tega se lahko uporabljajo v diagnostične namene, saj odziv na različne epitope pokaže, katera protitelesa bolnik ima, iz tega pa je razvidno, katero bolezen ima.

Epitope je načeloma mogoče določiti z laboratorijskimi poizkusi, je pa strojno učenje kajpak cenejše; tudi če ne more povsem nadomestiti laboratorijskih poizkusov, lahko vsaj zoža nabor možnih epitopov. V našem delu smo uporabili podatke, pridobljene s peptidnimi mikromrežami, ki so povedali, kateri peptidi (delci beljakovine) vsebujejo epitope. Peptidne mikromreže omogočajo dokaj enostavno ugotavljanje imunskega odziva na veliko število peptidov. Iz teh podatkov smo s strojnim učenjem zgradili klasifikator za prepoznavanje epitopov.

Peptide, ki smo jih klasificirali, smo opisali z osmimi nabori atributov. Za vsak nabor smo izbrali algoritem za strojno učenje, ki je dal najboljši klasifikator na tistem naboru. Te klasifikatorje smo nato uporabili kot ansambel: njihove klasifikacije so služile kot atributi za meta-klasifikator, ki je dal končno odločitev o tem, ali peptid vsebuje epitop. Rezultate ansambla smo primerjali z metodo za prepoznavanje epitopov, ki je sodeč po literaturi trenutno najboljša.

2 PODATKI

Peptid je delec beljakovine, ki ga sestavljajo zaporedno vezane aminokisline. Večina peptidov v naših podatkih je bila sestavljena iz 15 aminokislin. Različnih standardnih aminokislin je 20, kar pomeni, da si vsak peptid lahko predstavljamo kot besedo, sestavljeno iz 15 znakov, pri čemer ima abeceda 20 različnih znakov. Peptidna mikromreža je ploščica, na katero se nanesejo vzorci peptidov. V našem delu je bil na mikromrežo nato nanesen IVIg, ki je mešanica protiteles zdravih darovalcev. Ta protitelesa so se vezala predvsem na peptide, ki so vsebovali epitope. Iz izmerjenega števila protiteles na vzorcu je bilo tako mogoče določiti verjetnost, da peptid vsebuje epitop.

Moč vezave protiteles, izražena s številom med 1 in 65.536, je bila izmerjena za 75.534 epitopov. Za strojno učenje smo uporabili le tiste, za katere smo bili gotovi, da res vsebujejo epitope (moč vezave nad 10.000) ali da jih res ne vsebujejo (moč vezave pod 1.000). Takih je bilo 27.278 in razdelili smo jih na učno in testno množico. Učna množica je bila sestavljena iz 3.420 peptidov z epitopi (pozitivnih) in 10.218 peptidov brez epitopov (negativnih). Testna množica je bila sestavljena iz 3.421 pozitivnih in 10.219 negativnih peptidov.

Učna in testna množica sta vsebovali trikrat več negativnih kot pozitivnih peptidov. Tako neuravnoteženi podatki lahko kvarno vplivajo na strojno učenje, zato smo uporabili dve metodi za uravnoteženje: nad- in podvzorčenje [6]. Prva metoda naredi kopije naključno izbranih primerov manj številčnega razreda (v našem primeru pozitivnih peptidov), druga pa odstrani primere bolj številčnega razreda (negativnih peptidov). Pri učenju končnega klasifikatorja smo uporabili nadvzorčenje (oversampling), ker nismo hoteli zavreči nobenih podatkov. Pri primerjanju algoritmov za strojno učenje in nastavljanju parametrov pa smo uporabili podvzorčenje (undersampling), saj da manj podatkov in je zato učenje hitrejše, poleg tega pa nadvzorčenje povzroča težave pri algoritmih za strojno učenje, ki uporabljajo notranje prečno preverjanje (saj se enak primer lahko znajde v notranje učni in testni množici).

3 POSTOPEK STROJNEGA UČENJA

Učenje ansambla klasifikatorjev za prepoznavanje epitopov je potekalo v dveh korakih. V prvem koraku smo vsak peptid p_i opisali z osmimi različnimi atributnimi vektorji $a_1(p_i)$, ..., $a_8(p_i)$, ki so opisani v podrazdelku 3.2. Na teh vektorjih smo naučili osem osnovnih klasifikatorjev C_1, ..., C_8 (vsakega z najboljšim algoritmom za strojno učenje – več o tem v podrazdelku 3.1), od katerih je vsak vrnil verjetnost, da peptid p_i vsebuje epitop. Te verjetnosti označimo s $P_1(p_i)$, ..., $P_8(p_i)$.

V drugem koraku smo iz verjetnosti, pridobljenih v prvem koraku, tvorili atributni vektor $a_M(p_i) = [P_1(p_i), ..., P_8(p_i)]$. Pomagali smo si s petkratnim prečnim preverjanjem: osnovne klasifikatorje C_1, ..., C_8 smo naučili na štirih petinah učne množice, nakar smo z njimi klasificirali peto petino in tako zanjo dobili a_M. To smo ponovili petkrat, pri čemer smo vsakič klasificirali drugo petino, kar nam je dalo a_M za celotno učno množico. Na koncu smo na atributnih vektorjih a_M naučili meta-klasifikator C_M, ki je vrnil končno verjetnost, da peptid vsebuje epitop.

Testiranje ansambla klasifikatorjev je potekalo podobno kot učenje. Vsak peptid iz testne množice q_i smo predstavili z osmimi atributnimi vektorji $a_1(q_i)$, ..., $a_8(q_i)$ in ga klasificirali z osmimi osnovnimi klasifikatorji C_1, ..., C_8. Njihove klasifikacije smo združili v nov atributni vektor $a_M(q_i)$, ki smo ga klasificirali z meta-klasifikatorjem C_M.

3.1 Izbira algoritmov in nastavljanje parametrov

Naš ansambel klasifikatorjev je uporabljal več algoritmov za strojno učenje, od katerih ima vsak svoje parametre. Poleg tega smo lahko na različne načine izračunali tudi vsakega od atributnih vektorjev, kar je bilo določeno z atributnimi parametri, opisanimi v podrazdelku 3.2. Celoten prostor algoritmov in parametrov je bil prevelik za izčrpno preiskovanje, zato smo jih izbrali in nastavili v šestih zaporednih korakih, izvedenih na učni množici.

V prvem koraku smo preizkusili 41 algoritmov iz orodja Weka [4] na atributnih vektorjih, sestavljenih iz frekvenc aminokislin v peptidu. Pri vseh algoritmih smo uporabili privzete vrednosti parametrov. Izbrali smo osem najboljših algoritmov z metodo podpornih vektorjev (SVM) na čelu. V drugem koraku smo z SVM določili začetne vrednosti atributnih parametrov za vseh osem atributnih vektorjev.

V tretjem koraku smo primerjali osem algoritmov iz prvega koraka na atributnih vektorjih z začetnimi vrednostmi parametrov iz drugega koraka. Pri tem smo nabor najboljših algoritmov zožali na tri: SVM, logistično regresijo in klasifikacijo z regresijo [3]. V četrtem koraku smo te tri algoritme preizkusili z vsemi smiselnimi vrednostmi atributnih parametrov. Algoritme smo uporabili same zase, vse tri združene s skladanjem klasifikatorjev in boljša dva (SVM in logistična regresija) združena s skladanjem. Tako smo dobili najboljši algoritem in najboljše atributne parametre za vsak atributni vektor. V petem koraku smo za vsak atributni vektor določili najbolje parametre za algoritem, izbran v četrtem koraku. Izbrani algoritmi, njihovi parametri in atributni parametri so navedeni pri opisih atributnih vektorjev v podrazdelku 3.2.

V šestem koraku smo izbrali najboljši algoritem za meta-klasifikator in najboljše vrednosti parametrov zanj. Ta algoritem je bil različica linearne regresije PACE [7] s privzetimi vrednostmi parametrov.

3.2 Atributi

V nadaljevanju je opisanih osem atributnih vektorjev, s katerimi smo predstavili vsak peptid.

Frekvence. Ta atributni vektor je sestavljen iz frekvenc aminokislin v peptidu. Peptid je razdeljen na p delov enake dolžine in frekvence se izračunajo za vsak del posebej. Vektor ima obliko:

$[A_1, C_1, ..., Z_1; ...; A_p, C_p, ..., Z_p]$.

A_i je frekvenca aminokisline A v i-tem delu peptida (vsaka aminokislina se označuje z eno črko, A pomeni alanin).

Najboljša vrednost $p = 3$. Najboljši algoritem za strojno učenje je bil SVM s $C = 1$ ter jedrom PUK z $\omega = 0.5$ in $\sigma = 2{,}5$.

Razlike frekvenc. Ta atributni vektor je sestavljen iz razlik med frekvencami aminokislin v peptidu. Te razlike so sicer razvidne že iz prejšnjega vektorja, vendar ni nujno, da jih algoritmi za strojno učenje znajo izkoristiti, če niso podane izrecno. Peptid je spet razdeljen na p delov enake dolžine. Klasifikacija je bila boljša, če smo razlikam v frekvencah dodali tudi same frekvence (velja tudi pri nekaterih naslednjih vektorjih), tako da ima vektor obliko:

$[A–C_1, ..., A–Z_1, ..., Z–Z_1; ...; A–C_p, ..., A–Z_p, ..., X–Z_p; A, ..., Z]$.

$A–C_i$ je razlika med frekvencama aminokislin A in C v i-tem delu peptida. A je frekvenca aminokisline A v celotnem peptidu.

Najboljša vrednost $p = 2$. Najboljši algoritem za strojno učenje je bil logistična regresija z $\lambda = 10^{-8}$.

Frekvence podzaporedij. Ta atributni vektor je sestavljen iz frekvenc podzaporedij v peptidu dolžin do l. Ker se malo podzaporedij ponavlja, podzaporedje definiramo ohlapno: vanje se sme vriniti do g aminokislin, ki podzaporedju ne pripadajo. Podzaporedje z vrinjenimi aminokislinami se ne šteje kot celo podzaporedje, ampak kot w^g podzaporedja, pri čemer $w \leq 1$. Upoštevamo le c najbolj pogostih podzaporedij vsake dolžine. Vektor ima obliko:

$[S_{11}, S_{12}, ..., S_{1c}; S_{21}, S_{22}, ..., S_{2c}; ...; S_{l1}, S_{l2}, ..., S_{lc}]$.

S_{ij} je frekvenca j-tega najbolj pogostega podzaporedja dolžine i.

Najboljše vrednosti parametrov so bile $l = 5$, $g = 0$ in $c = 25$. Najboljši algoritem za strojno učenje je bil SVM s $C = 1$ ter polinomskim jedrom s $p = 2$.

Lastnosti aminokislin. Ta atributni vektor je sestavljen iz povprečnih vrednosti 19 fiziokemičnih lastnosti aminokislin, npr. kislosti, prožnosti in velikosti. Peptid je razdeljen na p delov enake dolžine in lastnosti so povprečene po vsakem delu. Vektor ima obliko:

[$prop_{1,1}$, $prop_{2,1}$, ..., $prop_{19,1}$; ...; $prop_{1,p}$, $prop_{2,p}$, ..., $prop_{19,p}$; A, ..., Z].

Vrednost $prop_{i,j}$ je povprečna vrednost lastnosti i v j-tem delu peptida.

Najboljša vrednost p = 2. Najboljši algoritem za strojno učenje je bil skladanje klasifikatorjev, zgrajenih z SVM in logistično regresijo. SVM je uporabljal C = 1 ter jedro PUK z ω = 1 in σ = 2. Logistična regresija je imela λ = 0.1.

Frekvence razredov. Ta atributni vektor je sestavljen iz frekvenc razredov aminokislin glede na njihove fiziokemične lastnosti. Tak razred je npr. razred kislih aminokislin. Zaradi združitve treh izmed lastnosti, uporabljenih v prejšnjem vektorju, aminokisline razvrstimo v razrede na 17 načinov. Vektor ima obliko:

[*freq* $(\text{prop}_1)_1$, ..., *freq* $(\text{prop}_{17})_1$; ...;
freq $(\text{prop}_1)_p$, ..., *freq* $(\text{prop}_{17})_p$; A, ..., Z]

freq $(\text{prop}_i)_j$ = [$prop_i$-low_j, $prop_i$-med_j, $prop_i$-$high_j$]

Vrednosti $prop_i$-low_j, $prop_i$-med_j in $prop_i$-$high_j$ so frekvence razredov aminokislin, razvrščenih po lastnosti i, v j-tem delu peptida.

Najboljša vrednost p = 3. Najboljši algoritem za strojno učenje je bil skladanje klasifikatorjev, zgrajenih z SVM in logistično regresijo. SVM je uporabljal C = 1 in linearno jedro. Logistična regresija je imela λ = 1.

Frekvence podzaporedij razredov. Ta atributni vektor je podoben frekvencam podzaporedij, le da so podzaporedja namesto iz posamičnih aminokislin sestavljena iz razredov, kakršni nastopajo v prejšnjem vektorju. Tako npr. podzaporedje "EADC" nadomestimo z "anan", kjer "a" pomeni kislo in "n" nevtralno aminokislino. Podzaporedje ima spet lahko dolžino do l, do g vrinjenih aminokislin in ima težo w^g. Upoštevamo le c najbolj pogostih zaporedij vsake dolžine. Vektorju smo dodali tudi frekvence navadnih podzaporedij, tako da ima obliko:

[*subseq* ($prop_1$), ..., *subseq* (prop_{17});
S_{11}, ..., S_{1c}; ...; S_{l1}, S_{l2}, ..., S_{lc}]

subseq ($prop_i$) = [$prop_{i11}$, ..., $prop_{i1c}$; ...; $prop_{i\,l1}$, ..., $prop_{i\,lc}$]

Vrednost $prop_{ijk}$ je frekvenca k-tega najbolj pogostega podzaporedja dolžine j, ko so aminokisline razvrščene v razrede po lastnosti i. S_{ij} je frekvenca j-tega najbolj pogostega podzaporedja aminokislin dolžine i.

Najboljše vrednosti parametrov so bile l = 2, g = 5, w = 0.5 in c = 25. Najboljši algoritem za strojno učenje je bil skladanje klasifikatorjev, zgrajenih z SVM in logistično regresijo. SVM je uporabljal C = 0.5 in linearno jedro. Logistična regresija je imela λ = 50.

Pari. Ta atributni vektor je sestavljen iz frekvenc parov aminokislin z določeno razdaljo med pripadnikoma para. Taka frekvenca je npr. frekvenca para (A, B) z razdaljo 3. Razlog za ta atributni vektor je, da se protitelesa na epitope utegnejo vezati na dveh mestih. Ker je takih atributov zelo veliko, njihovo število zmanjšamo na dva načina. Prvi je, da s sosednjih razdalj združimo v eno. Drugi pa je, da eno ali obe aminokislini v paru nadomestimo z razredom aminokislin. Vektor ima obliko:

[*pair* (A_1, A_2), ..., *pair* (A_1, A_{n2}); ...;
pair (A_{n1}, A_1), ..., *pair* (A_{n1}, A_{n2})]

pair (A_i, A_j) = [(A_i, A_j) z d_1, ..., (A_i, A_j) z $d_{\max}$].

A_i je i-ta aminokislina ali i-ti razred aminokislin, $n1$ in $n2$ sta števili različnih aminokislin ali razredov prvega in drugega pripadnika para, d_k je k-ta razdalja ali skupina razdalj med aminokislinama ali razredoma v paru in $d_{\max}$ je največja razdalja.

Najboljša izbira za prvega pripadnika para so bile aminokisline, za drugega pa razred po aromatičnosti. Najboljša vrednost s = 5. Najboljši algoritem za strojno učenje je bil skladanje klasifikatorjev, zgrajenih z SVM in logistično regresijo. SVM je uporabljal C = 0.1 in polinomsko jedro s p = 2. Logistična regresija je imela λ = 200.

Pričvrščeni pari. Ta atributni vektor je podoben parom, le da je eden izmed pripadnikov para pričvrščen na prvo mesto peptida. Razlog za to je, da je na peptidni mikromreži zgornje (prvo) mesto peptida najlaže dostopno, zato je najbolj verjetno, da se protitelo veže nanj. Spet lahko s sosednjih razdalj združimo v eno in aminokislino nadomestimo z razredom. Vektor ima obliko:

[A_1 na d_1, ..., A_1 na $d_{\max}$; ...; A_n na d_1,..., A_n na $d_{\max}$; *first*].

A_i je i-ta aminokislina ali i-ti razred aminokislin, n je število različnih aminokislin ali razredov, d_j je j-ta razdalja ali skupina razdalj od prvega mesta, $d_{\max}$ je največja razdalja in *first* je prva aminokislina v peptidu.

Najboljša izbira so bile aminokisline (ne razredi) in s = 5. Najboljši algoritem za strojno učenje je bil SVM s C = 5 in linearnim jedrom.

4 REZULTATI

Osnovne klasifikatorje in celoten ansambel klasifikatorjev smo najprej z desetkratnim prečnim preverjanjem testirali na učni množici. Uporabili smo s podvzorčenjem uravnoteženo učno množico, saj smo algoritme za strojno učenje izbrali in jim nastavili parametre na takšni množici. Uspešnost klasifikacije smo primerjali glede na ploščino pod krivuljo ROC (AUC) in klasifikacijsko točnost. Prva odlika AUC je, da je neodvisna od praga verjetnosti, nad katero štejemo, da peptid vsebuje epitop (vsi naši klasifikatorji so namreč vračali verjetnost, ne enega od dveh razredov). Druga odlika pa je, da je pretežno neodvisna tudi od razmerja razredov v učni in testni množici. Točnost

potrebuje prag (uporabili smo 0,5) in je odvisna od razmerja razredov, je pa bolj intuitivna. Rezultati so prikazani v tabeli 1. Osnovni klasifikatorji so se izkazali za podobno dobre, a ker so delali napake na različnih primerih, je njihovo združevanje klasifikacijo opazno izboljšalo.

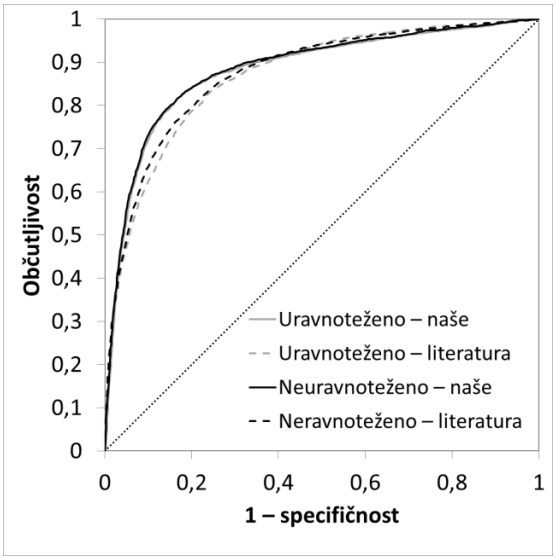
Klasifikator	AUC	Točnost
Frekvence	0,870	80,7 %
Razlike frekvenc	0,868	80,3 %
Frekvence podzaporedij	0,867	80,5 %
Lastnosti aminokislin	0,873	81,2 %
Frekvence razredov	0,866	80,5 %
Frekvence podzaporedij razredov	0,865	80,6 %
Pari	0,873	81,2 %
Pričvrščeni pari	0,863	80,3 %
Celoten ansambel	0,881	83,3 %

Tabela 1. *Klasifikacija na učni množici.*

Celoten ansambel klasifikatorjev smo testirali tudi na testni množici. Pri tem smo meta-klasifikator učili na z nadvzorčenjem uravnoteženi in na izvirni neuravnoteženi učni množici. Uravnotežena učna množica je smiselna, če ne poznamo razmerja razredov v testni množici, neuravnotežena pa, če je razmerje enako kot v učni množici. Rezultate našega ansambla klasifikatorjev smo primerjali z rezultati klasifikatorja, zgrajenega z SVM, ki je uporabljal znakovno jedro. Ta metoda je sodeč po literaturi [1][2] trenutno najboljša znana za prepoznavanje epitopov. Rezultati so prikazani v tabeli 2 in na sliki 1. Vidimo lahko, da se je naš ansambel v vseh pogledih obnesel bolje.

	Uravnoteženo		Neuravnoteženo	
	Naše	Literatura	Naše	Literatura
AUC	0,883	0,868	0,884	0,874
Točnost	83,7 %	82,8 %	85,9 %	83,9 %

Tabela 2. *Klasifikacija na testni množici.*



Slika 1. *Krivulja ROC na testni množici (naši krivulji se skoraj povsem prekrivata).*

5 ZAKLJUČEK

Razvili smo klasifikator za prepoznavanje epitopov, ki prekaša klasifikator, zgrajen z najboljšo znano metodo iz literature. Razlika v klasifikacijski točnosti znaša le 0,9 do 2 odstotni točki, a glede na naravo podatkov to ni malo. Med poizkusi smo namreč ugotovili, da že zelo enostavne metode dosežejo klasifikacijsko točnost nekoliko nad 80 %, vsakršno nadaljnje izboljšanje pa je zelo težavno.

Naš klasifikator je sestavljen iz ansambla osnovnih klasifikatorjev, ki jih – kot pri skladanju klasifikatorjev (stackingu) [5][8] – združuje meta-klasifikator, naučen z linearno regresijo. Naš postopek se od skladanja klasifikatorjev razlikuje po tem, da vsak osnovni klasifikator uporablja ne le svoj algoritem za strojno učenje, ampak tudi svoj atributni vektor. Tako lahko izkoristi različne algoritme za strojno učenje in različne predstavitve podatkov. Predstavili smo tudi temeljit in sistematičen način izbire algoritmov za strojno učenje in nastavljanje parametrov.

Zahvala

Zahvaljujem se prof. Hansu-Jürgenu Thiesnu, ki je priskrbel podatke, ter njemu, prof. Georgu Füllenu in prof. Michaelu Glockerju za nasvete pri delu.

Literatura

[1] EL-Manzalawy, Y., Dobbs, D., in Honavar, V. (2008). Predicting flexible length linear B-cell epitopes. Computational Systems Bioinformatics, str. 121–132.

[2] EL-Manzalawy, Y., Dobbs, D., in Honavar, V. (2008). Predicting linear B-cell epitopes using string kernels. Journal of Molecular Recognition 21 (4), str. 243–255.

[3] Frank, E., Wang., Y., Inglis, S., Holmes, G., in Witten, I. H. (1998). Using model trees for classification. Machine Learning 32, str. 63–76.

[4] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., in Witten, I. H. (2009). The WEKA data mining software: An update. ACM SIGKDD Explorations Newsletter 11(1), str. 10–18.

[5] Seewald, A. (2002). How to make Stacking better and faster while also taking care of an unknown weakness. ICML, 554–561.

[6] van Hulse, J., Khoshgoftaar, T. M., in Napolitano, A. (2007). Experimental perspectives on learning from imbalanced data. ICML, str. 935–942.

[7] Wang, H., in Witten, I. H. (1999) Pace regression. Working paper 99/12.

[8] Wolpert, D. H. (1992). Stacked generalization. Neural Networks 5, str. 241–259.

DETERMINING THE DAMAGE VALUE OF CRIME ACTS

Blaž Mahnič
Department of Intelligent Systems
Jozef Stefan Institute
Jamova 39, 1000 Ljubljana, Slovenia
e-mail: blaz.mahnic@ijs.si

ABSTRACT

This paper presents a data mining process of determining the damage value of a different crime acts. One of the most important tasks of crime analysis is to find pattens of crimes to evaluate where, when and why they happened. Before the 9/11 the major task of the governments was to deal with the crime after it has been committed. In my research I am trying to find some connections between attributes of the crimes. The dataset was form Slovenian police web page and did not contain the data about traffic incidents. The main goal was to preprocess the originaly dataset so that can be used in Weka software and try to classify the instances based on their descriptions. For the evaluation of the model I used 10-fold cross validation and percentage split validation. The best result that I get was 83% accuracy using Random forest alghoritm.

1 INTRODUCTION

Data mining a branch of computer science and artfical intelligence is a process of extracting patterns from data. Data mining is seen as an increasingly important tool by modern business to transform data into business intelligence giving an informational advantage. It is currently used in a wide range of profiling practices, such as marketing, surveillance, fraud detection, and scientific discovery [1]. The objective of this work is to find the attributes that are most relevant to the damage value of a crime act. From the given attributes I will try to discover which of them has the most influence to the damage value. I decided to do research on a dataset from Slovenian police department. There are several files for criminal acts in Slovenia for last ten years.
I have not found that similar data mining researches have been done till now in Slovenia so this is some kind of a motivation for me.
First I will obtain the data from web, then I will preprocess the data to get a dataset suitable for Weka software. In Weka software I will run different data minig alghoritms and evaluate the models. At the end I will make some conclusion about the results and possible further directives.

2 DATA

2.1 Original dataset

The dataset I used is from Slovenian police web page and is public, it is made of 18855 instances, two tables ("Crime acts" and "Persons involved") and 33 attributes of which 6 are numerical and 27 are nominal. The two tables are in relation "one to many" [2].

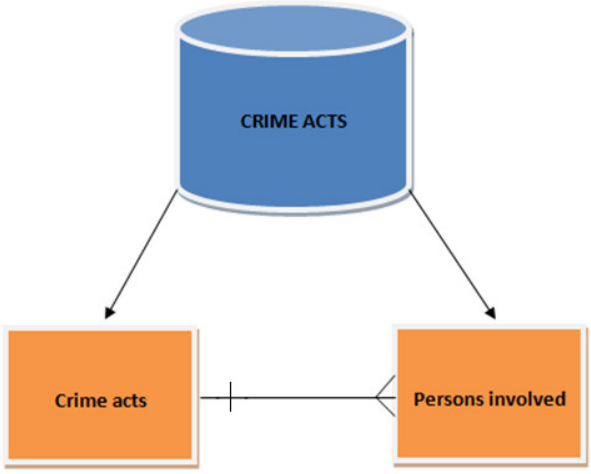


Figure 1: *Database structure.*

The original form of class variable – "Damage value" is nominal and made of intervals.
The intervals are:

- No damage,
- TO 40€,
- 40 € - 200 €,
- 200 € - 600 €,
- 600 € - 2.100 €,
- 2.100 € - 6.200 €,
- 6.200 € – 20.100 €,
- 20.100 € - 62.600 €,
- 62.600 € - 208.800 € and
- Above 208.800 €

Based on the structure and the distribution of instances of the target variable I decided to make three classes:

- No damage
- Medium damage
- High Damage

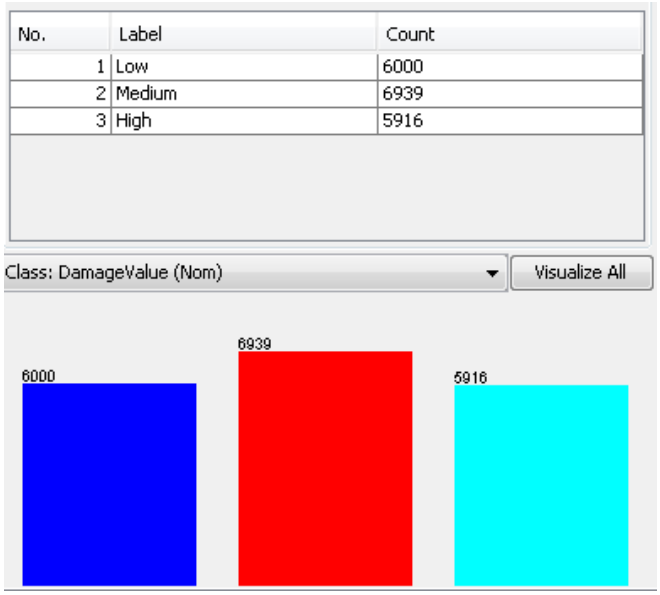


Figure 2: *Distribution of the instances over the class variable.*

2.2 Attributes

Data preprocessing was the most large task of my seminar work. It took about 80% of all the time that I spent for my research. There was a problem with some attributes that originally were maid of many values so I decided to made some manual abstraction of attribute values. I manually grouped together the values that derives from same group. For example the attribute values for an attribute Used subject: gun, semi automated gun, bomb, knife etc., were put into the same class - Weapons.
Because of the relation one to many between the dataset tables I have to inferred new agregated attributes.

2.2.1 Agregated attributes

- Age of a person: average age of all the persons involved,
- Sex of a persons: no. of man, woman and legal persons involved,
- Nationality of a person: no. Of slovenes and foreigns involved,
- Injury: if some involved people has submitted some injury.

2.2.2 Non agregated attributes

- Date: date when crime act occured,
- Day: day in a week,
- Time: time of occuration,
- Crime classification: four classes theft, position abuse, other, tax cheat and cheat,
- Crime type: general or economic,
- Used subject: weapons, documents, vehicles, technics, money, drugs and other,
- Scene: traffic, financial-Institution, bar-or-restaurant, living-place, public-place, company and other.

3 MACHINE LEARNING METHODS

Because the target variable was nominal I decided to use classification machine learning methods. Classification is a data mining (machine learning) technique used to predict group membership for data instances. For example, you may wish to use classification to predict whether the weather on a particular day will be “sunny”, “rainy” or “cloudy”.
The alghortims I decided to use are decision trees because of the model interpretability advantage. So when got the results I was able to explane how some attributes has affect to the results. I decided to run two data mining alghoritms in Weka software:

J 48: the alghoritm builds decision trees from a set of training data using the concept of information entropy. The training data is a set $S = s_1,s_2,...$ of already classified samples. Each sample $s_i = x_1,x_2,...$ is a vector where $x_1,x_2,...$represent attributes or features of the sample. The training data is augmented with a vector $C = c_1,c_2,...$ where $c_1,c_2,...$ represent the class to which each sample belongs [3],

Random forest: the random Forests alghoritm grows many classification trees. To classify a new object from an input vector, it puts the input vector down each of the trees in the forest. Each tree gives a classification, and we say the tree "votes" for that class. The forest chooses the classification having the most votes (over all the trees in the forest) [4]. I decided to use this alghoritm because of some advantages that is has:

- It runs efficiently on large data bases,
- It gives estimates of what variables are important in the classification,
- Generated forests can be saved for future use on other data.

4. EVALUATION

4.1 Decision tree alghoritm »J48«

4.1.1 Test 1 – default parameters

I first run a test on dataset without changing any parameter of the alghoritem. Using 10-fold cross validation the alghoritm achieved 78.33% accuracy. The results are shown in Fig. (3). The first thing I noticed was the problem of visualization. The size of the tree was 3409 so the visualization was impossible to interpret by a human. I decided to do another test where I modified the parameter »minNumObj« which sets the minimum number of instances per leaf in a tree.

4.1.2 Test 2 – pruning tree

The next test was to increase the number of instances in one leaf to 500. The accuracy using 10-fold cross validation decreased to 60.2% but I was able to read some rules from a decison tree because the size of tree has been reduced to 56:

- If crime type is economic and crime classification is tax-cheat or position abuse then the damage value is High
- If crime type is general and crime classification is tax-cheat then the damage value is high
- If crime type is general and crime classification is theft and date is winter and no.of legal persons involved is higher than zero and time is morning than the damage value is high.

Classifier output

```
=== Summary ===

Correctly Classified Instances       14769               78.3294 %
Incorrectly Classified Instances      4086               21.6706 %
Kappa statistic                          0.6737
Mean absolute error                      0.1843
Root mean squared error                  0.3272
Relative absolute error                 41.5853 %
Root relative squared error             69.4965 %
Total Number of Instances            18855

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.786     0.074      0.833     0.786     0.809      0.908    Low
                 0.783     0.151      0.752     0.783     0.767      0.898    Medium
                 0.78      0.104      0.775     0.78      0.778      0.93     High
Weighted Avg.    0.783     0.111      0.785     0.783     0.784      0.911

=== Confusion Matrix ===

    a    b    c   <-- classified as
 4716  810  474 |    a = Low
  634 5436  869 |    b = Medium
  312  987 4617 |    c = High
```

Figure 3: J48 10-fold cross validation

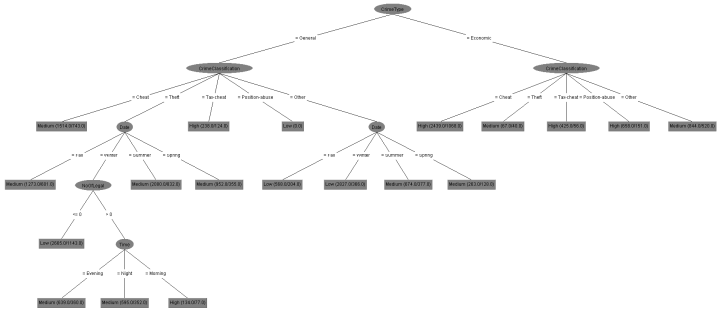


Figure 4: J48 – Decision tree after pruning

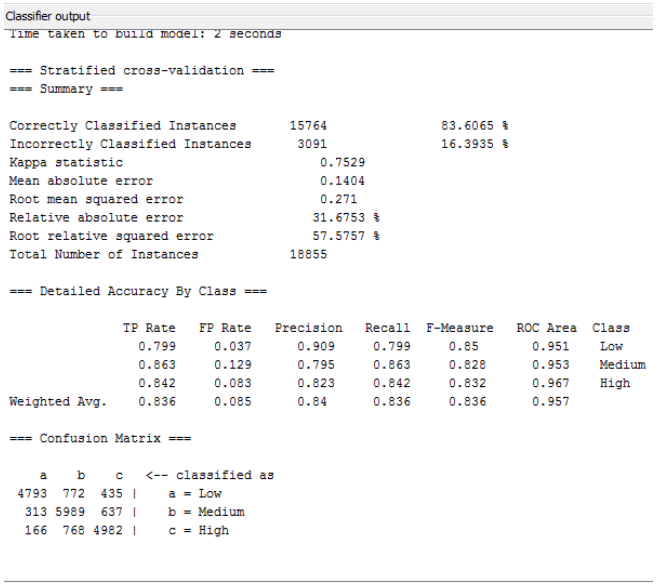
4.1.3 Test 3 - Reducing the number of attributes by using Weka attribute evaluator – J48

I decided to run another test with the attributes that Weka software has suggested me by using Weka attribute evaluator. I used CfsSubsetEval attrbute evaluator in combination with BestFirst search method, that suggested me to use only attributes Date, Crime classification, Crime type, No. Of legal persons and Injury. The accuracy achived was 60.7% and it has decreased for 18% comparing it to the accuracy of test 1.

4.2 Decision tree alghoritm »Random forest«

4.2.1 Test – 1 default parameters

In this test the alghoritm used was a decision tree alghoritm Random forest. Random Forests are easy to use, the only 2 parameters a user has to determine are the number of trees to be used and the number of variables (m) to be randomly selected from the available set of variables [5]. I first run an alghoritm with the default parameters that generates ten decision trees. This test gave me the best results of all the tests that I have run before. Using 10-fold cross validation the accuracy acchieved was 83.6% and It can be seen in the figure 6.



```
Classifier output
Time taken to build model: 2 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances       15764               83.6065 %
Incorrectly Classified Instances      3091               16.3935 %
Kappa statistic                          0.7529
Mean absolute error                      0.1404
Root mean squared error                  0.271
Relative absolute error                 31.6753 %
Root relative squared error             57.5757 %
Total Number of Instances            18855

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.799     0.037      0.909     0.799      0.85       0.951    Low
                 0.863     0.129      0.795     0.863      0.828      0.953    Medium
                 0.842     0.083      0.823     0.842      0.832      0.967    High
Weighted Avg.    0.836     0.085      0.84      0.836      0.836      0.957

=== Confusion Matrix ===

    a    b    c   <-- classified as
 4793  772  435 |    a = Low
  313 5989  637 |    b = Medium
  166  768 4982 |    c = High
```

Figure 6: 10-fold cross validation – Random forest alghoritm

Test 2 – Reducing the number of attributes by using Weka attribute evaluator – Random forest alghoritm

In this test I decided to use only attributes suggested by Weka attribute evaluator using CfsSubsetEval evaluator in combination with BestFirst search method. The suggested attributes were Date, Crime classification, Crime type, No. Of legal persons and Injury. The accuracy given was 60.48%.

Results comparison

ALGHORITM	ALL ATTRIBUTES	SELECTED ATTRIBUTES WITH WEKA
J48	78.33%	60.7%
Random forest	83.6%	60.48%

Table 1: Results comparisson

The best result was given by a Random forest alghoritm using the default parameters and all the attributes. The accuracy is for 5.3% better than the accuracy acchieved by an alghoritm J48 with same number of attributes. The comparisons between alghoritms can be seen in a table 1. The test 2 – Pruning decision tree J48 was just made for purposes of visualization and it is not included in result comparison table.

6. CONCLUSION

The goal of my seminar work was to predict the damage value for different crime acts and to find which attributes are the most relevant to the class attribute damage value. The results shows that the attributes with the biggest influence to the class value are Crime type, Crime classification and Date. The best results were acchived by alghoritm Random forest which reached 83.6% accuracy in comparison to the alghoritm J48 which reached 78.33% accuracy. The results are shown in table 1.
The results can be improved with the introduction of new attributes and with the improvement of domain knowledge. In the future more data mining alghoritms will be tested on the dataset.

REFERENCES

[1] *Investigative data mining for Security and Criminal Detection, 2003, Jesus Mena.*

[2] *About datamining. From web. 1.2.2011 http://en.wikipedia.org/wiki/Data_mining*

[3] *Slovenian police department. Fom web. 1.2.2011 http://www.policija.si/index.php/statistika/kriminaliteta*

[4] *About J48 decision tree alghoritm. From web. 1.2.2011 http://en.wikipedia.org/wiki/C4.5_algorithm*

[5] *About random forest alghoritm. From web. 1.2.2011 http://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm*

DOMÉNSKI SISTEMI ZA ODGOVARJANJE NA VPRAŠANJA

Domen Marinčič
Odsek za inteligentne sisteme
Institut "Jožef Stefan"
Jamova cesta 39, 1000 Ljubljana, Slovenija
domen.marincic@ijs.si

Povzetek

Prispevek opisuje splošne karakteristike doménskih sistemov za odgovarjanje na vprašanja. Opisuje potencialne načine uporabe sistemov, obravnava specifike vezane na uporabo sistemov znotraj omejenih domen. Opisuje, kako uporabiti domensko znanje in našteva pogoste probleme pri analizi uporabnikovih vprašanj ter podaja možne načine rešitev. Na koncu podaja še priporočila za razvoj doménskih sistemov OV.

1 Uvod

Sisteme za odgovarjanje na vprašanja (sisteme OV) lahko delimo na splošne sisteme in take, ki so vezani na omejeno domeno. Z vezavo na specifično domeno omejimo problem avtomatskega odgovarjanja vprašanja. Običajno lahko uporabimo poglobljeno formalno opisano predznanje o domeni. To je v splošnih sistemih precej teže in obsežneje, saj bi morali imeti na voljo znanje "o vsem".

Primeri področij, kjer lahko pridejo prav sistemi OV [6]:

- vmesniki za strojno berljive tehnične priročnike,
- sistemi za dostop do baz znanja, npr. medicinske baze znanja,
- sistemi za pomoč uporabnikom.

2 Splošno o sistemih OV

V splošnem lahko pristope k reševanju problema odgovarjanja na vprašanja delimo na dve skupini, konkretni sistemi lahko kombinirajo oba pristopa:

- *Iskanje odgovorov v prostem besedilu.* Tak pristop je primeren, ko imamo na voljo veliko besedila z redundantnimi informacijami, kjer lahko najdemo odgovor brez kompleksne analize besedila.
- *Strukturirani pristop z uporabo predznanja.* Tak pristop je primerno uporabiti, ko so viri informacij omejeni in jih moramo učinkovito izrabiti.

Nekaj prednosti sistemov OV pred drugimi načini iskanja informacij:

- *Uporaba naravnega jezika.* To uporabniku omogoča točneje formulirati svoje zahteve kot iskanje po ključnih besedah. V primerjavi s strukturiranimi umetnimi jeziki, ki jih uporabljamo za delo s podatkovnimi bazami odpade potreba po učenju sintakse umetnih jezikov.
- *Pogovor.* Sistemi, ki obravnavajo zaporedje vprašanj istega uporabnika kot pogovor, omogočajo opuščanje besed, uporabo osebnih zaimkov, kar skrajša čas pisanja vprašanj.

Nekatere slabosti sistemov OV:

- Uporabniki pričakujejo inteligenco v smislu človeškega sogovornika in zato postavljajo preveč kompleksna vprašanja.
- Kadar sistem OV ne pozna odgovora, običajno ni jasno, ali sistem nima dostopa do pravih informacij ali ni sposoben na pravi način interpretirati vprašanja.
- Potrebno je obsežno delo za pripravo sistema, priskrbitev primernih virov informacij, prilagajanje sistema določeni domeni ipd.

3 Lastnosti omejenih domen

Po [5] je primerno, da je domena sistema OV:

- *Dobro definirana.* To pomeni, da vemo, kakšna vprašanja lahko od uporabnikov pričakujemo. Primeri dobro definiranih domen: računalništvo, umetnost. Primer slabo definirane domene bi bili aktualni dogodki, saj vsebinsko lahko zajema praktično karkoli.
- *Primerno kompleksna.* Ni smiselno razvijati sistema za odgovarjanje na vprašanja, kjer bi zadoščal seznam pogosto zastavljenih vprašanj z odgovori.
- *Praktična.* Naj bo namenjena primerno velikemu krogu ljudi. Poleg koristnosti nam to omogoča analizo uporabe razvitega sistema OV v praksi.

4 Uporaba virov znanja v domenskih sistemih OV

Kakovostni sistemi OV poskušajo kar najbolje izkoristiti formalno zapisane domenske vire znanja, informacij. Viri so lahko preprosti, kot npr. seznami entitet, oseb, ali kompleksni, v obliki visokonivojskih ontologij, ki zagotavljajo nedvoumen zapis domenskega znanja [2, 4]. Pri izbiri ontologij je potrebna previdnost, še posebej če poskušamo uporabiti splošne ontologije, ki niso vezane na domeno. Pri tem lahko pride do sledečih problemov:

- *Splošne ontologije so premalo podrobne.* V omejenih domenah so pogosto v uporabi specifični izrazi, katerih ne najdemo drugje. V takem primeru je splošne ontologije treba dopolniti s specifičnim domenskim znanjem.
- *Splošne ontologije so preveč podrobne.* Splošne ontologije lahko definirajo relacije med besedami in pojmi (koncepti), pri čemer lahko večpomenske besede opisujejo več pojmov. To je v omejeni domeni lahko odveč, saj je terminologija ponavadi dobro definirana - ena beseda ima en pomen.
- *Splošne ontologije vsebujejo zavajajoče informacije.* V omejenih domenah imajo določeni izrazi pomen, ki je vezan zgolj na domeno. V splošni ontologiji so lahko navedeni samo drugi, bolj pogosto uporabljani pomeni tega izraza.

Učinkovita uporaba doménskih virov omogoča izboljšanje točnosti in praktičnosti sistemov OV. Pri tem je potrebno še ugotoviti, kakšne informacije bodo potencialni uporabniki iskali in na kakšen način bodo zastavljali vprašanja.

5 Zgradba tipičnega (doménskega) sistema OV

Tipičen (doménski) sistem OV je sestavljem iz sledečih sklopov: (i) analiza vprašanj, (ii) iskanje dokumentov, kjer se nahaja odgovor in (iii) luščenje odgovora. Če gre za iskanje odgovora v podatkovni bazi, potem v drugem koraku ne gre za dejansko iskanje dokumenta ampak za poizvedbo v podatkovi bazi.

Izkaže se, da je analiza vprašanj eden najpomembnejših korakov - če sistem tu naredi odločilno napako, bo zelo verjetno odgovor popolnoma napačen [3]. Pri analizi vprašanj lahko prihaja do možnosti razumevanja vprašanja na več načinov. V splošnem velja, da so ti problemi s primernimi viri predznanja znotraj domene lažje rešljivi, saj nekateri načini razumevanja znotraj domene niso smiselni. Če sistem OV vseeno ni sposoben ugotoviti pravega načina, je ena izmed rešitev, da uporabnika vpraša, kateri od načinov je pravi.

Opišimo nekaj najpogostejših težav, na katere lahko naletimo pri analizi vprašanja [1]:

- *Odvisnost predložne zveze.* Primer: *"Kateri so zaposleni instituta z licenco za delo z nevarnimi snovmi?"* V tem primeru brez predznanja ni jasno, ali uporabnik išče zaposlene z licenco ali institut z licenco.
- *Obseg kvantifikatorjev.* Primer: *"Ali ima vsak zaposleni dostop do strežnika?"* V tem primeru lahko vprašanje razumemo na dva načina: (i) ali obstaja en sam strežnik, do katerega ima vsak zaposleni dostop oziroma (ii) ali za vsakega zaposlenega obstaja nek strežnik, do katerega je možen dostop, pri čemer je strežnikov lahko več.
- *Konjukcija in disjunkcija.* Primer: *"Kdo so sodelujoči na olimpijskih igrah iz Slovenije in Hrvaške?"* V tem primeru ima veznik "in" pomen veznika "ali". V določenih primerih je možno ugotoviti, da je taka konjukcija konceptualno nemogoča in jo lahko sistem nadomesti disjunkcijo.
- *Pomen pridevnika v samostalniški frazi.* Primera: *"Kateri raziskovalni odseki so na institutu?"* in *"Kateri raziskovalni sistemi so na institutu?"*. V prvem primeru iščemo odseke v katerih se izvaja raziskave, v drugem primeru pa sistem s katerim se izvaja raziskave.
- *Anafora.* Primera zaporednih vprašanj: *"Kdo je avtor knjige? Kje jo najdemo?"* in *"Kdo je avtorica knjige? Kje jo najdemo?"*. Če uporabnik postavi več zaporednih vprašanj in se v kasnejših vprašanji sklicuje na prej omenjene entitete z zaimki, mora sistem OV zaimke povezati s pravimi entitetami. V prvem primeru je na podlagi ujemanja spolov možno ugotoviti, da se zaimek "jo" nanaša na knjigo. V drugem primeru je nemogoče ugotoviti, kaj je uporabnik hotel vprašati z drugim vprašanjem.
- *Opuščanje.* Primer zaporednih vprašanj: *"Kateri je največji odsek na institutu? Kateri pa najmanjši?"* Da lahko sistem OV ugotovi, da se drugo vprašanje ravno tako nanaša na institut, je treba vzdrževati model pogovora, s katerim sistem sledi temo pogovora
- *Slovnične napake.* S pomočjo črkovalnikov, slovarjev ipd. je možno uporabniku olajšati delo s sistemom OV, tako da sistem zazna oziroma popravi določene napake. Brez tega napačno napisanih vprašanj sistem OV ne bi znal odgovoriti.

6 Razvoj doménskega sistema OV

Pri razvoju doménskega sistema OV je priporočljivo narediti sledeče stvari:

- *Narediti analizo vprašanj.* Vprašanja, ki jih postavljajo uporabniki je treba analizirati, klasificirati, povezati z različnimi tipi informacij, ki jih uporabnik išče.

- *Izbrati primerno domensko znanje.* Določiti je treba formalno obliko zapisa znanja. Lahko so to, npr. nestrukturirani dokumenti z besedilom, delno strukturirani SGML dokumenti ali visokostrukturirano znanje zapisano v obliki podatkovnih baz, ontologij.

- *Določiti način pridobivanja in predstavitve domenskega znanja.* Zaradi običajno raznorodnih virov znanja je znotraj sistema potrebno vzpostaviti nek enoten način predstavitve. Interna predstavitev mora ustrezati operativnim zahtevam delovanja sistema.

- *Določiti tip uporabniškega vmesnika.* Po definiciji naj bi sistemi OV ponujali možnost vnosa vprašanja v naravnem jeziku. V primerih, ko ni mogoče zagotoviti zadosti visoke točnosti razumevanja, je možno to nadomestiti ali dopolniti z obrazci, kjer lahko uporabnik na nedvoumen način postavi vprašanje.

Literatura

[1] Androutsopoulos, I., Ritchie, G.D., Thanisch P. (2005). Natural Language Interfaces to Databases – An Introduction. *Natural Language Engineering*, 1(1), pages 29–81.

[2] Fellbaum, C. (1998). WordNet: Introduction. In *Christiane Fellbaum, editor, WordNet: An Electronic Lexical Database, Language, Speech, and Communication*, pages 1–19. MIT Press.

[3] Indurkhya, N., Damerau, F. J. (2010). Handbook of Natural Language Processing, Second Edition. CRC Press

[4] Lenat, D., Guha R. V (1990). Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project. Addison-Wesley.

[5] Minock, M. (2005). Where are the 'killer applications' of restricted domain question answering? In *Proceedings of the IJCAI Workshop on Knowledge Reasoning in Question Answering*, page 4.

[6] Mollá, D., Vicedo, J. (2007). Question Answering in Restricted Domains: An Overview. Computational linguistics, 33:1 (2007).

TOWARDS ROBUST FALL DETECTION

Violeta Mirchevska[1], Mitja Luštrek[2], Matjaž Gams[2]
[1]Result d.o.o., Bravničarjeva 11, 1000 Ljubljana, Slovenia
[2]Department of Intelligent Systems, Jožef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia
e-mail: violeta.mircevska@result.si

ABSTRACT

This paper presents a method for classifier development by combining domain knowledge and machine learning. The development is performed in two phases: (1) development of initial hypothesis using domain knowledge or interactive machine learning and (2) refinement of the initial hypothesis using genetic algorithms. The method is presented in the domain of fall detection.

1 INTRODUCTION

Recently, there has been a growing interest in innovative ICT solutions that would aid the elderly to live independently for longer and counteract reduced capabilities caused by age. In this respect, sustainable and personalized healthcare has become one of the strategic interests of the European society. The aging of the population in Europe causes immense pressure on the healthcare expenditures, which already account for 9% of the EU's GDP spending [1]. The development of personalized healthcare systems is one of the research focuses of the fifth challenge of the ICT Work Programme under FP7. These systems would provide more effective care of patients by monitoring patients' health conditions using wearable, portable and implantable systems, providing health professionals with comprehensive monitoring and diagnostic data.

The European FP7 project Confidence – Ubiquitous care system to support independent living [2] aims at developing a system that will monitor the health conditions of its elderly user in real-time. It encompasses detection of falls as well as changes in behavior, such as limping, slow moving and physical inactivity. In case a health problem is detected, the system issues a warning to the user and alerts a caregiver if necessary. This way the system would provide the elderly with confidence to continue to live independently at home as long as possible.

In this paper we focus on the fall detection part of Confidence [3]. Development of a fall detection classifier was challenging because of the following reasons: First of all, high reliability is needed since, on one hand, undetected falls may be disastrous for the user and, on the other, too many false positives may be disturbing for the user and may lead the user to be unwilling to use the system. Second, representative dataset for falls is difficult to obtain because of the variety of fall types, variations depending on the user, as well as ethical issues and injury danger which prevent collection of large amounts of data by healthy persons simulating falls. Non-representative dataset may cause classifier overfitting and poor performance in the general case. The problem that we address in this paper is how to create a robust fall detection classifier when only a limited amount of data from the domain of interest is available.

We addressed the problem of generation of a robust fall detection classifier by combining domain knowledge (DK) and machine learning (ML). DK and ML complement each other. ML algorithms can discover characteristic domain patterns which may be too subtle for humans to detect, but they can only discover patterns that are present in the training dataset. DK, on the other hand, may be related to examples not present in the available domain dataset. Therefore, their combination may improve the reliability of the developed classifiers, if done in a proper way.

In the following, we present preliminary study of a method of development of rule-based classifier by combining DK and ML and compare it to classifiers developed by known ML algorithms that are in the form of rules, or that may be converted to rules.

This paper is organized as follows: Section 2 presents related work on the topic of methods for classifier development by combining DK and ML. Section 3 presents the approach for development of a fall detection classifier by combining DK and ML. Its evaluation is presented in Section 4. Section 5 concludes the paper.

2 RELATED WORK

Different approaches for model development by combining DK and ML are found in the literature.

A comprehensive overview of methods for incorporating DK into inductive ML is presented by Yu [4], who categorizes these methods in four groups: (1) methods that use prior DK to prepare training examples [5][6], (2) methods that use prior DK to initiate the hypothesis or hypothesis space [7][8], (3) methods that use prior DK to alter the search objective [9][10] and (4) methods that use prior DK to augment search [11][12].

In addition to this, methods for combining DK and ML are found in the field of interactive ML. Interactive ML basically refers to an iterative process of classifier generation by human-computer interaction. Two strategies for model generation using interactive ML can be distinguished: (1) iterative improvement of a single model by refining the input information used during ML induction [13][14] and (2) generation of multiple models in order to select one or several that are the most relevant from the user point of view [15][16]. In principle, the combination of ML and DK improves the generality of the generated model and/or the efficiency of the learning process.

3 METHOD

The method for model generation by combining ML and DK, an extension of the method presented in [17], can be separated in two phases:

- Determination of initial hypothesis by DK or interactive ML
- Refinement of the initial hypothesis by genetic algorithms based on data of the domain of interest.

Detailed presentation of the approach for determination of initial hypothesis is presented in [17]. Basically, an expert determines the format of the rules in the classifier by DK. The expert may explore the data of the domain of interest by generation of ML models and extract patterns from these models to be included in the rule-based classifier.

The initial hypothesis is refined by genetic algorithms.

Genetic algorithms [18] are stochastic search algorithms, whose search strategy mimics evolution and natural selection. Genetic algorithms have been used for solving certain optimization subproblems in machine learning. Kononenko and Kukar [19] identify four machine learning subproblems that involve optimization in a large search space and which may be addressed by genetic algorithms: feature subset selection, parameter tuning, constructive induction and hypothesis learning. Genetic algorithms have been applied for solving such machine learning subproblems in rule induction, in decision tree induction, in neural networks as well as in support vector machines induction.

Genetic algorithms are suitable for refining the initial hypothesis because the structure of the rules in the initial hypothesis is fixed and standard heuristics used for induction of rules are not suitable in our case. In addition to this, constraints defining relations between rules as well as relations between parameters within a rule can be specified in a straightforward way when using genetic algorithms.

We have used the Pittsburg approach for rule discovery using genetic algorithms, meaning that each individual in the population represents one possible solution. The individual is basically a vector which contains the parameters of all rules in the rule-based classifier. The elements are real valued and take values in a predefined interval. The fitness function used for evaluating the quality of each individual is the accuracy of the solution on the training dataset; fitness values fall within the interval [0, 1]. Elitism was used, meaning that the best individual is always transferred in the new population. By this we want to tune the parameters of the rule-based classifier as good as possible to the training dataset. Overfitting should be avoided because the structure of the rule-based classifier is defined by domain expert.

4 EVALUATION

The evaluation of the presented method was done in the domain of fall detection. The rule-based classifier developed for the need of fall detection is based on the pattern that if an elderly is lying or sitting on the ground for long period of time, then there is high probability of a fall. It contains the following types of rules:

1. IF falling activity within $T1_{fall}$ seconds AND the user was lying/sitting on the ground $P1_{activity}$% in $T1_{activity}$ seconds AND the user was not moving $P1_{moving}$% in $T1_{moving}$ seconds THEN fall
2. IF falling activity within $T2_{fall}$ seconds AND the user was lying/sitting on the ground area afterwards $P2_{activity}$% in $T2_{activity}$ seconds THEN fall
3. IF the user was lying/sitting on the ground for $P3_{activity}$% in $T3_{activity}$ seconds AND the user was not moving $P3_{moving}$% in $T3_{moving}$ seconds THEN fall
4. IF the user was lying/sitting on the ground for $P4_{activity}$% in $T4_{activity}$ THEN fall

A test scenario that contains clear-case and complex events of falls was designed in order to test the reliability and the robustness of the developed classifier. The clear-case events contain typical fall and non-fall events, such as normal behavior (e.g. standing, sitting, lying in bed), tripping, falling from chair when trying to stand up and searching for something on the ground on all fours and lying. The complex events represent atypical falls and non-fall events which may be easily mistakenly classified as falls, such as falling slowly trying to hold onto furniture from standing, lying down quickly on the bed, sitting down quickly on the chair, falling slowly trying to hold onto furniture when standing up from the chair or sitting on a low chair. The fall examples of the complex events do not have high acceleration towards the ground during the falling activity, which is a characteristic feature of falls, whereas the non-fall events contain high acceleration towards the ground during going down activity and activities, such as sitting on a low chair, that may be misclassified as lying or sitting on the ground.

All events present in the test scenario were recorded in single recordings interspersed with short periods of walking. Each recording lasted around 20 minutes. The recordings were made by 5 healthy volunteers (3 male and 2 female), 5 times by each.

The classifiers were generated using the straightforward events only. We used leave-one-person-out evaluation, meaning that the classifier was generated on the examples from four persons and tested on the examples of the fifth person, which was excluded from the training dataset. Firstly, the initial classifier was specified by domain expert. Then genetic algorithms were used for tuning the initial classifier to suit system- and user-related characteristics creating the refined classifier using the examples in the clear-case events. The accuracy of the classifiers generated in this way was tested on both clear-case and complex events of the person excluded from the training dataset. The test on the clear-case events shows how well the classifier performs on types of events present in the training dataset. The test on the complex events, on the other hand, is used to test the generality and robustness of the generated classifier, since the complex events are not present in the learning process. For comparison, fall detection classifiers were induced using machine learning only with attributes equivalent to the parameters of the rules in the rule-based fall detection classifier. We used decision trees (J48 in Weka [20]) and rules (JRip in Weka [20]) because these algorithms create models that can be converted to or are a set of rules.

Table 1 presents a comparison of the performance of the obtained fall detection classifiers. J48 and JRip seem to be biased toward recognition of fall events. They detected all fall examples in the clear-case and complex events, at the same time raising false positives in many cases. The initial classifier is less biased towards recognition of falls, as experts introduce patterns for which they are sure that are relevant for the recognition of the class of interest. This, however, causes certain fall examples not to be recognized. Finally, the refined classifier significantly increases the precision of the classifier on the clear-case scenarios. The refinement of the initial classifier with genetic algorithms contributes to a slight increase in the performance of the classifier as compared to the initial classifier.

4 CONCLUSION

The combination of DK and ML may increase the credibility and performance of the developed classifier in the general case. On one hand, ML algorithms can discover characteristic domain patterns which may be too subtle for humans to detect, but they can only discover patterns that are present in the training dataset. DK, on the other hand, may be related to examples not present in the available domain dataset.

In this paper, we present a preliminary study of a method for generation of a classifier by combining DK and ML in the domain of fall detection. The presented method can be divided into two phases: (1) creation of initial hypothesis by DK or interactive ML and (2) refinement of the initial hypothesis using genetic algorithms. Preliminary tests show that the incorporation of DK in the learning process improves the performance of the classifier in the general case. Additionally, tests show that refinement by genetic algorithms may contribute to additional performance improvement.

As future work, we need to apply the presented method in other domains in order to test its generality. We need to explore different settings of the genetic algorithm in the refinement phase (e.g. different fitness functions, different recombination possibilities).

Table 1 Classifier comparison

	CLEAR-CASE EVENTS			*COMPLEX EVENTS*			*ALL EVENTS*		
	Precision	Recall	F-measure	Precision	Recall	F-measure	Precision	Recall	F-measure
J48	0.76	1.00	0.86	0.28	1.00	0.44	0.41	1.00	0.58
JRip	0.76	1.00	0.86	0.27	1.00	0.43	0.40	1.00	0.57
Initial classifier	0.84	0.98	0.91	0.33	0.98	0.49	0.48	0.98	0.64
Refined classifier	0.94	0.91	0.92	0.34	0.96	0.51	0.50	0.93	0.65

ACKNOWLEDGEMENTS

This research is partly financed by the European Union, European Social Fund and partly by the European Community's Framework Programme FP7/2007–2013 under grant agreement No. 214986.

REFERENCES

[1] European commission, ICT research in FP7, ICT challenge 5: Towards sustainable and personalized healthcare. http://cordis.europa.eu/fp7/ict/programme/overview5_en.html. 01.09.2011

[2] Confidence. http://www.confidence-eu.org. 01.09.2011

[3] B. Kaluža, V. Mirchevska, E. Dovgan, M. Luštrek, and M. Gams (2010) An agent-based approach to care in independent living, in Proceedings of AmI 2010, pp. 177-186

[4] T. Yu (2007) *Incorporating Prior Domain Knowledge into Inductive Machine Learning Its implementation in contemporary capital markets*, PhD thesis

[5] S. Kambar (2005) *Generating synthetic data by morphing transformation for handwritten numeral recognition (with v-svm)*, Master thesis

[6] P. Niyogi, F. Girosi, T. Poggio (1998) Incorporating prior information in machine learning by creating virtual examples, in Proceedings of the IEEE, pp. 2196-2209

[7] Z. Zhu, P. Liu (2010) Feasibility research of text information filtering based on genetic algorithm, *Scientific Research and Essays* 5(22), pp. 3405-3410

[8] B. D. Burns, A. P. Danyluk (2000) Feature Selection vs Theory Reformulation: A Study of Genetic Refinement of Knowledge-based Neural Networks, *Machine Learning - Special issue on multistrategy learning*, pp. 89-107

[9] I. Davidson, S. S. Ravi (2005) Hierarchical clustering with constraints: Theory and practice, in Proceedings of the Ninth European Principles and Practice of KDD (PKDD), pp. 59-70

[10] X. Wu, R. Srihari (2004) Incorporating prior knowledge with weighted margin support vector machines, in Proceedings of the tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.326-333

[11] D. Decoste, B. Scholkopf (2002) Training invariant support vector machines machine learning, *Machine Learning 46*, pp. 161-190

[12] M. Pazzani, C. Brunk, G. Silverstein (1991) A knowledge-intensive approach to learning relational concepts, in Proceedings of the Eighth International Workshop on Machine Learning, pp. 432-436

[13] S. Sun, D. R. Hardoon (2010) Active learning with extremely sparse labeled examples, *Neurocomputing 73*, pp. 2980-2988

[14] S. Stumpf, V. Rajaram, L. Li, W.-K. Wong, M. Burnett, T. Dietterich, E. Sullivan, J. Herlocker, (2009) Interacting meaningfully with machine systems: Three experiments, *International journal of human-computer studies 67*, pp.639-662

[15] V. Vidulin, M. Gams, (2011) Impact of higher-level knowledge on economic welfare through interactive data mining, *Applied artificial intelligence*, pp. 267-291

[16] K.-M. Osei-Bryson, (2004) Evaluation of decision trees: a multi-criteria approach, *Computers and Operations Research*, pp. 1933-1945

[17] V. Mirchevska, M. Gams, (2009) Towards robust rule engine for classifying human posture, in Proceedings of the 12th international multiconference IS 2009

[18] A. E. Eiben, J. E. Smith, (2003) Introduction to Evolutionary Computing

[19] I. Kononenko, M. Kukar, (2007) Machine learning and data mining

[20] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I. H. Witten, (2009) The WEKA Data Mining Software: An Update, *SIGKDD Explorations11(1)*, pp. 10-18

Gradnja točnih in razumljivih hibridnih klasifikatorjev na osnovi hierarhičnega gručenja

Rok Piltaver
Odsek za inteligentne sisteme
Institut "Jožef Stefan"
Jamova 39, 1000 Ljubljana, Slovenija
e-mail: rok.piltaver@ijs.si

POVZETEK

V prispevku je predstavljen algoritem, ki sprejme učno množico v atributnem opisu in zgradi množico nedominiranih hibridnih klasifikatorjev od čim bolj razumljivih do čim bolj točnih. Hibridni klasifikator deluje tako, da atributni prostor razdeli v podprostore (gruče učnih piremerov) in vsakemu podprostoru določi razumljiv (odločitveno drevo) ali točen (black-box) klasifikator. Delitev v podprostore poteka s pomočjo algoritma za hierarhično gručenje, ki uporablja posebni meri za razdaljo med piremeroma in med gručama, ki primerno usmerjata iterativni postopek gručenja. V zadnji fazi algoritem za iskanje nedominiranih hibridnih klasifikatorjev uporablja poseben algoritem, ki močno omeji število klasifikatorjev, ki jih je treba zgraditi, oceniti in primerjati po kvaliteti.

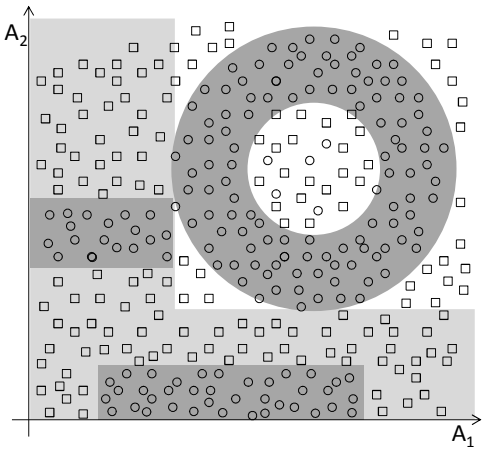
1. Uvod

Na algoritme strojnega učenja za gradnjo klasifikatorjev lahko gledamo kot na algoritme za iskanje po prostoru klasifikatorjev. Ker so ti prostori običajno preveliki, da bi jih lahko izčrpno preiskali, se za usmerjanje algoritma uporabljajo določene hevristike. Pristranskost hevristike (ang. search-bias) vpliva na vrstni red in smer preiskovanja. Ker je cilj najti klasifikator, ki se čim bolj sklada z učnimi primeri in čim bolje napoveduje razred še ne videnih primerov, se najpogosteje uporabljajo hevristike, ki strmijo k čim večji točnosti klasifikatorja.

Za uporabnika je poleg točnosti pomembna tudi razumljivost klasifikatorja, ker ljudje razumljivim klasifikatorjem zaupajo bolj kot nerazumljivim. Prav tako je razumljivost klasifikatorja zelo pomembna, kadar uporabnik želi razumeti učne podatke s pomočjo klasifikatorja, ki opisuje relacije v učnih podatkih.

Razumljivost in točnost sta kriterija, ki si nasprotujeta: bolj kot je klasifikator razumljiv slabšo točnost ima in obratno. V tem prispevku je predstavljen algoritem, ki ob gradnji klasifikatorjev upošteva oba kriterija hkrati in uporabniku ponudi paleto hibridnih klasifikatorjev od čim bolj točnih do čim bolj razumljivih. Cilj opisanega algoritma je razdeliti atributni prostor (oz. učno množico) v: podprostore (oz. gruče) znotraj katerih so si učni primeri zelo podobni in jih je mogoče klasificirati z lahko razumljivimi klasifikatorji; in podprostore (oz. gruče) znotraj katerih razumljivi kalsifikatorji ne dosegajo visoke točnosti ter je za njih edino smiselno uporabiti kompleksne klasifikatorje, ki dosežejo visoko klasifikacijsko točnost.

Za primer si oglejmo učno množico na sliki 1, kjer vsaka od osi predstavlja vrednost enega od dveh atributov, znak pa predstavlja razred, ki ga želimo napovedati (krog ali kvadrat). Vidimo, da je levi in spodnji del učne množice na sliki 1 mogoče pojasniti oz. klasificirati npr. z razumljivimi odločitvenimi drevesi. Klasifikacija desnega zgornjega dela učne množice na sliki 1 z relativno majhnim odločitvenim drevesom ali podobno razumljivim klasifikatorjem pa povzroči nizko klasifikacijsko točnost, zato je na tem delu atributnega prostora smiselno uporabiti kompleksnejši klasifikator in na račun tega izgubiti nekaj razumljivosti. Algoritem, ki je predstavljen v tem prispevku, skuša samodejno razdeliti atributni prostor v smiselne podporstore (kot v zgornjem primeru) in vsakemu od njih prirediti primeren klasifikator: razumljiv, če ima dovolj visoko točnost, sicer pa black-box.



Slika 1: *Primer učne množice*

V drugem razdelku je predstavljen predlagan algoritem, ki učinkovito dosega zgoraj opisan cilj. Razdelek 2.1 vsebuje kratek oris algoritma, naslednji podrazdelki pa podrobno opisujejo algoritem in razloge za izbiro hevristik. V tretjem razdelku so predstavljeni prvi eksperimentalni rezultati opisanega algoritma. V zadnjem razdelku so podani zaključki in ideje za nadaljnjo delo.

2. Algoritem

Ta razdelek opisuje delovanje predlaganega algoritma za gradnjo množice hibridnih klasifikatorjev s pomočjo

hierarhičnega gručenja. Prvi podrazdelek oriše celoten algoritem, drugi podrazdelek opisuje podrobnosti v zvezi s hierarhičnim gručenjem in merami podobnosti med primeri in gručami, tretji podrazdelek opisuje podrobnosti ostalih korakov algoritma.

2.1. Oris algoritma

Psevdokoda algoritma za gradnjo množice hibridnih klasifikatorjev na osnovi hierarhičnega gručenja je prikazana na sliki 2. V prvem koraku algoritem uporabi aglomerativno hierarhično gručenje učnih primerov za gradnjo hierarhije gruč (slika 4), ki delijo atributni prostor v posamezne podprostore. Pri združevanju gruč algoritem skuša najprej združiti gruče, ki so si blizu v atributnem prostoru (učni primeri imajo podobne vrednosti atributov) in je primere v združenih gručah mogoče klasificirati z razumljivim klasifikatorjem z visoko točnostjo.

```
1 S hierarhičnim gručenjem razdeli atributni prostor.
2 Za velike gruče zgradi klasifikatorja (drevo, BB).
3 Hierarhijo prereži na smiselnih višinah.
4 Za vsako višino: {
      zgradi vse možne klasifikatorje tako da v listih
      izbereš drevo ali BB. }
5 Obdrži le nedominirane klasifikatorje.
```

Slika 2: *Psevdokoda predlaganega algoritma*

V drugem koraku algoritem vsaki gruči, ki vsebuje dovolj veliko število primerov, priredi dva klasifikatorja: enega, ki strmi k čim večji razumljivosti (klasifikacijsko drevo), in drugega, ki strmi k čim večji točnosti (black-box klasifikator (BB)). Za gradnjo razumljivih klasifikatorjev trenutna različica algoritma uporablja algoritem C4.5, ki gradi odločitvena drevesa. Za gradnjo točnih klasifikatorjev lahko uporabnik izbere katerikoli drug algoritem iz programskega paketa Weka [5], ki je primeren za klasifikacijo zahtevnejših delov atributnega prostora.

V tretjem koraku algoritem zgrajeno hierarhijo prereže na smiselnih višinah, tako da obdrži le vrh hierarhije, ki predstavlja delitev atributnega prostora v posamezne podprostore. Slika 5 kaže enega od možnih prerezov hierarhije iz slike 4. Hierarhije ni smiselno prerezati prenizko, ker s tem dobimo delitev, ki vsebuje veliko število podprostorov za katere so klasifikatorji naučeni na premajhnih učnih množicah, kar povzroči preveliko prileganje učni množici (ang. overfitting), preveliko kompleksnost in s tem premajhno razumljivost hibridnega klasifikatorja in veliko verjetnost napačne klasifikacije zaradi prevelikega števila podprostorov (oz. gruč).

V četrtem koraku algoritem za vsako smiselno delitev atributnega prostora v podprostore (gruče) pridobljeno v tretjem koraku zgradi vse možne hibridne klasifikatorje, tako da uporabi vse možne kombinacije prireditve razumljivega ali točenega klasifikatorja vsaki od gruč. Slika 6 prikazuje vseh 16 možnih kombinacij za delitev v 4 gruče.

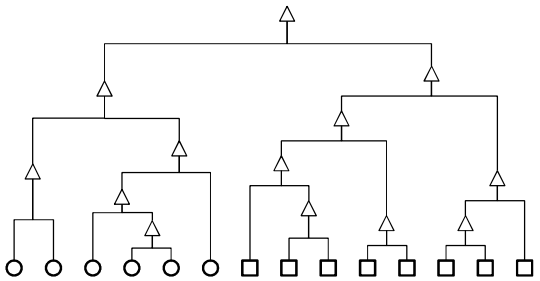
V zadnjem koraku algoritem oceni točnost in razumljivost vseh najdenih hibridnih klasifikatorjev in kot rezultat vrne le klasifikatorje, ki so nedominirani v prostoru razumljivosti in točnosti. Nedominiran klasifikator C je tak, za katerega ne obstaja noben drug klasifikator, ki bi bil hkrati bolj točen in bolj razumljiv od C [1].

2.2. Hierarhično gručenje

Za hierarhično gručenje učnih primerov se uporablja algoritem za aglomerativno hierarhično gručenje prikazan na sliki 3. Algoritem uporablja matriko D v kateri vsak elemnt $D_{i,j}$ predstavlja razdaljo med gručama i in j ter seznam C v katerem so shranjene trenutne gruče. Algoritem inicializira seznam C tako, da vanj doda vsak učni primer kot ločeno gručo. V matriki D so po inicializaciji razdalje med posameznimi primeri. V vsakem koraku algoritem poišče gruči C_i in C_j, ki sta si najbliže ($d(C_i,C_j) \leq d(C_k,C_l)$, $k,l \in \{1..N\}\setminus\{i,j\}$), ter ju združi v novo gručo C_{nov}. Nato iz množice C odstrani gruči C_i in C_j ter doda gručo C_{nov}. Iz matrike D odstrani i-to in j-to vrstico in stolpec ter doda novo vrstico in stolpec, ki predstavlja razdalje med novo gručo C_{nov} in ostalimi gručami v seznamu C. Algoritem se izvaja, dokler ne ostane ena sama gruča v kateri so združeni vsi učni primeri. Primer rezultat hierarhičnega gručenja je prikazan na sliki 4.

```
D_{i,j} = d(I_i,I_j);        //v tabelo razdalj dodaj razdalje gruč
C = {new Gruča(I_i); i = 1 .. N},        //seznam gruč
št_G = N;                                //trenutno št. gruč
while (št_G >1) {
    (C_i,C_j) =najdiMin(D);        //najdi najbližji gruči
    C_nov = združi(C_i,C_j);    //združi gruči C_i in C_j v C_nov
    C = C − C_i − C_j + C_nov;        //posodobi množico C
    D = posodobi(D);          //izračunaj nove razdalje
}
```

Slika 3: *Algoritma za hierarhično gručenje*



Slika 4: *Rezultat gručenja je hierarhija učnih primerov*

Pri delovanje algoritma za gručenje sta ključnega pomena meri razdalje med dvema primeroma in razdalje med dvema gručama. V naslednjih dveh podrazdelkih so opisane mere, ki se najpogosteje uporabljajo pri hierarhičnem gručenju, ter njihove slabosti in novi meri razdalj, ki jih uporablja opisani algoritem.

2.2.1. Razdalja med primeri

Vsak primer lahko opišemo kot vektor v atributnem prostoru. Elementi vektorja so realna števila, ko gre za numerične atribute, ali elementi neurejene končne množice, ko gre za nominalne atribute. Razdalja med primeroma a in b je torej razdalja med vektorjema (a_1, a_2, …, a_n) in (b_1, b_2, …, b_n) , ki opisujeta primera. Najpogosteje uporabljene mere razdalje med dvema vektorjema so npr. Euklidska, Manhattanska in Hammingova.

Euklidska razdalja med vektorjema *a* in *b* je definirana kot $\|a-b\|_2 = \sqrt{\sum_i (a_i - b_i)^2}$. Za naše potrebe ni primerna zaradi naslednjih treh razlogov. Numerični atributi z večjim razponom vrednosti k skupni razdalji prispevajo bolj kot numerični atributi z manjšim razponom vrednosti. Mera ne definira razdalje med nominalnimi atributi; preprosta rešitev je, da za razdaljo med nominalnima atributoma z enakima vrednostma določimo 0, med atributoma z različnima vrednostma pa 1. Prav tako je standardno definicijo Evklidske razdalje potrebno dopolniti za primer v katerem vrednost atributa vsaj enega primera ni znana: v tem primeru lahko postopamo, kot da atributa sploh ne bi bilo, vendar so potem razdalje med primeri, ki imajo manjkajoče vrednosti, manjše kot razdalje med primerljivimi primeri brez manjkajočih vrednosti.

Manhattansak razdalja med vektorjema *a* in *b* je definirana kot $\|a-b\|_1 = \sum_i |a_i - b_i|$ in ima iste pomanjkljivosi kot Evklidska razdalja. Razlikuje se le v tem, kako so kombinirane razdalje med komponentami vektorjev: Manhattanska razdalja uporablja vsoto absolutnih razlik, Evklidska pa koren iz vsote kvadratov razlik. Evklidska mera zato v primerjavi z Manhattansko upošteva večje razlike med vrednostma istoležnih atributov močneje ($x^2 > |x|$, $|x|>1$).

Hammingova razdalja med vektorjema *a* in *b* je definirana kot $\|a-b\|_h = \sum_i h(a_i, b_i)$, kjer je $h(x,y) = 0$ če je $x = y$; 1 sicer. Hammingova razdalja je najbolj primerna za ocenjevanje razdalje med primeri z nominalnimi atributi. Za ocenjevanje razdalje med primeri z numeričnimi atributi pa ni primerna, ker so po tej meri vse vrednosti atributov, ki niso povsem enake, med seboj oddaljene za 1. V praksi to pomeni, da že sam šum, ki je prisoten v učni množici, močno prispeva k razdalji med numeričnimi atributi. Prav tako ima mera enake težave z manjkajočimi vrednostmi, kot ostali dve predstavljeni meri.

Poleg predstavljenih mer se pogosto uporabljajo tudi druge, kot so maksimalna, cosinusna ali Levenshteinova razdalja. Ker imajo pogosto uporabljene mere naštete slabosti, smo definirali novo mero razdalje, ki primerno obravnava numerične in nominalne atribute ter manjkajoče vrednosti in ki ni odvisna od razpona vrednosti numeričnih atributov. Mero razdalje med učnima primeroma *a* in *b* smo definirana kot :

$$\|a-b\|_r = \frac{1}{\#znanih\ atr} \sum_{i \in \{znarn\ atr\}} f(a_i, b_i)$$

kjer je za numerične atribute *i*

$$f(a_i, b_i) = min\left\{\frac{|a_i - b_i|}{c_i}, 1\right\}$$

in za nominalne atribute *i*

$$f(a_i, b_i) = h(a_i, b_i).$$

Uporabnik sam določi mejno vrednost oz. normalizacijskjo konstanto c_i za vsak numeričen atribut. Smiselne vrednosti za c_i so na primer varianca atributa *i* v učni množici ali povprečna absolutna razdalja vrednosti atributa *i* od povprečne vrednosti atributa *i* v učni množici.

2.2.2. Razdalja med razredi

Poleg mere razdalje med primeri je potrebno definirati tudi postopek za izračun razdalje med grupami, ki vsebujejo več kot en primer. V literaturi se najpogosteje uporabljajo min, max in average mere s katerimi dosežemo t.i. single-, complete, in average-linkage gručenje.

Minimalna razdalja med grupama *A* in *B* je definirana kot $min(A,B) = \min_{a \in A, b \in B} \{d(a,b)\}$, kjer je *d*(*a*,*b*) neka mera razdalje med primeroma *a* in *b*. Minimalna razdalja za naše potrebe ni primerna, ker povzroči gradnjo "dolgih in ozkih" gruč. Maksimalna razdalja med grupama *A* in *B* je definirana kot $max(A,B) = \max_{a \in A, b \in B} \{d(a,b)\}$. Slabost te mere je, da daje prevelik pomen primerom, ki ležijo daleč od večine ostalih primerov v gruči. Najbolj primerna mera, ki se pogosto uporablja, je tako povprečna razdalja, ki je definirana kot

$$m(A,B) = \frac{1}{|A \cup B|} \sum_{a \in A} \sum_{b \in B} d(a,b)$$

Povprečna razdalja sicer teži k gradnji lokalno omejenih gruč, vendar pri združevanju gruč ne upošteva drugega želenega kriterija: podobnosti gruč v smislu delovanja klasifikatorja oz. razumljivosti klasifikatorja v združeni gruči. Zaradi tega opisan algoritem za združevanje gruč uporablja naslednjo mero razdalje med grupama *A* in *B*:

$$dist(A,B) = \begin{cases} m(A,B)\ ; |A \cup B| < c \\ m(A,B)\ ; m(A,B) \geq w(A,B) \\ 0.5\ w(A,B) + 0.5\ m(A,B);\ \text{sicer} \end{cases}$$

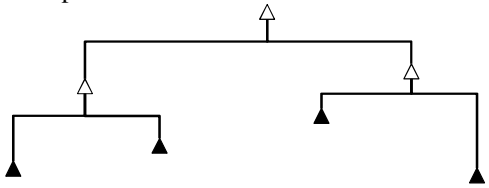
Kjer je

$$w(A,B) = 1 - \frac{\#corr\left(C_B(A)\right) + \#corr\left(C_A(B)\right)}{|A \cup B|}$$

Predlgana mera za združevanje gruč z majhnim številom primerov uporablja povprečno razdaljo, za združevanje gruč, ki so dovolj velike, da je na njih mogoče zanesljivo naučiti klasifikatorje, pa ji prišteje povprečno klasifikacijsko napako klasifikatorja za grupo *A* na primerih iz grupe *B* in klasifikatorja za grupo *B* na primerih iz grupe *A*.

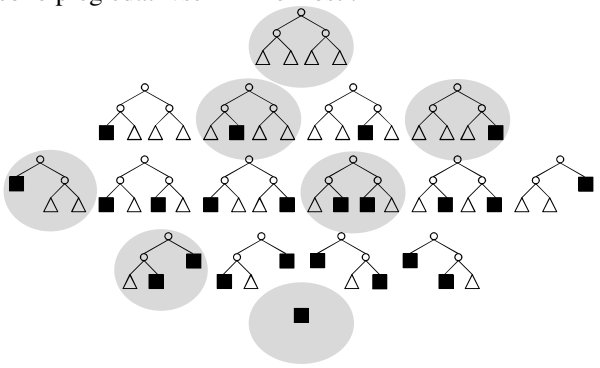
2.3. Iskanje nedominirahin klasifikatorjev

V drugem koraku algoritma se hierarhija gruč prereže na smiselnih višinah, tako da iz celotne hierarhije vzamemo le vrhnji del, kot je npr. prikazan na sliki 5. Hierarhije ni smiselno odrezati prenizko, ker bi tako dobili preveč gruč, kar bi pomenilo preveliko kompleksnosti hibridnega klasifikatorja, hkrati pa bi bilo veliko gruče vsebovalo premalo učnih primerov, da bi v njih lahko zgradili zanesljive klasifikatorje. Maksimalno število gruč naj bo manjše od ~10 in v večini gruči naj bo vsaj približno 10 % ali več učnih primerov iz celotne učne množice.



Slika 5: *Primer prereza hierarhije iz slike 4*

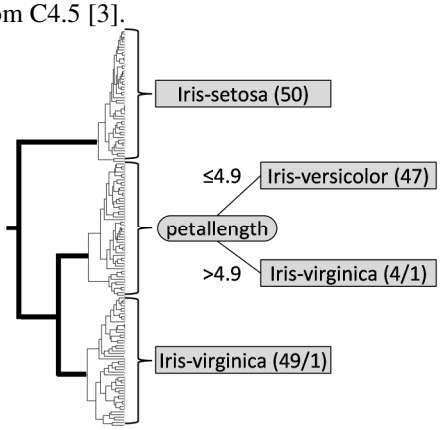
Ko s prerezom hierarhije razdelimo učno množico v gruče, lahko vsaki gruči priredimo razumljiv ali točen klasifikator. Če je v razdelitvi n gruč je število vseh možnih hibridnih kalsifikatorjev za dano razdelitev atributnega prostora 2^n, kar ni zanemarljivo. Na sliki 6 je prikazana množica hibridnih klasifikatorjev za primer, ko je atributni prostor razdeljen v 4 gruče. Na sliki trikotnik predstavlja odločitveno drevo, črn kvadrat predstavlja t.i. black-box klasifikator, sivo ozadje pa označuje nedominiran kalsifikator. Za iskanje nedominiranih hibridnih klasifikatorjev uporabimo algoritem [2], ki na učinkovit način generira nedominirane klasifikatorje in mu zato ni potrebno pregledati vseh 2^n možnosti.



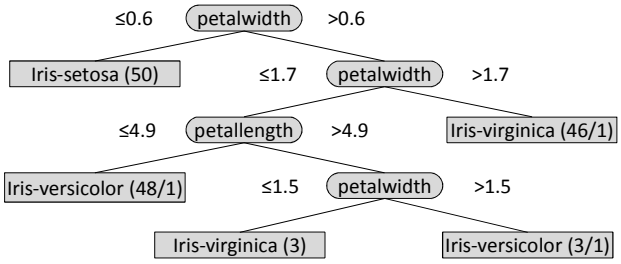
Slika 6: M*nožica hibridnih klasifikatorjev*

3. Rezultati

Opisan algoritem smo preizkusili na štirih standardnih učnih množicah Iris, Lenses, Weather in Labor iz UCI ML repozitorija [4]. Za vsako od štirih učnih množic je algoritem odkril hibridne klasifikatorje, ki se po točnosti lahko primerjajo s pogosto uporabljenimi algoritmi, kot so naivni Bayes, odločitvena drevesa in metoda podpornih vektorjev. Razumljivost hibridnih klasifikatorjev je bila vsaj taka kot (pogosto pa je celo prekašala) razumljivost odločitvenih dreves s primerljivo točnostjo dobljenih z algoritmom C4.5 [3].



Slika 7: *Hibridni klasifikator s tremi podprostori za učno množico Iris*



Slika 8: *Odločitveno drevo (C4.5) za učno množico Iris*

Na sliki 7 je prikazan hibridni kalsifikator za učno množico Iris, ki atributni prostor razdeli v tri gruče. V dveh gručah se uporablja odločitveno drevo z enim samim listom, v tretji gruči pa odločitveno drevo z le dvema listoma. Skupno število listov je torej manjše od števila listov odločitvenega drevesa, ki ga zgradi algoritem C4.5 (slika 8), pri tem pa je klasifikacijska točnost hibridnega klasifikatorja višja od točnosti odločitvenega drevesa.

4. Zaključek

V prispevku je predstavljen algoritem za gradnjo množice hibridnih klasifikatorjev na osnovi hierarhičnega gručenja. Cilj algoritma je uporabniku ponuditi paleto kalsifikatorjev od čim bolj točnih do čim bolj razumljivih. To doseže z delitvijo atributnega prostora v podporstore in izbiranjem primernega klasifikatorja za posamezen podprostor: klasifikator za določen podporostor je lahko razumljiv ali točen.

Za določanje primernih podprostorov algoritem uporablja aglomerativno hierarhično gručenje. H gradnji primernih gruč ga usmerjata posebej za ta namen določeni meri razdalje med primeroma in med gručama primerov. Za izbiro med razumljivim in točnim klasifikatorjem podporstora v zadnji fazi algoritma je uporabljen učinkovit algoritem, ki se izogne pregledovanju vseh možnih kombinacij [2].

Prvi testi algoritma na standardnih učnih množicah kažejo, da algoritem dosega točnost, ki je primerljiva z najpogosteje uporabljenimi algoritmi za strojno učenje, in razumljivost, ki je pogosto boljša od razumljivosti odločitvenih dreves s primerljivo točnostjo zgrajenih z algoritmom C4.5. Poleg tega algoritem uporabniku omogoča lažje razumevanje zakonitosti v učnih podatkih, saj jih razdeli v podskupine, v katerih je mogoče hitro opaziti povezavo med vrednostmi atributov in razredom.

V nadaljnjem delu nameravamo algoritem bolj temeljito preizkusiti na večjem številu učnih množic ter odkriti morebitne pomanjkljivosti in jih odpraviti. Prav tako se ponuja obilo možnosti za preizkušanje različnih mer razdalj med primeroma in med gručama ter odpravo morebitnega prevelikega prileganja zgrajenih klasifikatorjev učnim podatkom. Predvidena je tudi izboljšava uporabniškega vmesnika in izboljšanje integracije z obstoječimi programskimi paketi za strojno učenje kot je npr. Weak.

Reference

[1] K. Deb. Multi-Objective Optimisation using Evolutionary Algorithms, Wiley, 2001.

[2] R. Piltaver. "Iskanje točnih in razumljivih hibridnih klasifikacijskih dreves", *Zbornik 13. mednarodne multikonference Informacijska družba - IS 2010*, Zv. A, str. 82-85, 2010.

[3] R. Quinlan. "*C4.5: Programs for Machine Learning,*" Morgan Kaufmann Publishers, San Mateo, CA, 1993.

[4] UC Irvine Machine Learning Repository, dostopno na: http://archive.ics.uci.edu/ml/

[5] I. H. Witten, E. Frank. Data Mining: Practical Machine Learning Tools and Techniques (2nd edition), Morgan Kaufmann, 2005.

SEMANTIČNI IN SPLOŠNI PRISTOP K PREPOZNAVANJU ZDRAVSTVENIH TEŽAV STAREJŠIH

Bogdan Pogorelc
Department of Intelligent Systems
Jožef Stefan Institute
Jamova 39, 1000 Ljubljana, Slovenia

in

Špica International d.o.o.
Pot k sejmišču 33
1231 Ljubljana, Slovenia
Tel: +386 1 477 32 30
e-mail: bogdan.pogorelc@ijs.si

POVZETEK

Članek predlaga semantični in splošni pristop k prepoznavanju zdravstvenih težav starejših. Gibanje oseb je zajeto s sistemom za zajem gibanja in izhodne časovne vrste koordinat so modelirane z obema predlaganima pristopoma. Semantični pristop uporablja atribute na osnovi medicinskega znanja (semantične atribute) in tipično klasifikacijsko metodo podpornih vektorjev. Splošni pristop ima za atribute vse izmerljive kote sklepov in klasificira s kombinacijo algoritma k-najbližjih sosedov in modifikacijo algoritma dinamičnega ukrivljanja časa (DTW). Kljub temu da je drugi pristop splošnejši in uporaben tudi za druge klasifikacijske metode, doseže primerljivo visoko klasifikacijsko točnost kot semantični pristop.

1 UVOD

Razvite države se soočajo s hitro rastjo svojega prebivalstva. Do leta 2050 naj bi se število ljudi nad 60 let v Evropi podvojilo na 40% celotnega prebivalstva oz. na 60% aktivnega prebivalstva [1]. Starejši navadno živijo izolirani od potomcev, zato v primeru bolezni težko dobijo pravočasno pomoč. Namen te študije je razviti tehnologije, ki bi olajšale njihovo samostojno življenje.

Predlagamo dva pristopa podatkovnega rudarjenja k inteligentnemu in vseprisotnemu sistemu nadzora zdravja z namenom razpoznati nekaj najpogostejših in najpomembnejših bolezni starejših, ki so lahko razpoznane preko opazovanja in analize karakteristik njihovega gibanja. Semantični pristop uporablja atribute na osnovi medicinskega znanja (semantične atribute) in metodo podpornih vektorjev. Splošni pristop ima za atribute vse izmerljive kote sklepov v kombinaciji z algoritmom k-najbližjih sosedov in dinamičnega ukrivljanja časa (DTW). Naloga je klasificirati vzorce hoje v pet različnih zdravstvenih stanj, eno zdravo in štiri bolezenska.

Gibanje uporabnika je zajeto s sistemom za zajem gibanja, ki sestoji iz značk, pritrjenih na telo in senzorjev, nameščenih v stanovanju. Izhodne časovne vrste koordinat so obdelane s predlaganima pristopoma, da bi razpoznali specifično zdravstveno težavo.

2 SORODNO DELO

V sorodnem delu je zajem gibanja navadno narejen z inercialnimi senzorji [2, 5], s strojnim vidom in tudi s specifičnim senzorjem za merjenje kota upognjenosti sklepa [3] ali z elektromiografijo [4]. V naši študiji smo uporabili sistem (infrardeče) IR kamere z značkami pritrjenimi na telo. Ne naslavljamo samo razpoznave značilnih aktivnosti, kot je hoja, sedenje, ležanje, itd., kot je realizirano npr. v [6, 8], ampak tudi razpoznavamo zdravstvene težave. Z uporabo podobnega sistema za zajem podatkov so v [7] ločevali med hemiplegijo in diplegijo.

Bolj pogosti pristop iz sorodnega dela je zajem podatkov s sistemom za zajem gibanja in kasnejša ročna analiza podatkov [3, 4, 9]. Tak pristop ima pomanjkljivost v primerjavi z našim, da zahteva stalno pregledovanje strokovnjakov.

3 MATERIALI IN METODE

3.1 Ciljne zdravstvene težave za detekcijo

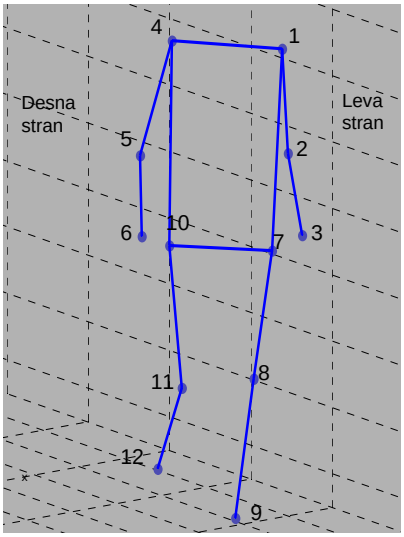
Vse zdravstvene težave, ki jih prepoznavamo, so bile predlagane s strani sodelujočega medicinskega strokovnjaka, na osnovi pogostosti nad 65 let starosti, medicinske pomembnosti in možnosti razpoznavanja iz gibanja. Sistem hojo klasificira kot:

- hemiplegijo (navadno po možganski kapi),

- Parkinsonovo bolezen,
- bolečino v nogi,
- bolečino v hrbtu in
- referenčno zdravo hojo.

3.2 Atributi za podatkovno rudarjenje

Meritve sestavljajo pozicije koordinat v *x*,*y*,*z* za 12 značk nošenih na ramenih, komolcih, zapestjih, kolkih, kolenih in gležnjih, zajete s sistemom za zajem gibanja Smart z 10 Hz, kot prikazuje Slika 1. Primeren prikaz uporabnikovega gibanja je bil pomemben del naše študije.



Slika 1: Pozicije značk na telesu

Semantični pristop smo zasnovali na osnovi dejstva, da zdravnik diagnosticira obravnavane zdravstvene probleme iz opazovanja hoje [10]. Ker so si vzorci podobni, mora biti pozoren na veliko detajlov, ki smo jih poskušali zapisati z merljivimi spremenljivkami. Za nalogo avtomatske prepoznave bolezni smo predlagali in testirali uporabo naslednjih 13 značilk (predvidevajo, da je oseba prizadeta po desni strani, ker če bi bila po levi, bi se strani zamenjale):

1. Absolutna razlika povprečnih razdalj desni komolec – desni kolk in desno zapestje – levi kolk.
2. Povprečni kot desnega komolca.
3. Kvocient med maksimalnim kotom levega in maksimalnim kotom desnega kolena.
4. Razlika med maksimalnim in minimalnim kotom desnega kolena.
5. Razlika med maksimalno in minimalno višino levega ramena.
6. Razlika med maksimalno in minimalno višino desnega ramena.
7. kvocient {razlike med maksimalno in minimalno višino levega in maksimalno in minimalno višino desnega gležnja}
8. Absolutna razlika {razlike maksimalne in minimalne hitrosti levega in razlike maksimalne in minimalne hitrosti desnega gležnja}
9. Absolutna razlika povprečne razdalje desna rama – levi komolec in leva rama – desno zapestje.
10. Povprečna hitrost desnega zapestja.
11. Pogostost desnega komolca prečkanja svoje povprečne vrednosti kota.
12. Povprečni kot med vektorjem {desna rama – desni kolk} in vektorjem {desna rama – desno zapestje}
13. Absolutna razlika povprečnih višin desnega in levega ramena.

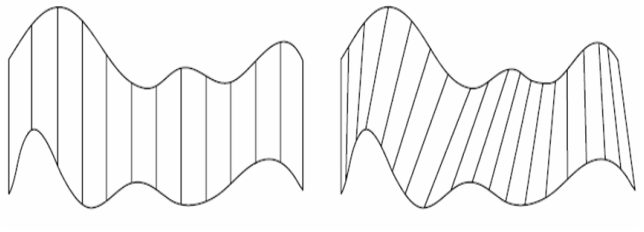
Pri splošnem pristopu je gibanje predstavljeno z enostavnimi in splošnimi atributi, da bo klasifikator s temi atributi delal dobro tudi na drugačnih gibanjih, saj zajamemo le majhen del vseh možnih gibanj. Upoštevajoč našteto smo zasnovali atribute kot kote med sosednjimi deli telesa:

- kot levega in desnega ramena glede na zgornji del trupa v trenutku *t*:
$$q^t_{SL} \text{ in } q^t_{SL},$$
- kot levega in desnega kolka glede na spodnji del trupa
$$q^t_{HL} \text{ in } q^t_{HR},$$
- kot med spodnjim in zgornjim delom trupa:
$$q^t_T \text{ ter}$$
- levi in desni komolčni ter levi in desni kolenski kot:
$$\alpha^t_{EL}, \alpha^t_{ER}, \alpha^t_{KL} \text{ in } \alpha^t_{KR}.$$

Koti med deli telesa, ki rotirajo v več kot eno smer, so izraženi s kvaternioni.

3.3 Dinamično ukrivljanje časa

Dinamično ukrivljanje časa (DTW) poravna 2 časovni vrsti na način, da minimizira neko mero. Optimalna poravnava je dobljena s preslikavo več zaporednih vrednosti ene časovne vrste v eno vrednost druge časovne vrste in tako je lahko DTW računan tudi na časovnih vrstah različnih dolžin. V nasprotju z Evklidsko razdaljo, DTW lahko najde podobnosti med vzorcema dveh časovnih vrst, tudi če ta vzorca nista časovno poravnana ali pa sta vzorca različnih dolžin (Slika 2).



Slika 2: Primer dveh časovnih vrst. Črte med časovnima vrstama predstavljajo poravnavo ob uporabi evklidske razdalje (levo) in DTW (desno).

3.4 Modifikacija: multivariantno dinamično ukrivljanje časa

DTW algoritem, ki je navadno opisan v literaturi, je uporabljen le za poravnavo univariantnih časovnih vrst. Splošni pristop te študije pa poravnava multivariantne časovne vrste. Najprej je vsaka točka zajete časovne vrste pretvorjena v prostor kotnih atributov, kjer bo izvedena klasifikacija.
Imamo testno meritev, ki jo želimo poravnati z učno meritvijo (kjer je bil klasifikator naučen) in najprej izračunamo matriko lokalnih razdalj *d(i,j)*, v katerem vsak element (*i*,*j*) predstavlja lokalno razdaljo med *j*-to časovno točko učne in *i*-to časovno točko testne meritve. Naj bo L_{jf} element generičnega atributnega vektorja glede na učno meritev in T_{if} naj bo element atributnega vektorja, relativno na novo testno meritev za razpoznavo, kjer je

$$1 \le f \le N$$

upoštevani atribut. Za definicijo lokalne razdalje je bila uporabljena Evklidska razdalja, definirana kot

$$d(i,j) = \sum_{f=1}^{N} \left(L_{jf} - T_{if} \right) .$$

Na osnovi matrike lokalnih razdalj je zgrajena matrika globalnih razdalj *D*. Končni izhod algoritma je vrednost minimalne globalne razdalje za celotno poravnavo DTW in je najdena v zadnji vrstici in stolpcu, $D(R_l,C_l)$.

4 EKSPERIMENTI IN REZULTATI

Uporabili smo 256 meritev zdravih posameznikov in posameznikov z določenimi zdravstvenimi težavami, pri čemer je bil vsak posameznik 4-5 krat posnet z različnimi hitrostmi izvajanja aktivnosti.
Pri splošnem pristopu klasifikacijski proces upošteva eno vhodno testno časovno vrsto, ki jo primerja z vsemi ostalimi, da najde minimalno globalno razdaljo za vsako poravnavo in sklepa, da je vhodna meritev istega razreda kot učna meritev, ki ima najmanjšo razdaljo do te vhodne meritve.
Metoda "izpusti-enega" da klasifikacijski točnosti 97.9% oz. 97.6 % za semantični oz. splošni pristop.
Za dejansko izvedbo lahko matrike zamenjav (Tabela 1 in Tabela 2) uporabimo za:

- **Napačni pozitivi (napačni alarmi):** Koliko jih lahko pričakujemo. Če v realnosti sistem javi napačni alarm, npr. zdrava hoja je klasificirana kot zdravstvena težava, bi moral rešilni avto po nepotrebnem priti po uporabnika, kar bi predstavljalo nepotrebno nadlogo in strošek.
- **Napačni negativi:** Koliko jih lahko pričakujemo. Predstavljajo potencialno tvegano situacijo, saj zdravstvena težava ni razpoznana.
- **Napake:** Med katerimi zdravstveni stanji se pojavijo zamenjave. Da jih odpravimo, lahko dodamo dodatne značilke za te razrede. So zelo redke.

		klasificiran kot				
		H	**L**	**N**	**P**	**B**
pravi razred	**H**	45	0	0	0	0
	L	1	24	0	0	0
	N	0	0	25	0	0
	P	2	0	0	23	0
	B	0	0	0	0	21

Tabela 1. Matrika zamenjav za semantični pristop, kjer je H=hemiplegia, L=bolečina v nogi, N=normalna (zdrava) hoja, P=Parkinsonova bolezen and B=bolečina v hrbtu. Številke ponazarjajo količino klasificiranih primerov.

		klasificiran kot				
		H	**L**	**N**	**P**	**B**
pravi razred	**H**	42	2	1	0	0
	L	0	25	0	0	0
	N	1	0	24	0	0
	P	0	0	0	25	0
	B	0	0	0	0	21

Tabela 2. Matrika zamenjav za splošni pristop, kjer je H=hemiplegia, L=bolečina v nogi, N=normalna (zdrava) hoja, P=Parkinsonova bolezen and B=bolečina v hrbtu. Številke ponazarjajo količino klasificiranih primerov.

Napačnih pozitivov, napačnih negativov in napak je zelo malo v obeh pristopih, kar je pomembno za dejansko implementacijo.

5 ZAKLJUČEK

Članek predstavlja semantični in splošni pristop k detekciji zdravstvenih težav za namen podaljšanja

samostojnega življenja starejših. Metoda "izpusti-enega" da klasifikacijski točnosti 97.9 % oz. 97.6 % za semantični oz. splošni pristop. Semantični pristop je zahtevnejši za izvedbo zaradi gradnje specifičnih semantičnih atributov, ki zahtevajo medicinsko znanje. Kljub temu, da je drugi pristop bolj splošen in lahko razpozna tudi nove vrste gibanj, dosega visoke klasifikacijske točnosti, podobne semantičnemu pristopu.

ZAHVALA

Operacije, ki so pripeljale do tega članka, je delno sofinancirala Evropska Unija, Evropski socialni sklad. Avtor članka se zahvaljuje mentorju prof dr. Matjažu Gamsu za pomoč.

REFERENCE

1. Toyne S., Ageing: Europe's growing problem, BBC News.
2. Strle D., Kempe V., "MEMS-based inertial systems", MIDEM 37(2007)4, pp. 199-209.
3. Ribarič S., Rozman J., "Sensors for measurement of tremor type joint movements", MIDEM 37(2007)2, pp. 98-104.
4. Trontelj J., et al., "Safety Margin at mammalian neuromuscular junction – an example of the significance of fine time measurements in neurobiology", MIDEM 38(2008)3, 155-160.
5. Bourke A.K et al. An optimum accelerometer configuration and simple algorithm for accurately detecting falls. In Proc. BioMed 2006 (2006), 156–160.
6. Confidence. http://www.confidence-eu.org.
7. H. Lakany, Extracting a diagnostic gait signature. Patt. recognition 41(2008), 1627–1637.
8. Luštrek, M., and Kaluža, B. Fall detection and activity recognition with machine learning. Informatica 33, 2 (2009).
9. Moore ST, et al., Long-term monitoring of gait in Parkinson's disease, Gait Posture (2006).
10. Perry J. Gait Analysis: Normal and Pathological Function. McGraw-Hill, Inc., 1992.

Zasnova govorne zbirke za sintetizator slovenskega govora Amebis Govorec

Tomaž Šef [1], *Miro Romih* [2]
[1] Odsek za inteligentne sisteme, Institut "Jožef Stefan", Jamova cesta 39, 1000 Ljubljana
[2] Amebis, d.o.o., Kamnik, Bakovnik 3, 1241 Kamnik
tomaz.sef@ijs.si, miro.romih@amebis.si

POVZETEK

V članku je predstavljena zasnova nove govorne zbirke za korpusno sintezo slovenskega govora, ki jo razvijata Institut Jožef Stefan in podjetje Amebis. Razvit je bil postopek, ki ob vnaprej določenem številu povedi izbere takšne, ki so fonetično čim bolj bogate oz. vsebujejo čim več različnih glasovnih nizov kot so monofoni, difoni, trifoni ter ostali polifoni.

1 UVOD

Govorne tehnologije, predvsem sinteza in razpoznavanje govora ter razpoznavanje govorcev, nezadržno prodirajo v naše življenje [1]. Na tržišču se je do sedaj pojavilo že kar nekaj zelo dobrih sintetizatorjev govora; predvsem za »večje« svetovne jezike. Na razpolago so tudi sintetizatorji govora, ki podpirajo slovenski jezik, vendar je njihova kvaliteta zaenkrat še preslaba, da bi bili sprejemljivi za najširši krog uporabnikov. Sintetizatorji govora se uporabljajo kot pomoč slepim in slabovidnim osebam, v samodejnih informacijskih centrih, govornih portalih ipd.

Najbolj naravno zveneči sintetizatorji govora temeljijo na korpusni sintezi. Metoda temelji na preiskovanju vnaprej posnete in označene govorne zbirke. Išče se zaporedja tistih posnetih glasov pri katerih se želene lastnosti čim bolj ujemajo. Kvaliteta takšnih sintetizatorjev govora je predvsem odvisna od zasnove govorne zbirke na kateri temeljijo. V splošnem velja, da je sintetizator govora kvalitetnejši, če uporabljamo za sintezo daljše osnovne segmente s čim manj spremembami prozodičnih parametrov. Daljši kot so osnovni segmenti, večje je njihovo število. Število segmentov veča tudi želja po minimalnih spremembah prozodičnih parametrov, saj te povzročajo dodatna popačenja sintetiziranega govora [2].

Govorne zbirke so nepogrešljive pri raziskovalnem delu na področju govornih tehnologij in predstavljajo pomemben člen osnovne infrastrukture za razvoj govornih tehnologij za posamezno jezikovno področje [3]. Vsebujejo predvsem računalniško berljive posnetke govora, ki so jim vedno priloženi še podatki, ki opisujejo posneti govor (govorni dejavniki, dejavniki govorcev ter zapisi in označitve posnetega govora) [2]. Vse več govornih zbirk in rezultatov na njih temelječih študij je dostopnih tudi v našem prostoru.

Najpomembnejša dejavnika pri snovanju govorne zbirke za potrebe korpusne sinteze govora sta izbira njene vsebine in označevanje posnetkov. Izbira velikosti govorne zbirke je posledica kompromisa med želenim številom variacij glasov oz. njihovim pokritjem na eni strani ter časom in stroški vezanimi na razvoj na drugi strani. Upoštevati je potrebno tudi čas za kasnejše preiskovanje govorne zbirke in potreben prostor za njeno hranjenje [4]. Kakovostna korpusna sinteza zahteva, da ima govorna zbirka pravilno označeno tako identiteto posamezni govornih segmentov kot njihov natančen položaj znotraj zbirke. Običajno avtomatskim metodam in postopkom sledi »ročno« popravljanje oznak, ki ga je ne glede na hiter razvoj tehnologije še vedno zelo veliko.

Strošek razvoja korpusnih sintetizatorjev je izredno visok, zato je večinoma na razpolago le omejeno število glasov.

2 AMEBIS GOVOREC sintetizator slovenskega govora

Sistem Amebis Govorec za sintezo neomejenega slovenskega govora sestavlja več med seboj povezanih in hierarhično urejenih modulov (slika 1):

- analiza besedila (predobdelava besedila, grafemsko fonemska pretvorba),
- nastavljanje prozodičnih parametrov (trajanje, osnovna frekvenca, amplituda, premori) in
- generiranje govornega signala (izbira osnovne enote, lepljenje, sprememba govornih parametrov).

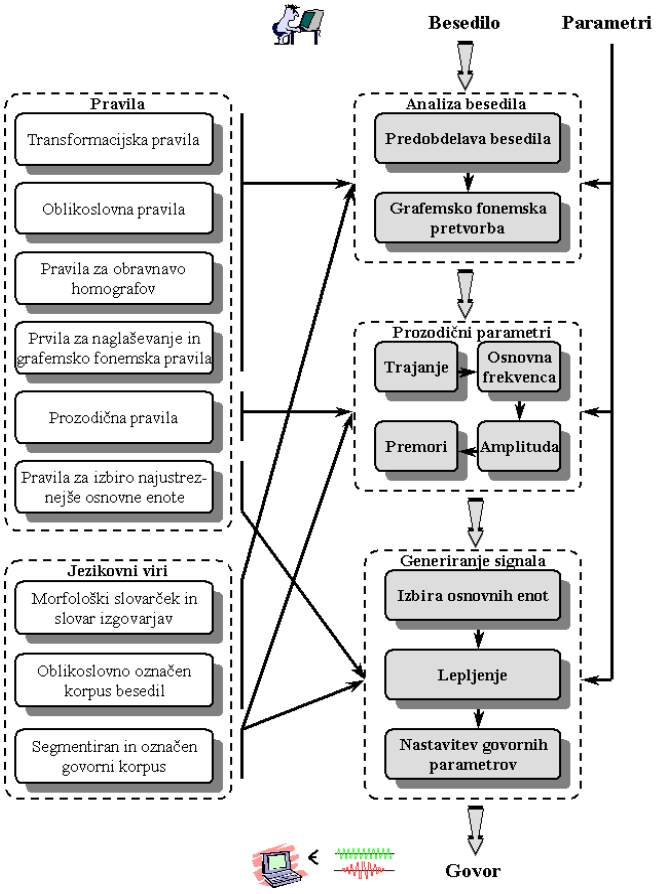
Modularnost ni omejena le na zgornje tri kategorije, ampak se v veliki meri kaže tudi znotraj njih samih. To omogoča preprosto popravljanje in dopolnjevanje posameznih delov sistema [5].

Sistem je zasnovan na obsežnih označenih tekstovnih in govornih korpusih ter je rezultat dolgoletnega raziskovalnega in aplikativnega dela Odseka za inteligentne sisteme na Institutu "Jožef Stefan" in podjetja Amebis.

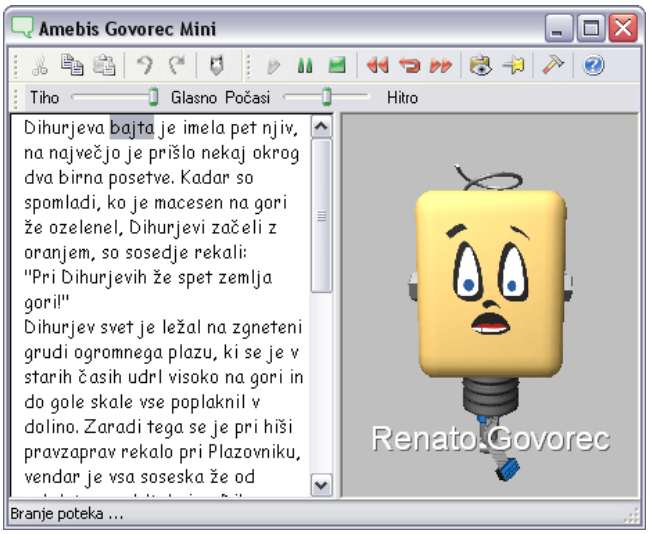
Vhod v sintetizator je poljubno slovensko besedilo v digitalni obliki, shranjeno v različnih formatih. Med samim besedilom lahko nastajajo različne kontrolne oznake, ki služijo krmiljenju sintetizatorja ali pa so mu v pomoč pri doseganju večje naravnosti umetnega govora. Slednjemu služi tudi vrsta nastavljivih parametrov. Tako lahko

spreminjamo način predobdelave besedila, in sicer glede na vrsto besedila, ki ga sintetiziramo (splošna besedila, kratka sporočila, matematični zapisi, časopisni članki, internet, elektronska pošta, črkovanje itd.). Pri nastavljanju prozodičnih parametrov določamo hitrost govora (normalna, hitra, počasna ali katerakoli vmesna stopnja), osnovno frekvenco, pogostnost in trajanje premorov, način naglaševanja besed, izbiramo lahko med različnimi modeli nastavljanja. V modulu za generiranje govornega signala lahko spreminjamo govorca, način določanja optimalnih enot lepljenja, izbiramo metode in algoritme, ki se bodo izvedli (TD-PSOLA, HNM+, linearna interpolacija, spektralno glajenje, normiranje amplitud na mestih lepljenja, kompresija, oblika izhodnega zapisa). Na izhodu dobimo sintetiziran govor, ki ga predvajamo prek zvočne kartice računalnika. Ker je sintetizator v celoti realiziran v obliki računalniškega programa, ni potreb po dodatni strojni opremi.

Cilj vseh trenutnih raziskav s področja govornih tehnologij na Institutu Jožef Stefan je ustvariti sintetizator slovenskega govora, ki bi bil primeren za čim širši krog uporabnikov. K temu bo največ pripomogla prav obsežna govorna zbirka katere zasnova in pridobivanje je podrobneje opisano v nadaljevanju.



Slika 1: *Zgradba sistema Amebis Govorec za sintezo slovenskega govora*



Slika 2: *Uporabniški vmesnik Amebis Govorec Mini*

Sistem Govorec je uporabljen v več aplikacijah:

- **Amebis Govorec** (http://govorec.amebis.si/) – razvijalci programske opreme lahko dodajo govor v svoje programe z le nekaj vrsticami programske kode, priložen pa je tudi enostaven programski vmesnik Amebis Govorec Mini (slika 2),
- državni portal **E-uprava** (http://e-uprava.gov.si),
- pomoč slepim in slabovidnim osebam (uporablja ga **Zveza društev slepih in slabovidnih Slovenije**) - prva nagrada Sklada za nagrajevanje inovacij na področju usposabljanja, življenja in dela invalidov,
- umetniški projekt Saše Sedlačka "**Žicar, robot za socialno ogrožene**",
- razvedrilno-informativna **oddaja Terminal na TV Slovenija**,
- **testira ga več podjetij** za potrebe izvedbe dinamičnih govornih portalov, vgradnje v programe in avtomate v proizvodnih procesih ipd.

3 ZASNOVA GOVORNE ZBIRKE ZA KORPUSNO SINTEZO GOVORA

Razvoj govorne zbirke za korpusno sintezo govora obsega več korakov:

- ustvari se obsežno tekstovno zbirko besedil, ki pokriva različne zvrsti (dnevni časopis, revije, leposlovje ipd.),
- iz zbirke besedil se odstrani vse oznake vezane na oblikovno podobo (glava besedila, tabele ipd.),
- okrajšave, števila ipd. se pretvori v polno besedno obliko (normalizacija besedil),
- besedila se pretvori v predvideni fonetični prepis (grafemsko-fonemska pretvorba),
- optimizira se obseg zbirke glede na vnaprej pripravljene kriterije (metoda požrešnega iskanja); doseči želimo statistično ustrezno vzorčenje izbranega področja govorjenega jezika,

- izbrane stavke se posname (ali pa se izlušči del obstoječih zvočnih zapisov),
- posneto govorno gradivo se fonetično in prozodično označi (samodejno grobo označevanje, fino ročno popravljanje).

Postopek za čim optimalnejšo izbiro povedi:

1. *Statistična obdelava besedil:*
   - Statistično obdelamo celoten besedni korpus in določimo pogostost pojavljanja posameznih glasov in glasovnih nizov v besedilu. Pri tem razlikujemo še med naglašenimi in nenaglašenimi glasovi ter glasovi, ki se pojavljajo na koncu stavka (oz. na mestih zajema zraka - ločila). Presledke na drugih mestih lahko ignoriramo oz. odstranimo.
   - Vključimo vse stavke (povedne, velelne, vprašalne itd.) in izdelamo statistiko posameznih vrst povedi oz. stavkov.
2. *Izdelava spiska glasovnih nizov z oceno zaželenosti posameznega niza:*
   - V spisek vključimo nabor vseh teoretično možnih kombinacij difonov; tudi tiste na katere pri statistični obdelavi nismo naleteli (zaradi robustnosti sintetizatorja govora).
   - V spisek vključimo vse trifone, štirifone in (po potrebi) ostale zaželene (najpogostejše) polifone, na katere smo naleteli pri statistični obdelavi besedil.
   - Utež oz. ocena zaželenosti niza je odvisna od pogostosti njegovega pojavljanja v besedilu.
3. *Postopek izbire povedi:*
   - Ocenimo doprinos glasovnih nizov za vsako poved iz tekstovnega korpusa.
   - Doprinos povedi je enak vsoti vseh ocen zaželenosti nizov (iz spiska), ki se v povedi pojavijo.
   - Doprinos posamezne povedi normiramo z dolžino povedi (št. besed v povedi ali št. fonemov v povedi).
   - Določimo takšno utež, da bodo dolžine izbranih stavkov čim bolj ustrezale statistični porazdelitvi dolžin stavkov iz korpusa.
   - Izberemo poved z najvišjim normiranim doprinosom.
   - Iz spiska odstranimo vse glasovne nize, ki jih izbrana poved vsebuje.
   - Ponovno ocenimo vsako poved in izberemo najboljšo (glede na novi spisek v katerem so izločeni tisti glasovni nizi, ki smo jih že pokrili) ter popravimo spisek.
   - Postopek ponavljamo dokler ne izberemo želenega števila povedi.
4. *Ovrednotenje rezultatov:*
   - Vsakih 1000 povedi izdelamo statistiko difonov, trifonov, štirifonov in drugih polifonov, ki jih že pokrivamo (gre za glasovne nize, ki smo jih do takrat že izločili iz zgoraj omenjenega spiska).
5. *Dodatne izboljšave algoritma:*
   - Ker mora zbirka vsebovati vse možne kombinacije difonov, algoritem popravimo tako, da difone dodatno utežimo glede na ostale polifone. Na takšen način bo algoritem na začetku dajal prednost povedim, ki bodo pokrile čim več novih difonov. Predvidoma se vsi difoni pokrijejo že po ca. 100 stavkih.
   - Pri trifonih in štirifonih upoštevamo pri robnih glasovih tudi podatek o glasovni skupini, ki ji pripadajo (npr. štirifon "krak" ne bo doprinesel prav dosti novega v našo zbirko, če ta že vsebuje štirifon "krat"; zato oceno koristnosti takega štirifona popravimo navzdol). To lahko naredimo preprosto tako, da v spisek vnesemo dodatne nize skupaj z njihovimi frekvencami pojavljanja v korpusu (primer takega štirifona: "k"+"r"+"a"+"pripornik").
   - Algoritem z različnim uteževanjem izboljšamo tako, da končni nabor vsebuje različne povedi (povedne, vprašalne, velelne, enostavne, sestavljene, naštevanje, itd.). Tako lahko isti korpus učinkovito uporabimo tudi za generiranje prozodičnih parametrov pri sintezi govora.

Drugi podobni do sedaj razviti postopki:

1. *Algoritem razvit na Fakulteti za elektrotehniko, računalništvo in informatiko (FERI), Univerza v Ljubljana [6]:*
   - Iz celotnega besedila se izloči vse povedi, ki so krajše od 15 besed in daljše od 25 besed.
   - Generira se štiri manjše zbirke s po 5.000 povedmi ob upoštevanju postavljenih omejitev
   - Vsakemu glasovnemu nizu se določi pogostost pojavljanja v vsaki od štirih majhnih zbirk,
   - Izloči se povedi, ki vsebujejo podvojene glasovne nize. Pri tem se pazi, da število ponavljanj določenega trifona ne pade pod vnaprej določen prag.
   - Prag se znižuje toliko časa, dokler ni v govorni zbirki samo 1200 povedi.
   - Izmed vseh štirih zreduciranih zbirk se izbere najboljšo.
   - Povedi, ki najmanj prinesejo k raznolikosti zbirke se zamenja z najboljšimi povedmi iz preostalih treh zbirk.
2. *Algoritem razvit v podjetju Alpineon ter na Fakulteti za elektrotehniko (FE), Univerza v Ljubljani [2,7]:*
   - Statistično se obdela celotno vhodno besedilo in določi pogostost pojavljanja posameznih glasovnih nizov v besedilu.
   - Izbere se število najpogostejših trifonov in štirifonov, ki se jih želi imeti v zbirki. Odločili so se za 500 najpogostejših trifonov in 300 štirifonov.
   - Oceni se doprinos glasovnih nizov za vsako poved in izbere tisto z najvišjim doprinosom. Iz spiska želenih glasovnih nizov se odstrani vse, ki jih izbrana poved vsebuje.
   - Postopek se ponavlja, dokler besedilo zbirke ne vsebuje vseh želenih glasovnih nizov.

Prva metoda se od naše predlagane metode razlikuje po tem, da so dolžine izbranih povedi omejene na 15-25 besed.

Nikjer ni poskrbljeno, da bi bili izbrani različni tipi povedi. Končni rezultat je precej odvisen od naključno generiranih zbirk s po 5.000 povedmi. Podobno kot pri naši metodi pa se končno število stavkov določi v naprej.

Druga metoda se od naše razlikuje že v zastavljenem cilju. Tudi predvideni končni obseg korpusa je za kvalitetno sintezo govora občutno premajhen. Uporabnost posameznih polifonov, ki jih ni na spisku, ni ocenjena. Vse povedi, ki so doprinesle enako število novih glasovnih nizov iz spiska, so ocenjene enako. Zavrženih je lahko veliko povedi, ki bi vsebovale koristne nove polifone, ki jih ni na spisku; po drugi strani so lahko sprejete povedi z enakimi dodatnimi polifoni, ki jih ni na spisku, kot jih vsebujejo že predhodno sprejete povedi. Prav tako ni upoštevana sorodnost nekaterih glasov (npr. »p«, »k«, »t«).

4 SNEMANJE GOVORNE ZBIRKE

Snemanje govorne zbirke je potekalo v studiu RTV Slovenija ob prisotnosti izkušenega tonskega tehnika. Med 10 profesionalnimi govorci smo izbrali najustreznejši moški in ženski glas. Med branjem besedila so govorci imeli nameščene elektrode Laryngographa, s katerimi smo spremljali nihanje glasilk za lažje kasnejše označevanje period govornega signala. Samo snemanje je zaradi obsežnosti besedila, ki ga je bilo potrebno prebrati trajalo več mesecev. Pri tem so nastavitve opreme ves čas ostale nespremenjene. Pred vsakim snemanjem je govorec poslušal svoje predhodne posnetke, s čimer se je skušalo zagotoviti čim bolj enak način govora, z enako intonacijo ipd.

5 STATISTIČNI PODATKI O GOVORNIH ZBIRKAH

V tabeli 1 so podani osnovni statistični podatki o govorni zbirki za korpusni sintetizator slovenskega govora Amebis Govorec. Podana je tudi primerjava z drugimi podobnimi zbirkami (tabela 2 in 3).

Velikost besednega korpusa	7.145.345 povedi 77 milijonov besed
Obseg govorne zbirke	**4.000 povedi** (46.785 besed)
Število različnih difonov	1.883
Število različnih trifonov (št. kombinacij v korpusu)	21.369 (24.702)

Tabela 1: *Statistični podatki o govorni zbirki Amebis Govorec*

Velikost besednega korpusa	2 milijona povedi 31 milijonov besed
Obseg govorne zbirke	**1.200 povedi** (dolžine 15-25 besed)
Število različnih difonov	1.030
Število različnih trifonov (št. kombinacij v korpusu)	9.391 (11.398)

Tabela 2: *Statistični podatki o govorni zbirki FERI [6]*

Velikost besednega korpusa	200.000 povedi
Obseg govorne zbirke	**297 povedi** (10.218 alofonov)
Število različnih difonov	1.132
Število različnih trifonov	17.784

Tabela 3: *Statistični podatki o govorni zbirki FE [2,7]*

6 SKLEP

Zasnovali smo govorno zbirko za kvalitetno korpusno sintezo slovenskega govora. Zbirka pokriva skoraj vse možne kombinacije difonov in trifonov na katere smo naleteli pri analizi obsežnega besednega korpusa s preko 7 milijoni povedi. Snemanje govorne zbirke (moški in ženski glas) je potekalo več mesecev. Prebranih je bilo 4.000 povedi povprečne dolžine 11 besed. Za lažje označevanje zbirke smo poleg govornega signala posneli še signal Laryngographa, ki prikazuje nihanje glasilk. Sledil je ročni pregled posnetega gradiva in grobo samodejno označevanje; temu sledi še fino popravljanje napak.

Gre za najobsežnejšo izdelano govorno zbirko namenjeno sintezi slovenskega govora do sedaj. Pričakujemo, da bo nova verzija sintetizatorja Amebis Govorec primerljiva z boljšimi sintetizatorji za druge večje jezike in tako sprejemljiva za najširši krog uporabnikov.

Literatura

[1] J. Žganec Gros, F. Mihelič, S. Dobrišek: Govorne tehnologije: pridobivanje in pregled govornih zbirk za slovenski jezik, Jezik in slovstvo, let. 48 (2003), št. 3-4, str. 47-59.

[2] A. Mihelič, J. Gros, N. Pavešič, M. Žganec: Pridobivanje govorne zbirke za korpusni sintetizator govora Phonectic, Zbornik konference Jezikovne tehnologije, str. 45-49, 2000.

[3] Z. Kačič, B. Horvat: Izgradnja infrastrukture potrebne za razvoj govorne tehnologije za slovenski jezik, Zbornik konference Jezikovne tehnologije za slovenski jezik, str. 100-104, 1998.

[4] I. Amdal, T. Svendsen: Unit selection Synthesis Database Development Using Utterance Verification, Zbornik INTERSPEECH 2005, str. 2553-2556, 2005.

[5] T. Šef, Analiza besedila v postopku sinteze slovenskega govora, *doktorska disertacija*, Fakulteta za računalništvo in informatiko, Univerza v Ljubljani, 2001.

[6] M. Rojc, Z. Kačič: Design of optimal Slovenian speech corpus for use in the concatenative speech synthesis system, Proceedings of the Second international conference on language resources an evaluation, Athens, Greece, str. 321-325, 2000.

[7] J. Žganec Gros, A. Mihelič, N. Pavešič, M. Žganec, S. Gruden: AlpSynth – Concatenation-based Speech Synthesis for the Slovenian Language, 47th International Symposium ELMAR-2005, Zadar, Hrvaška, str. 213-216, 2005.

HEVRISTIČNO ODLOČANJE

Aleš Tavčar, Matjaž Gams
Odsek za inteligentne sisteme
Institut "Jozef Stefan"
Jamova 39, 1000 Ljubljana, Slovenija
e-mail: {ales.tavcar, matjaz.gams}@ijs.si

POVZETEK

Hevristnične ocenitvene funkcije se pogosto uporabljajo pri reševanju preiskovalnih problemov. S poenostavljanjem in namernim izpuščanjem informacij omogočajo hitro in uspešno rangiranje potez po kvaliteti rešitve. Z opuščanjem informacij pa se v oceno hevristične funkcije vnese določena napaka. V večini primerov je ta napaka nezaželjena, izkaže pa se, da v določenih primerih manjša napaka blagodejno vpliva na kvaliteto izbrane poteze. V tem prispevku bomo najprej omenili, kako ljudje z uporabo hevrtistik uspešno rešujejo določene probleme. V nadaljevanju bomo analizirali, kako napaka hevristične funkcije vpliva na reševanje problemov preiskovanja.

1 UVOD

Hevristične funkcije se uporabljajo za reševanje velikega števila problemov. Najbolj pogosto se uporabljajo pri preiskovalnih problemih, kjer je prostor možnih stanj izjemno obsežen in izčrpno preiskovanje ne pride v poštev. Z opuščanjem določenih informacij o problemu in poenostavljanjem, hevristične funkcije večinoma uspešno usmerjajo preiskovalni algoritem proti rešitvi. Poenostavitve pa prinašajo tudi napake pri oceni kvalitete potez. Prevelike napake botrujejo pojavu t.i. patologije preiskovanja (Beal 1980, Nau 1979). Patologija se pojavi, ko globlje preiskovanje ojača napake hevurističnih ocen in zato globlje preiskovanje privede do slabše odločitve.
Uporaba hevristik pa ni omejena samo na reševanje preiskovalnih problemov, saj tudi ljudje v določeni obliki uporabljamo hevristike kot pomoč pri reševanju nekaterih problemov. Pokazali bomo, da v določenih situacijah uporaba manj informacij, manj preračunavanja in manj časa izboljša izvajanje določenih nalog.

V nadaljevanju prispevka analiziramo kako napake hevristične funkcije vplivajo na uspešnost minimin preiskovalnega algoritma (Korf 1990). Analizo smo opravili na umetno zgrajenih preiskovalnih drevesih in na realnem problemu iskanja poti v labirintu.

2 SORODNA DELA

S preučevanjem enoagentnega preiskovanja so se ukvarjali številni raziskovalci. Bulitko idr. (Bulitko 2003) je patologijo enoagentnega preiskovanja odkril šele leta 2003. Patologijo je pokazal na enostavnem binarnem drevesu, ni pa ponudil razlage, zakaj do nje pride.
Podrobnejšo razlago poda Luštrek (Lustrek 2008), kjer razloge razdeli v dve skupini:

- Domena problema, na katero težko vplivamo.
- Uporabljena hevristična funkcija, na katero imamo vpliv.

Piltaver idr. (Piltaver 2011) so preučevali patologijo v igri osmih ploščic. Igro so z dodajanjem novih potez spreminjali in tako omogočili spreminjanje vejitve in podobnosti vozlišč. Preučevali so tudi vpliv različnih tipov hevristične funkcije in zrnatosti na pojav patologije.
S patologijo v igri osmih ploščic sta se ukvarjala tudi Sadikov in Bratko (Sadikov 2006) , kjer sta ugotovila, da je pesimistična hevristična funkcija učinkovitejša od optimistične, saj patologijo oslabi.

3 PROCES KOGNITIVNEGA ODLOČANJA

Eden od načinov, kako se ljudje odločamo, ko nimamo na razpolago vseh informacij je z uporabo hevristik. Po definiciji, uporaba hevristike zahteva določene poenostavitve problema in zavestno ignorira del informacije.
Na primer, izkušeni igralci nogometa zavestno vedo, kako ujeti prihajajočo visoko žogo. Eden od možnih načinov bi bilo predvidevanje trajektorije žoge in računanje lokacije pristanka. Taka metoda je seveda prekompleksa in bi zahtevala ogromno časa, zato se igralci poslužujejo t.i. Gaze hevristike (Gigerenzer 2007). Igralci fiksirajo pogled na žogo in se tako premikajo, da ohranjajo kot pogleda konstanten, oziroma v določenem intervalu. Taki premiki zagotavljajo, da bo žoga zadela igralca. V tem primeru ljudje zanemarjajo vse potrebne podatke za računanje trajektorije (začetni kot, hitrost, lokacijo) in se osredotočajo le na kot med pogledom in tlemi.

Gigerenzer (Gigerenzer 2011) je zagovornik teorije o uporabi preprostih hevristik za reševanje problemov. Skupaj z Goldsteinom (Goldstein & Gigerenzer 1999) sta

analizirala, kakšne hevristike ljudje uporabljajo pri razpoznavanju stvari. V njihovem delu trdita, da če se posameznik odloča med dvema objektoma in enega od teh posameznik pozna, drugega pa ne, ljudje pripišejo poznanemu predmetu višjo vrednost glede na nek kriterij.
Kot primer navajata rezultate anketiranja dveh skupin ljudi. Američanom in Nemcem sta postavila vprašanje katero mesto ima več prebivalcev, Detroir ali Milwaukee. Izkazalo se je, da je 60% Američanov pravilno odgovorilo na vprašanje, medtem ko je bilo pravilnih odgovorov med Nemci 90 %. Avtorji razlagajo, da so Nemci pri odločitvi uporabljali enostavno hevristiko: Slišali so že za Detroit, za Milwaukee pa ne, torej so sklepali, da je prvo mesto večje. Američani pa so slišali za obe mesti in so uporabljali priklic za ugotavljanje rešitve. Torej so vedeli preveč.
Avtorja poudarjata, da uporabljena hevristika mora biti ustrezno poenostavljena. Pretirano kompleksne hevristike, z majhno napako se v praksi ne obnesejo dobro, prav tako so neuporabne preveč enostavne hevristike s preveliko napako. To teorijo smo želeli preveriti tudi z računalniškimi modeli.

4 MINIMIN MODEL PREISKOVANJA

Vpliv napake hevristične funkcije na kvaliteto rešitve preiskovalnega algoritma smo najprej preverili na umetno zgrajenem minimin modelu preiskovanja. Gradnja minimin modela se zgleduje po modelu opisanem v (Tavčar 2009, 2009). Model omogoča uporabo različnega števila pravih in hevurističnih vrednosti v vozliščih drevesa, omogoča spreminjanje vejitve drevesa in stopnjo podobnosti.

V preiskovalnem drevesu je prava vrednost vozlišča enaka minimalnemu številu potez, ki so potrebne za dosego cilja. Razlika med pravo vrednostjo in hevristično oceno neke pozicije se imenuje statična napaka hevristične funkcije. Statično napako generične hevristične funkcije smo modelirali z Gaussovim šumom pri katerem smo varirali standardni odklon σ. Tako lahko aproksimiramo bilokatero hevristično funkcijo, ki se uporablja v praksi.

Pri ocenjevanju kvalitete rešitve smo se osredotočili na delež izbranih potez v korenu, ki ne vodijo v sina z najboljšo pravo vrednostjo. Izbrana poteza je napačna, če obstaja druga poteza, ki vodi v vozlišče z večjo pravo vrednostjo. Verjetnost napake, pri določeni globini preiskovanja d in hevristični funkciji h, označimo z E(h,d).
Patološkost preiskovanega drevesa se običajno definira z izrazom:

$$p = \frac{E(h,d_i)}{E(h,d_j)},\ \ 1 \le d_j \le d_i$$

Iz zgornje enačbe lahko sklepamo, da vrednost p > 1 pomeni prisotnost patologije. Torej, večja kot je patologija slebše rezultate dobimo s preiskovanjem.
V pričujočem delu smo želeli ugotoviti, ali spreminjanje napake hevristične funkcije vpliva na uspešnost preiskovalnega algoritma. V ta namen smo uporabili mero *Gain*. To mero je uporabljal Piltaver (Piltaver 2011) pri analizi preiskovanja v igri osmih ploščic. Definirana je podobno kot patologija:

$$gain_{i,j} = \frac{1-E(h,d_i)}{1-E(h,d_j)},\ \ 1 \le d_j \le d_i$$

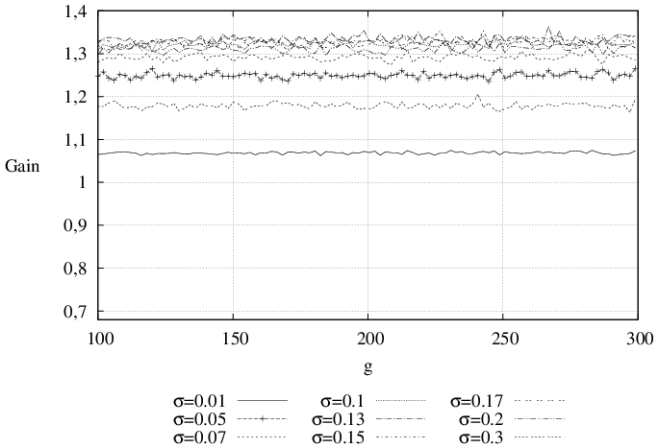
Globlje preiskovanje (do globine i, namesto j) se splača, če $gain_{i,j} > 1$.
Gain nam pove, kolikšen je delež ugodnejših potez, ko preiskujemo do večje globine v primerjavi s plitkejšim preiskovanjem.

Vpliv velikosti napake na stopnjo mere Gain, pri različnih stopnjah zrnatosti, smo analizirali s simulacijo Monte Carlo.

Fiksirali smo vejitev na b=2 in varirali stopnjo stopnjo zrnatosti na intervalu [100,300]. Za stopnje podobnosti iz množice {0.4, 0.5, 0.6} smo tvorili 50.000 dreves globine d_{max}=5.
Napako poteze na ustrezni globini smo izračunali za vsako zgrajeno drevo in jo povprečili čez vseh 50.000 dreves
Iz dobljenih povprečnih napak smo nato izračunali Gain.

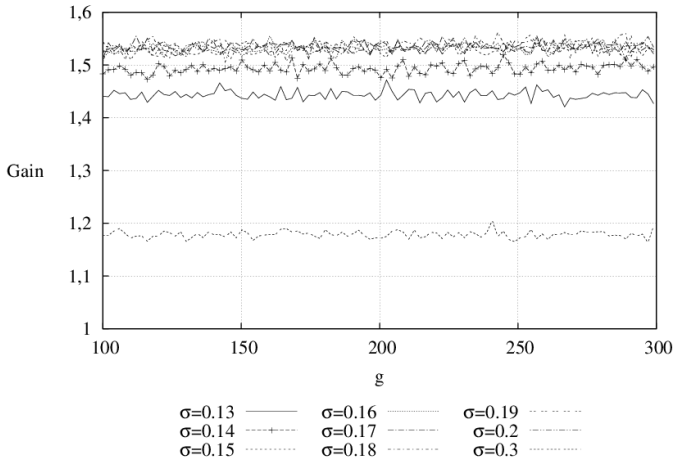
Ena taka meritev je predstavljena z eno krivuljo na sliki 1.

S spreminjanjem standardne deviacije Gausovega šuma smo simulirali različne napake hevritične funkcije. Simulacijo smo pognali večkrat za različne vrednostni standardne deviacije. Na sliki 1 so prikazani rezultati meritev vrednosti mere Gain za devet različnih vrednosti standardne deviacije. Opazimo lahko, da se zelo nizka vrednost napake odraža v vrednostih mere Gain blizu 1 in se z večanjem napake povečuje do nekje 1.35. Iz meritev lahko tudi opazimo, da se pri večjih napakah mera Gain znova zmanjša in za σ =0.3 znaša približno 1.18. Vrednost sigme σ =1.15 daje največjo vrednost mere Gain.



Slika 1: Mera Gain za različne vrednosti napake in stopnji podobnosti 0.4

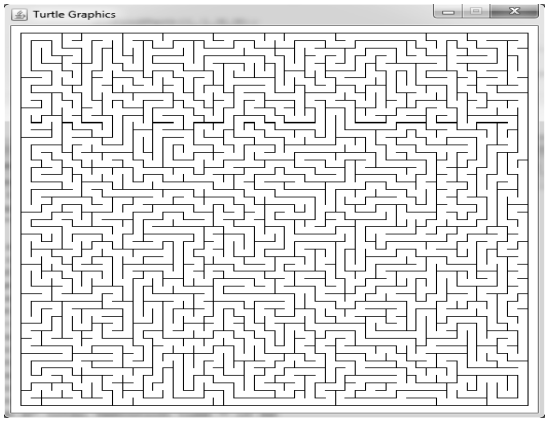
Podobne meritve smo izvedli še ze stopnjo podobnosti 0.6. Rezultati so prikazani na sliki 2. Opazimo lahko podobno obnašanje krivulje Gain. Z večanjem standardnega odklona se njena povprečna vrednost povečuje. Pri σ=1.19 doseže višek ($\overline{Gain} \approx 1.55$) in se nato zopet zmanjša. Iz spodnjega grafa lahko še opazimo, da so krivulje za različne vrednosti standarnega odklona precej skupaj in iz tega lahko predvidevamo, da se pri večjih stopnjah odvisnosti ta pojav še ojača. Stopnja podobnost torej izniči vpliv napake hevristične funkcije na mero Gain. Pri primerjavi obeh grafov lahko še opazimo, da se povprečna vrednost mere Gain povečuje z večanjem stopnje podobnosti. Ta ugotovitev je v skladu z ugotovitvami v (Kaluža 2007, Tavčar 2009, Piltaver 2011), kjer se je izkazalo, da večja stopnja podobnosti zmanjša patologijo, oziroma jo odpravi.



Slika 2: Mera Gain za različne vrednosti napake in stopnji podobnosti 0.6

5 ANALIZA PROBLEMA ISKANJA POTI V LABIRINTU

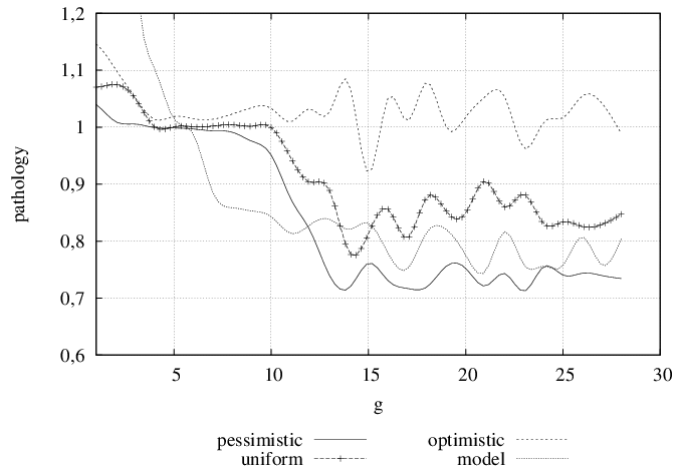
Naslednja domena, kjer smo preverili veljavnost hipoteze je iskanje poti v labirintu. Implementirali smo metodo, ki po principu naključnega Kruskalovega algoritma generira labirinte poljubnih velikosti. Pri realnih problemih sta običajno vejitev in stopnja podobnosti podani z domeno. Povprečna vrednost vejitve v avtomatsko zgrajenih labirintih je 2. Na sliki 3 lahko vidimo grafični prikaz avtomatsko zgrajenega labirinta velikost 50×50 polj.



Slika 3: Avtomatsko zgrajeni labirint

Meritve izkoristka smo začeli z retrogradno analizo (Thompson 1986), kjer izračunamo prave vrednosti za vse pozicije v labirintu. To izvedemo tako, da začnemo v končnem vozlišču in s pomikanjem proti začetni poziciji vsakemu vozlišču pripišemo najkrajšo pot do končnega vozlišča. Podobno kot v modelu minimina, pravim vrednostim vozlišč dodamo Gaussov šum in tako dobimo hevristične vrednosti. Nato smo z uporabo minimin algoritma izračunali napako poteze na prvem in petem nivoju. Na koncu pa smo iz obeh napak izračunali vrednosti patologije in izkoristka.

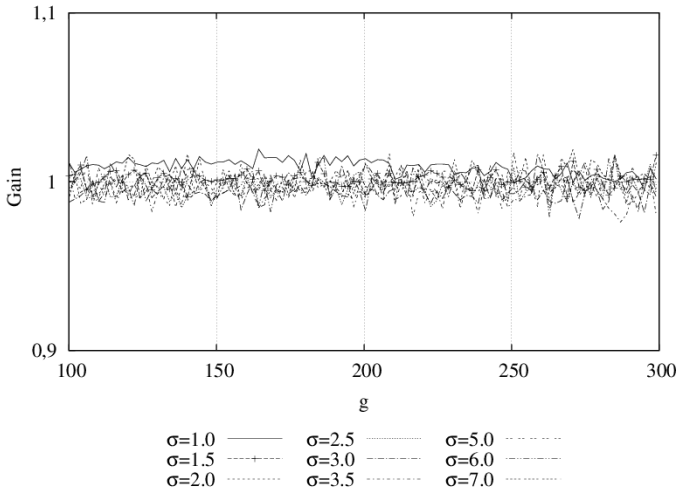
Najprej smo želeli preveriti skladnost meritev patologije na modelu in v realnem problemu. Meritve v labirintu smo izvedli s tremi različnimi hevrističnimi funkcijami: pesimistično, optimistično in uravnoteženo. Optimistična hevristična funkcija vedno podcenjuje ceno do rešitve. Ta lastnost je pri popolnih preiskvalnih algoritmih (A*) zaželjena, saj zagotavlja optimalno rešitev. Pesimistična hevristična funkcija pa precenjuje razdaljo do cilja. Nedavne raziskave so pokazale, da so pesimistične hevristične funkcije bolj uspešne od optimističnih pri nepopolnem preiskovanju (Sadikov in Bratko 2006). Na sliki 4 so prikazane meritve patologije za tri različne tipe hevrističnih funkcij na problemu labirinta in meritve patologije na minimin modelu. Vidimo lahko, da je pesimistična hevristična funkcija na problemu labirinta bolj uspešna od optimistične. Uporaba uravnotežene funkcije pa se obnese bolje od optimistične, a slabše od pesimistične. Iz slike lahko tudi vidimo, da se meritve na modelu skladajo z meritvami na realnem problem. Do razlike pride le pri manjših stopnjah zrnatosti, kjer je stopnja patologije v modelu precej večja od stopnje patologije v labirintu.



Slika 4:Primerjava različnih hevrističnih funkcij

Pri naslednjih meritvah smo želeli preveriti vpliv napake na izkoristek preiskovalnega algoritma pri globljem preiskovanju. Spreminjanje napake smo izvedli podobno kot v modelu, s spreminjanjem standardne deviacije Gaussovega šuma.

Na sliki 3 so prikazani rezultati meritev. Na sliki najprej opazimo, da so krivulje mere Gain tesno skupaj in težje sklepamo o veljavnosti hipoteze. Razlog temu je najverjetneje višja stopnja podobnosti, ki izniči vpliv napake hevristične funkcije. V tem pogledu se ugotovitve na modelu skladajo z ugotovitvami v realni domeni. Opazimo lahko tudi, da je gain precej manjši kot v modelu in se giblje okoli vrednosti 1. Torej preiskovanje v globino ne prinese bistvenih izboljšav.



Slika 5: Mera Gain za različne vrednosti napake

6 ZAKLJUČEK

V pričujočem delu smo analizirali vpliv hevristične napake na kvaliteto potez, izbranih po principu minimin. Pri postavitvi hipoteze smo se nanašali na ugotovitve Gigerenzera, ki trdi, da se ljudje pri sprejemanju odločitev poslužujemo enostavnih hevristik, ki se v večini primerov obnašajo izjemno učinkovito. Njegovo hipotezo smo želeli preveriti na minimin modelu in problemu iskanja poti v labirintu.

Definirali smo generično hevristično funkcijo, katere napaka je modelirana z Gaussovim šumom. Funkcija vrača prave vrednosti vozlišč katerim je dodan Gaussov šum. Različne stopnje napake dobimo s spreminjanjem standardne deviacije σ.

Z Monte Carlo simulacijami smo potrdili hipotezo, da je določena vrednost napake koristna in pri globljem preiskovanju pripomore k izbiri kvalitetnejše poteze.

References

[1] Beal, D. F. (1980) : An analysis of minimax. V zborniku Advances in Computer Chess 2, str. 103-109, Edinburg University Press

[2] Bulitko, V., Li, L., Greiner, R., Levner, I., (2003). Lookahead pathologies for single agent search. V zborniku 18th International Joint Conference on Artificial Intelligence, str. 1531-1533

[3] Gigerenzer, G. (2007): "Gut Feelings" (The intelligence of the Unconscious)

[4] Gigerenzer, G. and Gaissmaier, W. (2011): Heuristic decision making, Annual Review of Psychology, vol. 62, 451-482

[5] Goldstein, D. G., & Gigerenzer, G. (1999). The recognition heuristic: How ignorance makes us smart. In G. Gigerenzer, & P. M. Todd, (Eds.). *Simple heuristics that make us smart.* Oxford: Oxford University Press.

[6] Kaluža, B., Luštrek, M., Gams, M. and Tavčar, A. (2007) Pathology in Minimax Searching. V zborniku šestnajste Mednarodne Elektrotehniške in Računalniške konference – ERK 2007, vol. B, str. 107-110.

[7] Korf, R. E. (1990): Real-time heuristic search. Artificial intelligence, 42(2,3), str. 100-107

[8] Lustrek, M., Bulitko, V. (2008): Thinking Too Much: Pathology in Pathfinding. V zborniku ECAI 2008: 18th European Conference on Artificial Intelligence.

[9] Nau, D. S. (1979): Quality of decision versus depth of search on game trees, Doktorska disertacija, Duke University

[10] Piltaver, R., Luštrek, M., Gams, M. (2011): The pathology of heuristic search in the 8-puzzle, Journal of Experimental & Theoretical Artificial Intelligence

[11] Sadikov, A., Bratko, I. (2006) Pessimistic heuristics beat optimistic ones in real-time search. V zborniku European Conference on Artificial Intelligence (ECAI), str. 148-152

[12] Tavčar, A. (2009): Patologija minimin preiskovanja. Diplomsko delo, Fakulteta za računalništvo in informatiko, Univerza v Ljubljani.

[13] Tavčar, A., Luštrek, M., and Gams, M. (2009). Patologija minimin preiskovanja. V zborniku konference Informacijska družba.

SCHEDULING OF FLEXIBLE ELECTRICITY PRODUCTION AND CONSUMPTION IN A FUTURE ENERGY DATA MANAGEMENT SYSTEM: PROBLEM FORMULATION

Tea Tušar, Erik Dovgan, Bogdan Filipič

Department of Intelligent Systems, Jožef Stefan Institute
Jamova cesta 39, SI-1000 Ljubljana, Slovenia
Tel: +386 1 477 34 62, fax: +386 1 477 31 31
e-mail: {tea.tusar, erik.dovgan, bogdan.filipic}@ijs.si

ABSTRACT

Rapidly changing electricity markets call for innovative solutions to support balancing of energy production and consumption, and utilize the increasing amount of energy from renewable sources. MIRABEL is a future energy data management system based on flexible offers (flex-offers) for energy production and consumption. One of its core functionalities is scheduling of aggregated flex-offers to minimize the costs of the balance responsible party. This paper presents a formulation of this scheduling problem in terms of decision variables, constraints and the objective function, and discusses the problem characteristics.

1 INTRODUCTION

Electrical energy markets are rapidly changing. Their deregulation is redefining the roles and activities of the involved parties. In addition, the urgency of environmental sustainability calls for reduction of carbon emissions and higher utilization of renewable energy sources (RES). However, RES, like solar panels and windmills, make it hard for electricity distributors to include their production into daily schedules because of their dependence on external factors, such as weather conditions. Finally, smart metering is being increasingly adopted in electricity consumption. Under these conditions, new solutions are sought to support flexibility on electricity markets, ensure reliable supply, and balance the costs and benefits of the involved parties.

In striving for these goals, information and communication technology (ICT) is of crucial importance. An ICT system to serve the needs of a deregulated electricity market and enable the integration of a higher rate of energy from distributed and renewable sources into the electricity grid is being developed in the European Seventh Framework Programme project MIRABEL (Micro-Request-Based Aggregation, Forecasting and Scheduling of Energy Demand, Supply and Distribution) [1]. The project proposes a conceptual and infrastructural approach to supply and demand side management where electricity producers and consumers issue flexible offers (termed *flex-offers*), indicating flexibilities in start time and energy amount. These flex-offers are then processed by the MIRABEL system to balance electricity supply and demand.

As electricity market regulations vary across the countries, a common platform was identified first to build upon in MIRABEL. For this purpose the Harmonized Electricity Market Role Model [2] defined by the European Network of Transmission System Operators for Electricity (ENTSO-E) and cooperating institutions was selected. Despite still being refined, the model provides a coherent view of the electricity markets in Europe, represented by roles, domains, and their interactions. In this model, the elementary role is *party connected to the grid* that contracts for the right to produce or consume electricity at a metering point. Types of this party are producers and consumers, sometimes denoted by a common term *prosumers*. A collection of metering points (related to prosumers) for imbalance settlement is a domain called *balance group*. The role providing balance responsibility and financial security for a balance group is *balance responsible party* (BRP).

Balance group is the basic domain where the MIRABEL system will be applied. To assist the BRP in equalizing the inflows and outflows of electricity at the balance group endpoints, i.e., producers, consumers and connections to the external network, the system provides:

- handling of the novel concept of flex-offers for electricity production and consumption,
- forecasting of electricity production and consumption,
- aggregation of flex-offers on a regional level, scheduling of electricity production and consumption based on aggregated flex-offers, and disaggregation of the scheduled flex-offers for the purpose of their contracting,
- a distributed, decentralized and scalable computer infrastructure to handle the data load from the prosumers.

The overall MIRABEL architecture and functionalities are described in [3]. This paper focuses on the flex-offer scheduling problem as faced in this system. It presents a formal definition of the scheduling problem, upgrading its draft version given in [4].

The paper is further organized as follows. First, the concepts needed to formulate the problem are explained. Next, the scheduling problem is formulated as an optimization problem in terms of decision variables, constraints and the objective function. The paper then discusses the characteristics that make the problem highly specific and complex. It concludes with a summary of the presented work.

2 CONCEPT DESCRIPTION

2.1 Time intervals

In the MIRABEL scheduling problem, time is discretized into intervals (usually 15 minutes long). Each such interval is called a *time step interval*, and every time related concept is defined as a multiple of time step intervals. *Scheduling interval* is the interval for which scheduling needs to be performed and is also a multiple of time step intervals.

2.2 Mismatch and imbalance prices

For each time step interval, a *mismatch* amount is given. Mismatch represents the difference between all produced and consumed energy that is forecast for the corresponding time interval. Mismatch is positive when forecasts imply more produced than consumed energy, and negative when the consumed energy is forecast to exceed the produced energy. Mismatch is merely a prediction of imbalances that are about to happen in reality when the time in question has passed and the producers and consumers of energy will adhere (or not) to the forecast behavior. The BRP has to pay penalties for any imbalance. Their price is called the *imbalance price*.

2.3 Flex-offers

A *flex-offer* represents an offer of a consumer to buy energy from the BRP or an offer of a producer to sell energy to the BRP. Each flex-offer is defined with:

- start time flexibilities,
- energy intervals, where each interval is defined with its duration, price, and energy flexibilities, and
- total energy constraint.

The start time flexibilities denote on which time step intervals the execution of the flex-offer can start. For example, the production flex-offer from Figure 1 can start on four time step intervals, while the consumption flex-offer can start from the second to the eleventh time step interval. This means that the consumption flex-offer has greater time flexibility than the production one. The flex-offer's energy intervals are shown as boxes in Figure 1 (the production flex-offer has four energy intervals, while the consumption flex-offer two). Each energy interval has its duration expressed in multiples of time-step intervals, a price per energy amount and flexibilities, i.e., the minimum and maximum energy that can be assigned to the flex-offer for that energy interval (indicated with arrows in Figure 1). The energy of production flex-offers is regarded as positive, and the energy of consumption flex-offers as negative.

2.4 Market energy and prices

The mismatch that remains after all flex-offers have been scheduled can sometimes be bought (negative mismatch) or sold (positive mismatch) on the *energy market* at a price called the *market price*.

3 PROBLEM FORMULATION

The MIRABEL scheduling problem is defined with:

- the scheduling interval,
- mismatch and imbalance prices, which are given for every time step interval in the scheduling interval,
- market prices, which are given for some time step intervals in the scheduling interval, and
- (aggregated) flex-offers with all their defining information.

The task is to fix time and energy flexibilities of all given flex-offers and establish the amount of energy to be bought (sold) on the market so that all constraints are satisfied and the cost for the BRP is minimized.

In the continuation of this section, the problem is formally defined in terms of decision variables, constraints and the objective function, using the notation from Table 1.

Table 1: *Notation used in the problem formulation.*

Variable	Meaning
n	number of flex-offers
m	number of time step intervals in the scheduling interval
E^i_{I}	remaining mismatch amount for time step interval i
$p^i_{\mathrm{I}+}$, $p^i_{\mathrm{I}-}$	price of positive/negative imbalance for time step interval i
E^i_{M}	market energy amount for time step interval i
$p^i_{\mathrm{M}+}$, $p^i_{\mathrm{M}-}$	price of energy that can be sold/bought on the market for time step interval i
s_k	schedule for the k-th flex-offer
t_k	start time of the k-th flex-offer
E^j_k	energy amount of the j-th energy interval of the k-th flex-offer
p^j_k	price of the j-th energy interval of the k-th flex-offer

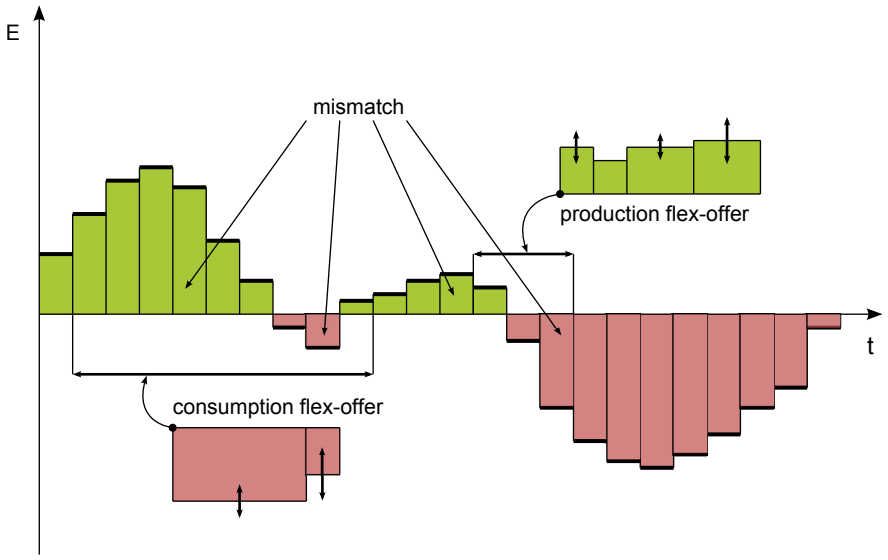


Figure 1: *An example of mismatch and two flex-offers (one production and one consumption).*

3.1 Decision variables

The decision variables of the scheduling problem are represented with the pair (S, M), where S is a vector of flex-offer schedules, and M is a vector of market energy amounts:

$$S = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix}, \; M = \begin{bmatrix} E^1_{\mathrm{M}} \\ E^2_{\mathrm{M}} \\ \vdots \\ E^m_{\mathrm{M}} \end{bmatrix},$$

where the notation from Table 1 is used. Each flex-offer schedule s_k is defined with the flex-offer start time t_k and energy amounts for each energy interval of the flex-offer:

$$s_k = (t_k, E^1_k, E^2_k, \ldots, E^{n_k}_k),$$

where n_k is the number of energy intervals of the k-th flex-offer. The pair (S, M) of flex-offer schedules and market energy amounts is a solution to the scheduling problem.

3.2 Constraints

Flex-offer flexibilities are in fact constraints on flex-offer schedules. Each flex-offer schedule $s_k = (t_k, E^1_k, E^2_k, \ldots, E^{n_k}_k)$ is subject to the following constraints:

$$t_k \in \left[\underline{t_k}, \overline{t_k}\right],$$

where $\underline{t_k}$ and $\overline{t_k}$ are the earliest and latest start times for t_k, and

$$E^j_k \in \left[\underline{E^j_k}, \overline{E^j_k}\right] \text{ for all } j = 1, \ldots, n_k,$$

where $\underline{E^j_k}$ and $\overline{E^j_k}$ are the minimum and maximum energy amounts for E^j_k. In addition to providing time and energy flexibilities, a prosumer may provide a total energy constraint that determines the sum of energy that needs to be produced (or consumed) in the entire flex-offer over all its energy intervals:

$$\sum_{j=1}^{n_k} E^j_k \in \left[\underline{E_k}, \overline{E_k}\right],$$

where $\underline{E_k}$ and $\overline{E_k}$ are the minimum and maximum total energy constraint amounts for the k-th flex-offer.

Any of the specified constraint intervals can degenerate into a point, leaving no flexibility. For example, the production flex-offer from Figure 1 has no flexibility in the second energy interval.

Market prices that are part of the given problem instance can have an arbitrary form. For any time step interval i, the following combinations are possible:

- only the price $p^i_{\mathrm{I}+}$ for selling energy on the market is defined,
- only the price $p^i_{\mathrm{I}-}$ for buying energy on the market is defined,
- both prices $p^i_{\mathrm{I}+}$ and $p^i_{\mathrm{I}-}$ are defined, or
- none of the prices is defined.

The vector of market energy amounts is constrained according to the defined prices:

$$E^i_{\mathrm{M}} \in \begin{cases} \mathbb{R}^+_0 & \text{if only } p^i_{\mathrm{I}+} \text{ is defined,} \\ \mathbb{R}^-_0 & \text{if only } p^i_{\mathrm{I}-} \text{ is defined,} \\ \mathbb{R} & \text{if both prices are defined,} \\ \{0\} & \text{if none of the prices is defined.} \end{cases}$$

This means, for example, that if only the price $p^i_{\mathrm{I}+}$ for selling energy on the market is defined, the market energy amount for interval i must not be negative, since it is not possible to buy energy on the market.

A schedule is considered feasible if its start time is fixed within the given flexibilities, the energy amounts are fixed within the given flexibilities, and all the energy constraints are satisfied. A market energy amount is considered feasible if it is set according to its constraints. A *solution is feasible* if all of its schedules and market energy amounts are feasible.

3.3 Objective function

The objective of the scheduling problem is minimization of cost for the BRP. The cost for the BRP c of a solution (S, M) consists of the cost of remaining negative imbalances $c_{\mathrm{I}-}$, the cost of remaining positive imbalances $c_{\mathrm{I}+}$, the cost of flex-offers c_{FO} and the cost of the energy bought on the market $c_{\mathrm{M}-}$ minus the profit from the energy sold on the market $c_{\mathrm{M}+}$:

$$c(S,M) = \underbrace{\sum_{\substack{i=1 \\ E^i_{\mathrm{I}}<0}}^{m} p^i_{\mathrm{I}-} \left|E^i_{\mathrm{I}}\right|}_{c_{\mathrm{I}-}} + \underbrace{\sum_{\substack{i=1 \\ E^i_{\mathrm{I}}>0}}^{m} p^i_{\mathrm{I}+} E^i_{\mathrm{I}}}_{c_{\mathrm{I}+}} + \underbrace{\sum_{k=1}^{n} \left(\sum_{j=1}^{n_k} p^j_k E^j_k \right)}_{c_{\mathrm{FO}}}$$
$$+ \underbrace{\sum_{\substack{i=1 \\ E^i_{\mathrm{M}}<0}}^{m} p^i_{\mathrm{M}-} \left|E^i_{\mathrm{M}}\right|}_{c_{\mathrm{M}-}} - \underbrace{\sum_{\substack{i=1 \\ E^i_{\mathrm{M}}>0}}^{m} p^i_{\mathrm{M}+} E^i_{\mathrm{M}}}_{c_{\mathrm{M}+}},$$

where the notation from Table 1 is used. Note that while buying energy from the market increases the cost for the BRP, selling energy decreases it (hence minus in the formula above). Similarly, the BRP must buy the energy produced by the production flex-offers (which increases the total cost), while the energy consumed by the consumption flex-offers represents profit for the BRP and decreases the total cost. This is expressed by the sign of the energy amount E^j_k.

If there are more solutions that minimize the cost c, the one which results in the smallest amount of remaining mismatch is preferred.

The decision variables t_k denoting start times of flex-offers are present in the calculation of the objective function only implicitly. They influence the amounts of remaining mismatch and market energy amounts, i.e., E^i_{I} and E^i_{M} values, and therefore indirectly affect the total sum of costs for the BRP.

In short, the presented scheduling problem consists of finding a feasible solution (S, M) which minimizes the cost function $c(S, M)$.

4 DISCUSSION

The scheduling problem formulated in this work differs from the scheduling problems treated in the literature either in the context of production systems (e.g., [5]) or energy sector (e.g., [6]). Unlike the usually scheduled activities, flex-offers are structured, consisting of several energy intervals, each interval with its own properties. Regarding the decision variables, in addition to start time, flex-offer scheduling involves determining energy amount for each energy interval of every flex-offer, and the market energy amounts. This substantially increases the problem complexity in terms of the number of candidate solutions. Finally, the objective function is not related to a time measure, but is rather a composed cost function.

These characteristics and the expected large number of flex-offers to be processed make the MIRABEL scheduling problem non-standard and highly complex. Known scheduling heuristics are therefore very unlikely to be directly applicable. The approach followed in solving this problem will therefore be metaheuristic algorithms, possibly hybridized with local optimization.

5 CONCLUSION

Scheduling of electricity production and consumption based on prosumer flex-offers was formulated as an optimization problem. This formulation is a prerequisite for implementing suitable optimization algorithms. Solving the problem will be a functionality of the MIRABEL energy data management system that is currently under development.

ACKNOWLEDGEMENT

The work presented in this paper was carried out in the project *Micro-Request-Based Aggregation, Forecasting and Scheduling of Energy Demand, Supply and Distribution* (MIRABEL) funded by the European Commission under grant agreement number 248195, and co-funded by the Slovenian Research Agency.

REFERENCES

[1] MIRABEL Project. Available from `http://www.mirabel-project.eu/`, retrieved on September 10, 2011.

[2] European Transmission System Operators. The Harmonized Electricity Market Role Model, version 2009-01. Available from `http://www.scribd.com/doc/48492720/Role-Model-v2009-01`, retrieved on September 6, 2011.

[3] Z. Marinšek, G. Černe, H. Berthold, C. Nychtis, F. J. Rumph, M. Konsman, H. Frey, L. Šikšnys, T. B. Pedersen, D. Kaulakiene, B. Filipič, M. Böhm. D1.2 Final role model and process specification. Technical report, MIRACLE Consortium, December 2010.

[4] T. Tušar, E. Dovgan, B. Filipič, A. Savinov. D5.3 Initial draft of the scheduling and negotiation framework. Technical report, MIRACLE Consortium, May 2011.

[5] M. L. Pinedo. *Scheduling: Theory, Algorithms, and Systems*. Third edition. Springer, New York, 2008.

[6] H. Y. Yamin. Review on methods of generation scheduling in electric power systems. *Electric Power Systems Research*, Vol. 69, No. 2–3, pp. 227–248, 2004.

Agent Architecture for Smart Home Management

Domen Zupančič*
Robotina d.o.o.
OIC-Hrpelje 38, SI-6240 Kozina, Slovenia
email: domen.zupancic@ijs.si

Abstract

Agent approach by smart home management enables flexible, open, distributed and modular architecture. This facts favorize agent systems by contemporary implementations of smart spaces. System ontology is of high importance due to its structurized representation of system units. There is general ontology, similar to all units, and domain specific ontology, which represents particular unit in details. Ontologies represents a context, which is the basis for further reasoning. Context are time and space labeled data from environment, acquired from sensor readings, actuator states or from other sources, acquired through web servers. The article presents a proposal for smart home implementation, based on agent architecture.

1 Introduction

Agent approach in building automation system (BAS) engineering is an research issue for last several years. Distributed, flexible, open and modular BAS favorize agent architecture in contrast to classic monitoring and control systems [2]. Multi agent system is a set of individual entities - agnets. Agents differs among each other by operation and layers. Agent layers in MavHome example [3, 4] are decison, information, communication and physical layer. Agent modelling is consequently of very important and complex task, while agent groups can be implied for common goal achievement with communication and cooperation.

The idea for modern BAS is simple and common in literature: BAS should maximize user preferences and minimize energy consumption[1, 2, 4, 5, 8]. As these facts are contradictory, the solution of how to achieve such goal is not trivial. The agent, which represents actuator (heater, liht, air-conditioner) needs to acquire data from higher level. Control system should dictate the temperature, illuminance and other setpoints on one hand, and conditions, when the particular actuator can swithc on or off, on the other. The task for information layer agents (in heating domain) are so to get data about (i) current weather state, (ii) weather forecast, (iii) current hot water demand, (iv) future hot water demand and many others. The more factors included, the more comples the problem. Agents are able to make offers, and by processing of all offers, the best are chosen to be realised.

The complexity level increases from low level to high level agents. Sensor agents acquire data and can also extract sensor data (temperature, presence, illuminance, humidity,..) and send them to database. Communication agents include software and drivers for data transfer and arrange communication betwen agents, between agnets and users and between users and between systems and data from other sources (web). Information layer arrange database - collects and saves data and generate knowledge. Decision layer is in charge to make decisions when and how some particular system should work. For that task, decision layer uses data from information layer and give orders to physical agents - actuators, which accomplish the task.

The article proposes an implementation of smart home, based on multi agent system architecture. The implementation will first be done with simple rules and expert knowledge. Later it will be expanded with observation of user behavior and system behaviour (data trends), which will be used for reasoning to adapt schedules and rules for system operation.

*Operation part financed by the European Union, European Social Fund. Operation implemented in the framework of the Operational Programme for Human Resources Development for the Period 2007-2013, Priority axis 1: Promoting entrepreneurship and adaptability, Main type of activity 1.1.: Experts and researchers for competitive enterprises.

2 Multi Agent Systems

Multi agent systems can be represented as a set of agents. Each agent has its own task, and should accomplish in such a way, to reach the common goal. Agents for lighting or heating control, for example, tends to optimize the heating service and switch on/off or dim the lights.

Services in smart home are distributed on sets of subsystems, which controls actuators. Users supplement the system by adding new entities, which influences existent services and subsystems. Agent platform gives an oppurtunity to continiously learn, acquire and process data, negotiate about actions, which are appropriate to be taken in future.

2.1 Agent system properties

Important agent properties are [1, 2, 7]:

- distributed: each entity (sensor, actuator, database,...) has its own agent representative, and they are spatially distributed in smart house. Agent architecture enables communication and cooperation among these entities.
- flexibile: enables dynamic adaptation of schedules. By observation of human behavior and enviroment states, the schedule of home appliance activities is being adapted [11].
- open: enables adding new or removing old entities without system reconfiguration.

There are several research activities: MavHome [3] is testbed environment for multi agent system development in BAS, iHome[6] is another testbed, where white appliances and HVAC system entities are represented by agents, ThinkHome [9] where smart home learns from inhabitants and is able of self (re)configuration.

2.2 Agent architecture

Smart home includes several subsystems, spaces, appliances. It depends on ones interest, how to structure the system. It is possible to distribute it:

- functional: heating, cooling, lighting, safety,...
- spatial: floor, sanitation, living room, WC, backyard,...
- according to appliances: boiler, owen, window,...
  - sensors: temperature, presence,...
  - actuators: heating valve, light switch,...
- and others.

When implementing slightly sophisticated smart home control, the line of separation between upper mentioned distributions is not always clear. For example: outdoor temperature data can be used for heating, dimming, windows closing/opening, ventilation control. Furthermore the user has interest to read such data from user interfaces at home or through internet. Agents tends to become a member of a group or service. So an outdoor temperature sensor can be a member of heating, ventilating or air-conditioning service. There are several agent architectures proposed in [8–10].

Agent architectures in literature usually consists of environment, devices, which have influence on environment, sensors for data acquisition, communication units, units for local regulation, control center with databases, servers, and learning capabilities and finally structure agents, which manages adding new and removing old entities.

2.3 Agent platform

There is a number of available frameworks for multi agent research, simulation and development. JADE[1] is a java agent development framework, which implements FIPA specifiactions[2]. FIPA is a IEEE computer society standards organisation that promotes agent-based technology and the interoperability of its standards with other technologies. JADE also provides LEAP extension, which is developed for agent on mobile entities with limited processing and memory resources. JADE enables agent platform, distributed across several processing units, which is in favour for our system proposal.

The proposed platform is shown on figure 1. There are three layers, where several type of agents are defined and knowledge base.

- Physical layer includes programmable logic controllers, with appurtenant PLC agents. These agents provides sensor readings and actuator states to history data trends and to appropriate service agents. When PLC agent collects temperature from temperature sensor, than it forward current readings to heating service agent.

[1] http://jade.tilab.com/; acquired 13.9.2011
[2] http://www.fipa.org/; acquired 13.9.2011

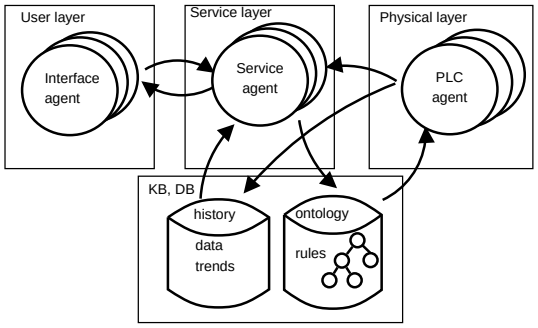


Figure 1: Intelligent Building Control

- Service layer includes service agents, which manages services as heating, lighting, blinds, air conditioning etc. Figure 2 represents the service agent structure in interaction with other platform entities. Service agent generates rules according to history data trends and system ontology, managed in knowledge base and data base. Using rules, current sensor readings and actuator states, service agent predicts user action. Example for inhabitant action prediction algorythm is in [3]. Using reward and punishment policy, the service agent than modify existing rules in knowledge base.

- User layer includes interface agents, which manages user interaction with elements in building automation. Data representation, reports generation, profiling and scheduling, manual control and system configuration activities are included in user layer.

- Knowledge base and data base layer includes storage entities. History trends for services, sensors and actuators are equipped with time stamps. System ontology and rule implementation is being created using web ontology language (OWL) where software like Protégé[3] is used and rule inference using reasoning engine, for example Pellet[4]. Figure 3 represents an example of ontology, created with Protégé.

3 Anticipation, action, reaction and negotiation

There is several research projects, where MAS approach in building automation is being studied. There are many

[3]http://protege.stanford.edu/; acquired 13.9.2011
[4]http://clarkparsia.com/pellet/; acquired 13.9.2011

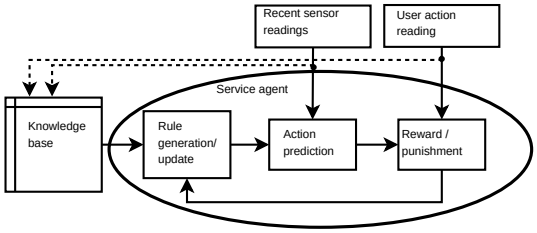


Figure 2: Service agent

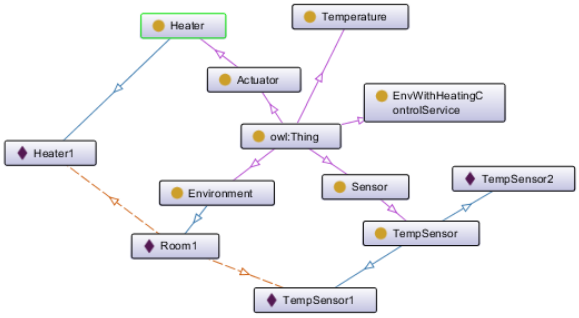


Figure 3: Example of ontology with concepts, individuals and relations

reasons, why these intelligent approaches have not become commercial yet. One of those is, that when system makes a mistake, it forces inhabitant, to repair or discard that action. As a consequence user comfort is being downgraded.

In literature there has been proposed anticipative [1] and reactive [2] mechanism, which are long term and short term system behavior mechanisms respectively. Reactive mechanism manages system operation and arises, when some constraints are being violated, which is expressed as critical situation. Anticipative mechanism is trying to prevent such critical situations with long term loads/consumption forecasting. Anticipative mechanism is useful in such cases, where a system has devices for energy storage and devices, which consume energy and has not fixed timeplan for operation. If there is a possibility to forecast energy consumption of a device, than is also possible to optimize the operation timeplan of that device.

Agent systems enables negotiation. Negotiation among agents whether to perform an action is being managed by mediator, which evaluates agents proposals. If these proposals pass defined treshold, than mediator inform agents to realize proposals.

4 Conclusion

The article summarizes part of research work concerning agent systems in building automation technology and gives a proposal for research work. There are two main goals, which needs to be satisfied: to assure user comfort on one hand, and to minimize operation costs on the other. As these are two contradictory goals, the multi agent system task is to continiously optimize device parameters and rules for regulation. Context information represented using ontologies and derived rules are important for system realization and relies on structural system representation in form of semantic web with concept, entities and relations.

The research work will continue with searching for connections between human activities in environment and past activities. When the smart environment recognizes typical event sequences, than it can influence the building automation system in the way to improve user comfort and optimize energy costs.

References

[1] S. Abras, S. Pesty, S. Ploix, and M. Jacomino. An anticipation mechanism for power management in a smart home using multi-agent systems. In *Information and Communication Technologies: From Theory to Applications, 2008. ICTTA 2008. 3rd International Conference on*, pages 1 –6, april 2008. doi: 10.1109/ICTTA.2008.4530305.

[2] Shadi Abras, Stéphane Ploix, Sylvie Pesty, and Mireille Jacomino. A multi-agent home automation system for power management. In Juan Andrade Cetto, Jean-Louis Ferrier, José Miguel Costa dias Pereira, and Joaquim Filipe, editors, *Informatics in Control Automation and Robotics*, volume 15 of *Lecture Notes in Electrical Engineering*, pages 59–68. Springer Berlin Heidelberg, 2008. ISBN 978-3-540-79142-3.

[3] D.J. Cook, M. Youngblood, III Heierman, E.O., K. Gopalratnam, S. Rao, A. Litvin, and F. Khawaja. Mavhome: an agent-based smart home. In *Pervasive Computing and Communications, 2003. (PerCom 2003). Proceedings of the First IEEE International Conference on*, pages 521 – 524, march 2003. doi: 10.1109/PERCOM.2003.1192783.

[4] S.K. Das, D.J. Cook, A. Battacharya, III Heierman, E.O., and Tze-Yun Lin. The role of prediction algorithms in the mavhome smart home architecture. *Wireless Communications, IEEE*, 9 (6):77 – 84, dec. 2002. ISSN 1536-1284. doi: 10.1109/MWC.2002.1160085.

[5] W. Kastner, G. Neugschwandtner, S. Soucek, and H.M. Newmann. Communication systems for building automation and control. *Proceedings of the IEEE*, 93(6):1178 –1203, june 2005. ISSN 0018-9219. doi: 10.1109/JPROC.2005.849726.

[6] Victor Lesser, Michael Atighetchi, Brett Benyo, Bryan Horling, Victor Lesser Michael Atighetchi, Anita Raja, Regis Vincent, Ping Xuan, Shelley XQ. Zhang, Thomas Wagner, Ping Xuan, and Shelley Xq. Zhang. The intelligent home testbed. In *Proceedings of the Autonomy Control Software Workshop*, 1999.

[7] C.M. Macal and M.J. North. Tutorial on agent-based modeling and simulation. In *Simulation Conference, 2005 Proceedings of the Winter*, page 14 pp., dec. 2005. doi: 10.1109/WSC.2005.1574234.

[8] Bing Qiao, Kecheng Liu, and Chris Guy. A multi-agent system for building control. *Intelligent Agent Technology, IEEE / WIC / ACM International Conference on*, 0:653–659, 2006. doi: http://doi.ieeecomputersociety.org/10.1109/IAT.2006.17.

[9] C. Reinisch, M.J. Kofler, and W. Kastner. Thinkhome: A smart home as digital ecosystem. In *Digital Ecosystems and Technologies (DEST), 2010 4th IEEE International Conference on*, pages 256 –261, april 2010. doi: 10.1109/DEST.2010.5610636.

[10] U. Rutishauser, J. Joller, and R. Douglas. Control and learning of ambience by an intelligent building. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 35 (1):121 – 132, jan. 2005. ISSN 1083-4427. doi: 10.1109/TSMCA.2004.838459.

[11] Weiming Shen, Lihui Wang, and Qi Hao. Agent-based distributed manufacturing process planning and scheduling: a state-of-the-art survey. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 36(4): 563 –577, july 2006. ISSN 1094-6977. doi: 10.1109/TSMCC.2006.874022.

Zbornik 14. mednarodne multikonference

INFORMACIJSKA DRUŽBA – IS 2011

Proceedings of the 14th International Multiconference

INFORMATION SOCIETY – IS 2011

Izkopavanje znanja in podatkovna skladišča (SiKDD 2011)

Data Mining and Data Warehouses (SiKDD 2011)

Uredila / Edited by

Dunja Mladenić, Marko Grobelnik

http://is.ijs.si

10. oktober 2011 / October 10th, 2011
Ljubljana, Slovenia

Preface / Predgovor

Data Mining and Data Warehouses (SiKDD 2011)

Data handling technologies have significantly progressed in the 90's. The first phases mainly dealing with storing and efficiently accessing the data, resulted in the development of industry delivering tools for handling large databases, standardization of related processes, queering languages, etc. When the data storage was not a primary problem any more the need for improving the database organization resulted in the databases supporting not only transactions but also analytical views of the data. At this point data warehousing with On-Line-Analytical-Processing entered as a usual part of a company information system, requiring from the user to set well defined questions which is not always easy and possible. On the other hand, Data Mining offers automatic data analysis trying to obtain some new information from the existing data and enabling the user some new insights in the data. The Slovenian KDD conference covers a broad area including Statistical Data Analysis, Data, Text and Multimedia Mining, Semantic Technologies, Link Detection and Link Analysis, Social Network Analysis, Data Warehouses.

Odkrivanje znanja in podatkovna skladišča

Tehnologije, ki se ukvarjajo s podatki so v devetdesetih letih močno napredovale. Iz prve faze, kjer je šlo predvsem zato kako podatke shraniti in kako do njih učinkovito dostopati, se je razvila industrija za izdelavo orodij za delo s podatkovnimi bazami, prišlo je do standardizacije procesov, povpraševalnih jezikov itd. Ko shranjevanje podatkov ni bil več poseben problem se je pojavila potreba po bolj urejenih podatkovnih bazah, ki bi služile ne le transakcijskem procesiranju ampak tudi bolj analitskim pogledom v podatke – pojavilo se je skladiščenje podatkov (data warehousing), ki postaja vse bolj standarden del informacijskih sistemov v podjetjih. Paradigma OLAP (On-Line-Analytical-Processing) zahteva od uporabnika, da še vedno sam postavlja sistemu vprašanja in dobiva nanje odgovore in na vizualen način preverja in išče izstopajoče situacije. Ker seveda to vedno ni mogoče, se je pojavila potreba po avtomatski analizi teh podatkov oz. z drugimi besedami to, da tehnologija sama pove, kaj bi utegnilo biti zanimivo za človeka – to prinašajo tehnike odkrivanja znanja (data mining), ki iz obstoječih podatkov skušajo pridobiti novo znanje in tako uporabniku ponudi novo razumevanje dogajanj zajetih v podatkih.

Slovenska KDD konferenca pokriva vsebine, ki se ukvarjajo z analizo podatkov in odkrivanjem zakonitosti v podatkih: pristope, orodja, probleme in rešitve.

Marko Grobelnik and Dunja Mladenić, Program Chairs / Urednika

Exploring the Space of Coding Matrix Classifiers for Hierarchical Multiclass Text Categorization

Janez Brank
Artificial Intelligence Laboratory
Jozef Stefan Institute
Jamova 39, 1000 Ljubljana, Slovenia
janez.brank@ijs.si

Abstract

One of the ways of approaching a multiclass classification problem is to transform it into several two-class (binary) classification problems. An ensemble of binary classifiers is trained for these tasks and their predictions are combined using a voting method into predictions for the original multiclass problem. Each of the new binary problems uses some of the original classes as positive training data, some classes as negative training data and the remaining classes (if any) are not used at all. The relationship between classes (of the original problem) and binary classifiers can be concisely represented by a matrix called the *coding matrix*. In this paper we explore some of the statistical properties of the space of coding matrix based classifiers in the context of a small hierarchical multiclass learning problem.

1 Introduction

Many machine learning methods were initially designed to handle binary (two-class) classification problems. If we want to use such a method to deal with a multi-class problem, one way to do this is to convert the multi-class problem into several binary problems, train binary classifiers for these new problems and then combine their predictions via a voting scheme to obtain predictions for the original multi-class problem.

Each of the new binary classification problems is defined by labelling some of the classes of the original multi-class problem as positive, some as negative, and the rest as unused. The union of instances of the positive (resp. negative) classes from the original multi-class problem then forms the positive (resp. negative) class of the new binary problem.

This mapping between classes of the original problem and the new binary problems can be concisely represented by a *coding matrix*. Let k be the number of classes in the original problem and let m be the number of new binary problems. Then the coding matrix M is a matrix from $\{-1, 0, 1\}^{k \times m}$, where the element M_{cj} tells us how class c of the original problem is used in the j'th new binary problem: $M_{cj} = 1$ means that instances from this class are used as positive examples, $M_{cj} = -1$ that they are used as negative examples and $M_{cj} = 0$ means that they aren't used at all when training the j'th binary classifier.

Thus, the coding matrix has one row for each class of the original multi-class problem. The c'th row of the matrix tells us how class c has been used while training the individual binary classifiers. Consider an instance x belonging to class c; let $y_j(x) \in \{-1, +1\}$ be the prediction of the j'th classifier; then, for those j where $M_{cj} \neq 0$, we would expect $y_j(x)$ to be equal to M_{cj} (unless the classifier y_j is making an error on this particular x), whereas for those j where $M_{cj} = 0$ we cannot make any advance judgments about what the predictions of y_j will be, because the j'th classifier hasn't seen members of class c during training. Thus, if x belongs to c, we would expect the sum $\sum_j y_j(x) M_{cj}$ to be high, and it's reasonable to take $z(x) := \arg\max_c \sum_j y_j(x) M_{cj}$ as the final prediction of the ensemble as a whole with regard to the original multi-class problem.

Coding matrices can be seen as a generalization of various traditional approaches to transforming multi-class problems into binary ones, such as the one-vs-rest approach (corresponding to $k = m$ and a matrix in which $M_{cj} = 1$ if $c = j$, and $M_{cj} = -1$ otherwise) and the one-vs-one approach (corresponding to $k = \binom{m}{2}$ and a matrix in which each column has exactly two nonzero entries).

The space of coding matrices is exponentially large in terms of k and m. Various approaches to constructing coding matrices have been considered in the literature, e.g. based on error correcting output codes [1] or by a greedy search through the space [2]. However, for small values of k and m, the space of coding matrices is tractable and we can afford to consider all possible coding matrices. In this paper we present some statistical properties of the space of coding matrix classifiers and their performance for a small dataset consisting of seven classes organized into a three-level hierarchy.

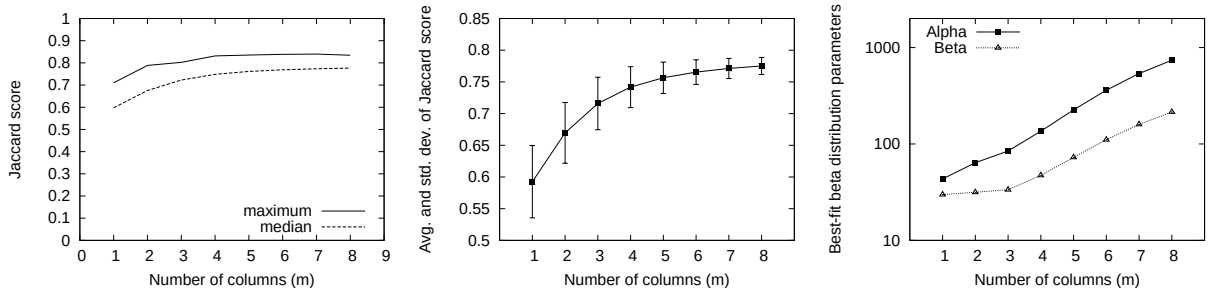


Figure 1. (a) Median and maximum Jaccard score over all m-column matrices; (b) average and std. dev. of the Jaccard score; (c) best-fit beta distribution parameters, α_m and β_m, as functions of m (note the logarithmic scale on the y-axis).

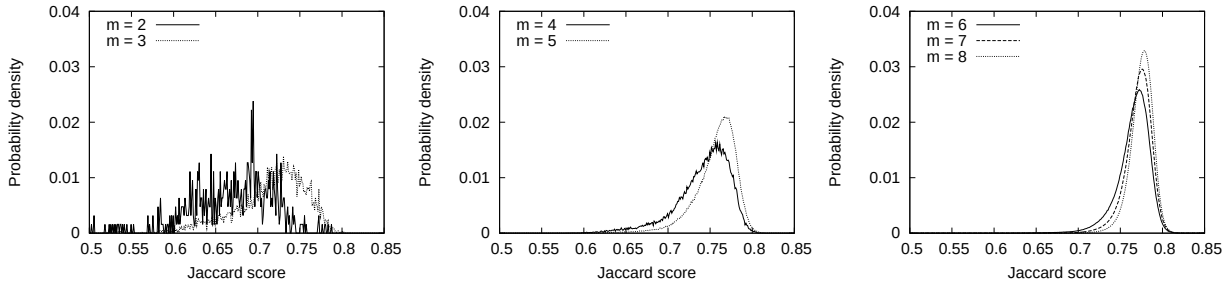


Figure 2. Distribution of Jaccard scores of m-column matrices, for various m.

2 Coding matrix space

In principle, since the matrix has $k \times m$ entries and each entry has three possible values, we could say that there exist 3^{km} possible coding matrices for the decomposition of a k-class problem into m binary problems. But this is in fact only a loose upper bound; it is not useful to distinguish matrices that differ only in the order of columns, nor to permit columns that do not have at least one $+1$ entry and at least one -1 entry. This leaves us with $u := 3^k - 2 \cdot 2^k + 1$ possible states of a column. If we furthermore require that all columns in the matrix be distinct (the effects of allowing multiple identical columns would be better achieved by introducing weighted voting when combining the predictions of binary classifiers), the number of distinct coding matrices (up to the reordering of columns) is $\binom{u}{m}$.

Sometimes the classes are organized hierarchically (a typical example are topical categories such as dmoz.org or Yahoo, but this also occurs in traditional text categorization datasets such as Reuters [4]); if a class p is a parent of class c in the hierarchy, this means that every instance of c is also an instance of p, and this in turn means that if $M_{pj} \neq 0$, then M_{cj} should be equal to M_{pj}, otherwise the j'th classifier would see class c as simultaneously positive and negative.

With this additional constraint, we can derive the following bound on the number of distinct coding matrices. Suppose that our class hierarchy consist of h levels, that all leaves are at level h and that all internal nodes have exactly b subtrees. Each column of the matrix now corresponds to a labelling of this hierarchy with the labels -1, $+1$ and 0, subject to the constraints described in the preceding paragraph. Let u_h be the number of distinct labellings for a h-level hierarchy; it can be shown that $u_h = 2 + (u_{h-1})^b$ and $u_1 = 3$. On the other hand let v_h be the number of distinct labellings that use only the labels $+1$ and 0 but not -1; it can be shown that there are v_h such labellings, where $v_h = 1 + (v_{h-1})^b$ and $v_1 = 2$. Thus the total number of different valid nontrivial labellings is $u := u_h - 2v_h + 1$, and the number of distinct coding matrices is again $\binom{u}{m}$ for this new value of u.

3 Experiments

3.1 Experimental setup

We use a small hierarchy of classes extracted from the large topic hierarchy of the Open Directory Project (ODP; see `dmoz.org`). There is a root topic with 2 subtopics, each of which has another 2 subtopics; thus we have 7 classes arranged in a 3-level hierarchy. Section 2 tells us (taking $h = 3$, $b = 2$) that there are 36 possible states of each column of the coding matrix; thus there are $\binom{36}{m}$ distinct matrices of m columns (up to the reordering of columns). Each topic except the root had 100 training documents and 100 test documens.

m	1	2	3	4	5	6
$\binom{36}{m}$	36	630	7 140	58 905	376 992	1 947 792

This means that for small values of m the number of all m-column matrices is small enough that we can afford to investigate them all; for higher values of m we can at least examine a considerable subset of all m-column matrices by random sampling. We use linear SVM [3] to train binary classification models corresponding to individual columns of the matrix.

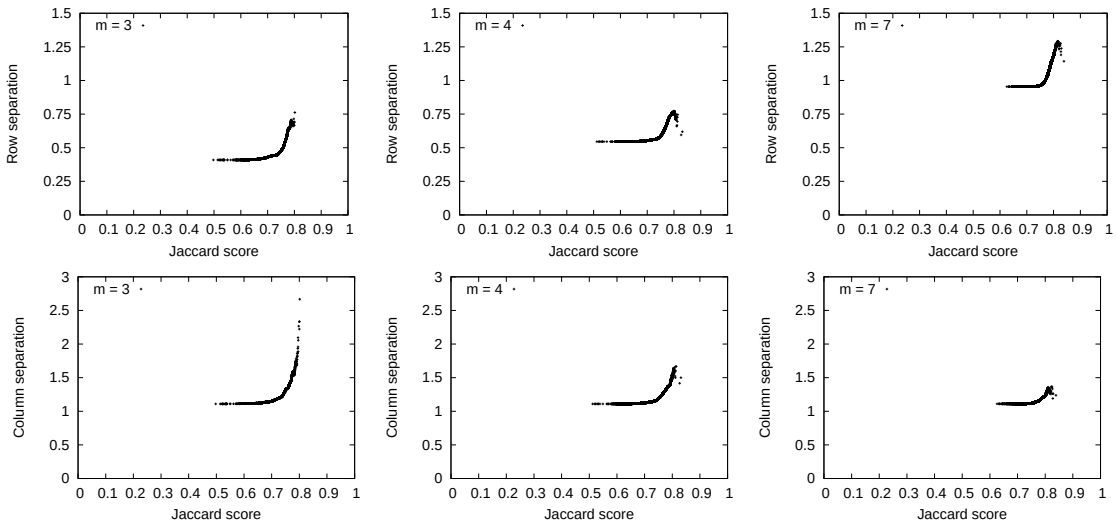


Figure 3. Scatterplots showing the relation between average row/column separation of a matrix and its Jaccard score, for $m = 3, 4, 7$.

To evaluate the classification performance of a coding matrix, we use an evaluation measure based on the Jaccard coefficient. For any class c, let $A(c)$ be the set consisting of c and all its ancestors. If an instance belonging to class c is predicted as belonging into the class c', we define the Jaccard score of this prediction to be $|A(c) \cap A(c')|/|A(c) \cup A(c')|$. We define the Jaccard score of the matrix as the average Jaccard score over all instances from the test set. This evaluation measure lies in the range $[0, 1]$, with higher values indicating better performance.

3.2 Distribution of matrix scores

A frequently used approach in the construction of coding matrices is to fill the matrix at random. If the matrix is effectively a random variable, then so is its Jaccard score, and it is interesting to investigate its probability distribution. Figure 1 shows the median, maximum and average Jaccard scores over all m-column matrices, as a function of m. The median tells us that if we choose a random coding matrix, we have a 50 % probability of getting a Jaccard score greater than or equal to the median; the maximum tells us the best performance we could get if we could examine all possible matrices.

As we can see from Fig. 1, the maximum score grows very slowly from $m = 4$ onwards, while the median keeps on growing. In other words, good matrices with a small number of columns do exist, but there are few of them and it's therefore harder to find them.

We can also describe the distribution of Jaccard scores with a histogram. We divided the range $[0, 1]$ into 1000 equal subintervals and counted the percentage of matrices whose score falls into each particular subinterval. The resulting histograms are shown on Fig. 2. As we can see from the charts there, the maximum score remains roughly the same from $m = 4$ to $m = 8$, but the mode of the distribution gradually moves upwards (i.e. as m increases, more and more matrices have higher scores).

The shape of these distributions approximately resembles that of the well-known beta distribution, $B(\alpha, \beta)$, whose probability density function is $f(x) = x^{\alpha-1}(1-x)^{\beta-1}/B(\alpha, \beta)$. For each m, we can fit a beta distribution over the histogram of Jaccard scores of m-column matrices and obtain a new pair of parameters, α_m and β_m. Fig. 4 shows some comparisons between the original distribution of Jaccard scores and the best-fit beta distribution; the mani difference is that the original distribution leans more towards higher values and has a higher mode than the best-fit beta distribution. Fig. 1(c) shows the value of the best-fit parameters α_m and β_m as a function of m. A very interesting relationship can be seen to emerge, as both α_m and β_m are approximately exponential functions of m.

3.3 Matrix score vs. other properties

An interesting question regarding coding matrices is whether the classification performance of (an ensemble based on) a matrix (e.g. as measured by its Jaccard score) is correlated to some other more easily measurable and controllable properties of the matrix; knowledge of such relationships could be used to construct good coding matrices more easily, or to guide a search through the space of coding matrices.

An example of such property is the *row* (or *column*) *separation*, which is defined as the average Hamming distance between all pairs of rows (or columns). Informal arguments can be constructed why it is desirable for the matrix to have high row and column separation. Figure 3 explores the relationship between row/column separation and Jaccard score empirically, by plotting one symbol for each matrix, with the Jaccard score being used as its x-coordinate and the row or column separation as the y-coordinate. We can see that, in general, better matrices (as measured by the Jaccard score) have high row/column separation. But on the other hand,

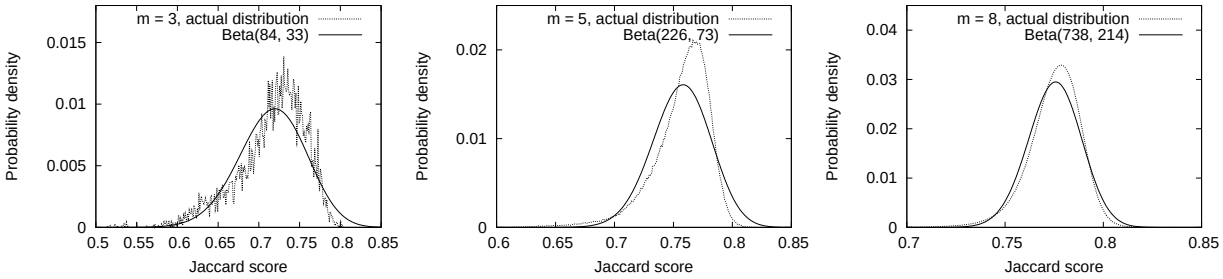


Figure 4. The distribution of Jaccard scores of m-column matrices, with a best-fit beta distribution superimposed on the same chart.

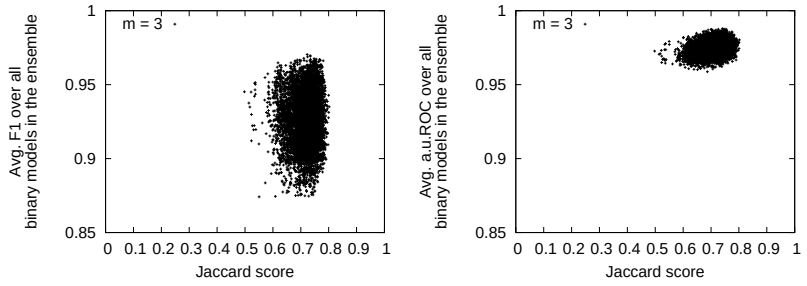


Figure 5. Scatterplots showing the relation between the average performance (F_1 on the left chart, a.u.ROC on the right one) of binary classifiers in the ensemble, and the Jaccard score of the ensemble as a whole.

the very best matrices do not actually have the highest row/column separation, which shows that merely maxmizing these two measures is not necessarily the best way to construct coding matrices.

Another interesting question is whether the performance of the ensemble is related to the performance of the individual binary classifiers in it. We evaluated each individual binary classifier from the point of view of its binary classification problem, and computed measures such as F_1 and the area under the ROC curve. We can then compute the average F_1 or a.u.ROC over all the columns of a matrix, and see whether this average is correlated to the Jaccard score of the matrix. Fig. 5 shows the results of this experiment; the correlation coefficients are 0.015 for the F_1 measure and 0.25 for the a.u.ROC, which shows that the performance of an ensemble isn't strongly correlated to the performance of the individual binary classifiers in it.

4 Conclusions and future work

We have investigated the performance of coding-matrix based ensembles on a small 7-class hierarchical text classification problem using exhaustive experiments on matrices of up to 8 columns. (1) Our experiments show that matrices with good performance can be found even with a small number of columns, but they are much more rare than if we allow more columns. (2) We have shown that the matrix performance scores are distributed approximately following a beta distribution, whose parameters are exponential in the number of columns. (3) We have shown that row/column separation is correlated with matrix performance, but maximizing separation does not lead to maximal performance. (4) We have shown that matrix performance is not correlated with the performance of the individual binary classifiers in the corresponding ensemble.

In the future, the experiments could be extended to a higher number of columns, where it would be necessary to use random sampling as the number of possible matrices is too high. Similarly it would be interesting to extend this research to problems with a larger number of classes, and to involve other easily computable matrix properties in addition to row/column separation.

References

[1] A. Berger. Error-correcting output coding for text classification. *Proc.* IJCAI, Stockholm, 1999.

[2] J. Brank, D. Mladenić, M. Grobelnik. *Generation of ontologies from very large databases*. SEKT report D1.13.1, August 2006.

[3] C. Cortes, V. Vapnik. Support-Vector Networks. *Mach. Learning*, 20(3):273–297, Sept. 1995.

[4] *Reuters Corpus, Volume 1*, English language, 1996-08-20 to 1997-08-19. Release date 2000-11-03, *http://trec.nist.gov*.

SEMANTIC ROLE FRAMES GRAPH-BASED MULTIDOCUMENT SUMMARIZATION

Ercan Canhasi, Igor Kononenko
Laboratory for Cognitive Modeling
University of Ljubljana, Faculty of Computer and Information Science
Tržaška cesta 25, SI-1000 Ljubljana, Slovenia

ABSTRACT

Multi-document summarization is a process of automatic creation of a compressed version of the given collection of documents. Recently, the graph-based models and ranking algorithms have been extensively researched by the extractive document summarization community. While most work to date focuses on sentence-level relations in this paper we present graph model that emphasizes not only sentence level relations but also the influence of under sentence level relations (e.g. a part of sentence similarity). By using the proven cognitive psychology model (the Event-Indexing model) and semantic role parsing for generating the frame graph, we establish the bases for distinguishing the sentence level relations. Based on this model, we developed an iterative frame and sentence ranking algorithm, based on the existing well known PageRank algorithm. Experiments are conducted on the DUC 2004 data sets and the ROUGE (Recall-Oriented Understudy for Gisting Evaluation) evaluation results demonstrate the advantages of the proposed approach.

1 INTRODUCTION

Multi-document summarization (MDS) aims to filter the most important information from a set of documents to generate a compressed summary. Recently, the graph-based models and sentence ranking algorithms based on these models have been extensively researched.

Conventionally, they model a document or a group of documents as a text graph composed by taking a text unit, such as a term or a sentence as a node, and similarity or affinity between text units as edges. The significance of a node in a graph is then estimated by graph-based ranking algorithms which take into account global information recursively computed from the entire graph. Sentences in document(s) are ranked according to the computed node significance and the most important ones are selected to form an extractive summary.

While most work to date focuses on the sentence and the document level relations, in this work, considering importance of the intra sentence relations and being inspired by Event-Indexing model [10], a well known cognitive model for text understanding and representation, we present a new graph model, the frame graph. A frame graph is composed by taking semantic role frames [2] as nodes and similarities between frames as edges. The significance of a node in a frame graph is estimated by graph-based ranking algorithms.

The remainder of this paper is organized as follows. Section 2 reviews existing graph-based summarization models. Sections 3 and 4 introduce the proposed summarization method. After that, Section 5 reports experiments and evaluation results. Finally, Section 6 concludes the paper.

2 RELATED WORK

The graph-based models have drawn considerable attention from the extractive document summarization community in the past years [3,6].

So far, the most popular graph-based ranking algorithms applied in document summarization are PageRank [1] and HITS [4] and their variations. Those algorithms make use of "voting" or "recommendations" between sentences to evaluate the importance of sentences in the documents.

Erkan and Radev [3] represented documents as a weighted undirected graph by taking sentences as nodes and the cosine similarity between sentences as the edge weight function. An algorithm called LexRank, adapted from PageRank, was applied to calculate the sentence significance, which was then used as the criterion to rank and select summary sentences. Meanwhile, Mihalcea and Tarau [6] presented their PageRank variation, called TextRank, in the same year.

All above mentioned exemplary work was concerned with generic summarization. Later on, graph-based ranking algorithms were introduced in query-oriented summarization too, when this new challenge became a popular research topic recently. For example, a topic-sensitive version of PageRank was proposed in [7]. A variety of other graph-based methods have been proposed for topic-focused multi-document summarization [8,9]. Different from generic summarization, a query-oriented summarization is necessarily driven by queries.

3 MDS VIA SEMANTIC ROLE GRAPHS ARGUMENTS

The purpose of this study is to show that it is possible to improve the efficiency of summarization using a semantically richer representation. Here by richer representation we mean the semantic graph of a document or a set of documents. The document set $D = d_1, d_2, \dots, d_n$

is represented as a weighted undirected text graph G by taking parts of sentences (semantic frames) in D as vertices and adding an edge to connect the two vertices if the two frames of concerned sentences are similar enough.

3.1 Motivation for using semantic role graphs

Summarization task requires understanding the document and presenting the important parts. In extractive summarization, this task is achived by selecting the sentences to be included in the summary. The most common method to solve this problem is to rank the sentences according to their informativeness.

Since humans tend to include sentences containing most frequent words in their summaries, the word-based frequency calculations for sentence scoring are baseborn approachs for MDS. However, this approach is semantically incomplete, because words alone usually do not carry semantic information. On the other hand, even if humans do not always agree on the content to be added to a summary, they perform very well on this task. Therefore our goal should be to find a way of mimicking the cognition behind the human like summarization process.

Our motivation for using SRL frames in sentence scoring for MDS originates from given concerns. Instead of using individual terms for sentence scoring, we exploit semantic arguments and relations between them by using the psychology cognitive situation model, namely the Event-Indexing model.

3.2 Event Indexing Model and Semantic Role Labeler

According to Event-indexing model a human-like system should keep track of five indices while reading the document. Those indices are *protagonist*, *temporality*, *spatiality*, *causality* and *intention*, with the given descending order of importance. One can also show that the semantic role parser's output can be mapped to the above proposed cognitive model.

Semantic roles are defined as the relationships between syntactic constituents and the predicates. Most sentence components have semantic connections with the predicate, carrying answers to the questions such as who, what, when, where etc. From the aspect of the semantic parser, frame arguments can be mapped to cognitive model indices as follows:

- A protagonist can be found in an answer to question "who", or more precisely in arguments A0 or A1 or A2. Argument A0 is the subject of the frame, as shown in Table 1, A1 is the object and A2 is the indirect object. Although in original work the protagonist is defined as a person around whom the story takes place, we see it reasonable to expand the notion of protagonist to the subject or object that can be everything, from a person to an organization or some abstract concept.
- Temporality is the temporal information in each frame and can be extracted from the frame argument AM_{TMP}.
- Spatiality is the space or location information of each frame and is equal to argument AM_{LOC}.
- Causality indexing is concerned with actions of frames so it can be mapped to the frame predicate.
- The intentionality-indexing is quite vague but since its weight of significance is less than of the others, as defined in the original work, we decided to omit it in this early versions of the system.

The SRL parser takes each sentence in the document set and properly labels the semantic word phrases. We refer to these phrases as semantic arguments or shortly arguments. There is an issue related to the SRL parsing process that we should take into account. For each verb in a sentence, the SRL parser provides a different frame. It considers the verb as the predicate of the sentence and tries to label the remaining part of the sentence as proper arguments. However, if the selected verb is not the actual predicate, the parser fails to identify most of the words as a part of an argument. Therefore, we consider the frame that leaves the least number of terms unlabeled as the complete parse of the sentence. In our calculations we use also the rest of frames but we treat them as incomplete. Since we don't want to lose information that can be brought to the resulting graph, instead of eliminating partially parsed frames we use them, but with lower weight in the similarity calculation.

Arguments labeling		Arguments modifier	
rel	verb	AM_{ADV}	Adverb mod.
A0	Subject	AM_{DIR}	Direction
A1	Object	AM_{DIS}	Discourse mrk
A2	Ind. object	AM_{LOC}	Location
A3	Start point	AM_{MNR}	Manner
A4	End point	AM_{NEG}	Negation
A5	Direction	AM_{PRD}	Sec. Predicate
		AM_{PRP}	Purpose
		AM_{TMP}	Temporal mrk.

Table1: Representation of label arguments and modifiers.

4 PROPOSED METHOD

The summarization method, we propose, works in the following way, as illustrated in Fig.1. First, the documents are given to the SRL parser where the semantic arguments from each parsed sentence are extracted. We calculate the composite similarity between all semantic frames based on the event-indexing. Then we generate a semantic graph where nodes are semantic frames and edges are the composite similarity values.

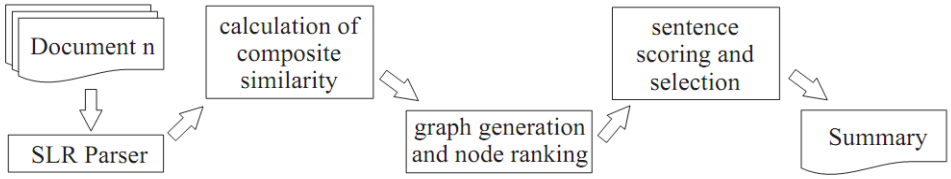


Figure 1: MDS System

Next we use the modified version of PageRank for identifying the significant edges in the graph. Later, we sum the PageRank scores of semantic frames, originating from the same sentence, and we use it as a score for sentence scoring. The next step aims to further remove redundant information in the summary by penalizing the sentences largely overlapping with other highly ranked sentences. Based on the text graph and the obtained rank scores, a greedy algorithm is applied to inflict the diversity penalty and compute the final rank scores of the sentences. Subsequently, the top scoring sentences are selected one-by-one and put into the summary.

4.1 Graph modeling

In this section we present our novel graph model, which is used in the frame ranking algorithm, presented in the next section. Let a set of documents D be a text similarity graph $G = (V, E, \alpha_v, \alpha_e)$ where V represents the frame vertex set, $E \subseteq V \times V$ is a frame edge set. $\alpha: V \longrightarrow \Re_+$ is defined to label frame vertices, while $\beta: E \longrightarrow \Re_+$ is a function for labeling frame edges.

Vertex Function	$\alpha(f_i) = \begin{cases} \frac{1}{N} \; if \; frame \; complete \\ \frac{1}{2N} \; otherwise \end{cases}$
Edge Function	$\beta(f_i, f_j) = sim_{composite}(f_i, f_j)$

Table 2: Different frames of a sentence.

The SRL frame edge function is formulated as the composite similarity function of two frames f_i and f_j. Let N be the total number of frames in a documents set. The frame vertex function assigns to frame vertices the value of $\frac{1}{N}$ or $\frac{1}{2N}$, depending on their completeness, where incomplete frames have lower weight. In Table 2, N is the total number of the frames in a document set, and the similarity between any two frames is defined by the composite similarity function, which will be detailed in the next paragraphs.

Our goal is to capture the similarity and redundancy between sentences, but at a lower structural and a higher semantic level. To accomplish this, we use the event-indexing model as the base for calculations of semantic similarity between frames of semantic role parser outputs, namely frames. According to this model we need to define the similarity measure for protagonist (prt), temporality (tmp), spatiality (spt), and causality(cst).

$$sim_{prt}(f_i, f_j) = \alpha_1 sim(A0_i, AO_j) + \alpha_2 sim(A1_i, A1_j) + \alpha_3 sim(A2_i, A2_j) + \alpha_4 sim(A0_i, A1_j) + \alpha_5 sim(A0_i, A2_j) + \alpha_6 sim(A1_i, A2_j)$$

$$sim_{tmp}(f_i, f_j) = sim\left(AM_{TMP_i}, AM_{TMP_j}\right)$$

$$sim_{spt}(f_i, f_j) = sim\left(AM_{LOC_i}, AM_{LOC_j}\right)$$

$$sim_{cst}(f_i, f_j) = sim(Predicate_i, Predicate_j)$$

In order to have the flexible weighting scheme we use coefficients $\alpha_1 = \alpha_2 = \alpha_3 = 0.25$; $\alpha_4 = \alpha_5 = 0.10$; $\alpha_6 = 0.5$;

The compose similarity is defined as:

$$sim_{cmp}(f_i, f_j) = [\beta_1 sim_{prt}(f_i, f_j) + \beta_2 sim_{tmp}(f_i, f_j) + \beta_3 sim_{spt}(f_i, f_j) + \beta_4 sim_{cst}(f_i, f_j)] / \#arguments$$

where $\beta_1 = 0.4$; β_2=0.3; $\beta_3 = 0.2$; $\beta_4 = 0.1$. The values for coefficients are chosen based on the cognitive model which gives emphasis in the decreasing order to the protagonist, temporality, spatiality and causality. We also normalize the composite similarity value with the number of arguments used in the calculation of similarity.

4.1 Frame graph-based ranking algorithm

In previous section the idea of frame similarity graph is presented. Based on it, in this section we present a modified iterative graph-based sentence ranking algorithm. Our algorithm is extended from those existing PageRank-like algorithms reported in the literature that calculate the graph only in the sentence level [3,6].

In the summary, PageRank method (in matrix notation), as described in the original paper [1], is

$$\pi^{(k+1)T} = \alpha \pi^{(k)T} \boldsymbol{H} + \left(\alpha \pi^{(k)T} a + 1 - \alpha\right) \boldsymbol{v}^T$$

where $\boldsymbol{H}$ is a very sparse, raw sub stochastic frame similarity matrix, α is a scaling parameter between 0 and 1, π^T is the stationary row vector of $\boldsymbol{H}$ called the PageRank vector, $\boldsymbol{v}^T$ is a complete dense, rank-one teleportation vector and a is a binary dangling node vector. In terms of the sentence ranking the matrix $\boldsymbol{H}$ is an adjacency matrix of frames, $\boldsymbol{v}^T$ is the preference vector of frames and the resulting π^T is the frame ranking vector.

5 PLEMINARY RESULTS

The DUC[1] 2004 data set from DUC was tested to analyze the efficiency of the proposed summarization method. The Task 2 in the DUC 2004 is to generate a short summary (665 bytes) of an input set of topic-related news articles. The total number of document groups is 50, with each group containing 10 articles on average.

For each group, four NIST assessors were asked to create a brief summary. The machine-generated summaries are evaluated using ROUGE [5], the automatic n-gram matching which measures performance based on the number of co-occurrences between machine-generated and ideal summaries in different word units. The 1-gram ROUGE score (a.k.a.ROUGE-1) has been found to correlate very well with human judgments at a confidence level of 95%, based on various statistical metrics.

Even though in this version of method we did not consider sentence positions or other summary quality improvement techniques such as sentence reduction, its overall performance is promising, see Table 3. The use of the frame graph model in summarization can make considerable improvements even though the results presented here do not report a significant difference.

System	ROUGE-1[2]
Avg. of human assessors	0.403[0.383,0.424]
Best Machine (SYSID=65)	0.382[0.369,0.395]
Median Machine (SYSID=138)	0.343[0.328,0.358]
Worst machine (SYSID=111)	0.242[0.230,0.253]
Our model	0.374[0.359,0.389]
LexRank	0.369[0.354,0.382]
2 (NIST Baseline) (Rank 25/35)	0.324[0.309,0.339]
Random baseline:	0.315[0.303,0.328]

Table 3: ROUGE-1 scores of the DUC 2004 and evaluation of our model.

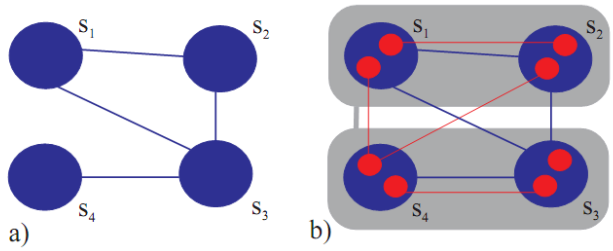


Figure 2: Summarization graph model (a) before and (b) after introducing multilayered model.

6 DISCUSSION AND FUTURE WORK

We have presented a frame graph model and a ranking algorithm for generic MDS. The main contribution of our work is introducing the concept of the frame graph model. The results of applying this model in extractive summarization are quite promising. There is work still left to be done, however. While most work to date focuses on homogeneous connectedness of sentences and heterogeneous connectedness of documents and sentences (e.g. sentence similarity weighted by document importance), in the future we hope to be able to present a novel 3-layered graph model that emphasizes not only sentence and document level relations but also the influence of under sentence level relations (e.g. a part of sentence similarity). By using an intelligent weighting scheme we plan to add two more layers, namely the sentence and the document layers, to frame graph model presented in this work (see Figure 2), which will yield to a richer multilayered graph model with the inter and intra sentence and the documental level relations. Currently, we are also working on further improvements of the model, and its adaptation to other summarization tasks, such as query and update summarizations.

[1] Document Understanding Conference(http://duc.nist.gov)

[2] 95% confidence interval

References

[1] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. Computer Networks, 30(1-7):107-117, 1998.

[2] Xavier Carreras and Lluis Marque. Introduction to the conll-2004 shared task: Semantic role labeling. In CoNLL, pages 89-97, 2004.

[3] Gunes Erkan and Dragomir R. Radev. Lexrank: Graph-based lexical centrality as salience in text summarization. J. Artif. Intell. Res. (JAIR), 22:457-479, 2004.

[4] Jon M. Kleinberg. Authoritative sources in a hyperlinked environment. J. ACM, 46(5):604-632, 1999.

[5] Chin-Yew Lin and Eduard H. Hovy. Automatic evaluation of summaries using n-gram co-occurrence statistics. In HLT-NAACL, 2003.

[6] Rada Mihalcea and Paul Tarau. Textrank: Bringing order into text. In EMNLP, pages 404-411, 2004.

[7] Jahna Otterbacher, Gunes Erkan, and Dragomir R. Radev. Biased lexrank: Passage retrieval using random walks with question-based priors. Inf. Process. Manage., 45(1):42-54, 2009.

[8] Xiaojun Wan. Document-based hits model for multi-document summarization. In Tu Bao Ho and Zhi-Hua Zhou, editors, PRICAI, volume 5351 of Lecture Notes in Computer Science, pages 454-465. Springer, 2008.

[9] Furu Wei, Wenjie Li, Qin Lu, and Yanxiang He. A document-sensitive graph model for multi-document summarization. Knowl. Inf. Syst., 22(2):245-259, 2010.

[10] R.A. Zwaan, M.C. Langston, and A.C. Graesser. The construction of situation models in narrative comprehension: an event-indexing model.Psychological Science, 6(5) : 292-297, 1995.

On the complementarity of OLAP and rich associations mining

David Chudán, Vojtěch Svátek
Faculty of Informatics and Statistics, University of Economics, Prague
Nám W. Churchilla 4, 130 67 Prague, Czech Republic
e-mail: xchud01@vse.cz, svatek@vse.cz

ABSTRACT

The paper presents a comparison of possibilities and proposal for complementary use of OLAP and the rich variant of association rule mining based on the GUHA method. The rationale is to determine the point when it is useful for the analyst to proceed from OLAP to descriptive data mining, as well as the point of return from data mining results to OLAP in order to see them in a broader view.

1 INTRODUCTION

In large datasets (or warehouse environments) it is possible to use many types of analysis. OLAP analysis and data mining (DM) are the most frequent. They are based on completely different techniques and algorithms and often applied on different kinds of analytical problems; yet they are often part of a single business intelligence (BI) solution, see Fig. 1. With a certain simplification we can say that the regular use of OLAP analysis can solve analytical questions like "what happened last year with the profit in the Northern region", while DM is trying to answer the question "why it happened". The task of (especially, the predictive type of) DM is very often complementary to OLAP analysis, when OLAP only pinpoints some problem while DM can provide an insight into the problem and estimate its reason, which may eventually lead to finding a solution to the problem.

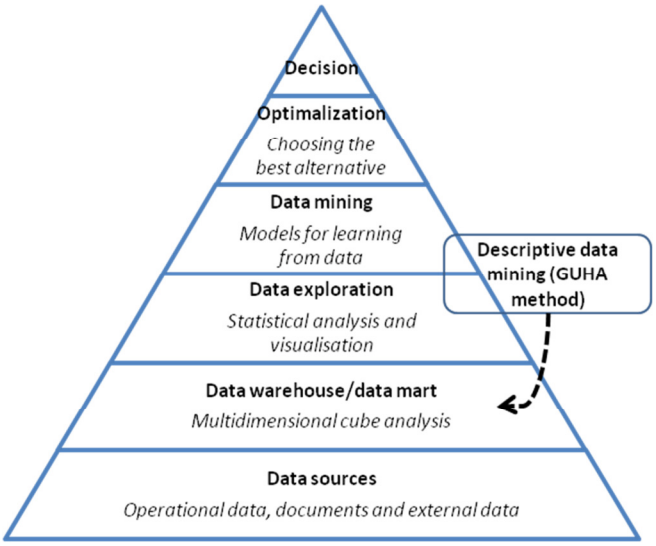


***Figure 1** – The main components of BI [1] with the loop from data mining back to OLAP analysis*

In this paper we however advocate a somewhat different scenario, suitable for descriptive DM such as the GUHA method detailed in Section 3. Descriptive DM represents a transitional analysis type, which exhibits higher representational power than statistical analysis while preserving its exploratory nature (in contrast to predictive DM model building). The crucial novelty of the scenario can be outlined as follows. Let us imagine that we found some interesting rules using DM and we would like to examine them in a broader perspective. Is it possible and helpful at this point to switch back to OLAP analysis and see the surroundings of the rules (the loop in *Figure 1*)?
Sections 2, 3 and 4 briefly present and verbally compare OLAP and the DM method, Section 5 explains the data set and task, section 6 shows the course of complementary analysis and finally, Section 7 wraps up the paper.

2 OLAP (MULTIDIMENSIONAL) ANALYSIS

The objective of OLAP (multidimensional) analysis is to gain insight into the meaning contained in databases [2]. OLAP analysis is based on the OLAP cube, a data structure that overcomes several limitations of two-dimensional relational databases. (Despite the word "cube", there is no limitation of three dimensions. The number of dimensions can be tens and depends on the used BI platform.)
The basic operations in OLAP analysis are [2]:

- **Drill-down/up** – Analytical technique that enables to navigate among levels of data granularity from the most summarized to the most detailed.
- **Slice/dice –** A slice is a subset of a multi-dimensional array corresponding to a single value for one or more members of the dimensions not in the subset. Slicing enables to see different slices of information presented in the OLAP cube. Dice means to slice a data cube in more than one dimension.
- **Pivot (rotate) –** Pivoting means the changing of the dimensional orientation of the data view.

The examples of OLAP used throughout this paper were provided using the Pentaho BI suite. It is an open source suite that integrates ETL (extract, transform, load), dashboard, reporting, workflow and data mining capabilities. Due to the Open Source Business Intelligence and Reporting survey [3], open source solutions are more preferable than commercial ones, and Pentaho is the open source leader with market share of 30%.

3 RICH ASSOCIATION RULES MINING

Association rules are one of the data mining techniques (sixth most used DM method[1]) used to discover interesting relations between variables in large datasets. Although the idea is much older, it has been widely popularized in 90s with market basket analysis where the aim of is to determine which items in a supermarket are bought together [4].

In the context of this paper we however do not consider traditional associations but rich associations, in particular those produced by the LISp-Miner data mining software [5] with its underlying GUHA method [6], the original Czech method of exploration analysis. GUHA provably overtakes the market basket analysis [4] and also the a-priori algorithm [7] in terms of representational power, while its recent implementations guarantee high throughput. The rich representational features of GUHA, such as (runtime-generated) value groupings in rule literals, make it particularly suitable for alignment with OLAP analysis.[2]

4 MAJOR DIFFERENCES

Both described methods, OLAP analysis and rich association rules mining, can be used to analyze the same kind of tabular data. The main difference lies in the level of model granularity and the degree of automation. The GUHA attributes are not a priori labeled for a role in the analysis (antecedent / consequent) as the dimensions / measures in OLAP. In OLAP analysis the (numerical) *measure* fields are displayed in terms of aggregations such as sums or averages; the internal structure of *dimensions* is not altered during the analysis. GUHA, in contrast, relies on pre-processing of numerical data into *discrete* categories, and these can be further amalgamated at *runtime*.

Furthermore, OLAP analysis is carried out *manually*, as the analyst "browses" through data and inspects the view of data s/he currently needs. GUHA, on the other hand, generates many individual hypotheses in one shot [8], even if some of them are very similar to others.

5 DATASET AND TASK

The dataset used in this demonstrative case study is an updated version of the Financial Data Set first introduced in the PKDD'99 Discovery Challenge [9]. The Financial Dataset consists of 8 tables describing the operations of bank customers. For this task only one table is used, the Loans table, which contains the following columns: *loan_id, birth number, district, salary, amount, payments, duration* and *status.*

The task is to identify, between 6181 customers available in the dataset, subgroups with high occurrences of bad loans. From the available columns, for better clarity of the analysis and inserted figures only four attributes are used: *district* (the location where the client lives), *amount* (overall amount of the loan), *duration* (the time period for which the loan is granted) and *status***.** Status is the key indicator of the bad loans and potentially unreliable clients. Status has four possible values, A, B, C and D [9].

- **A** stands for finished contracts with no problem.
- **B** stands for finished contracts with loan not payed.
- **C** stands for running contracts that are OK so far
- **D** stands for running contracts, where the client is in debt and it is probable than s/he will have problems with paying the loan

The task considered here is to describe groups of clients that tend to belong to groups B and D.

6 SAMPLE COURSE OF ANALYSIS

6.1 OLAP analysis

OLAP analysis is easy to use, as it is very similar to contingency tables known from spreadsheet processors. We choose individual columns from the dataset, thus creating the dimensions of the cube, the data element that categorizes each item into non-overlapping regions. Let the dimensions in this analysis be *district*, *amount* and *duration*. The measure can generally be a variety of key performance indicators in business environment, e.g. days, amount of money or more complex ratios as 'cost per person per week'. In our analysis let the measure simply be the count of status, more precisely the count of status that indicates bad loan quality (status B and D). Fig. 2 shows the absolute numbers of aggregated status B and D drilled through all dimensions (for the example of Brno district).

		Measures						
		Status						
		Duration						
District	Amount	All Durations	12	13	24	36	48	60
All Districts	All Amounts	727	111	1	151	158	156	150
Beroun	All Amounts	18	3		3	5	2	5
Blansko	All Amounts	9	1		2		3	3
Breclav	All Amounts	9	1		2	2	2	2
Brno	All Amounts	45	5		9	10	13	8
	75000	5	1		3	1		
	80000	2			1	1		
	82000	1			1			
	84600	1				1		
	85000	3			1	2		
	90000	3				3		
	100000	3			1	2		
	200000	4	1				3	
	210000	2					2	
	220000	1	1					
	235000	1					1	
	350000	2					1	1
	360000	5	1		1		3	
	380000	1			1			
	386000	1					1	
	450000	1					1	
	473280	1						1
	475000	2						2
	480000	2					1	1
	500000	3						3
	535000	1	1					

***Figure 2** - Analysis view in Pentaho BI suite*

[1] http://www.kdnuggets.com/polls/2007/data_mining_methods.htm

[2] We however do not show these features here for simplicity.

As we can see, three dimensions are on the border of clarity, and with further dimensions the table would be rather confusing.
Based on this view the analyst can conclude that no matter of the amount of the loan, only few clients from Brno region have problem with paying short-term loans (for one year). But finding such rules manually is very time consuming, as the analyst has to drill down and roll up through the individual districts and individual amounts.
For more dimensions this task becomes impossible, hence more complex relationships are beyond the reach of OLAP.

6.2 Associations mining with high data granularity

In the DM process and important task is to prepare the data for analysis via data pre-processing. The task of data pre-processing usually consists in grouping data into subsets with common characteristics, and in 'binning' numerical values into intervals. In the first DM task shown here only basic pre-processing was done, which consists in creating a group 'good loan quality' (A and C) and 'bad loan quality' (B and D) for the *Status* attribute.

The task is formulated as a template rule with three elements in antecedent, such that every antecedent attribute can be valued by a subset of its domain of size exactly (i.e. minimum as well as maximum) 1:

District(subset 1-1) & Duration(subset 1 - 1) & & Amount(subset 1-1) => Quality (Bad)

The interest measures (confidence and support) thresholds were set as follows: *minConf=0.9* and *support=10.* Detailed explanation of the DM setting in LISp-Miner is out of the scope of this paper, details can be found at [10].

The DM tool finds 16 rules in data, all of which have *Conf=1* (see Fig. 3).

1	2	1.000	Amount(30276) & Duration(12) & District(Tabor) III Status(Bad)
2	4	1.000	Amount(96396) & Duration(12) & District(Sokolov) III Status(Bad)
3	6	1.000	Amount(103680) & Duration(36) & District(Havlickuv Brod) III Status(Bad)
4	7	1.000	Amount(103680) & Duration(60) & District(Havlickuv Brod) III Status(Bad)
5	9	1.000	Amount(215388) & Duration(36) & District(Most) III Status(Bad)
6	11	1.000	Amount(250000) & Duration(48) & District(Bruntal) III Status(Bad)
7	1	1.000	Amount(30276) & District(Tabor) III Status(Bad)
8	3	1.000	Amount(96396) & District(Sokolov) III Status(Bad)
9	5	1.000	Amount(103680) & District(Havlickuv Brod) III Status(Bad)
10	8	1.000	Amount(215388) & District(Most) III Status(Bad)
11	10	1.000	Amount(250000) & District(Bruntal) III Status(Bad)
12	12	1.000	Duration(12) & District(Sokolov) III Status(Bad)
13	13	1.000	Duration(36) & District(Havlickuv Brod) III Status(Bad)
14	14	1.000	Duration(48) & District(Bruntal) III Status(Bad)
15	15	1.000	Duration(48) & District(Jesenik) III Status(Bad)
16	16	1.000	Duration(60) & District(Havlickuv Brod) III Status(Bad)

***Figure 3** –The results in LISp-Miner system*

These rules describe 16 groups of clients with bad loans quality. From the list of rules we can easily identify problematical districts (Sokolov, Havlickuv Brod, Bruntal etc.) and problematical amounts of loan. However, the bank analyst can, for example, ask the question (referring to the highlighted rule in Fig. 3): *Is it really dangerous for us to provide a loan to a client from Bruntal or is it only bound to a certain amount of money or to certain duration?*

We can resolve the question in several ways. One option is to change the confidence and support thresholds in the DM task itself. Another option is to exclude all districts except Bruntal (cf. Fig. 4) in the task (but even in this case it is necessary to change the thresholds, to *minConf=0.5* and *support=7*). Now we can see that the probability that a client from Bruntal will have problems with paying his/her loan is 67,3% (see the highlighted rule in Fig. 4).

1	3	1.000	Amount(250000) & Duration(48) & District(Bruntal) III Status(Bad)
2	5	1.000	Amount(300000) & Duration(36) & District(Bruntal) III Status(Bad)
3	2	1.000	Amount(250000) & District(Bruntal) III Status(Bad)
4	4	1.000	Amount(300000) & District(Bruntal) III Status(Bad)
5	7	1.000	Duration(48) & District(Bruntal) III Status(Bad)
6	1	0.673	District(Bruntal) III Status(Bad)
7	6	0.500	Duration(36) & District(Bruntal) III Status(Bad)

***Figure 4** - Results of data mining task, all disctricts except Bruntal excluded*

However, another way is to go back to OLAP analysis, drill down to district Bruntal, and display all variants of *amount* and *duration.* As we can see in Fig. 6, [3] the bad status in not related to lower amounts of the loan. So, these clients, based on this dataset, are credible for the loan.

District	Amount	Status	Duration	% of Status
Bruntal	120000	A	24	1,82%
	125000	A	24	3,64%
	130000	A	24	5,45%
		C	36	1,82%
	140000	A	24	1,82%
	150000	A	24	1,82%
	220000	A	24	1,82%
		C	36	7,27%
	225000	C	36	3,64%
	230000	C	36	1,82%
	240000	C	36	1,82%
	245000	D	48	1,82%
	250000	D	48	21,82%
	299088	B	36	1,82%
	300000	B	36	14,55%
		D	48	1,82%
	325000	D	48	1,82%
	330000	D	48	1,82%
	335000	D	48	5,45%
	350000	D	48	9,09%
	375000	D	48	3,64%
	400000	D	48	3,64%

***Figure 5** – Filtered analyzer report, only district Bruntal included*

6.3 Associations mining with lowered data granularity

In the second DM task more preprocessing was done. The granularity of the field *district* (77 values) was decreased by grouping them into *regions* [11] (14 values) and the *amount* was divided into four intervals, starting from (0:100000> and ending by 300000+.

[3] This is a different view on data in Pentaho BI Suite, the „Analyzer report", which is designed to create reports from data, which enables more options than the standard analysis view.

The task is formulated analogously:

Region(subset 1-1 & Duration(subset 1 - 1) & & Amount(subset 1-1) => Quality (Bad)

The interest measures were set as *minConf=0.6* and *support=20.* With this setting, one rule was found:

Region(Karlovarsky) & Amount(0;100000> & Duration(12) => Quality(Bad) with Conf=0.69

This rule is potentially very interesting for the bank, because it tells that clients from the Karlovarsky region (which contains three districts: Cheb, Sokolov and Karlovy Vary), with relatively low loan amount (up to 100000) and duration of the loan for one year, are unreliable to pay their commitments, with probability of nearly 70%. As in the previous task we can now return to OLAP and check the data there (see Fig. 6).

District	Amount	Duration	Status	Count of Status
Cheb	75000	12	A	8
	90000	12	A	1
Karlovy Vary	35000	12	B	1
	50000	12	B	4
	60000	12	B	1
	75000	12	B	2
	90000	12	B	1
Sokolov	60000	12	B	1
	75000	12	B	2
	80000	12	B	3
	95000	12	B	2
	100000	12	B	3

Figure 6 *– Filtered analyzer report, districts from Karlovarsky region with duration of 12 month included*

Now it is clear that association rules mining could lead the bank management to incorrect conclusions. The rate of unreliable clients in this region is very high (status B) and, depending on the bank's risk management policy, they could deny all loans from the whole region. However, as we can see in Fig. 7, the Cheb district has no bad loans for the duration 12 months at all, so this decision could eliminate many reliable clients.

7 CONCLUSIONS

This paper shows that using both analysis methods, OLAP and rich association rules mining, on the same dataset as complements can be useful and may prevent premature, inaccurate conclusions. OLAP analysis is very time-consuming and with an increasing number of attributes it becomes impossible to manually discover all interesting relationships between data. On the other hand, association rules mining is (with appropriate knowledge) fast and gives us many clear and accurate results. But it is usually at the high data granularity level, so it is possible to lose the complex view to data and arrive to premature conclusions.

The imminent future work consists in a more substantial case study and in construction of a formal apparatus for relating OLAP and GUHA hypotheses. In longer term, implementation of a software facility allowing for system interoperation and user interface integration is envisaged.

8 ACKNOWLEDGEMENTS

The research is partially supported by the CSF project no. 201/08/0802.

References

[1] C. Vercellis: Business intelligence: data mining and optimization for decision making. pp. 10. John Wiley and Sons, 2009. ISBN: 0470511389

[2] OLAP and OLAP server definition. The OLAP Coouncil. 1995. Available at: *http://www.olapcouncil.org/research/glossaryly.htm*

[3] J. C. Diaz: Adoption and Usage Survey: Open Source and Business Intelligence and Reporting, 2009, BeyeNetwork

[4] R. Agrawal; T. Imielinski; A. Swami: Mining Association Rules Between Sets of Items in Large Databases", SIGMOD Conference 1993: pp. 207-216

[5] M. Šimůnek: Systém LISp-Miner — akademický systém pro dobývání znalostí z databází, Historie vývoje a popis ovládání. skripta VŠE, Praha, Oeconomica, 2010, 106 stran, ISBN: 978-80-245-1699-8.

[6] P. Hájek; T. Havránek: Mechanising Hypothesis Formation – Mathematical Foundations for a General Theory. Berlin – Heidelberg – New York, Springer-Verlag, 1978, 396 pp.

[7] R. Agrawal; R. Srikant: Fast Algorithms for Mining Association Rules. Proc. 20th Int. Conf. Very Large Data Bases. 1994.

[8] T. Kliegr; D. Chudán; A. Hazucha; J. Rauch: SEWEBAR-CMS: A System for Postprocessing Association Rule Models. Alexandria 21.10.2010 – 23.10.2010. In: *RuleML-2010 Challenge.* Washington : CEUR-WS, 2010, s. 1–8. ISSN 1613-0073.

[9] http://lisp.vse.cz/pkdd99/

[10] J. Rauch; M. Šimůnek: An Alternative Approach to Mining Association Rules. In:Lin T Y, Ohsuga S, Liau C J, and Tsumoto S (eds): Foundations of Data Mining and Knowledge Discovery. Springer-Verlag, 2005, pp. 219-239.

[11] http://obce.sweb.cz/

Nonlinear Principal Component Analysis for Compression of Spectral Data

*Syamak Farajikhah, Faraz Madanchi and Seyed Hossein Amirshahi**
Department of Textile Engineering
Amirkabir University of Technology (Tehran Polytechnic)
Tehran, Iran
Tel: +98 21 64542642; Fax: +98 21 66400245
*email: hamirsha@aut.ac.ir

ABSTRACT

Abstract— In this study, the principal component analysis (PCA) technique and its nonlinear version (NLPCA) are employed for the compression and reconstruction of spectral data. The reflectance spectra of 1269 matt Munsell color chips are used as original dataset in 400 to 700 nm with 10 nm intervals. The hidden patterns of spectral data are determined by employing the classical PCA as well as its nonlinear version. Different numbers of feature vectors are used in both methods and the results compared by using the root mean square error (RMS), the goodness fit coefficients (GFC) as well as the color difference values under D65 illuminant and 1964 standard observer. Results show the priority of NLPCA over the PCA in low-dimensional spaces i.e. up to 4 basic functions, while different results are observed in higher-dimensional spaces.

1 INTRODUCTION

While the spectral data, like reflectance spectra of surfaces, provide full information about the visual properties of objects in different viewing conditions, they suffer from the sizes of information that would be transferred. In fact, opposed to colorimetric tristimulus values the spectral data are composed of several channels, e.g. 31 points in visible spectrum from 400 to 700 nm with 10 nm interval. To clarify the differences between the spectral and colorimetric data, it should be emphasized that the spectral data is unique for each object and is known as the "fingerprint" of sample while different objects with different spectral properties could provide identical color in a given set of viewing condition. This issue is known as metamerism and is evident in most color reproducing systems.

In the recent decades, some methods have been presented to reduce the sizes of spectral data while the main features of data have been kept. In fact, it was shown that the three dimensional colorimetric data are not fairly enough to convey the adequate information about the objects in different viewing conditions and the 31 dimensional behavior of the reflectance spectra would prepare surplus information and lead to some problems in data transformation. Hence, some mathematical techniques have been employed to extract the basic functions of spectral information to present them in the lower dimensional spaces with the minimum loss of information. One of the most applicable techniques which have been used in the field of spectral data reducing method is the principal component analysis technique abbreviated by PCA.

PCA has been widely used in compression and reconstruction of reflectance spectra of surface colors [1-6]. Fairman and Brill [2] explained the application of classical PCA method for the compression of spectral reflectance as well as the reconstruction of spectra from the corresponding CIEXYX tristimulus values. Different methods were also introduced to increase the efficiency of compression as well as the reconstruction techniques by choosing the suitable sets in the learning step and/or weighing the samples prior to extraction of principal directions [5].

Recently, the application of nonlinear version of PCA (denotes by NLPCA) has been reported [7]. Opposed to the classical method, the extracted features by NLPCA are not limited to the orthonormal vectors.

In this study, the classical and nonlinear versions of PCA are employed to reduce the spectral reflectances of 1269 samples of Munsell color chips. Compression process is performed by using different number of principle components i.e. 1 to 10 bases, and the effects of selected sizes on the reconstructed spectra have been reported. Results of employing of NLPCA and PCA methods in different employed sizes are evaluated by the values of root mean square errors (RMS) and the Goodness Fit Coefficient (GFC) between the original and the reconstructed spectra. The color difference values, i.e. ΔE, between the actual and synthesized spectra are also reported under D65 illuminant and 1964 standard observer.

2 Theoretical background

2.1 Principal Component Analysis

The main aim of PCA is to reduce the dimensions of the data set. To fulfill this goal, the hidden patterns of data becomes uncover and a small set of underlying basic functions are extracted and used as the projection space for the original data. By this way, samples are redefined in the new reduced space. The size of the reduced space depends on the nature of the data and the expected accuracy in reconstruction process. Since the mean centered data is

usually used in the classical PCA method, the mean vector should be added for the reconstruction of data from compressed information. Equation (1) mathematically shows this procedure for the spectral compression-reconstruction purpose.

$$\hat{R} \approx V_0 + \sum_{j=1}^{k} C_j V_j \qquad (1)$$

Where $\hat{R}$ is the reconstructed spectral reflectance, V_0 is the mean vector of spectral reflectance of dataset, C_j shows the specification of sample for the j[th] principle component, while V_j illustrates the j[th] eigenvector. k is the size of selected reduced space i.e. the numbers of chosen principle directions which are used for compression purpose.

The column vector *C* that weights the columns of *V* can be derived from Equation (2).

$$C = V^T (R - V_0) \qquad (2)$$

Where *V* is the selected eigenvectors, V_0 is the mean vector of spectral reflectance, and R is the spectral reflectance of dataset.

2.2 Non Linear Principle Component Analysis

Similar to classical PCA, the nonlinear principle component analysis recognizes and removes correlation among variables for dimensionally reduction while opposed to PCA it is not limited to linear correlation among the variables.

Let R is an m×n matrix where m and n represent the number of observations and the number of variables respectively. By using a nonlinear vector function, the nonlinear principle components are achieved as illustrated in Equation 3.

$$T = G(R) \qquad (3)$$

Where G is nonlinear vector function composed of f nonlinear functions G = {G1, G2,…, Gf } and T is the nonlinear principle components.

For reconstruction process, the second nonlinear vector function is employed and data could be reconstructed as showed in Equation 4.

$$\hat{R} = H(T) \qquad (4)$$

Where H is nonlinear vector function composed of m nonlinear functions H= {H1, H2,…, Hm} and $\hat{R}$ is the reconstructed spectral reflectance. By using an artificial neural network (ANN) the functions G and H are selected in a manner to minimize the $\left\| R - \hat{R} \right\|$.

It was already shown that the functions which is shown in Equation (5) could fit any nonlinear functions *v=f(u)* to an arbitrary degree of accuracy.

$$u_k = \sum_{j=1}^{N_2} w_{jk2} \sigma \left(\sum_{i=1}^{N_1} w_{ij1} u_i + \theta_j \right) \qquad (5)$$

Where $\sigma(x)$ is any continuous and monotonically increasing function with $\sigma(x) \to 1$ as $x \to +\infty$ and $\sigma(x) \to 0$ as $x \to -\infty$. As shown in Equation (6), a sigmoidal function is a suitable one.

$$\sigma(x) = \frac{1}{1 + e^{-x}} \qquad (6)$$

Clearly, Equations 5-6 are the describing equations for a feedforward artificial neural network (ANN). The nets consisted of N1 inputs, a hidden layer of N2 nodes as well as a linear output node for each nod. The weight for the connection of node i in layer k to node j in layer k + 1 is shown by w_{ijk} in Equation 5. Nodes of same layers are fully interconnected while the intralayer links are not allowed. The nodal biases that are adjustable parameters like the weights introduced by θ [8].

3 EXPERIMENTS

In this study, we borrowed an ensemble of 1269 reflectance spectra of the chips in the Munsell Book of Color–Matt Finish Collection [9]. The spectral reflectances of samples were measured with Perkin Elmer Lambda 9 spectrophotometer and the wavelength range was from 380 to 800 nm with 1 nm interval. In this research, the reflectance data were fixed between 400 to 700 nm at 10 nm intervals.

4 RESULTS

To compare the PCA and NLPCA techniques, the reflectance spectra of 1269 color chips of Munsell Color System were used and compressed and compressed in reduced spaces by using the classical PCA as well as the NLPCA techniques. Different numbers of principle components from 1 to 10 were used for both methods.

To quantitatively compare the results of different methods, the RMS, GFC and CIELAB color difference (ΔE) values were calculated between the reconstructed and the original spectra. Equation 6 was used to calculate the GFC value and the results were evaluated as unacceptable (GFC<0.9990), acceptable (GFC≥0.9990), good (GFC≥0.9995) and excellent (GFC≥0.9999).

$$GFC = \frac{\sum R(\lambda)\hat{R}(\lambda)}{\sqrt{\sum [R(\lambda)]^2}\sqrt{\sum [\hat{R}(\lambda)]^2}} \qquad (6)$$

Where $R(\lambda)$ is the actual reflectance value of sample in the λ^{th} wavelength and $\hat{R}(\lambda)$ shows the reflectance value of the reconstructed spectra in the same wavelength.

To show the effect of the sizes of compressed spaces, the fluctuations of the mean value of RMSs against the number of principle components was plotted for both PCA and NLPCA methods and illustrated in Figure1.

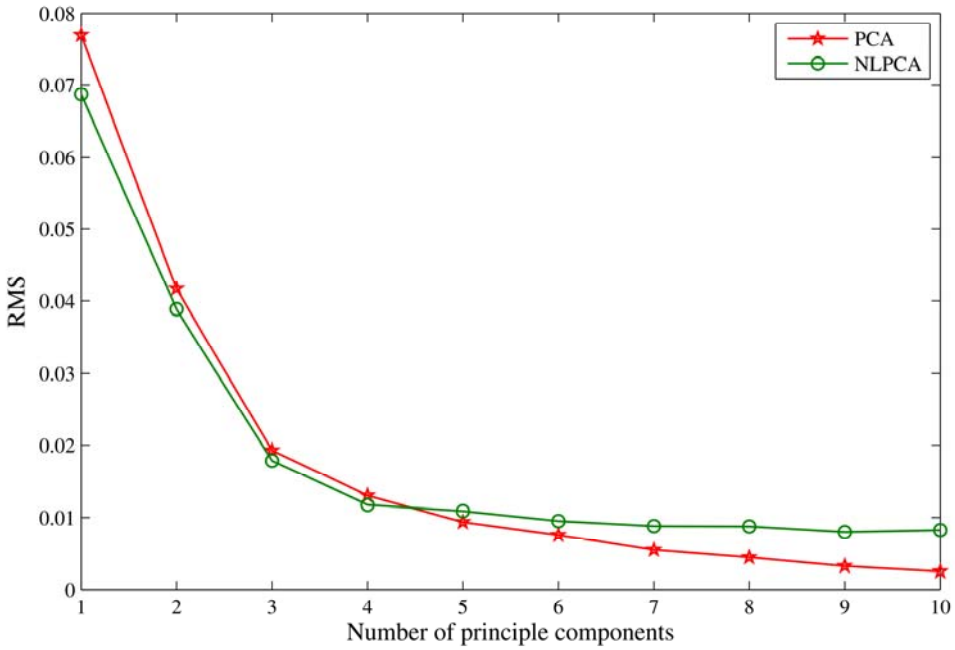


Figure 1- The fluctuations of the mean value of RMSs against the number of chosen basic functions.

Table 1 shows the mean of RMS values, the frequencies of different grades of GFC and the mean of color difference values for different numbers of basic functions obtained by employing of PCA and NLPCA techniques.

Table 1- The values of mean of RMS and ΔE as well as the frequencies of different grades of GFC obtained by PCA and NLPCA compression and reconstruction techniques using different numbers of principal components.

	PCA						NLPCA					
No. PC	RMS	GFC			ΔE		RMS	GFC			ΔE	
		Acceptable	Good	Excellent	mean	max.		Acceptable	Good	Excellent	mean	max.
1	0.0769	124	7	0	22.05	69.35	0.0687	189	20	0	19.68	65.44
2	0.0417	445	91	0	15.05	75.42	0.0389	520	112	0	13.80	55.95
3	0.0192	942	342	5	3.41	27.32	0.0178	977	412	5	3.65	39.03
4	0.0130	1123	702	41	1.39	11.66	0.0117	1164	721	29	2.02	15.59
5	0.0094	1216	887	81	0.82	5.17	0.0109	1216	753	61	1.98	24.38
6	0.0076	1236	1011	123	0.80	5.21	0.0095	1232	905	94	1.51	20.38
7	0.0055	1264	1116	417	0.18	2.93	0.0088	1228	922	143	1.25	13.57
8	0.0045	1267	1148	589	0.14	1.52	0.0088	1235	960	133	0.94	6.06
9	0.0032	1268	1248	810	0.14	1.24	0.0080	1246	994	193	1.22	17.07
10	0.0025	1269	1261	972	0.09	1.13	0.0083	1232	948	180	1.21	19.13

As Table 1 and Figure 3 show, the mean of RMS values for both PCA and NLPCA totally decrease by increasing the number of employed basic functions. However, the decreasing continues for classic PCA while the rate rapidly decreases for NLPCA method. As the results show, the NLPCA exhibits better performance in the lower dimensions, let say up to 4, while the classic PCA leads to superior results when higher dimensions were employed. The achievements are totally reconfirmed by the GFC and ΔE values. As the results show while the errors converge to zero in the PCA, they do not totally meet the lower values for NLPCA and remain constant in higher dimensional spaces.

5 CONCLUSION

In this paper, the NLPCA method was employed for the compression and reconstruction of spectral data and its performance was compared with classical PCA routine. The spectral reflectances of 1269 samples of Munsell colored chips were used for compression and reconstruction purposes. The compression and reconstruction were conducted with different number of principle components, i.e. 1 to 10. To make comparison between the PCA and NLPCA performances, the mean of RMS, the GFC and ΔE values between the original and the reconstructed spectra were utilized. Results showed that, the NLPCA performed better than PCA while less than or equal to 4 bases were employed. By contrast, when more than 4 principle components were used, the PCA showed better performances than NLPCA.

5 REFERENCES

[1] J. B. Cohen, "Dependency of the spectral reflectance curves of the Munsell color chips, Psychon", Sci. 1, 369–370(1964).

[2] H. S. Fairman, and M. H. Brill, "The principal components of reflectances", Color Res. Appl. 29, 104–110 (2004).

[3] A. Garcia-Beltran, J. L. Nieves, J. Hernandez-Andres., and J. Romero, "Linear bases for spectral reflectance functions of acrylic paints", Color Res. Appl. 23, 39–45 (1998).

[4] L. T. Maloney, "Evaluation of linear models of surface spectral reflectance with small number of parameters", J. Opt. Soc. Am. A. 3, 29–33 (1986).

[5] F. Agahian, S. A. Amirshahi, and S. H. Amirshahi, "Reconstruction of reflectance spectra using weighted principal component analysis", Color Res. & Appl. J., 33, 360-371 (2008).

[6] D. Tzeng, and R. S. Berns, "A review of principal component analysis and its applications to color technology", Color Res. Appl. 30, 84–98 (2005).

[7] F. Del Frate, and G. Schiavon, "Nonlinear Principal Component Analysis for the Radiometric Inversion of Atmospheric Profile by Using Neural Networks", IEEE T Geosci Remote. 37, 2335–2342 (1999).

[8] M. A. Kramer, "Nonlinear Principal Component Analysis Using Autoassociative Neural Networks", AIChE journal, Vol. 37, No.2 , 233-243 (1991).

[9] University of Joensuu Color Group. Spectral Database, Available at: http://spectral.joensuu.fi/.

COMPARISON OF DISTANCES FOR MULTI-LABEL CLASSIFICATION WITH PCTs

Valentin Gjorgjioski, Dragi Kocev, Sašo Džeroski
Jozef Stefan Institute
Jamova cesta 39, 1000 Ljubljana, Slovenia
e-mail: valentin.gjorgjioski@ijs.si

ABSTRACT

Multi-label classification has received significant attention in the research community over the past few years: this has resulted in the development of a variety of multi-label classification methods. These methods either transform the multi-label dataset to several simpler datasets or adapt the learning algorithm so it can handle the multiple labels. In this paper, we consider the latter approach. Namely, we use predictive clustering trees to perform multi-label classification. Furthermore, we perform an experimental comparison of four distance measures used to select the splits in the nodes of the trees. The experimental evaluation was conducted on 6 benchmark datasets using 6 different evaluation measures. The results show that, averaged overall, the Euclidean distance and the Hamming loss yield the best predictive performance.

1 INTRODUCTION

Traditionally, binary classification is concerned with deciding whether a given example has (or doesn't have) a single given target property/class. Multi-class classification involves the labeling of a given example with a single label/class λ_i from a finite set of disjoint labels $L = \{ \lambda_1, \lambda_2,\ldots, \lambda_Q\}$, Q>2. In contrast, multi-label classification learns a mapping from an example in the input space ($x \in X$) to a set of labels ($Y \subseteq L$) from the output space L. Note that, unlike in multi-class classification, in multi-label classification the labels are not mutually exclusive, i.e., a single example can be labeled with multiple labels. The labels that belong to the output Y are called relevant labels, while those from $L \setminus Y$ are called irrelevant for a given example.

The machine learning task of multi-label classification data has lately received significant attention from the research community [1], which has resulted in development of many methods that tackle this task. The developed methods can be generally divided into two categories: problem transformation and algorithm adaptation. Problem transformation methods transform problem into one or more single-label classification problems. These problems are then solved using a commonly used method for single-label classification and, afterwards, the output is transformed back into a multi-label representation. Algorithm adaptation methods adapt the learning algorithms to handle the multi-label data directly. In this work, we focus on algorithm adaptation methods. Specifically, we use predictive clustering trees (PCTs) [2] as classifiers and extend the distance function used when learning the tree. PCTs are a generalization of decision trees that are capable of predicting structured outputs. Namely, PCTs can handle multiple continuous targets, multiple discrete targets, time-series [3] and hierarchies of classes [4]. In the context of multi-label classification, we employ the PCTs for multiple discrete targets where a weighted Euclidean distance is used to generate the tests in the internal nodes of the tree. Here, we extend the PCTs with three distance measures: Hamming distance, Jaccard distance and a matching distance. These distances will provide additional flexibility for the users when they apply PCTs to different domains.

We compare the predictive performances of the PCTs obtained using different distance measures. The predictive performance was assessed on several benchmark datasets from multi-label classification. The predictive performance was measured with six evaluation measures: Hamming loss, accuracy, precision, recall, F1 score and subset accuracy.

The remainder of this paper is organized as follows. In Section 2, we present the predictive clustering trees for multiple discrete targets. We define the distances that we use in Section 3. We give the experimental design and in Section 4 and the results in Section 5. Section 6 concludes.

2 PREDICTIVE CLUSTERING TREES

The Predictive Clustering Trees (PCTs) framework sees a decision tree as a hierarchy of clusters: the top-node corresponds to one cluster containing all data, which is recursively partitioned into smaller clusters while moving down the tree. The PCT framework is implemented in the CLUS system, which is available for download at http://www.cs.kuleuven.be/~dtai/clus.

PCTs can be induced with a standard top-down induction of decision trees (TDIDT) algorithm. The algorithm takes as input a set of examples and outputs a tree. The heuristic that is used for selecting the tests is the reduction in variance caused by partitioning the instances. By maximizing the variance reduction the cluster homogeneity is maximized and it improves the predictive

performance. If no acceptable test can be found, that is, if the test does not significantly reduces the variance, then the algorithm creates a leaf and computes the prototype of the instances belonging to that leaf. The main difference between the algorithm for learning PCTs and a standard decision tree learner is that the former considers the variance function and the prototype function, that computes a label for each leaf, as parameters that can be instantiated for a given learning task. So far, the PCTs have been instantiated for the following tasks: multiple targets prediction [5], hierarchical-multi label classification [4] and prediction of time-series [3].

In this paper, we focus on the first task. PCTs that are able to predict a tuple of discrete variables are called multi-target classification trees (MTCTs). An example of a MTCT is shown in Figure 1. This MTCT presents a habitat model for 14 bioindicator species [6]. The internal nodes of the tree contain tests on the descriptive variables (in this case, chemical parameters of the water samples) and the leaves store the predictions (in this case, which species are encountered and which not in a given water sample).

The variance is calculated as the sum of the squared pairwise distances between the instances, i.e.,

$$Var(E) = \frac{1}{2|E|^2} \sum_{X \in E} \sum_{Y \in E} d^2(X, Y)$$

The function used to calculate the prototype is then $m = \arg\min_q \sum_{X \in E} d^2(X, q)$. In this case, the prototype is an instance from the dataset and is called medoid. Different distances can be used depending on the application domain. By default, PCTs use the Euclidean distance.

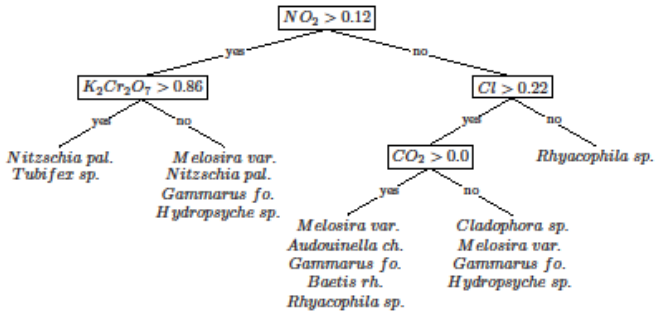


Figure 1: An example of a predictive clustering tree for predicting multiple discrete targets. The leaves predict the presence or absence for each bioindicator species.

3 DISTANCES FOR MULTI-LABEL LEARNING

In a multi-label learning setup, the target variable is a set of labels. Therefore, we can readily use distances over sets. Another approach to the problem is to see the multi-label classification problem as a predicting tuples of discrete targets and use distances over tuples. PCTs (and decision trees) have been previously used in the later context [2]. The focus of this study is the former approach to multi-label classification. In the remainder of this section, we present the distances over sets that can be used for extension of PCTs for multi-label classification.

3.1 Euclidean distance

The target in multi-label classification can be represented as a tuple of 0/1 values. The length of the target tuple is the number of all labels in the dataset. In this case, the Euclidean distance between two sets of labels C_i and C_j is defined as the Euclidean distance between their vector representations.

3.2 Hamming distance

The Hamming distance between two strings (i.e., bit-vectors) of equal length is the number of positions at which the corresponding symbols are different. In other words, it measures the minimum number of substitutions required to change the first string into the second. In terms of sets, the Hamming distance between two sets C_i and C_j is defined as:

$$h(C_i, C_j) = |C_i \cup C_j| - |C_i \cap C_j|$$

3.3 Jaccard distance

The Jaccard distance measures the dissimilarity between two sets by dividing the difference of the sizes of the union and the intersection of the two sets with the size of the union. The Jaccard distance can be calculated as follows.

$$j(C_i, C_j) = \frac{|C_i \cup C_j| - |C_i \cap C_j|}{|C_i \cup C_j|}$$

3.4 Matching distance (MD)

Motivated by a recently introduced distance on sets of structured objects, this distance is based on the matching between object from the sets. The matched objects do not contribute to the distance, which has the value of the unmatched part of the larger dataset, as defined below

$$md(C_i, C_j) = \max(|C_i|, |C_j|) - |C_i \cap C_j|$$

4 EXPERIMENTAL DESIGN

We begin by describing the benchmark datasets used in this study. Next, we present the most typically used evaluation measures for multi-label classification. We then give the experimental setup for the data analysis.

4.1 Datasets

We use 6 multi-label classification benchmark problems. Parts of the selected problems were used in various studies and evaluations of methods for multi-label learning. In the process of selection of problems, we opted to include benchmark datasets with different scale and from various application domains. Table 1 presents the basic statistics of the datasets. The datasets vary in size: from 391 up to 5318 training examples, from 202 up to 2635 testing examples, from 16 up to 1449 features, from 5 to 53 labels, and from 1.20 to 6.34 average number of labels per example.

	domain	N/T	D	Q	l_c
water quality	ecology	721/339	16	14	5.07
emotions	music	391/202	72	6	1.87
mediana	text	5318/2635	79	5	1.20
soil quality	ecology	1308/636	54	39	6.34
medical	text	645/333	1449	45	1.25
enron	text	1123/579	1001	53	3.38

Table 1. Description of the datasets in terms of application *domain*, number of training (*N*) and test (*T*) examples, the number of features (*D*), the total number of labels (*Q*) and label cardinality (l_c). The problems are ordered by their overall complexity roughly calculated as *N* x *D* x *Q*.

4.2 Evaluation measures

The evaluation of the predictive performance for multi-label learning systems differs from that of classical single-label learning systems. In any multi-label experiment, it is essential to include multiple and contrasting measures because of the additional degrees of freedom that the multi-label setting introduces. In our experiments, we used various evaluation measures that have been suggested by Tsoumakas et al [1]. In particular, we used six example-based evaluation measures: Hamming loss, accuracy, precision, recall, F1 score and subset accuracy.

In the definitions below, *Yi* denotes the set of true labels of example xi and $h(x_i)$ denotes the set of predicted labels for the same examples. All definitions refer to the multi-label setting.

Hamming loss evaluates how many times an example-label pair is misclassified, i.e., label not belonging to the example is predicted or a label belonging to the example is not predicted. The smaller the value of *hamming_loss*(*h*), the better the performance. The performance is perfect when *hamming_loss*(*h*) = 0. This metric is defined as:

$$hamming\_loss(h) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{Q}\left|h(x_i)\Delta y_i\right|$$

where Δ stands for the symmetric difference between the two sets, *N* is the number of examples and *Q* is the total number of possible class labels.

Accuracy for a single example x_i is defined by the Jaccard similarity coefficients between the label sets $h(x_i)$ and y_i. Accuracy is micro-averaged across all examples.

$$accuracy(h) = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|h(x_i)\cap y_i\right|}{\left|h(x_i)\cup y_i\right|}$$

Precision is defined as:

$$precision(h) = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|h(x_i)\cap y_i\right|}{\left|y_i\right|}$$

Recall is defined as:

$$recall(h) = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|h(x_i)\cap y_i\right|}{\left|h(x_i)\right|}$$

F1 score is the harmonic mean between precision and recall and is defined as:

$$F_1(h) = \frac{1}{N}\sum_{i=1}^{N}\frac{2\times\left|h(x_i)\cap y_i\right|}{\left|h(x_i)\right|+\left|y_i\right|}$$

F_1 score is an example based metric and its value is an average over all examples in the dataset. F_1 score reaches its best value at 1 and worst at 0.

Subset Accuracy is defined as follows:

$$subset\_accuracy(h) = \frac{1}{N}\sum_{i=1}^{N} I\left(h(x_i) = y_i\right)$$

where *I*(*true*) = 1 and *I*(*false*) = 0. This is a very strict evaluation measure as it requires the predicted set of labels to be an exact match of the true set of labels.

4.3 Experimental setup

We used the predictive clustering framework implemented in the CLUS system to investigate the performance of the different distance measures. To this end, we constructed single PCTs.

The PCTs were pruned with the F-test pruning method. This method checks whether a given test statistically significantly reduces the intra-cluster variance at a given significance level. An optimal significance level was selected by using internal 3-fold cross validation, from the following values: 0.01, 0.02, 0.03, 0.04 and 0.05.

5 RESULTS

Tables 2, 3, 4, 5, 6 and 7 show the results from the experimental evaluation of the distance measures. In the following, we briefly discuss the results for each evaluation measure. The Hamming distance has best predictive performance according to the Hamming loss measure. This is expected, since the trees with this distance are set to optimize that measure. Furthermore, since the Euclidean and Hamming distance are quite similar for vectors with 1/0 values, the Euclidean distance also has good predictive performance. On average, the Jaccard distance has the lowest predictive performance.

	Euc.	Ham.	Jac.	MD
water quality	0.314	**0.309**	0.528	0.312
emotions	**0.249**	0.272	0.274	0.253
mediana	**0.157**	0.165	0.355	0.203
soil quality	0.106	**0.099**	0.169	0.100
medical	**0.013**	**0.013**	0.014	**0.013**
enron	0.058	**0.055**	0.062	0.057

Table 2. The Hamming loss measure for different distances

In terms of accuracy, the Euclidean, Hamming and MD distance have similar predictive performance on average, while the Euclidean distance has the best performance on three datasets. The Jaccard distance, on the other hand, has the worst performance on average.

	Euc.	Ham.	Jac.	MD
water quality	0.298	0.315	**0.370**	0.317
emotions	**0.496**	0.469	0.488	0.493
mediana	**0.589**	0.588	0.302	0.505
soil quality	0.481	0.502	0.347	**0.504**
medical	**0.733**	0.731	0.718	0.727
enron	0.413	**0.435**	0.427	0.425

Table 3. The accuracy for the different distances

The precision and recall have inverted values. In the case of precision, Jaccard distance is the best performing, while for recall it is the worst performing. The distance to the other methods is large in the both cases. This means that the labels produced with Jaccard distance are reliable (low false positive rate); however, they do not cover all relevant labels for a given example (high false negative rate). The other three distances have similar performances to each other.

	Euc.	Ham.	Jac.	MD
water quality	0.352	0.382	**0.860**	0.390
emotions	0.583	0.561	**0.635**	0.580
mediana	0.605	**0.641**	0.465	0.602
soil quality	0.595	0.606	0.556	**0.618**
medical	0.755	**0.761**	0.746	0.755
enron	0.502	0.524	**0.558**	0.523

Table 4. The precision for the different distances

	Euc.	Ham.	Jac.	MD
water quality	**0.625**	0.623	0.397	0.614
emotions	**0.613**	0.592	0.571	0.600
mediana	**0.722**	0.704	0.359	0.595
soil quality	0.719	**0.730**	0.492	0.712
medical	0.779	**0.787**	0.771	0.776
enron	0.568	**0.600**	0.552	0.572

Table 5. The recall for the different distances

The F_1 score balances the performance measured by the precision and the recall. On average, the Jaccard distance has the lowest performance (because of the weak results for recall). The Hamming distance is slightly better than the remaining two distances.

	Euc.	Ham.	Jac.	MD
water quality	0.423	0.441	**0.523**	0.444
emotions	0.574	0.551	**0.575**	0.568
mediana	0.634	**0.642**	0.385	0.567
soil quality	0.617	0.634	0.491	**0.635**
medical	**0.757**	0.760	0.746	0.753
enron	0.515	**0.543**	0.535	0.530

Table 6. The F_1 scores for the different distances

The subset accuracy measures the fraction of the complete and accurate predictions. In this regard, the Euclidean distance has the best average performance, while MD is the best performing distance on four datasets. The worst performing distance is the Jaccard distance.

	Euc.	Ham.	Jac.	MD
water quality	0.009	0.012	0.000	**0.018**
emotions	0.262	0.233	0.223	**0.272**
mediana	**0.468**	0.440	0.063	0.327
soil quality	0.036	0.041	0.003	**0.044**
medical	**0.661**	0.640	0.631	0.646
enron	0.145	0.149	0.149	**0.150**

Table 7. The subset accuracy for the different distances

6 CONCLUSIONS

In this paper, we have presented an experimental evaluation of four distance measures for multi-label classification. The evaluation was performed on 6 benchmark datasets using 6 evaluation measures.

The results show that there is no overall best distance measure. The best choice for a distance measure is the one that optimizes a selected evaluation measure. For example, the Hamming distance works the best when optimizing the Hamming loss, while the best according to precision is the Jaccard distance (since there is a strong connection between precision and the Jaccard coefficient). All in all, the Euclidean distance and Hamming loss perform the best averaged across all evaluation measures.

References

1. G. Tsoumakas, I. Katakis. Multi Label Classification: An Overview. International Journal of Data Warehouse and Mining 3(3). 2007. pp. 1–13.
2. H. Blockeel, L. D. Raedt, J. Ramon. Top-down induction of clustering trees. In Proceedings of the 15th International Conference on Machine Learning. 1998. pp. 55–63.
3. I. Slavkov, V. Gjorgjioski, J. Struyf, and S. Dzeroski. Finding explained groups of time-course gene expression profiles with predictive clustering trees. Molecular Biosystems 6. 2010. pp. 729-740.
4. C. Vens, J. Struyf, L. Schietgat, S. Dˇzeroski, H. Blockeel. Decision trees for hierarchical multi-label classification. Machine Learning 73 (2). 2008. pp. 185–214.
5. J. Struyf and S. Dzeroski. Constraint based induction of multi-objective regression trees. In proceedings of the 4th International Workshop on Knowledge Discovery in Inductive Databases. 2005. pp. 110-121.
6. H. Blockeel, S. Dzeroski, J. Grbovic, Simultaneous prediction of multiple chemical parameters of river water quality with TILDE. In Proceedings of the 3rd European Conference on PKDD - LNAI 1704. 1999. pp. 32-40. Springer.

VISUALIZATIONS OF SLOVENIAN SCIENTIFIC COMMUNITY

Mario Karlovčec, Dunja Mladenić, Marko Grobelnik, Mitja Jermol
Jozef Stefan Institute
Jamova 39, 1000 Ljubljana, Slovenia
Tel: +386 1 4773419; fax: +386 1 4251038
e-mail: mario.karlovcec@ijs.si

ABSTRACT

Using advanced analysis techniques new useful insight into data can be achieved. This paper addresses a problem of gaining insights in the data on scientific collaboration on a National level, where data can be seen as a graph with researchers and research content. Two existing visualization techniques were applied on data about scientific community in Slovenia: collaboration diagram and competence map. Collaboration diagram gives a clear overview of collaborations for a selected researcher, while competence map shows semantically grouped research content the researcher has worked on.

1 INTRODUCTION

Different types of data about scientific community are being collected on National level in many countries. This includes data about researchers, their projects and organizations. With more advanced analysis techniques user can gain deeper understanding of different aspects of scientific community. There are at least three types of users of such data: (1) researchers – they are interested in finding possibilities for collaboration, comparing achievements of colleagues, observing profiles of organizations; (2) investors – they want compare investments and results of projects, directions of development in science, etc. (3) companies – want to see latest achievements in science which they can commercialize, search for experts in particular fields, etc.

In the next chapter detail description of the data is given, following two different visualization techniques: collaboration diagram and competence map.

2 DATA

The data that we have used for evaluation was provided by Slovenian Research Agency (ARRS), and can be publicly accessed with SICRIS (SlovenIan Current Research Information System , 2000) information system managed by Institute of Information Science, Maribor. Main entities contained in the data are: researchers, research organizations and research projects. In 2011, there are 13897 researchers who have taken part in Slovenian Research Agency projects since 1998, or whose active status has been either registered or confirmed by the research organizations. Researchers/experts not being funded by the Slovenian Research Agency are also able to join SICRIS if they provide the data. Second entity – research organizations contains list of all (901) research organizations inside the country carrying out projects partly financed by Ministry of science from 1995 onwards. The complete data, however, is available only for the organizations that responded to the invitation of 14 June 1999 and submitted the required data. Third entity is research projects and it contains data on 5389 projects partly financed by Slovenian Research Agency from 1998 onwards. In the future it will also include data on any other research projects that the researchers from the database will wish to present.

3 COLLABORATION DIAGRAM

Collaboration diagram[1] is a type of sociogram that visualizes collaboration between people. **Sociograms** serve as simple visual illustrations in helping people to make sense of the underlying network information (Stuart K. Card, 1999); they are the most important instrument, basis for visualization of social networks and analysis of social ties structure (Wouter de Nooy, 2005).A **social network** is a collection of actors (such as people, organizations or other social entities) and relationships among the actors, indicating the way in which they are connected socially (such as friendship, trade or information exchange). **Social network analysis** is a methodological approach to understanding the structure of such networks, by means of mapping and measuring these relationships (Wasserman S., 1994).

3.1 Main Characteristics of The Diagram

Proposed collaboration diagram is different from other sociograms that represent collaboration between people. The main difference is focusing on a single actor (researcher in our case). This means that the chosen researcher is placed in the middle of the diagram (central researcher) and all others, with which he is collaborating, are placed radially around

[1] The usage of term collaboration diagram in the context of SNA should be differentiated with one type of UML diagram the same name

him with the distance relative to intensity of collaboration. With this drawing convention, we lose information about collaboration between researchers that are not central, but increase readability for a single researcher. The argument for this approach is assumption that user is not interested in exploring collaboration of multiple researcher in the same time, but is always focused on a single researcher. Even if this assumption is wrong and user wants insight into collaboration of multiple researchers in the same time, because of too many nodes and connections between them, the diagram would not be clear and the information would not be communicated effectively. Furthermore the diagram would not clearly show the collaboration of the single researcher, which was maybe in primary focus of the user.

As given in (Weidong Huang, 2005) the major concern in network visualization is effectiveness, it includes two issues: readability and communication. Readability can be strongly affected by network layout. Among many factors, edge crossing has been widely accepted as a major aesthetic affecting the ease of reading. In proposed approach for drawing collaboration diagram, crossing edges are completely avoided. Second issue of effectiveness – communication, refers to communicating exactly the same information which is intended to convey with the visual representation of the underlying network (Ulrik Brandes, 1999). Radial layout with a central actor is inteded to clearly communicate structure of collaboration.

In (Weidong Huang, 2005) authors compare five sociogram drawing conventions: circular, radial, hierarchical, group and free. The radial layout (figure 1) had moderate user ratings and task performance. But the radial layout used in these experiments differed from our approach for drawing collaboration diagram. Single actor was placed in the center, but the radial layout of the others was not clear. To emphasis this layout, a circle was drawn in the edge level of the network. Observations of results in referenced research show that the subjects did not realized nodes were actually arranged radially. A big circle around the diagram did not make viewers read radially, it caused additional confusion.

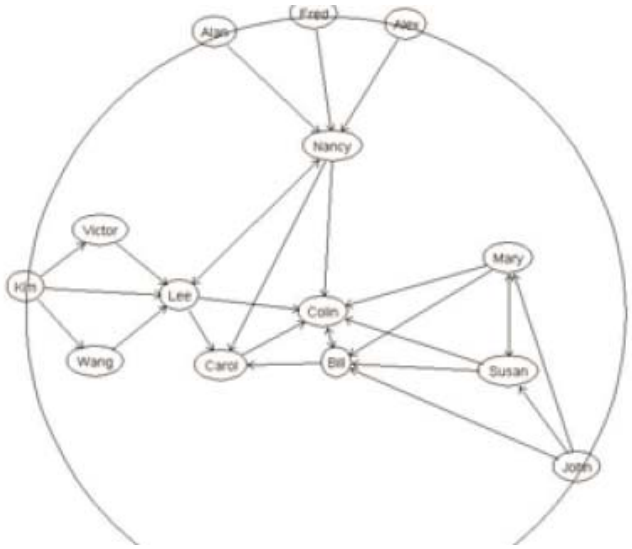


Figure 1: *Radial layout used in experiments in* (Weidong Huang, 2005)

3.2 Process of Creating Diagram

Proposed approach for drawing collaboration diagram can be performed in the following steps (Figure 2):

1. Place the chosen actor in the center of the network
2. Divide the maximal allowed radius number of different groups of collaboration count to get the distance between circles
3. Draw circles around the central actor so that each circle is a multiple of distance calculated in previous step
4. Draw actors of each group of collaboration count on a corresponding circle. The angle of the first actor of a group is determined randomly, while the others are evenly spaced on the circle.

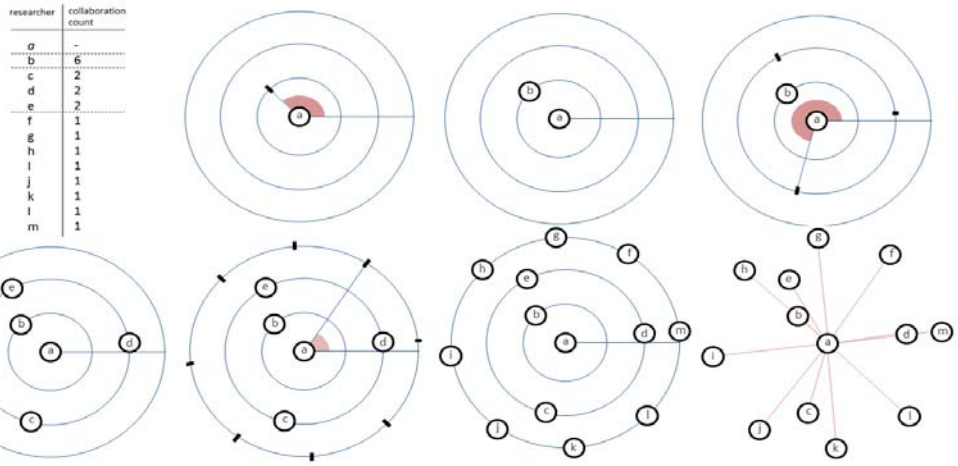


Figure 2: *Steps in creating collaboration diagram*

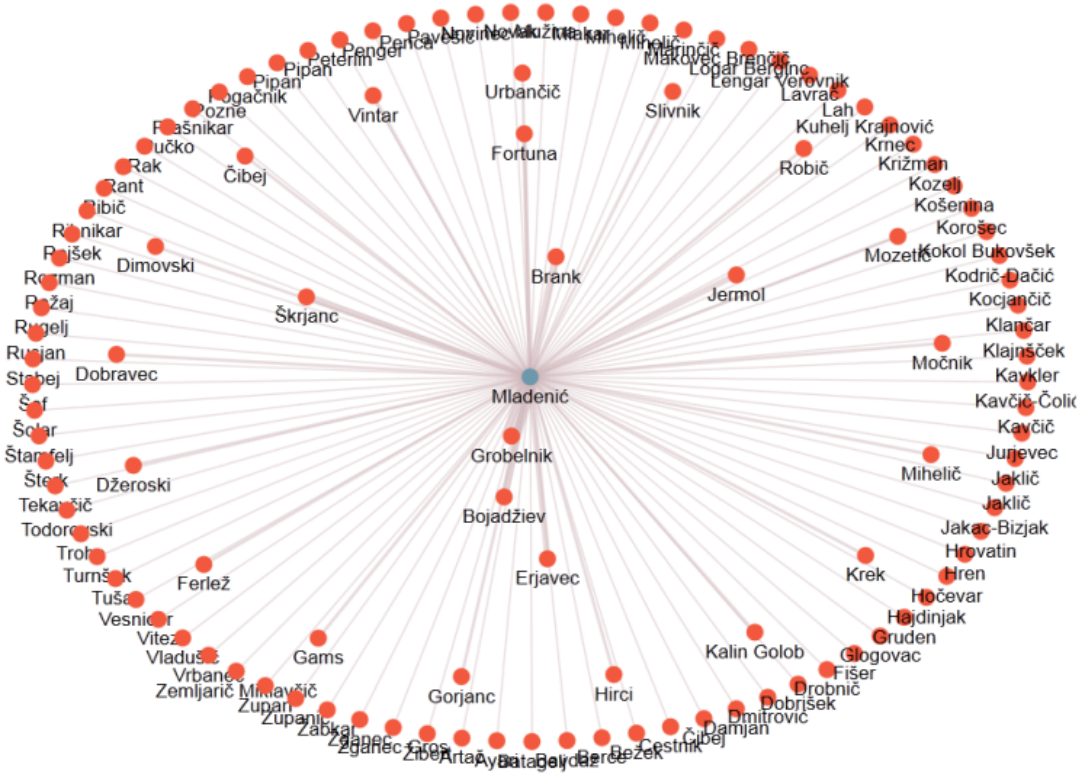


Figure 3 - *Collaboration diagram of a researcher*

An example of collaboration diagram on our data is shown in Figure 3. Selected researcher (Mladenić) is located in the center of the diagram and marked with different color than the other researchers she collaborates with. The diagram shows that the selected researcher collaborated with a number of other researchers on some projects. The greater the number of common projects with the selected researcher, the greater is the proximity to the center. For instance, Grobelnik, Brank, Bojadžiev, Erjavec, Jermol have been the most intensive collaborators and are thus places closer to the center of collaboration diagram.

4 COMPETENCE MAP

Competence map is a diagram used to represent intellectual achievements, skills or knowledge of a person, organization or some other entity. Proposed competence map is directed to a single person, which is in our case a researcher. Competence map of a researcher is based on scientific projects on which she has worked on. To build a competence map we have used Document Atlas (Blaz Fortuna, 2011). Document Atlas is a utility for visualizing large corpora of text documents using Latent Semantic Indexing and multidimensional scaling for positioning the corpus on a 2D plane. Input for creating competence map of a researcher are titles and when available, abstracts and keywords of projects on which he has worked on. The output are points which represent project, which are positioned on the 2D plane in such a way that it is easily to recognize most important groups of competences researcher is working on.

An example of a competence map on our data is shown in Figure 4. It can be seen that the researcher has collaborated on several projects, six of them (in the middle cluster) involving data analysis, two projects targeting Slovenia (the left up cluster), two involving semantic technologies (the right most cluster), three projects related to military topics and two projects not belonging to bigger clusters.

Figure 4: *Competence map of a researcher*

5 CONCLUSION

This paper describes visualization techniques applied on data about scientific community in Slovenia. We have shown how these visualizations enable new insights into the data. Two visualization techniques were used: collaboration diagram and competence map.

Collaboration diagram clearly communicates collaboration of a researcher. Novel approach, which focuses on a single actor, is used to construct the diagram. In this way, information contained in the diagram is decreased, but its effectiveness and ability to amplify users' cognition is increased.

Competence map shows projects of a researcher, which are visualized in the way that the user can recognize most important groups of competences researcher is working on. It is build using Document Atlas utility. Described visualization techniques enable deeper understanding of different aspects of scientific community in Slovenia.

Further visualizations will be developed in the future work, in order to provide additional knowledge to researchers, investors into scientific work and users of scientific achievements in practice.

6 ACKNOWLEDGMENTS

This work was supported by the Slovenian Research Agency, the IST Programme of the EC under PASCAL2 (IST-NoE-216886) and MetaNet (ICT-NoE-249119).

References

[1] Blaz Fortuna, D. M. (2011). *Document Atlas*. Retrieved 8 4, 2011, from http://docatlas.ijs.si/

[2] Stuart K. Card, J. D. (1999). *Information Visualization Using Vision to Think.* San Francisco: Morgan Kaufmann Publishers, Inc.

[3] Ulrik Brandes, P. K. (1999). Centrality in Policy Network Drawings. *Proc. 7th Intl. Symp. Graph Drawing*, (pp. 250-258).

[4] Wasserman S., F. K. (1994). *Social Network Analysis: Methods and Applications.* Cambridge University Press.

[5] Weidong Huang, S.-H. H. (2005). *Layout Effects: Comparison of Sociogram Drawing Conventions.* Sydney: University of Sydney.

[6] Wouter de Nooy, A. M. (2005). *Exploratory Network.* Cambridge: Cambridge University Press.

A FRAMEWORK FOR A MULTILINGUAL CONTEXTUAL AND BEHAVIORAL ONLINE ADVERTISING NETWORK: A CASE STUDY

Domen Košir[1,2], Zoran Bosnić[1], Igor Kononenko[1]
[1] Faculty of Computer and Information Science, University of Ljubljana, Tržaška 25, 1000 Ljubljana, Slovenia
[2] Httpool Ltd., Cesta v Gorice 8, 1000 Ljubljana, Slovenia
e-mail: domen.kosir@fri.uni-lj.si

ABSTRACT

An online advertising network connects web content providers and advertisers enabling the providers to monetize their content and the advertisers to reach online consumers. In this paper we study the workings of a successful state-of-the-art online advertising network with branch offices all over the world. The framework is designed to support different targeting strategies and advertising in multiple languages. We describe the implementation of contextual and behavioral targeting, and also discuss different methods to evaluating these strategies. Statistics show that by employing contextual targeting instead of random targeting we can achieve significantly higher CTR values.

1 INTRODUCTION

Online advertising is an increasingly popular form of advertising that uses the internet to deliver advertisments to consumers. An online advertising network acts as intermediary between advertisers and website managers (publishers). The goal of an online advertising network is a better utilization of the available advertising space.

In so-called "block web advertising" websites typically have some parts reserved for advertisments - these are called advertisment blocks. Each advertisment block is usually of a fixed size and can display one or more advertisments at a time.

The publisher can choose to maintain control over the selection of advertisments which will be displayed to the website visitors or he can give control over the advertisment blocks to an advertising network. This way the publisher does not have to deal with the advertisers in order to ensure that the advertisments shown on his website are interesting and diverse.

Modern advertising networks are often specialized for a specific targeting strategy of which most common are:

- contextual targeting [6][2] and
- behavioral targeting [1].

In contrast to random targeting, both mentioned strategies make an effort to find the best possible advertisment for the user who is currently viewing the website. Contextual advertising is based on web content analysis. If the website currently viewed has already been analyzed by the network then advertisments most similar to the website content are displayed. Behavioral targeting on the other hand focuses the user, his habits and previously viewed content. This type of advertising obviously is not possible without some form of user tracking (e.g. via HTTP cookies or IP addresses). The tracking of users itself can be a privacy issue but it will not be addressed here.

This paper is a case study of a successful international state-of-the-art advertising network that supports both contextual and behavioral targeting. The network operates in many countries around the world and is still expanding. So far, it supports more than 10 Indo-European languages.

In Section 2 we describe the network's architecture and its main components. Sections 3 and 4 describe how contextual and behavioral targeting are implemented. In Section 5 we describe different approaches to evaluation of targeting strategies or an advertising network as a whole. WeWe conclude our work in Section 7.

2 FRAMEWORK ARCHITECTURE

As it is shown in Figure 1, the framework consists of three main parts:

- Advertisment server – This is the front-end component which recieves requests from websites and responds by sending back advertisments.
- Web crawler – A background process that browses the web and analyzes the web content.
- Database – It contains information about the ongoing advertising campaigns, analyzed web content and web users. This distributed document-oriented database is exposed as a web service and is highly scalable.

When a user visits a website which is a part of the advertising network:

1. A request is sent to the advertisment server. Each request, among other pieces of information, contains the URL address of the website and user identifiable information.
2. The server then searches the database for information about the website and/or user (depending on the

targeting strategy used). If no information about the website is found the server sends this URL address to the web crawler to be inserted into its priority queue.

3. Using all the available information, the server then chooses the most appropriate advertisments and sends them back to be displayed on the website.

The request times and the origin URL addresses are sent to the web crawler. Its job is to constantly fetch web content, analyse it and update the framework's database.

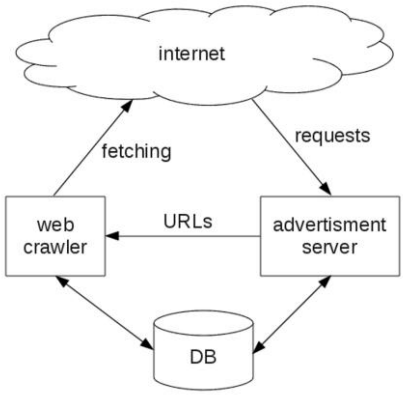


Figure 1: *The architecture of the advertising network.*

2.1 Web Crawler

Any kind of advanced advertisment targeting requires some information about the website content. Network delays and request peaks make fetching of the website content at request-time infeasible. Fetching of website content is done in a separated process using a web crawler. This is an independent software package that visits certain webpages and saves their contents. The framework uses a very straightforward URL-oriented approach using:

- a priority queue used to visit URLs in an orderly fashion,
- a categorization system to process the website content and
- a database to save all the information about each URL.

The priority queue contains all the URL addresses from which requests for advertisments originated. The priority of each URL is affected by:

- Age of the URL – New URL addresses (from the advertising network's point of view) have a higher priority while older have lower priority.
- Frequency of changes in the website's content – More frequently changing websites are fetched more often.
- Number of requests – More popular websites have a higher priority than the unpopular ones.

After the crawler fetches the content of a website the content must be processed (see Figure 2).

1. HTML code cleanup – The fetched web content are usually HTML documents. Since the quality of HTML code varies greatly it must be cleaned up before further processing.
2. Text extraction – The HTML document is parsed and text is extracted. We focus mainly on bigger parts of text with few HTML tags in order to reduce the noise caused by menus, sidebars, footers, etc.
3. Language identification – This step is optional but may be needed for the categorization process. A very robust approach to language identification is with the usage of n-grams [3][4]. N-grams are fast, consume very little space and work reliably in spite of textual errors.
4. Text categorization - The categorization process maps each document to one or more categories based on its content. For the purpose of providing high-level overview of the system, in the following we will consider the text categorization process as a black box function that takes text as input and produces a list of categories with corresponding biases.

After the web document analysis, the contents and categorization results are saved in the network's database to be used by the advertisment server.

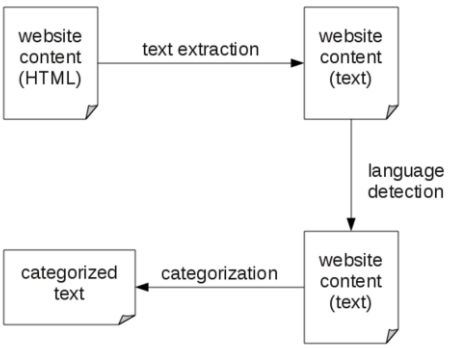


Figure 2: *The process of website content analysis.*

2.2 Advertisment Server

The server's job is to recieve the requests from the internet, select advertisments and send them back for display. Each request comes from a specific advertisment block on the website. The request contains the following information:

- the URL address of the website,
- the size of the advertisment block,
- the publisher's preferences regarding the targeting strategies used to select advertisments,
- the publisher's preferences regarding the types (e.g. textual, animated graphics, Flash banners) and colors of advertisments to be displayed on his website,
- the viewer's language preferences,
- the viewer's IP address and possibly other user identifiable information.

The process of selecting the appropriate advertisments can be a little tricky. For example, there is no obvious rule specifying in which language the displayed advertisments should be. If a German-speaking user is viewing an English website then the language of the web content probably does not match the user's language preferences.

We can either decide to match the language of advertisments to the website's language and show him English advertisments or we can consider the user's preferences and show him advertisments in German language.
This is of course only an issue when we have an international network with advertisers and publishers from different countries.

3 CONTEXTUAL TARGETING

The goal of contextual targeting is to display advertisments that match the website content. If we assume that our web crawler has already fecthed and analyzed the content, currently viewed by the user, we can select the advertisments in the following manner:

1. Like the web content all the advertisments should also be categorized. The advertisments can be categorized based on their content (using advertisment title and text or image analysis) or they can be categorized by manually assigning them to specific categories. In either case, it is important that the same set of categories is used as in the web content analysis.
2. Depending on the size of the block we select a number of advertisments. The selection of contextual advertisments is illustrated in Figure 3. A categorized web document and all the advertisments are placed in a n-dimensional vector space where each dimension represents a single category. The web document and the advertisments are vectors with the initial point in the origin and the terminal point defined by the categorization results. We use cosine similarity [4][7] to compare the vector representing the web document to the vectors representing the advertisments. We select the advertisments closest to the web document.

Cosine similarity is used in order to reduce the impact of fluctuations of categorization results which may occur because of:

- varying amount of text in web documents,
- varying number of meaningful words in a web documents (e.g. spam sites typically contain many keywords while other websites do not) and
- the difference in size of a web document and an advertisment.

4 BEHAVIORAL TARGETING

Behavioral targeting is based on the analysis of the user's behavior. This is a very broad definition and there are numerous strategies that fall under it.
A very common approach in web user profiling is building of two user models [5]. One model represents the user's short-term interests and is derived from the users activities in the past few days. The other model is derived from user's older web browsing history and represents his long-term interests (see Figure 4). By building two models to describe the user's interests we can separate his true interests from the interests influenced by events like birthdays, holidays, car purchase, interesting news, etc.

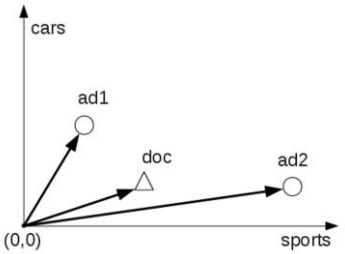


Figure 3: *In this example of selecting advertisments using contextual targeting we limit ourselves to only two categories: cars and sports. We are trying to select the most appropriate advertisment (1-NN) for the website "doc". For the purpose of contextual advertising we are more interested in the ratio of biases in the categorization results than in their absolute values. That is why we use the cosine similarity which calculates the cosine of the angle between two vectors. By using the cosine similarity we get "ad2" as the best match for "doc".*

There are several other attributes that influence the user's behavior and can be taken into account when modeling the user:

- The time of day – We can expect that many users will show different interests during various activities like working, studying, playing games, etc. These activities are often performed at the same time of day.
- The day of the week – The first thing we want to do is to distinguish between the workdays and the weekend. A more comprehensive analysis of the user's behavior is possible by looking at his day-to-day activities.
- The user's IP address – Given that we are able to reliably identify a user without using his IP address (e.g. using web cookies), the address can still provide some information about the user's interests. A user with a laptop computer will always be identified by the same web cookie but by using his IP address we can distinguish among his interests related to his workplace, home, favourite internet cafe, etc.

These attributes are somewhat problematic – we can only use them in the user behavior analysis after we have tracked the user for an extended period of time.

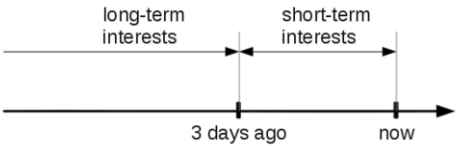


Figure 4: *The division between the user's short-term and long-term interests. The boundary at "3 days ago" is just illustrative.*

5 EVALUATING TARGETING STRATEGIES

Advertising campaigns are usually planned toward a specific goal (e.g. trademark promotion, new product promotion, increasing sales). Depending on the goal a pricing model is chosen. The most common pricing models in the online advertising are:

- PPI (Pay Per Impression) - The advertiser pays for each appearance of his advertisement on a website.
- PPC (Pay Per Click) - An online advertising pricing model, where the advertiser pays for each click on their advertisement. The click usually redirects the user to the advertiser's website.
- PPA (Pay Per Action) - The advertiser pays to the publisher for each specified action (e.g. a purchase or a form submission) that results from his advertisment.

One way to measure the success of a targeting strategy is to sum up all the profits made using this strategy. This value is undoubtedly very informative but it is not an objective measure as it is influenced by the varying advertisment prices. Because we do not want the prices to influence our evaluations, we only monitor the users' engagement. We do this by looking at all the clicks and other actions users performed.

Click-through rate (CTR) is a standard way to measure the success of an online advertising campaign or an individual advertisment. It is calculated as the ratio between the number of clicks on the advertisment and the number of times it was shown to the users. It tells us the probability a user will click on the advertisment when he sees it.

We have to be careful to remember whether we are dividing the number of clicks with impressions or displays when calculating the CTR. Most advertising networks do not distinguish between these two events. An *impression* happens when the server delivers an advertisment to the webpage and HTML code of the advertisment is rendered by the user's computer. A *display* happens when the advertisment is actually shown on the user's monitor for a minimum of 2 seconds. If the webpage is larger than the user's monitor and if the advertisment block is not located near the top of the page where the user can see it, then a display does not happen until the user scrolls down to it. In other words: a display cannot occur without an impression. This is an important distinction, because in our experience the number of displays is typically 20-30% smaller that the number of impressions. This, of course, affects the value of CTR.

6 LOG ANALYSIS

We analyzed the advertising networks server logs for the first 6 months of 2011. In the analysis we limited ourselves to the Slovenian subnetwork – this is the primary and one of the largest subnetworks under Httpool's management. We took all advertising campaigns that were in circulation during the period of 6 months and calculated average CTR values per advertisement type, distinguishing campaigns with random targeting from those with contextual targeting. Behavioral targering is is not yet used in production.

We found that CTR values tend to be bigger in campaigns that use contextual targeting. We saw a 28% increase in CTR for textual advertisements, a +12% increase for shop advertisements (text with a picture) and a 74% increase for rich media advertisements (Flash, animated GIF).

7 CONCLUSION AND FUTURE WORK

In this paper we described a framework for an online advertising network capable of analyzing web content and displaying advertisment using contextual and behavioral targeting.

We found that CTR values tend to be higher when we are using contextual targeting instead of random targeting.

Behavioral targeting is not yet used in production. We expect that it will prove to be an effective advertising tool.

So far, only Indo-European languages are supported by the framework. As the advertising network grows, we expect that other languages will have to be supported as well. Languages like Mandarin and Arabic will undoubtedly be a challenge.

Acknowledgment

Operation part financed by the European Union, European Social Fund. Operation implemented in the framework of the Operational Programme for Human Resources Development for the Period 2007-2013, Priority axis 1: Promoting entrepreneurship and adaptability, Main type of activity 1.1.: Experts and researchers for competitive enterprises.

References

[1] A. Addis, G. Armano and E. Vargiu. Profiling users to perform contextual advertising. In *Proc. WOA 2009.*

[2] A. Broder, M. Fontoura, V. Josifovski and L. Riedel. A semantic approach to contextual advertising. In *Proc. SIGIR 2007*, pp. 559-566, 2007.

[3] W. B. Cavnar and J. M. Trenkle. N-gram based text categorization. *In Proc. SDAIR-94*, pp. 161-175, 1994.

[4] Z. Elberrichi and B. Aljohar. N-grams in Texts Categorization. *Scientific Journal of King Faisal University (Basic and Applied Sciences)*, 8(2):1428H, 2007.

[5] S.E. Middleton, N.R. Shadbolt and D.C. De Roure. Ontological user profiling in recommender systems. *ACM Transactions on Information Systems (TOIS)*, 22(1):54-88, 2004.

[6] B. Ribeiro-Neto, M. Cristo, P. B. Golgher and E. S. de Moura. Impedance coupling in content-targeted advertising. In *Proc. SIGIR 2005*, pp. 496-503, 2005

[7] G. Salton and M. McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill, New York, 1983.

ANALYSIS AND PREDICTION OF BUG DUPLICATES IN KDE BUG TRACKING SYSTEM

Gregor Leban
Artificial Intelligence Laboratory
Jozef Stefan Institute
Jamova 39, 1000 Ljubljana, Slovenia
Tel: +386 1 477-53-61
e-mail: gregor.leban@ijs.si

ABSTRACT

Bug tracking systems (BTS) are systems that allow users of some software to report to developers bugs they encountered while using it. Common problem of BTS are duplicated reports of the same bug. Since identifying bug duplicates is a time consuming task we show in this paper an approach to automatically identifying duplicates using text-mining methods. We demonstrate the usability of our method on KDE Bugzilla BTS which contains 249,083 bug reports of which 47,093 are duplicates.

1 INTRODUCTION

Software developing companies and organizations very often use bug/issue tracking systems (BTS) such as Bugzilla[1], Mantis[2] or LaunchPad[3]. Using such systems, users can report to the developers the issues they encountered while using the software. Bug reports consist of a structured and unstructured part. The structured part of the report contains the name of the product where the problem occurred, the component's name, status of the bug, priority and severity. The unstructured part contains the summary of the bug and a description. The provided information should ideally be enough for the developer to identify the problem and fix it. People can also post comments to bug reports where they can clarify the problem or provide additional information or solution to the problem. It is common for a bug to have tens of comments.

One of the difficulties with bug reports is that they are reported by users who don't have an overview of the existing bug reports. As a consequence, many reports describe an issue that has already been reported sometime in the past. Such redundant reports are called bug duplicates. The reason they are problematic is that a lot of time is wasted because of them. They falsely appear to provide information about a new issue and can cause different developers to not-knowingly work on fixing the same bug. To avoid such loss of time a bug is often first inspected if it is a duplicate. If a bug is identified as such, bug tracking systems allow users/developers to explicitly mark them as duplicates to let others know that they can be disregarded. Identifying a bug as a duplicate is however a difficult and imprecise task which requires a lot of time.

In order to alleviate the problem of manually identifying bug duplicates ALERT[4], an EU project that aims to help open source communities to more efficiently manage projects, plans to implement automatic methods that will be able to determine if a given bug is a duplicate. In this paper we will present our first experiments in this area which show promising results.

We will start by introducing the bug tracking system that we used in our experiments and describe the way in which the data was processed. Next we will present the algorithm used to identify the duplicates and show its performance. We will conclude with a summary and some ideas for future work.

2 KDE BUG TRACKING SYSTEM

KDE[5] is an international software community that is developing a set of free, cross-platform applications. They have more than 1,800 developers who have created more than 6 million lines of code. We selected KDE as our case study because they are partners in the ALERT project.

KDE uses Bugzilla BTS to track bugs. KDE started using Bugzilla in 1999 and until August 2010 249,083 bugs were reported. On average there are almost 2,000 bugs reported per month. As for most projects, the number of duplicate reports represent a significant percentage of the repository. In the KDE repository, almost every fifth report is a duplicate. What is even worse is that the ratio of duplicates seems to increase over time and has in the past already reached the value of 0.42.

Out of 249,083 bugs 47,093 were manually marked as bug duplicates. In most cases there is only one duplicate of a bug. Figure 1 shows that there are more than 10,000 such reports. Similarly, there are 3,000 cases where there are two duplicates of one bug. As we can see, the numbers quickly decrease, although we can even find a bug that has 251 duplicated reports.

2.1 Importing the data

In order to be able to analyze the data and predict the duplicates we imported the content of KDE Bugzilla into

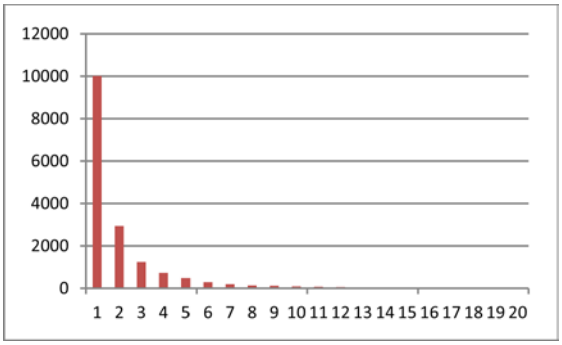


Figure 1: *Distribution of the number of times the same bug was reported. For example, in 3,000 cases there were two duplicate reports created for the same bug.*

Contextify[6]. For each bug report we treated the initial description of the bug and all the following comments as individual documents – in the same way as the data is also stored in Bugzilla. Each of these documents was stored in Contextify in the form of bag of words. As a part of the pre-processing we also ignored the stop words and stemmed the words using the Porter stemmer.

3 PREDICTING THE DUPLICATES

There are different tasks that we can identify regarding the prediction of bug duplicates. One task could be to build a model that would be able to classify a given bug report as a duplicate or non-duplicate. On the surface our problem seems like a typical binary classification problem. Each bug report in the repository represents one learning example, the words in the report are the attributes and the value of the class is 1 or 0, depending on whether the bug has a marked duplicate or not. However, if we think a bit about this definition of the problem we can see that it is not very sensible. Words themselves don't carry any valuable information that would enable us to separate reports into two classes. Reports that mention, for example, the word "kmail" are no more likely to be duplicates than the reports that mention "gnome". If such differences do exist they can only be accidental and using them would only lead to an over-fitted model. The classification model could potentially be improved by adding as attributes the available meta-data that is present in bug reports. By observing the author of the bug report, for example, the model could perhaps identify a group of people who more often than others submit bug duplicates. Since people usually don't intentionally submit bug duplicates such discoveries are unlikely and would not significantly improve the accuracy of the model.

As an alternative option for building a classification model we could also consider the following scenario. Each bug is again a learning example and the words in the report are the attributes. In this case, each bug has a different class value except the duplicates. All the duplicates of a particular bug would have the same class value as the original bug report. The classification model built on such data could predict that a new bug report is the duplicate of an existing bug. If the prediction would be probabilistic we could say that if the probability of the most likely class is below a certain threshold then the bug is not a duplicate. There are two related problems with this approach. As the number of reported bugs increases so does also the number of possible classes. Having 200,000 possible classes is unacceptable since there are no methods that could build a reliable model with so many classes. Also, in order for a method to build an accurate model it has to generalize the learning examples. In this scenario, however, most of the class values only have one learning example (exceptions are bug duplicates) which doesn't allow us any generalization. Classification models built in this way are consequently also bound to be inaccurate.

What is it therefore that we can do with this data? What we can is for a given bug to successfully identify other bug reports that are similar and get a numerical value of this similarity. In this case we can't say that the given bug *B1* is a duplicate but we can say, for example, that the most similar bug to *B1* is *B2* and that similarity between them is 0.56. Such result is not as useful as classification would be, but a list of most similar bugs would still be very helpful for the person who is about to commit a new bug report. Potentially we could also set a threshold for similarity and say that the bug is a duplicate if the similarity exceeds the selected threshold.

3.1 Computing similarities between bug reports

As stated before, each bug report contains an initial description of the bug and potentially any number of comments. Since we expect that the comments can contain valuable additional information about the bug we decided to concatenate the subject of the bug, the initial description and all the comments into one report and to treat this as a single document when we import bugs into Contextify. Documents are in Contextify represented using the vector space model. Each term in the document is weighted using the TF-IDF weighting scheme. We computed term frequency (TF) and inverse document frequency (IDF) for term t_i in document d_j as:

$$TF_{i,j} = f_{i,j} \qquad IDF_i = \log \frac{N}{n_i}$$

where $f_{i,j}$ is the frequency of term t_i in document d_j, N is the number of all documents and n_i is the number of documents that contain t_i. There are several variants of term-weighting and we decided to compute TF-IDF weight $w_{i,j}$ simply as:

$$w_{i,j} = TF_{i,j} \times IDF_i$$

In order to compute similarities between bug reports we also need a measure that would evaluate the correlation between any two reports. We used *cosine similarity* which is the standard measure for quantifying this correlation. For documents d_k and d_l we computed the similarity $sim(d_k, d_l)$ as:

$$d_k = [w_{1,k}, w_{2,k}, \dots, w_{M,k}]$$
$$d_l = [w_{1,l}, w_{2,l}, \dots, w_{M,l}]$$
$$sim(d_k, d_l) = \frac{\overrightarrow{d_k} \cdot \overrightarrow{d_l}}{|\overrightarrow{d_k}| \times |\overrightarrow{d_l}|}$$

3.2 Comparison of similarities between duplicates and non-duplicates

Using the described measure we can now compute similarity between any two bug reports. The question that now arises is how well can this similarity be used to detect bug duplicates – in other words, do bug duplicates really use more similar words in their descriptions than non-duplicates?

To answer this question we performed the following experiment. First we computed for all bugs that don't have duplicates and are not marked as duplicates what is the similarity of their most similar bug report. We then performed a similar computation on bugs that have duplicates (we'll call them original bugs) or are marked as a duplicate. We created sets of bugs where each set consisted of one original bug and all its duplicates. For each bug in the set we computed similarities with other bugs in the set and remembered the maximum value. A graph displaying probability density function of similarities for these two groups of bugs is displayed in Figure 2. As it can be seen, the curves are similar, although the curve for non-duplicates is shifted more to the left and has lower density for higher values of similarity. Based on the graph we can conclude that the similarities are higher between duplicated reports but there is no good threshold that would allow us to accurately classify the bug as duplicate or not based on the highest similarity.

3.3 Ranking bug reports based on similarity

Although classification based on similarity is not accurate we can still help the users to identify bug duplicates. For a selected bug report we can compute a ranked list of most similar bug reports. The user can then inspect the list and decide if the bug is a duplicate or not. If the ranking is good it would be enough for the user to check only the first few reports in order to decide if the report is a duplicate or not.

To test how much can the ranking help the users to identify the duplicates we performed an experiment. For each bug report that has duplicated reports we computed a ranked list of 100 most similar bug reports in the whole repository. We then checked how well are the duplicated reports of the bug ranked in this list. This is a standard information retrieval task where the query is the tested bug report and the answer set is the set of similar reports.

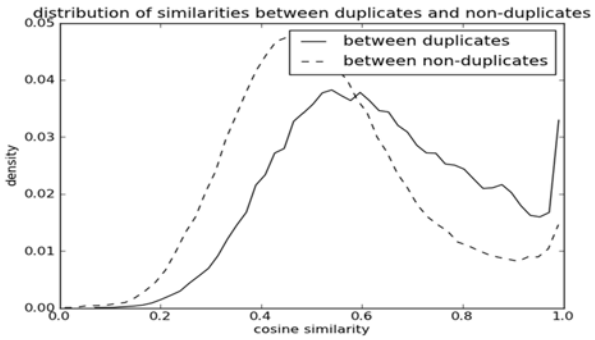


Figure 2: *The comparison of similarities between duplicated reports and reports without duplicates.*

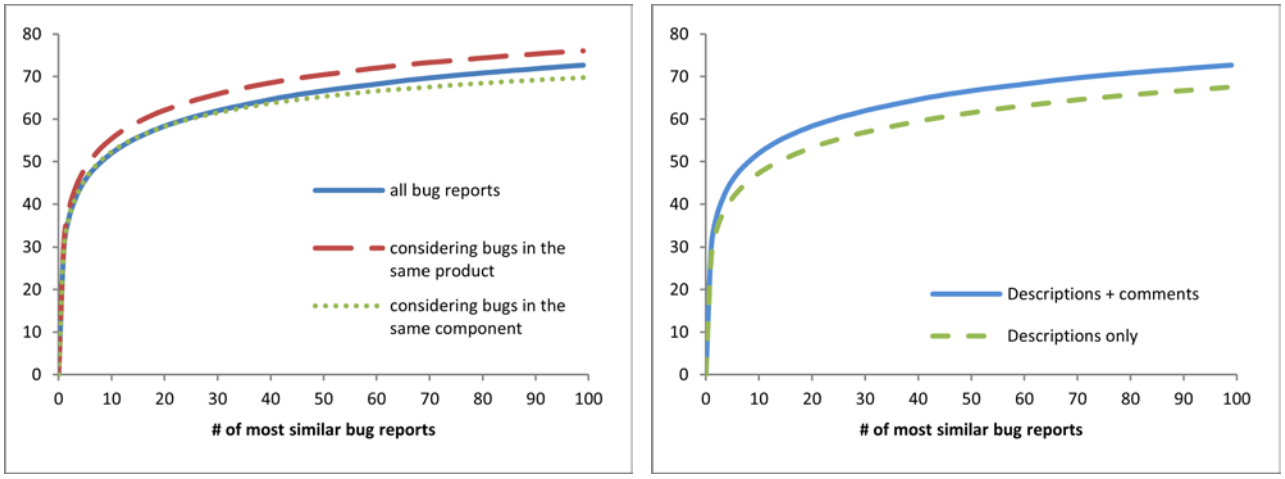
There are different metrics that can be used to evaluate the success of this task. Commonly used measures are precision and recall. They are not the most appropriate for us since we (1) are only interested where in the list is the *first* correct answer (duplicate) and (2) there is most often only one correct answer (most bugs have only one duplicate) which would automatically result in low precision. Instead we decided to use mean reciprocal rank that is often used in question answering systems. Reciprocal rank is the inverse of the rank of the first correct answer and mean reciprocal rank (MRR) is the average of the reciprocal ranks for a sample of queries *Q*:

$$MRR = \frac{1}{|Q|}\sum_{i=1}^{|Q|}\frac{1}{rank_i}$$

Our set of queries *Q* consisted of 63,861 reports (duplicated + original bug reports) and the computed MRR was 0,374. The Figure 3.a shows the percent of detected duplicates in relation to the number of inspected reports. The full line shows the results we obtained by comparing the test reports with all other bug reports in the repository. We can see that the curve is very steep and by looking at the first 5 most similar bug reports we can detect more than 45% of all duplicates.

In order to additionally improve the ranking we wanted to see if we can use some information from the meta-data of the reports. For each bug report, the user has to specify to which product it belongs and then even more specifically to which component inside the selected product. The information about the selected product and component is then stored as meta-data of the report. Our expectation is that we can use this information to improve the ranking by only considering those reports that are assigned to the same product/component. The influence of this information on ranking is also shown in Figure 3.a. Interestingly we can see that using the product information improves the ranking, while using the component information has a detrimental effect. The reason for this is that users sometimes assign the bug to the wrong product/component. In Bugzilla, there are almost 5,000 duplicates that are assigned to the wrong product and more than 12,000 that are assigned to the wrong component. If we only consider reports within the same product/component we therefore cannot locate the duplicate for these reports which in turn degrades the ranking quality.

We were also interested in knowing if the initial bug descriptions contain all the necessary information needed to identify the duplicates or do the following comments also contribute something valuable. For this we again for all duplicated reports computed 100 most similar reports in the whole repository, but this time the documents representing the reports consisted only of the bug subjects and the initial



(a) (b)

Figure 3. *The percent of detected duplicates in relation to the number of inspected reports. The figures show how the use of the meta-data (a) and the use of comments in the reports (b) influences the accuracy of the ranking.*

descriptions (without comments). Results are shown in Figure 3.b. It is evident that the ranking without using comments is worse which indicates that comments are valuable and should also be used when identifying duplicates.

4. Related work

The presented work does not present any new methods but only demonstrates the applicability of existing approaches on a special use case. The most related work was done by Hiew[7] who searched for similar bugs by first computing centroids of related bugs. Čubranić and Murphy's Hipikat[8] project determines which reports in a repository are similar to each other using an information retrieval algorithm. Wang describes another approach to detecting bug duplicates using natural language and execution information[9]. Automatic detection of duplicate documents has also been considered in other contexts. In large document collections, for example, the duplicates are identified to maintain the speed of search engines[10].

5 CONCLUSION AND FUTURE WORK

We have presented our preliminary work in the field of automatically identifying bug duplicates. First we have described the details of the KDE BTS that was used as our case study. Next, we described the preprocessing steps that we used to represent the bug reports in the vector space model. We defined the similarity measure used and evaluated how accurately we can rank duplicated bug reports.

We have several ideas for future work. First we plan to inspect the reports that have high similarity and are not marked as duplicates. We will identify what is the common property of these reports (for example, they might be very short reports, or might contain common phrases that are computer generated, like "no debugging symbols found") and try to take this information into account when weighting terms or computing similarity between reports. Another idea is to check how informative the date of the reported bug is. Since new bugs occur with new releases of the software it is probably more likely that two reports are duplicates if they are closer in time. We will also try to improve the use of the meta-data like product/component information. Instead of considering only reports in the same product we can soften the constraint to include also related products.

6 ACKNOWLEDGMENTS

This work was supported by the Slovenian Research Agency, European Social Fund and ALERT (ICT-2009.1.2).

References

[1] "Bugzilla," 2011. Available: http://www.bugzilla.org/.
[2] "Mantis," 2011. Available: http://www.mantisbt.org/.
[3] "LaunchPad," 2011. Available: https://launchpad.net/.
[4] "ALERT," 2011. Available: http://www.alert-project.eu/.
[5] "KDE," 2011. Available: http://www.kde.org/.
[6] G. Leban and M. Grobelnik, "Displaying email-related contextual information using Contextify," International Semantic Web Conference, Shanghai, China, 2010, pp. 181-184.
[7] L. Hiew, "Assisted detection of duplicate bug reports," The University Of British Columbia, 2006.
[8] D. Cubranic and G. C. Murphy, "Hipikat: Recommending pertinent software development artifacts," 2003.
[9] X. Wang, L. Zhang, T. Xie, J. Anvik, and J. Sun, "An approach to detecting duplicate bug reports using natural language and execution information," in *Proceedings of the 30th international conference on Software engineering*, 2008, pp. 461–470.
[10] A. Chowdhury, O. Frieder, D. Grossman, and M. C. McCabe, "Collection statistics for fast duplicate document detection," *ACM Transactions on Information Systems (TOIS)*, vol. 20, no. 2, pp. 171–191, 2002.

WIKImage: CORRELATED IMAGE AND TEXT DATASETS

Doni Pracner [1]*, Nenad Tomašev* [2]*, Miloš Radovanović* [1]*, Dunja Mladenić* [2]*, Mirjana Ivanović* [1]

[1]Department of Mathematics and Informatics
Faculty of Sciences, University of Novi Sad
Trg Dositeja Obradovića 4, 21000 Novi Sad, Serbia

[2]Jožef Stefan Institute
Jamova cesta 39, 1000 Ljubljana, Slovenia
doni.pracner@dmi.rs, nenad.tomasev@ijs.si, radacha@dmi.rs
dunja.mladenic@ijs.si, mira@dmi.rs

ABSTRACT

This paper presents work towards the creation of free and redistributable datasets of correlated images and text. Collections of free images and related text were extracted from Wikipedia with our new tool *WIKImage*. An additional tool – *WIKImage browser* – was introduced to visualize the resulting dataset, and was expanded into a manual labeling tool. The paper presents a starting dataset of 1007 images labeled with any combination of 14 tags.

The images were processed into a number of scale invariant (SIFT) and color histogram features, and the captions were transformed into a bag-of-words (BOW) representation. Experiments were then performed with the aim of classifying data with respect to each of the labels on dataset variants with just the image information, just the textual data, and both, in order to estimate the difficulty of the dataset in the context of different feature spaces. Results indicate improvements in precision, recall and the F-measure when using the combined representation with *support vector machines* as well as the *k-nearest neighbor* classifier with the cosine similarity measure.

1 INTRODUCTION

A large body of research exists on mining image [3], and textual data [2], making these topics mature areas of research and practice. In addition, there exist many sources where one can find pictures with short captions associated to them, which can offer better understanding of the context underlying the images. In this paper we describe initial efforts to create correlated image and text datasets that will facilitate further research into these areas. One of the main guidelines in the creation of such datasets is to compose them out of free and redistributable images, thus enabling their free use for research purposes.

One of the options for collecting image data is crawling the Web, fetching images and surrounding text. This approach can be somewhat problematic since web pages tend to be chaotic, with a general lack of adequate context information. There are also numerous copyright issues: pictures are rarely labeled, and even when they are, the manner of labeling is not standardized.

Our approach to getting around the aforementioned issues is by crawling Wikipedia. Images are used in articles, thus captions for them can easily be extracted. Also, there exist categorizing schemes in Wikipedia. Most importantly, a lot of Wikipedia material is free to use and is labeled adequately, since the project makes efforts to protect itself from copyright infringements.

The rest of the paper is organized as follows. In the next section the new tool *WIKImage* is presented. Section 3 contains the description of an example dataset, as well as the feature representation. The results of classification experiments that were performed on the dataset are discussed in Section 4. Finally, in Section 5 a short summary, conclusions and future work options are given.

2 WIKImage

Our project for collecting image and text data from Wikipedia is named *WIKImage* – as a combination of the terms *WIKIpedia* and *Image*. In this section we will describe the general approach and technical issues regarding the process of data extraction from Wikipedia.

The wiki package Mediawiki, used to create Wikipedia, has a complex web-based API for accessing and editing information, which enabled us to retrieve the images and text. Entry point is of the form `http://www.example.com/w/api.php`. It can return data as XML, YAML, JSON, WDDX, serialized PHP, or human readable PHP arrays. All of these can be viewed on-line in a browser by appending "fm" to the format name.

A list of options is given when the API is called without parameters. A number of actions can be performed. In the context of this work the most important was the action *query*. It gives information about pages, their templates, categories, revisions, internal and external links, image usage, etc. Being that we are interested in obtaining free images, the retrieval was image-centric. There are several page categories that contain only free images, which provided a good starting point.

The process of obtaining images and text is split into two steps that can be executed simultaneously: retrieval and pro-

cessing, since the latter is much faster, and can be performed in several ways. Retrieval starts from a category, and attempts to fetch (at least) a predetermined number of images that are actually used on at least one page. They are all downloaded to the *images* folder. The API is asked for a list of pages that use the image, limited to results in the 'articles' name-space (that is, no user pages, talk pages, templates...). In the retrieval step the whole article is saved in the *pages* folder. Data about the usage of the image is saved separately in the *data* folder. All file names use the unique IDs of pages (both images and articles), and in all cases the existing data will not be downloaded multiple times.

Once everything is downloaded, adequate pieces of the text are extracted for every image. One option is for complete paragraphs that contain the link to the image to be extracted, and all of them stored in the *text* folder under the image ID name. Another option is to extract the captions for the images, since most of them use the standard wiki markup for these purposes. An attempt is also made to recognize if the picture is a part of a gallery construct. These extracted lines are stored in the *captions* folder. More refinements are expected to be made, depending on what will be needed in future work.

For a visual presentation of the obtained information the *WIKImage dataset browser* was created (Figure 1). It is a PHP web page that shows the images and appropriate text pieces page by page, and can be used to explore the datasets on-line.

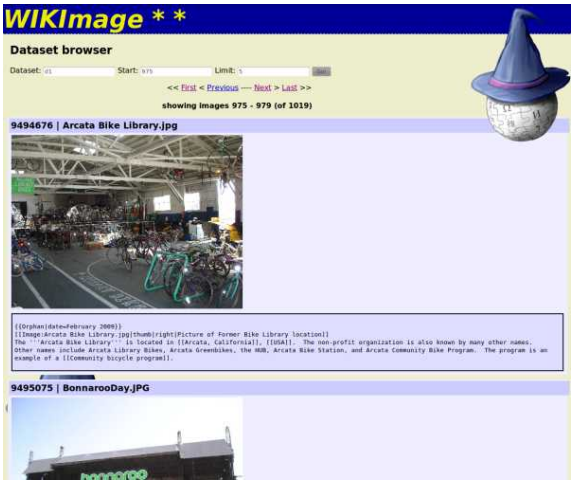


Figure 1: *WIKImage dataset browser.*

Our initial aim was to provide labels with all extracted data, to facilitate the task of automated classification. The original idea was to use Wikipedia's categories to automatically obtain class data about our images. But the problem with this is that the categories are chaotically applied to the images – some are not labeled at all, some are labeled with multiple categories. The categories themselves are very inconsistent in the level of detail. There are some hierarchies, but they are also problematic, since there are no central top-level categories. On the other hand, articles tend to be well categorized, but these categories are not necessarily relevant to the images used – for example in the article about Napoleon, there are images of him, but also of his tomb, the building where he lived on Elba, of the island itself, maps, battles, etc., all of which would not be appropriate to label with "people" or similar.

In the end, the decision was made to manually label the images into several binary categories, being that many images would be hard to classify into just one class. For this purpose the dataset browser was expanded into a dataset labeler to assist in the process.

3 FEATURE REPRESENTATION

Using the described tools, several datasets were obtained. We will present the dataset internally named "d1." For it's creation the English Wikipedia was used (API: `http://en.wikipedia.org/w/api.php`), with images from "Category: Creative Commons Attribution-ShareAlike 2.5 images." The dataset contains 1007 instances.

All instances were manually classified into binary labels shown in Table 1, along with the number of images in the dataset that were labeled with each of the tags. Many of the images had two or more tags applied to them. Some of these labels were introduced with the idea of being "secondary" – generally applied with another label, but with the potential for providing interesting results nevertheless. For instance, "sports" usually appears with "people" or "vehicles." Additionally, a few "special" labels were applied – for pictures with no captions, bad captions, and pictures that were found ambiguous.

Table 1: *Distribution of labels in the dataset.*

Label	Abbrv.	Instances
abstract and generated	abstract	6
animals	animals	89
art	art	53
buildings and constructions	buildings	375
documents and maps	documents	80
logos and flags	logos	43
machines, tools and tech	machines	23
misc nontech objects	misc	33
nature and scenic	nature	202
people	people	215
plants and fungi	plants	30
space	space	7
sports	sports	31
vehicles	vehicles	58

The image IDs, names, captions and label assignments (14 binary features) were exported to a comma-separated values (CSV) file for further processing.

Image analysis requires, first and foremost, an appropriate image representation which captures the information relevant for the task at hand. Many different feature types have been proposed, SIFT [5], GLOH [6] and LESH [9] being among the most frequently used. The basic idea is simple: detect some *keypoints* in images that are of interest for building the representation, and proceed by describing the local surrounding regions in a robust and detailed way. SIFT features in particular exhibit robustness to scaling, rotation and translation and are

hence very popular and used in many applications. This is why we opted for a SIFT-based representation of the acquired data.

More specifically, we used the *bag of features* representation, where groups of similar local features are treated as *visual words*. A *codebook* representing prototypical SIFT features was obtained by K-means clustering [4] on a large sample of features taken from all the images. The quantized representation is then obtained by mapping each individual feature to its closest prototype [11]. In our experiments (Section 4) we have used a 400-feature vocabulary. Since SIFT features are traditionally extracted from gray-scale images, we also appended a 16-bin color histogram to each SIFT representation [11].

As for the textual part of the representation, WEKA [10] software was used to transform the captions into a bag-of-words (BOW) representation. The *StringToWordVector* filter was used, applied with the alphabetic tokenizer which produces only "pure" words, resulting in a total of 3182 text attributes. No stemming was applied at this step, although it may be used in future experiments.

4 EXPERIMENTS

To demonstrate the difficulty of the constructed dataset "d1," and explore the influences of a combined representation, three datasets were created for each label – one with just the BOW representation (*text*), one with just the SIFT and histogram data (*image*), and one that contains both (*all*). All these representations are available for download at the WIKImage project home page [8].

The three created variants of data were tested in WEKA with the *k-nearest neighbor (kNN)* classifier (IBk implementation [1] with *Euclidean*, *Manhattan* and *cosine* distance/similarity measures) and a *support vector machine (SVM)* classifier (SMO implementation [7] with the default settings: $C = 1$ and the linear kernel), on binary classification problems corresponding to labels shown in Table 1.

The differences in percentage of accuracy between the image/text/all datasets were not great on average, with all of them scoring around 90%. There was a slight improvement when using all the features with SVM, as well as the cosine measure for kNN, but also a similar loss of accuracy with other distance/similarity measures.

What is more relevant for the observed imbalanced classification problems are the *precision*, *recall* and *F-measure* values, and here the differences were more notable, especially with SVM and kNN-cosine ($k = 1$) tests. Figure 2 overviews these results, showing averages of performance measures over all classification problems that were evaluated in 10 runs of 10-fold cross-validation. When using other distance metrics for kNN, the results for the text-only sets were much weaker than with the other sets. Likewise, usage of higher k values did not produce a notable difference in the results.

A more in-depth analysis of the results indicates that for most of the individual tags classifiers tend to be more successful with the combined datasets. Tables 2 and 3 show F-measure values (as the balance between precision and recall) for all the tags and types of datasets.

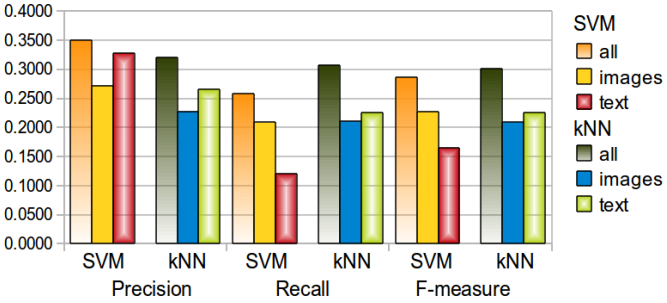


Figure 2: *Averaged results for SVM and 1NN (cosine).*

Table 2: *SVM – F-measure results.*

Dataset	All	Images	Text
abstract	0.0400±0.20	0.0000±0.00	**0.1000**±0.30
animals	0.2225±0.15	0.0825±0.10 •	**0.2659**±0.18
art	0.0961±0.14	0.0121±0.05	**0.1056**±0.15
buildings	**0.6807**±0.06	0.6262±0.06 •	0.4693±0.09 •
documents	**0.6334**±0.13	0.6161±0.14	0.2190±0.17 •
logos	**0.5844**±0.24	0.5807±0.26	0.0087±0.04 •
machines	0.0000±0.00	0.0000±0.00	0.0000±0.00
misc	**0.0999**±0.18	0.0925±0.16	0.0000±0.00
nature	**0.6603**±0.08	0.5976±0.09 •	0.3477±0.10 •
people	**0.5202**±0.08	0.4336±0.09 •	0.2394±0.10 •
plants	0.1253±0.22	0.0353±0.12	**0.1457**±0.23
space	**0.2000**±0.40	0.0900±0.28	**0.2000**±0.40
sports	0.0617±0.15	0.0000±0.00	**0.1272**±0.22
vehicles	0.0740±0.12	0.0217±0.07	**0.0752**±0.14
Average	**0.3014**	0.2091	0.2252

• statistically significant degradation compared to "all"

Table 3: *1NN (cosine) – F-measure results.*

Dataset	All	Images	Text
abstract	**0.2900**±0.45	0.0067±0.07	0.1800±0.39
animals	**0.4160**±0.14	0.1272±0.12 •	0.3346±0.20
art	**0.3216**±0.18	0.1222±0.14 •	0.2907±0.21
buildings	**0.5908**±0.05	0.5338±0.06 •	0.4990±0.19
documents	0.3891±0.15	**0.5151**±0.15	0.2444±0.17 •
logos	0.3118±0.18	**0.3359**±0.21	0.0241±0.07 •
machines	**0.0413**±0.14	0.0095±0.05	0.0375±0.13
misc	**0.0352**±0.11	0.0327±0.10	0.0029±0.03
nature	**0.4529**±0.09	0.4424±0.08	0.3748±0.16
people	**0.3856**±0.08	0.3666±0.08	0.3011±0.13 •
plants	**0.2491**±0.25	0.0020±0.02 •	0.2139±0.25
space	**0.3650**±0.46	0.1100±0.30	0.2900±0.45
sports	0.1457±0.17	0.1361±0.17	**0.1493**±0.19
vehicles	**0.2250**±0.17	0.1873±0.13	0.2110±0.19
Average	**0.2856**	0.2277	0.1646

• statistically significant degradation compared to "all"

Results shown in bold are the best for the corresponding row in the table, and the bullet signs next to the results denote statistically significant degradation compared to the combined representation, according to the corrected resampled t-test [10] at 0.05 significance level.

5 CONCLUSIONS

This paper presented work towards the creation of free and redistributable datasets of correlated images and text. Collections of free images and related text were gathered from Wikipedia by the *WIKImage* tool. An additional tool – *WIKImage browser* – was created to visualize the resulting dataset, and was expanded into a manual labeling tool. The paper presented an initial dataset of 1007 images labeled with combinations of 14 different tags.

The images were processed into a number of SIFT and histogram features, and the captions were transformed into a BOW representation. Binary classification tests were then performed on with respect to each of the labels, on dataset variants with just the image information, just the textual data, and the combination of both.

Initial experiments showed slight improvements in raw accuracy when using the combined sets for SVM and kNN with the cosine similarity measure. More importantly, there were notable improvements in *precision*, *recall* and *F-measure* values for these experiments, where on average the combined representation was the best and showed statistically significant improvements for several of the tags.

The datasets used in the presented experiments can be obtained from the WIKImage project home page [8].

Being that what this paper described are only the initial steps in the creation of the datasets, there is considerable space for future improvements.

The system for detecting captions and relevant text could be enhanced further. There are many examples of pictures used in info-boxes and similar templates, which (at this point) are ignored. Some of the more popular ones could be recognized automatically, producing more reliable captions.

Effort could be invested in searching for schemes for automatic assignments of classes to the images, based on the categories in Wikipedia, providing that suitable hierarchy schemes can be found. Similarly, the aforementioned Wikipedia info-boxes could be used to suggest classes for at least a part of the dataset, to reduce the manual effort in labeling. For instance, there are info-boxes for musicians, actors, etc., which could tag the pictures as "people," or info-boxes for states which could tag pictures as "logos and flags."

An interesting option for the future is to use the tool on other Mediawiki projects. The API we connect to can be changed in a single line – it can be used on any other Mediawiki page (for instance Serbian or Slovenian Wikipedia). This could make datasets of images with captions in different languages, producing various options for multi-language research.

The most important future steps concern dataset expansion. As of this writing, new and bigger sets are already being labeled. The labeling tool could be advanced to a cooperative project like Wikipedia, with tags submitted by different users.

Finally, more experiments with different algorithms for classification would be desirable, and also a more detailed comparison of the influence of various parameters. The BOW representation can also be tweaked in the future, for instance by using stemming rules.

Acknowledgments

This work was supported by bilateral project between Slovenia and Serbia "Correlating Images and Words: Enhancing Image Analysis Through Machine Learning and Semantic Technologies," the Slovenian Research Agency, the Serbian Ministry of Education and Science through project no. OI174023, "Intelligent Techniques and Their Integration into Wide-Spectrum Decision Support," and the ICT Programme of the EC under PASCAL2 (ICT-NoE-216886) and PlanetData (ICT-NoE-257641).

References

[1] D. W. Aha, D. Kibler, and M. K. Albert. Instance-based learning algorithms. *Machine Learning*, 6(1):37–66, 1991.

[2] R. Feldman and J. Sanger. *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press, 2007.

[3] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Prentice Hall, 3rd edition, 2007.

[4] S. P. Lloyd. Least squares quantization in pcm. *IEEE Transactions on Information Theory*, 28:129–137, 1982.

[5] D. Lowe. Object recognition from local scale-invariant features. In *Proceedings of the 7th IEEE International Conference on Computer Vision (ICCV)*, volume 2, pages 1150–1157, 1999.

[6] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, 2005.

[7] J. C. Platt. Fast training of support vector machines using sequential minimal optimization. In B. Schölkopf, C. J. C. Burges, and A. J. Smola, editors, *Advances in Kernel Methods – Support Vector Learning*, pages 185–208. MIT Press, 1999.

[8] D. Pracner. The WIKImage project home page. `http://perun.dmi.rs/pracner/wikimage/`, 2011.

[9] M. S. Sarfraz and O. Hellwich. Head pose estimation in face recognition across pose scenarios. In *Computer Vision Theory and Applications*, pages 235–242, 2008.

[10] I. H. Witten and E. Frank. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann Publishers, 3rd edition, 2010.

[11] Z. Zhang and R. Zhang. *Multimedia Data Mining: A Systematic Introduction to Concepts and Theory*. Chapman and Hall / CRC Press, 2008.

EXPLORING HISTORY THROUGH NEWSPAPER ARCHIVES

Jasna Škrbec, Marko Grobelnik, Blaž Fortuna, Boštjan Pajntar
Artificial Intelligence Laboratory
Jozef Stefan Institute
Jamova 39, 1000 Ljubljana, Slovenia
Tel: +386 1 4773127; fax: +386 1 4773851
E-mail: jasna.skrbec@ijs.si

ABSTRACT

This paper proposes a pipeline for searching and browsing through newspaper archives. It uses a combination of algorithms for extracting information from text and tools for visualizing different text structures capable of handling large amount of articles that are normally collected in archives. The proposed pipeline is implemented as a web application, illustrative results show appropriateness of the proposed pipeline for searching and browsing news archives.

1 INTRODUCTION

News publishers are collecting old articles into large archives with millions of articles. Even though the articles are typically annotated with additional meta data, the archives are hard to browse or used to discover larger stories. Typically search interfaces do not work well since they are not specialized for archives, and as such do not take advantage of the inherent structure.

This paper presents Archive Explorer, a system for browsing the archives, which combines text mining and visualization techniques. The goal is to go beyond typical search and browse interfaces, which focus on retrieval and visualization of articles, and try to shows the articles in context with the rest of the archive. For example, how does an article fit into larger story, which developed over longer time period, and is discussed in many articles, or how is the searched topic represented in the archive with respect to time, place, major events, important people, important keywords etc. The system is designed to get the user's attention and interest for browsing through other related issues.

2 ARCHITECTURE AND BASIC PIPELINE FOR DOCUMENTS IN ARCHIVE EXPLORER

Basic architecture is combined from a pre-processing articles and a live processing of a data from a database as can be seen on Figure 1.

In pre-processing level archives are imported in database from xml files. Xml file contains text and meta data which varies from archive to archive. Mostly all files have author or authors of article, publish date, title and which part of text is lead paragraph. Usually there are added pages or section where it was published, internal categories, place where it was written, etc. A pure text article and its basic meta data are taken from xml and stored in a database.

Second part of pre-processing is enriching a text with some context. The text from a database is sent to Enrycher application [1]. Enrycher is a service-oriented framework for extraction and representation of document content. Information is extracted from an unstructured document with different knowledge extraction techniques. The first result of Enrycher used by Archive Explorer is extracted list of named entities, such as people, organizations, cities, countries or other places and things that are commonly known. They are extracted from a text with two different techniques, a pattern-based and a supervised learning one [2]. The second result of Enrycher is classification of articles into categories and extraction of keywords. A hierarchical classifier is used for a taxonomy categorization. Relevant categories are defined by a word and a phrase similarity. Hierarchy for categories is taken from DMoz topic ontology [3] and consists of standard centroid vectors. Comparison with document is started by hierarchy's top category and then down through tree. Result is a list of categories with their whole hierarchy path which are ranked by similarity to a document. Only top ten most similar ones are used to classify an article.

All results given by Enrycher are inserted into a database and used to describe a context of an article. With a context it is possible to connect an article with other articles into stories and to place it into timeline.

Articles must be pre-processed and placed in a database before they are used in processing on the server. It is very important to have a rich context around some article if we want to have a good base for connecting and processing the whole archive. Since a typical newspaper archive easily contains millions of articles, the pre-processing time can quickly become a bottle neck of whole system. That is why articles are first saved into database and later they are sent to get a context. So an article can be found even if it has no context around it yet. Faster way to get a context is with

multithreading and speeding up the whole process of enrichment.

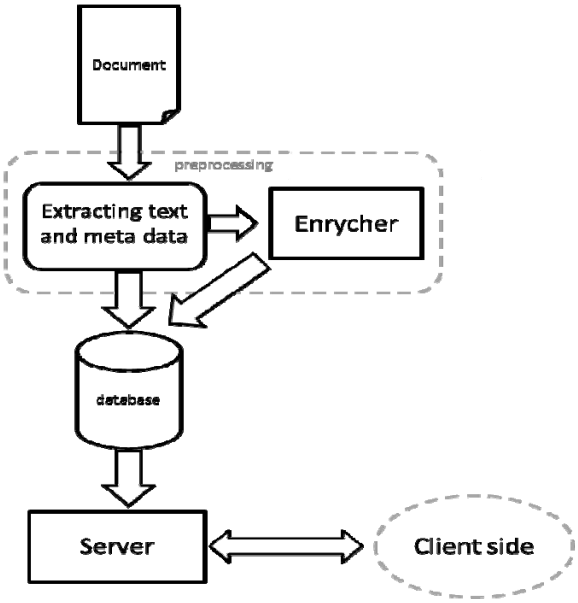


Figure 1: *Basic architecture of Archive Explorer.*

3 EXPLORING ARCHIVES

If we want to offer archives that are readable and can be explored in interesting and convenient ways, we need to offer more than just an interesting text. Instead of having just a regular search form it can be upgraded with a faceted search interface.

One way for user to start browsing around is to choose entity, keyword, year, author or one of the main categories which are put on first site of the application. The other way, which is most common, is that user type in his search query. In that purpose there are several fields so user can narrow down what he is looking for. Next step is showing user search results with some context around those results. Then a user can choose from given options from a context in which direction the search continues. With offering user options related with his search is not just helping him find what he wants, but also encouraging him to read and search more about this or related topics since there will be only topics he chose before; topics he is interested in.

In an archive with so many articles there are a lot of similar topics. Finding them is not easy especially when user is not sure what or when or who is he looking for. Some nice visualization can manage problems like that. One of goals of Archive Explorer is to put a power of the queries and advantages of the visualization together.

3.1 Browsing search results with Searchpoint

First example of cooperation between a context and a visualisation is Searchpoint [4]. It is an application for ranking and visualizing search results from an ordinary search engine. In our case Searchpoint is using entities, connections between entities and articles for ranking. Entities are divided into four different groups which are presented in four different parts of visualization.

Every part is presented in its own window where entities are illustrated with a text and spots of different colours. In the middle of every window there is a red dot that can be dragged around with a mouse. If the red dot is moved then order of articles will change. An order is changed in the way so articles most connected with entity or entities nearest to the red dot are on the top of search results. For example if we are interested in a specific person we should drag a red dot in a person window to cover a spot where a name of this person is illustrated. Articles pushed on the top are the ones most relevant for this person.

For testing purposes all examples are from part of New York Times archive. When we typed in search query ‘art’, Searchpoint returned entities illustrated on Figure 2. In location window we dragged red dot from centre up so the nearest entity/location is Brooklyn, in other windows we leave red dots at the centre position, which means that in search results articles concerning Brooklyn are at the top.

Figure 2: *Searchpoint visualization of entities. In location window red dot is dragged up to the Brooklyn.*

3.2 Visualization with graphs

All kind of graphs can be also very powerful visualization tool. Sometimes it is enough just to put all set of some keywords on the spot, just to see how many there are. It is very illustrative for articles if there is a picture with all the words used in it where words are bigger if they are used more commonly. Of course it is smart to remove all stop words.

With help of Searchpoint it is obviously which entities are related with which articles, but it is also very interesting how and which entities are related between themselves. So if someone is interested in some topic, he can see which person is connected with other person or with which city or

which city has the most connections in this topic, etc. JavaScript library called Arbor.js [5] is used for visualization of that kind of a graph. For now graphs are used to show connection between entities found in search results and to illustrate one entity's connections. In Figure 3 we can see example for first case. Data needed to realize graphs is provided by sql queries and server live processing.

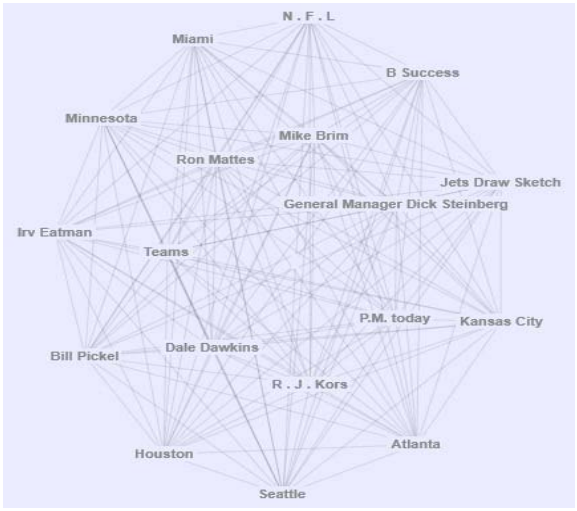


Figure 3: *Connections between entities for one article visualized with graph.*

3.3 Visualization of article's content

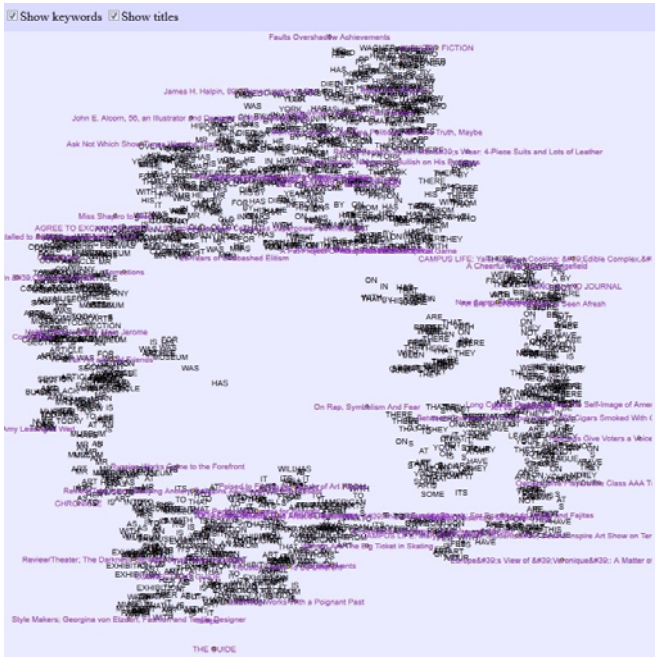


Figure 4: *Document Atlas visualization of articles.*

Visualizations described in last two sections are meant for showing how meta data is connected with either other meta data either with articles. Visualization in this section is intended for connections between contents of articles. It helps to understand, discover and summarize the topics in articles. This visualization is called Document Atlas [6] and example of use in our system is shown in Figure 4.

Documents pushed into an application are mapped onto a two-dimensional plane based on similarity between them. If documents are very similar by content, then their coordinates are closer than those of documents that are less similar. A set of different methods is used for a calculation of the semantic space which includes documents and named-entities from the text corpus [7]. In Archive Explorer Document Atlas is used to show the whole picture of search results. With that picture it is illustrated which articles belong together in same stories or in same topics. On picture there are article's titles and their keywords. From set of keywords we can guess topics on different spots.

3.4 Timeline

People, places, organizations, major events in history and others such things are covered with entities, keywords and other context, but time perspective is missing. Since we are working with archives a time component is quite important. For example, if someone is searching for a specific city in an archive, he would probably want to have some overview by its history over the years. One way to visualized search results is also by years or by months. Graph visualized with help of Raphel library [8] shows number of found articles by years. A set of articles in the chosen year is best shown with Document Atlas. The same kind of graph is also used for showing numbers of articles per month for one year (Figure 5). This is useful for spotting important events in a specific year.

Figure 5: *Number of articles per month for one year.*

4 EXAMPLE FROM USER POINT OF VIEW

Let's say we have a student who has to do a research for his history class. In research he has to decide how different events have influenced its history. He chooses a city where he was born and focuses on the time when he was born. So he decides to find some bigger newspaper which existed back than and has online archive. Since he is interested in specific date and location he type in his query, for example in location text field he type New York and in date field he type 1992-02-02.

When he gets results (Figure 6) he can immediately see from Searchpoint which people and organizations were the most important at that time in the city and he can see connections with other locations. He is interested if the

events were new or old, so he browses forward with the help of sidebars to find important categories, entities, keyword and authors related with the city in the specified time period.

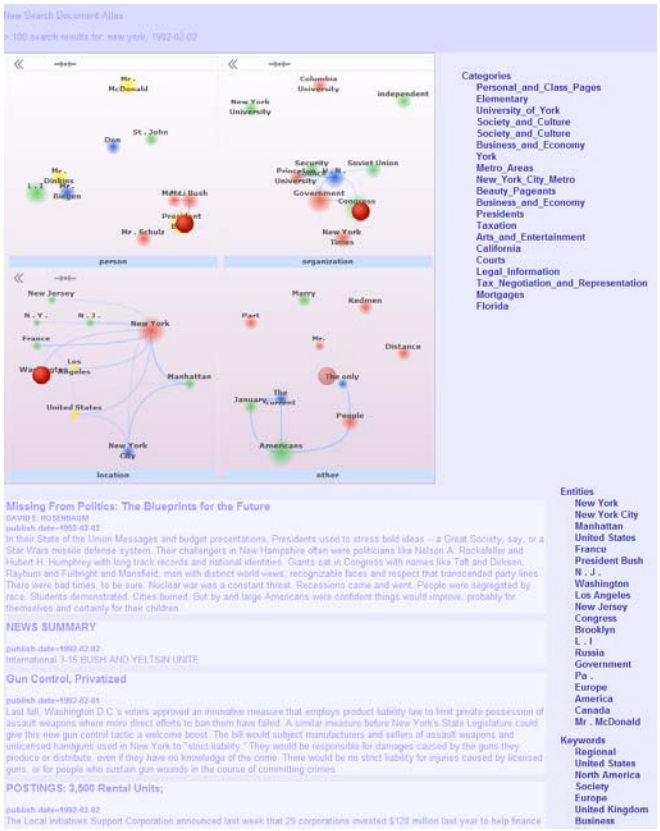


Figure 6: *Page with search results and Searchpoint.*

By browsing authors' pages he gets image of what kind of publisher is specific author. So if he found a lot of articles from author who was writing mostly about art, he knows there were a lot of movement in this area by that time.
Because he doesn't want to read everything he looks over Document Atlas to see which were the major topics at the time. There he discovers that in that time they were planning to build hydroelectric plan on border with Canada. Now he can go further browsing about that. The article about hydroelectric plan is on Figure 7.

5 CONCLUSIONS AND FUTURE WORK

At this point on one side Archive Explorer is a working system but on the other side it is just a basic scheme of possibilities that can be implemented in a system like that. It already shows the advantages of merging text mining and visualization.
A lot of ideas are still to be implemented. For example, some existing visualizations could be used in more areas of the system. Search of articles can be improved with narrowing criteria and query suggestions. We will also try to automatically connect articles into stories. The time component has also has a lot of space for improvement.

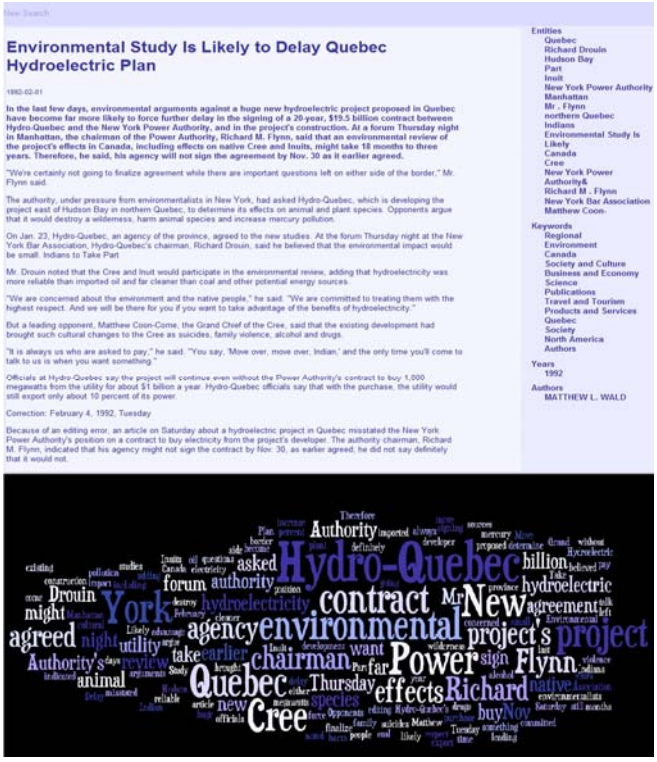


Figure 7: *Article's page.*

6 ACKNOWLEDGMENTS

This work was supported by the Slovenian Research Agency and the IST Programme of the EC under PASCAL2 (IST-NoE-216886) and RENDER (ICT-257790-STREP).

References

[1] Enrycher, http://enrycher.ijs.si.
[2] Stajner, T.; Rusu, D.; Dali, L,; Fortuna, B.; Mladenić, D.; Grobelnik, M. A service oriented framework for natural language text enrichment. Informatica 34, 3 (2010).
[3] DMoz, http://www.dmoz.org
[4] Searchpoint, http://searchpoint.si.
[5] Arbor.js, http://arborjs.org.
[6] Document Atlas, http://docatlas.ijs.si.
[7] Fortuna, B.; Mladenić, D.; Grobelnik, M. Visualization of Temporal Semantic Spaces. In: Davies, J. et al (ed.) Semantic Knowledge Management (Springer, 2008).
[8] Raphael, http://raphaeljs.com/

EXPLORING THE HUBNESS-RELATED PROPERTIES OF OCEANOGRAPHIC SENSOR DATA

Nenad Tomašev, Dunja Mladenić
Jožef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia
e-mail: nenad.tomasev@ijs.si, dunja.mladenic@ijs.si

ABSTRACT

In this paper we examine how the high dimensionality of oceanographic sensor data impacts the potential use of nearest-neighbor machine learning methods. We focus on one particular consequence of the curse of dimensionality – hubness. We examine the hubness of oceanographic data and show how it can be used to visualize and detect both prototypical sensors/locations, as well as ambiguous and potentially erroneous ones. We proceed to define an easy classification problem on the data, showing that the recently developed hubness-aware classification methods may help to overcome some of the hubness-related issues in sensor data.

1 INTRODUCTION

Various sensor arrays spread across the world have endowed us with a greater insight into the dynamics of many natural phenomena. Due to the sheer quantity of such data, semi-automatic analysis and interpretation is not an option, but a necessity. Machine learning methods are hence used for prediction, categorization, clustering, error-detection and other tasks which may prove beneficial within a certain problem context.

Nearest neighbor methods are frequently used in machine learning. They are based on an intuitive notion that similar instances (where similarity is measured by some appropriate metric) often share some common properties. Therefore, in order to deduce something about the current point of interest, its nearest neighbors can be examined and used to infer the desired property. If this property is in fact the instance label, we could use one of the many *k*-nearest neighbor classification methods proposed in literature.

The basic *k*-nearest neighbor method (*k*NN) was first introduced in [1]. It is a simple procedure, where a majority vote over the *k*-nearest neighbor set is taken in order to determine the label of the point of interest. This simple rule was shown to have some useful asymptotic properties [2][3] and hence became very popular – many extensions of the basic algorithm have been successfully applied to various problems. Some very robust general-purpose *k*-nearest neighbor methods have been proposed recently, such as [4] where the metric is learned from the data in such a way that in the new imposed metric space – the proportion of neighbors in *k*-neighbor sets which are of the same class as the observed points is maximized.

Most real world data these days is inherently high-dimensional, whether its images, text, medical data, or – time series data, such as sensor data addressed in this paper. It was shown that high-dimensional data are likely to express significant *hubness* [5]. In such cases, some very influential points emerge (*hubs*) which greatly impact all aspects of nearest-neighbor reasoning. Hubness will be discussed in more detail throughout Section 2.

Ever since the topic of climate change started gaining focus, analyzing such sensor data is becoming more important. Therefore, we wished to provide further insights into the applicability of nearest-neighbor reasoning in such data. Section 3 provides description of the data we have used in the experiments.

We examined the *k*-occurrence distributions based on several measured physical quantities. We used these results to map influential sensor nodes and to detect those which might pose severe difficulties in subsequent nearest-neighbor based inference. We also experimented with several recently developed hubness-aware classification methods and tested their applicability to the problem domain. Experimental results are presented in Section 4.

2 HUBNESS

The *curse of dimensionality* is a term commonly used to address the difficulties inherent in dealing with such data in various practical applications. One of these difficulties is known as *hubness*. Under hubness, different points occur in *k*-neighbor sets with increasingly unequal frequencies. Some points occur in many *k*NN sets, while others occur either very rarely or not at all. The former are referred to as *hubs* and the latter as *anti-hubs*. More specifically, hubness refers to an increasing *skewness* (the third central moment, which describes asymmetry) in the *k*-occurrence distribution in high-dimensional data [5]. This property of the *k*-occurrence distribution was successfully used in [10] to define a hubness-based clustering algorithm aimed specifically at clustering high-dimensional data. This shows that even such detrimental phenomena can be used to our advantage if understood properly.

2.1 BAD HUBNESS

When the data is labeled (i.e. meaningful categories exist), it is possible to distinguish between two sorts of k-occurrences: the *good* and the *bad* occurrences (implying *good hubs* and *bad hubs*, respectively). The distinction is made based on the nature of their influence on kNN classification. When a neighbor shares the same label as the observed point of interest, that is a *good occurrence*. Label mismatches define *bad occurrences* and add to the *bad hubness* of the neighbor point. Hence, the total number of k-occurrences of point x ($N_k(x)$) can be decomposed into the sum of its good and bad occurrences: $N_k(x) = GN_k(x) + BN_k(x)$. Bad hubness can be expected in border regions between different categories, as well noisy, erroneous, or otherwise mislabeled data. Some bad hubness, however, is no more than a consequence of high-data dimensionality and class imbalance.

Hubness-aware classification methods aim at diminishing the influence of bad hubness or otherwise exploiting it in other ways.

2.2 HUBNESS-AWARE CLASSIFICATION

There are several ways in which hubness in general and bad hubness specifically can be dealt with. The simplest approach was suggested in [5], where hubness-based weights were incorporated into the kNN rule in order to reduce the weight of votes from bad hubs, since these points were considered essentially unreliable. We will refer to this algorithm as hw-kNN.

This was taken a step further in [6], where *class-hubness* was considered instead. The algorithm itself was based on the fuzzy-nearest neighbor framework [7]. So, instead of decomposing the total of k-occurrences into good and bad hubness, it was deemed more beneficial to simply take all the occurrence information into account by treating these occurrence probabilities as fuzzy neighbor votes. A Bayesian alternative to the fuzzy approach was introduced in [8], where a simple, easily extensible Naïve Bayesian framework for probabilistic kNN classification was presented. An information-theoretic approach was the most recent among the algorithms relying on class-hubness for hubness-aware kNN classification [9]. We will refer to these algorithms as h-FNN, NHBNN and HIKNN, respectively.

2.3 HUBNESS IN TIME-SERIES DATA

The hubness phenomenon in time series data in general has recently received some attention [11]. It was shown that, even though time series data do not usually exhibit excessively high dimensionality, it often leads to some tangible hubness. The hubness-weighted kNN algorithm has been thoroughly tested on this data and shown to lead to more accurate classification when combined with the DTW (dynamic time warping) distance [12]. It will, however, become apparent later in this paper that hw-kNN may not be the best hubness-aware approach for nearest-neighbor time series classification, at least in the oceanographic domain.

3 THE OCEANOGRAPHIC SENSOR DATA

In our experiments, we were working with the Integrated Ocean Observing System data (http://www.ioos.gov/). We were analyzing a sample of measurements from many nodes and attached sensors in a period of 20 days in November 2010. Each sensor is monitoring some physical quantity. We analyzed 8 such quantities: air temperature, barometric pressure, wind observation, water level observation, water level prediction, salinity, water temperature and conductivity. The data came from sensors distributed across the coastlines of North America, so it was partly about the Pacific, partly about the Atlantic ocean and also partly about the Great Lakes. These three location profiles we used as the labels for the sensors, thereby dividing them into 3 location-categories. Each physical property was analyzed separately.

There were some missing values in the data, but not much. Out of the total 4801 time points, usually 50-100 was missing, sometimes none. The values were sampled once every six minutes. This means that there was essentially little difference between neighboring points, so we replaced the missing values by the means of the closest known values.

4 EXPERIMENTS

In our experiments we used two distance measures: the Manhattan distance (sum of absolute differences) and the variance of the difference between the two series over time. The basic hubness-related properties for the data are given in Table 1, separately for the two distance measures.

Two of the sensor types were only present in one region (so they were all of the same label) – conductivity and salinity and have hence not been included in subsequent classification tests. We see that both the k-occurrence distribution skewness and the bad hubness are quite similar in both metrics. Most sensor-type data sets exhibit medium skewness, which is consistent with observations from [11], while two measurement types also exhibited quite high hubness, as well as bad hubness in particular: wind and water temperature measurements. The two metrics producing similar results, we will only show the experiments on the Mahnhattan metric in the classifier tests.

Sensor type	size	S_{N3}	BN_3	S_{N5}	BN_5
air temperature	211	0.34	4.7%	0.14	6.7%
barometric pressure	214	0.26	3.4%	-0.06	4.2%
wind	205	3.8	23%	3.6	28%
water level obs.	238	0.6	8.1%	0.47	10%
water level pred.	218	0.34	8.7%	-0.03	11%
salinity	18	-0.13	-	-0.67	-
water temperature	183	0.81	22%	0.67	26%
conductivity	18	0	-	-0.73	-
air temperature	211	0.60	6%	0.55	7.9%
barometric pressure	214	0.11	3.9%	-0.05	4.3%
wind	205	5.2	20%	4.8	24%
water level obs.	238	0.92	9.5%	0.92	12%
water level pred.	218	0.27	6.6%	-0.03	8.9%
salinity	18	0.79	-	0.68	-
water temperature	183	1.16	26%	1.40	31%
conductivity	18	1.01	-	0.81	-

Table 1: *The summary of the data: the number of sensors of a given type, skewness of the 3-NN and 5-NN occurrence distribution (S_{N3}, S_{N5}) and bad hubness of the respective distributions (BN_3, BN_5). The upper part of the table represents results for the Manhattan metric, the lower part for the between-series difference variance.*

It is possible to visualize these bad hubs which are expected to exhibit detrimental influence on nearest-neighbor reasoning in this data. This is shown in Figures 1 and 2, where each sensor was mapped onto a world map according to its latitude/longitude. The size of the circles is proportional to sensor hubness (so, big circles correspond to prototypical, influential points) – and the shade/color to bad hubness (the darker circles corresponding to bad hubs). Figures 1 and 2 represent the two sensor types which were found to exhibit hubness and bad hubness in particular. Two things are apparent from these visualizations. First of all, there are some very big bad hubs. Not only do these sensors exhibit bad hubness, they exhibit it quite frequently. Also, in Figure 1 we see that these big bad hubs are located amidst some low-hubness *good* points, which exhibit no bad hubness. Since both the wind and the water temperature profiles are expected to be similar between these good and bad points, we can conclude that the reasons for bad hubness might even be artificial in this case, like some measurement equipment malfunctioning and producing noisy data. This suggests that bad hubness might also be used for potential error-detection (even though, obviously – not all bad hubs are erroneous data points – and not all erroneous points exhibit bad hubness).

We have tested all the existing hubness-aware classification methods described in Section 2.2 and have compared them to *k*NN as the baseline. The Manhattan metric was used in the experiments. The classification accuracies were obtained via 10-times-10-fold cross validation procedure. Corrected resampled *t*-test was used to test statistical significance. The results for two fixed *k*-values are shown in Table 2. Default parameter settings were used for each of the algorithms.

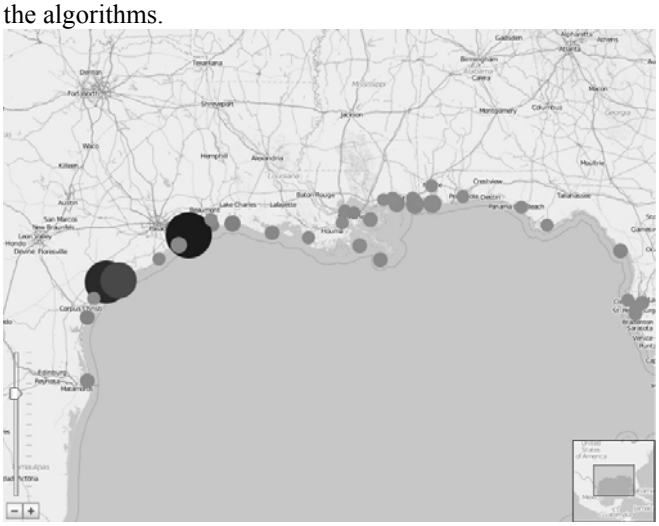


Figure 1: Good/bad hubness of sensors shown on a part of the coastline, based on wind measurements.

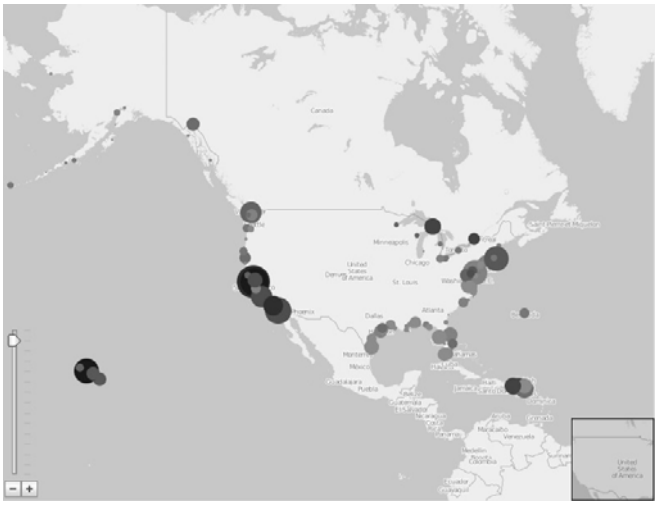


Figure 2: Good/bad hubness of sensors, based on water temperature.

By examining the results in Table 2 we see that the hubness-aware methods prove beneficial precisely in the two outlined high hubness and high bad hubness cases: wind observation and water temperature. HIKNN algorithm achieves the best result in 10 out of 12 cases with the best overall average. We see that hw-*k*NN algorithm did not achieve the best accuracy on any of the data sets. Of course, one has to have in mind that this classification setup is slightly artificial, since we do not really need to classify a sensor into a region – we know that beforehand. These tests were performed in other to test the capabilities of the listed algorithms on this type of measurement data. The results suggest that hubness-aware classification can indeed rectify

the misclassifications which occur as a consequence of the underlying bad hubness present in the data.

Sensor type	*k*NN	hw*k*NN	NHBNN	h-FNN	HIKNN
air temp.	96.8	96.7	96.0	**97.1**	96.9
bar. press.	96.8	97.0	97.0	97.0	**97.1**
wind	75.2	83.6 •	**86.0** •	84.1 •	83.2 •
wat. l. o.	92.6	91.4	90.8	91.8	**93.3**
wat. l. p.	93.3	93.1	92.8	93.7	**94.5**
wat. tmp.	78.6	80.7	81.9	82.0	**83.3** •
air temp.	96.2	96.0	94.0	95.9	**96.2**
bar. press.	96.9	96.5	97.1	97.2	**97.3**
wind	70.6	81.5 •	81.3 •	**82.0** •	**82.0** •
wat. l. o.	91.8	91.2	90.4	92.2	**92.8**
wat. l. p.	90.2	90.9	89.6	91.0	**91.7**
wat. tmp.	77.9	79.2	77.3	80.3	**82.6** •
AVG	88.1	89.8	89.5	90.4	**90.9**

Table 2: *Classification accuracy of kNN, hw-kNN, NHBNN, h-FNN and HIKNN on sensor measurements. The upper half of the table corresponds to k=3, the lower to k=5. The best result in each line is given in bold and the statistically significant results ($p < 0.05$) are marked by* •.

5 CONCLUSION

We have explored some basic *hubness*-related properties of sensor data measured by the Integrated Ocean Observing System. Most of this data was found to exhibit low-to medium hubness, but the wind observations and water temperature measurements were more prone to the emergence of hubs. *Bad hubness* was also present in this data. By visualizing the localization of these bad hubs, it was possible to see that some of the bad hubness might actually be a consequence of erroneous data. Bad sensor hubs of the different measured properties were located at different nodes, in different regions.

Several hubness-aware classification methods: hw-*k*NN, NHBNN, h-FNN and HIKNN were tested on this data and compared to the basic *k*NN method. An improvement in accuracy was observed on the sensor types exhibiting bad hubness. HIKNN seems to be the most promising approach.

6 ACKNOWLEDGEMENTS

This work was supported by the Slovenian Research Agency, the ICT Programme of the EC under PASCAL2 (IST-NoE-216886) and PlanetData (ICT-NoE-257641).

References

[1] E.Fix and J.Hodges. Discriminatory analysis, nonparametric discrimination: consistency properties, Technical report, USAF School of Aviation Medicine, Randolph Field. Texas. 1951.

[2] T.M. Cover and P.E. Hart. Nearest neighbor pattern classification. IEEE Transactions on Information Theory. vol. IT-13. no 1. pp. 21-27. 1967.

[3] L. Devroye, L. Gyorfi and G. Lugosi. On the strong universal consistency on nearest neighbor regression function estimates. Annals of Statistics. pp. 1371-1385. 1994.

[4] K.Q. Weinberger and J. Blitzer and L.K Saul. Distance metric learning for large margin nearest-neighbor classification. *Proceedings of the NIPS conference.* MIT Press. 2006.

[5] M. Radovanović and A. Nanopulous. Nearest-neighbors in high-dimensional data: the emergence and influence of hubs. *Proceedings of 26th International Conference on Machine Learning (ICML)* pp. 865-872. 2009.

[6] N. Tomašev and M. Radovanović and D. Mladenić and M. Ivanović. Hubness-based fuzzy measures for high-dimensional k-nearest-neighbor classification. *In Proc. MDLM 2011, 7th International Conf. on Machine Learning and Data Mining*. New York. 2011.

[7] J.E. Keller and M.R. Gray and J.A. Givens. A fuzzy k nearest-neighbor algorithm. In: IEEE Transactions on Systems, Man and Cybernetics. pp. 580–585. 1985.

[8] N. Tomašev and M. Radovanović and D. Mladenić and M. Ivanović. A Probabilistic approach to nearest-neighbor classification: Naive Hubness-Bayesian kNN. *In Proc. CIKM*. 2011.

[9] N. Tomašev and D. Mladenić. Nearest-neighbor voting in high dimensional data: learning from past occurrences. (under review) 2011.

[10] N. Tomašev and M.Radovanović and D. Mladenić and M. Ivanović. The Role of hubness in clustering high-dimensional data. *In Proc. of PAKDD*. Shenzhen. 2011.

[11] M.Radovanović, A. Nanopulous and M. Ivanović. Time Series Classification in Many Intrinsic Dimensions. *In Proc. SDM.* pp. 677-688. 2010.

[12] H. Ding, G. Trajcevski, P. Scheuermann, X.Wang and E.J. Keogh. Querying and mining of time series data: Experimental comparisons of representations and distance measures. *In Proc. 34th Int. Conf. on Very Large Databases (VLDB).* pp. 1542-1552. 2008.

THE INFLUENCE OF WEIGHTING THE K-OCCURRENCES ON HUBNESS-AWARE CLASSIFICATION METHODS

Nenad Tomašev, Dunja Mladenić
Jožef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia
e-mail: nenad.tomasev@ijs.si, dunja.mladenic@ijs.si

ABSTRACT

Hubness is a phenomenon present in many high-dimensional data sets. It is related to the skewness in the distribution of k-occurrences, i.e. occurrences of data points in k-neighbor sets of other data points. Several hubness-aware methods that focus on exploiting this phenomenon have recently been proposed. In this paper, we examine the potential impact of weighting the *k*-occurrences, by taking into account the distance between the respective data points, on hubness-aware nearest-neighbor methods, more specifically hw-*k*NN, h-FNN and HIKNN. We show that such distance-based weighting can be both advantageous and detrimental and that it influences different methods in different ways.

1 INTRODUCTION

Nearest-neighbor classification methods are often used in machine learning tasks, due to their inherent simplicity and good asymptotic properties. They are based on a notion that similar data points often share the same label. Let $D = \{x_1, x_2, .. x_N\}$ be the training set. 1-NN classification rule is quite simple: given the point of interest x, find the point from D that is closest to x according to some appropriate distance function $d(x, .)$, and denote this point by NN(x); assign the label of NN(x) to the new point x. Due to the common sparsity of data, as well as noisy measurements and records, a generalized *k*NN rule is usually used instead, where a label of a new instance is determined by a majority vote of its k nearest neighbors [1].

Many extensions to the basic algorithm have been proposed over the years, offering modifications of various stages of the classification process. Attribute weighting was successfully used in conjunction with nearest-neighbor classification [2]. Fuzzy approaches were also shown to be quite promising [3][4][5]. Recently, large margin *k*NN was introduced, which achieves accuracy comparable to other state of the art classification methods [6].

Hubness is a phenomenon attributed to high-dimensional data which has potentially severe consequences for nearest-neighbor methods [7]. Denote by $N_k(x)$ the number of k-occurrences of x, i.e. the number of times x appears in k-neighbor sets of other data points. It has been noted that the distribution of $N_k(x)$ exhibits high skewness when the inherent dimensionality of the data is high. This leads to the emergence of *hubs,* influential data points. Hubs have been shown to appear frequently in many kinds of high-dimensional data, like time series, music and images. Several approaches to exploiting the hubness of the data have recently been proposed [7][8][9][10]. Improvement over the basic *k*NN was frequently present.

2 A BRIEF OVERVIEW OF THE USED HUBNESS-BASED NEAREST-NEIGHBOR CLASSIFICATION METHODS

We can distinguish between two sorts of hubs: the *good hubs* and the *bad hubs*, based on the usefulness of their influence in *k*NN classification. Consequently, we define *good hubness* ($GN_k(x)$) and *bad hubness* ($BN_k(x)$) so that $N_k(x) = GN_k(x) + BN_k(x)$, where $BN_k(x)$ denotes the number of label mismatches between x and data points where x appears in k-neighbor sets. Also, denote by $N_{k,c}(x)$ the number of k-occurrences of x in neighbor sets of elements belonging to class c. We will refer to this quantity as *class hubness*. Three types of approaches to exploiting previous occurrences have been proposed voting by label, voting by class hubness and combined voting.

2.1 VOTING BY LABEL

In hw-*k*NN[7], the basic weighted k-nearest neighbor voting framework is retained. Each neighbor votes by its own label and the label weight is determined so as to minimize the influence of bad hubs on classification outcome. The weights are set as $e^{-h_b(x)}$, where $h_b(x)$ is standardized bad hubness. Even this simple weighting scheme was shown to often lead to significant improvements over the basic *k*NN algorithm.

2.2 VOTING BY CLASS HUBNESS

Instead of observing only good and bad hubness, it is possible to take into account class-specific previous k-occurrences, i.e. class hubness [8][9]. The h-FNN algorithm is based on this notion and it integrates class hubness information into a fuzzy k-nearest neighbor voting framework. It uses a threshold to distinguish between low-hubness points (*anti-hubs*) and medium-to-high hubness points where inference based on class hubness is

meaningful. Therefore, it requires a separate mechanism to deal with anti-hubs.

2.3 COMBINED VOTING

Hubness-information *k*-nearest neighbor (HIKNN) [10] is a robust algorithm which uses both the information contained in an instance label and the information contained in its previous occurrences. This approach is based on an information-theoretic perspective, so that the vote of x is shifted more towards using class hubness if $N_k(x)$ is high and more towards the label of x if $N_k(x)$ is low. The algorithm also weights all the individual fuzzy votes based on their total occurrence frequencies, so that more weight is given to anti-hubs, since they are considered more local to the point of interest and, therefore, more important when trying to determine its label.

3 WEIGHTING THE K-OCCURRENCES

Weighted voting in *k*NN helps in implicitly reshaping the neighborhood to give more emphasis to the closer neighbors. Choosing the proper k is far from trivial and sometimes no global neighborhood size gives satisfactory results.

Since this idea is commonly encountered, we wished to see what the effects would be if the same line of reasoning was applied when dealing with inverse neighbor sets. The final voting in h-FNN and HIKNN is also distance-weighted, so introducing some sort of weighting in class hubness calculations does seem somewhat reasonable.

On the other hand, hw-kNN does not employ distance-based weighting. It is based on the simple idea of weighting down the votes of bad hubs. Introducing some weights in the inverse neighbor sets might reduce the bad hubness estimates and increase the voting weights of bad hub points, which may in fact have an overall negative influence on the final classification accuracy. So, intuitively, we would expect to see differences in how these three algorithms change under weighted hubness scores.

We opted for testing a very simple distance-based weighting scheme for calculating class hubness scores. Denote by $NN(x)$ the nearest-neighbor of x. Let $D_k(x)$ be the k-neighborhood of x. We define weighted hubness score of x_i as:

$$WN_k(x_i) = \sum_{x : x_i \in D_k(x)} \frac{d(x, NN(x))}{d(x, x_i)}$$

Weighted good and bad hubness ($WGN_k(x)$ and $WBN_k(x)$), as well as weighted class hubness scores ($WN_{k,c}(x_i)$), are defined analogously to their non-weighted counterparts.

4 EXPERIMENTAL SETUP

For small neighborhood sizes, weighted class hubness calculations would have little to no effect, since all neighbors would be close to the points of interest. For larger neighborhoods, the tendency of some neighbors to be much further away than others is amplified. This is why we chose to run all the experiments for a fixed value of k=30.

In order to reduce the influence of a particular anti-hub handling method in h-FNN, we opted for eliminating any such separate case of neighbor handling by setting the threshold value θ in h-FNN to zero. This means that in h-FNN every instance votes purely by class hubness scores from its previous *k*-occurrences. Since every element is by default included in its own neighborhood, $N_k(x) > 0$ for every x, which avoids the pathological case of zero hubness.

We selected 15 publicly available datasets, 10 from the UCI data repository and 5 from ImageNet repository [13]. 10-times 10-fold cross validation was performed on every data sets for all the algorithms and the corrected resampled *t*-test was used to check for statistical significance. The results are given in Table 1. . The shortened UCI dataset names in the table correspond to colonTumor, vowel, ecoli, parkinsons, sonar, ionosphere, vehicle, segment, isolet, mfeat-factors. Since UCI data exhibits low-to-medium *k*-occurrence skewness [8][9][10], some high-dimensional and high-skewness data was also included in the experiments. Datasets I-s3 to I-s7 represent five different subsets of images and they are composed of 3,4,5,6 and 7 categories, respectively [8]. The images are represented in a hybrid way, by combining the 400-dimensional quantized SIFT representation [11] and a 16-dimensional color histogram representation. SIFT features [12] capture the local information contained in highly textured image parts and are calculated at certain interesting *keypoints* [14]. Color histograms, on the other hand, capture the global color information. In our experiments, these two representation parts are given equal weight in distance calculations.

The Manhattan metric (sum of absolute differences) was used in all the experiments.

It is clear from the results shown in Table 1 that, most of the time, there is no significant difference if the weighted hubness scores are used. On the other hand, when there is a difference between the weighted and non-weighted implementations, it tends to be detrimental in hw-*k*NN and beneficial in h-FNN and HIKNN as expected.

So, in which cases does using the weighted hubness scores improve the classification result in h-FNN and HIKNN? In

order to partially answer this question, we tested the algorithms on the vowel dataset for a range of different neighborhood sizes k = {2,3..30}. The resulting accuracies are shown in Figure 1 and Figure 2.

	*k*NN	hw-*k*NN	whw-*k*NN	h-FNN	wh-FNN	HI-*k*NN	wHI-*k*NN
cTum	72.3	65.4	63.1	62.7	63.4	64.1	64.7
vowel	84.3	57.4	60.3●	62.3	75.4●	78.4	85.4●
ecoli	82.0	85.6	84.4	86.5	86.3	86.3	86.0
psons	90.3	83.3	84.8	84.6	85.3	85.7	85.7
sonar	82.4	73.5	71.5	71.7	72.1	77.3	76.6
ionos	79.7	86.3	82.6	87.5	87.2	87.0	87.6
vehic	61.7	62.4	61.9	59.6	60.6	62.0	62.3
seg	86.4	78.6	78.8	79.6	82.7●	82.9	86.1●
isolet	74.2	85.4	87.5	84.5	85.1	87.2	87.8
mfact	94.5	94.2	93.6	94.0	94.3	94.9	94.9
I-s3	71.2	84.8	80◄	82.7	82.7	84.5	84.4
I-s4	55.4	68.4	63.7◄	63.9	64.1	67.4	67.4
I-s5	45.8	64.3	53.8◄	61.1	61.0	65.4	65.3
I-s6	58.9	70.7	70.2	68.0	68.0	70.9	70.9
I-s7	43.0	63.1	49.9◄	59.2	59.1	62.2	62.1
AVG	72.14	74.89	72.41	73.86	75.15	77.08	77.81

Table 1: *A comparison of classifier accuracies for neighborhood size of k=30. Algorithm implementations using the $WN_k(x_i)$ are given with prefix "w". A filled circle (●) marks those cases where the weighted implementations were significantly better. Inversely, a filled triangle (◄) shows when significant deterioration was observed. Significance level p=0.01 was used in all cases.*

Both vowel and segment (the two datasets where the improvement was observed) are datasets where the classification accuracy deteriorates with increasing neighborhood sizes. The rates of deterioration of *k*NN, h-FNN and HIKNN are not the same, though. The basic *k*NN falls to an accuracy plateu, while the accuracies of the two hubness-based algorithms continue to drop – more steeply in the case of h-FNN. The use of weighted hubness seems to reduce the deterioration rate, resulting in constantly better performance of the weighted implementations, over the entire *k*-range.

If it turns out that this is indeed the only case where one might improve by using the weighted hubness scores, then their usefulness is rather limited, since they do not, in fact, lead to an overall improvement on the dataset, given that the best results are obtained for k = 1 where there is no weighting.

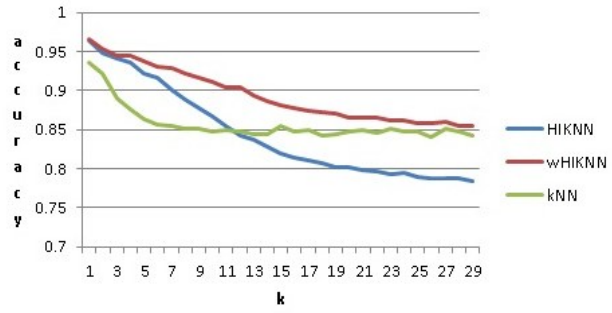


Figure 1: *Accuracies of weighted and non-weighted class hubness implementations of HIKNN for k = {2,3..30} on vowel dataset. The basic kNN is given as a baseline for comparison.*

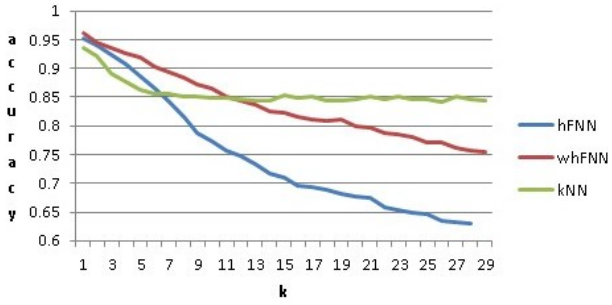


Figure 2: *Accuracies of weighted and non-weighted class hubness implementations of h-FNN for k = {2,3..30} on vowel dataset. The basic kNN is given as a baseline for comparison.*

As for the global influence of weighting on hubness scores, we compared the resulting skewness in *k*-occurrence distributions of weighted and non-weighted occurences. This is illustrated in Figure 3, where the difference between the two skewness values is shown for each of the used datasets. We see that in 14 out of 15 datasets there is a noticeable increase in the *k*-occurrence skewness when weighting is used.

Since any increase in the $N_k(x)$ distribution skewness entails higher hubness of the data, it is clear how this might occasionaly prove beneficial to hubness-based algorithms. On the other hand, even if these algorithms are designed so as to take data hubness into account, this does not imply that, for any specific dataset, they achieve the best performance when the hubness of the data is at the highest point. The observed increase in the skewness may even prove more useful in the unsupervised case, if used for hubness-proportional clustering (HPC) [15].

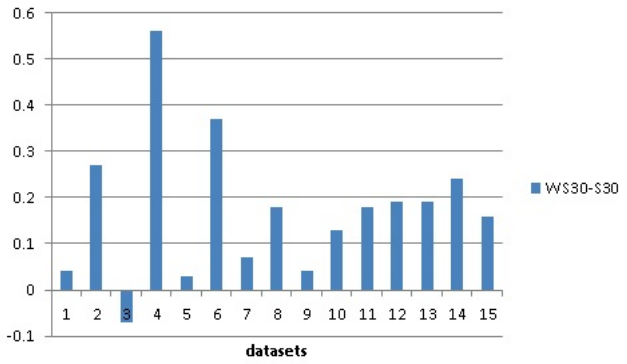


Figure 3: *The difference between weighted and non-weighted k-occurrence skewness for datasets from Table 1, given in the same order (1 – cTum, 2 – vowel, etc.).*

7 CONCLUSION

Data hubness, as a consequence of high inherent dimensionality, is a phenomenon of great importance for nearest-neighbor classification. We have explored how the potential weighting of class hubness scores affects several recently proposed hubness-based algorithms, namely hw-*k*NN, h-FNN and HIKNN. We observed occasional improvements in case of h-FNN and HIKNN, as well as performance deterioration in hw-*k*NN, which is in agreement with our starting hypothesis. We also detected a noticeable increase in *k*-occurrence skewness in the weighted case.

6 ACKNOWLEDGEMENTS

This work was supported by the Slovenian Research Agency, the ICT Programme of the EC under PASCAL2 (IST-NoE-216886) and PlanetData (ICT-NoE-257641).

References

[1] E.Fix and J.Hodges. Discriminatory analysis, nonparametric discrimination: consistency properties, Technical report, USAF School of Aviation Medicine, Randolph Field. Texas. 1951.

[2] E.H. Han and G. Karypis and V. Kumar. Text categorization using weight adjusted k-nearest neighbor classification. *Advances in Knowledge Discovery and Data Mining*. Springer. Berlin. pp. 53–65. 2001.

[3] J.E. Keller and M.R. Gray and J.A. Givens. A fuzzy k nearest-neighbor algorithm. In: IEEE Transactions on Systems, Man and Cybernetics. pp. 580–585. 1985.

[4] R. Jensen and C. Cornelis. A new approach to fuzzy-rough nearest neighbour classification. Rough Sets and Current Trends in Computing. Springer. Berlin. pp. 310–319. 2008.

[5] W. Shang and H. Huang and H. Zhu and Y. Lin and Y. Qu and H. Dong. An adaptive fuzzy knn text classifier. *Computational Science—ICCS* Springer. Berlin. pp. 216–223. 2006.

[6] K.Q. Weinberger and J. Blitzer and L.K Saul. Distance metric learning for large margin nearest-neighbor classification. *Proceedings of the NIPS conference.* MIT Press. 2006.

[7] M. Radovanović and A. Nanopulous. Nearest-neighbors in high-dimensional data: the emergence and influence of hubs. *Proceedings of 26th International Conference on Machine Learning (ICML)* pp. 865-872. 2009.

[8] N. Tomašev and M. Radovanović and D. Mladenić and M. Ivanović. Hubness-based fuzzy measures for high-dimensional k-nearest-neighbor classification. *In Proc. MDLM 2011, 7th International Conf. on Machine Learning and Data Mining*. New York. 2011.

[9] N. Tomašev and M. Radovanović and D. Mladenić and M. Ivanović. A Probabilistic approach to nearest-neighbor classification: Naive Hubness-Bayesian kNN. *In Proc. CIKM*. 2011.

[10] N. Tomašev and D. Mladenić. Nearest-neighbor voting in high dimensional data: learning from past occurrences. (under review)

[11] Z. Zhang and R. Zhang *Multimedia Data Mining*. Chapman & Hall. 2009.

[12] D. Lowe. Object recognition from local scale-invariant features, *Proceedings of the International Conference on Computer Vision.* pp. 1150-1157. 1999.

[13] IMAGENET. http://www.image-net.org/

[14] D. Lowe. SIFT Keypoint Detector. http://www.cs.ubc.ca/~lowe/keypoints/

[15] N. Tomašev and M.Radovanović and D. Mladenić and M. Ivanović. The Role of hubness in clustering high-dimensional data. *In Proc. of PAKDD*. Shenzhen. 2011.

HIGH-COVERAGE EXTRACTION OF SEMANTIC ASSERTIONS FROM TEXT

Mitja Trampuš, Dunja Mladenić
Artificial Intelligence Laboratory, Jozef Stefan Institute
Jamova 39, 1000 Ljubljana, Slovenia
e-mail: mitja.trampus@ijs.si, dunja.mladenic@ijs.si

ABSTRACT

We present an early version of a method for open-domain semantic assertion extraction from natural language texts. To combat the shortage of training data for the task, a two-stage pipeline is employed: we first perform semantic role labeling, then map the resulting frames onto predicate-form, ontology-aligned statements. We chose FrameNet and Cyc as the frame database and the ontology, respectively.

1 INTRODUCTION

In the majority of text mining tasks, algorithms operate at the syntactic level. Most often, the syntactic tokens are simply words. Such a representation is clearly limited in its expressive power: a lot of information is hidden in the interplay of the words. A standard method for accounting for this is the inclusion of word n-grams or other word co-occurrence structures into the data model. This approach, however, is not extensible indefinitely as it hits problems with sparsity.

A different approach to getting richer features from text is the use of parsers and the structural information they output. Thanks to advances in speed in accuracy (Charniak 2005; Collins 2003), this has been a particularly popular approach in the recent years. Because the approach has been proven advantageous in numerous applications, a natural next step is to increase the level of semanticity further and abstract the text into a purely logical form.

A big challenge in trying to perform such a mapping is the high cost of obtaining training data. Because data is further dependent on the choice of formalism, there is no large corpus which could be used for training. Data is, however, available for the task of semantic role labeling (SRL) (Toutanova 2007). At the same time, the FrameNet (Ruppenhofer 2008) collection of frames is quite semantic in nature, causing us to expect it can be mapped to an ontology reasonably well.

We therefore propose an approach for extracting ontology-based predicate assertions (in our case, Cyc) from plain text in two steps, using frame representation as a middle point. In this paper we describe both steps and show some results.

A quick note on notation: we use a sans serif typeface for sample sentences and *italics* for definitions of terms.

2 SEMANTIC ROLE LABELING

The task. Semantic role labeling (SRL) is a well-established text processing task in which the goal is to mark up text with a predefined set of *frames* and *frame elements,* also called *roles*. A *frame* is defined [Fillmore82] as any system of concepts (*roles*) related so that to understand any one concept it is necessary to understand the entire system. Examples of frames are Addiction, Annoyance, Attack, Drinking etc. The latter, for instance, consists of roles Drinker, Fluid, Quantity, Container and perhaps others. There are also some roles that can be included in any frame, e.g. Location, Time, Frequency, Purpose and Manner. Not every occurrence of a frame in natural text needs fill all the roles; for example, the sentence [Paul DRINKER] took a [sip TARGET] of [red wine FLUID] from [the tall glass CONTAINER] and nodded approvingly. omits the Quantity role as well as all target-nonspecific roles. Note that this example uses the standard bracket notation for marking up frames. Also seen in the example is the [… TARGET] role; this is a special role filled by the word that *evokes/triggers* the frame.

Another example of a frame would be BiologicalUrge: [He EXPERIENCER] gave me a [tired TARGET] [shrug EXPRESSOR]. Frames are meant to be language independent.

Resources. There are two large resources available for training automatic SRL systems: FrameNet (Ruppenhofer 2008) and PropBank (Palmer, Kingsbury 2005). We describe and use FrameNet here, though many of he claims and approaches generalize to PropBank as well. FrameNet is a collection of frames and, importantly, frame-annotated sentences from various domains. There are 1020 frames, of which 540 have at least 40 annotated examples and 180 have at least 200. Each frame is also tagged with a list of trigger words (e.g. drink.v, drink.n, sip.v etc. for the Drinking frame). Every frame and every role is defined with a short natural-language definition. Frames are loosely connected with several relations, most notably generalization/specialization. For each pair of connected frames, the mapping between their roles is given as well.

The three stages of SRL. The process of automatic SRL decomposes naturally into three stages: frame identification (which frame is evoked by the sentence?), boundary

detection (which sentence fragments are role fillers?) and role identification (what roles do the role fillers fill?). Although these problems can be solved jointly, it is easier and computationally much more efficient to approach them separately. This does not affect performance: it is intuitively clear that syntactic context should suffice for frame identification, but surprisingly, performing boundary detection and role identification jointly does not bring significant gains either (Gildea and Jurafsky 2002, Erk 2005). Our method thus performs each of the three stages separately as well.

Stage 1. For the frame identification task, we use a recall-oriented simplistic approach. First, we make the standard assumption that frames do not extend over more than one sentence. We then consider the lemmatized version of every word *w* in the sentence *s*. If, for any frame *f*, the lemma *w* occurs in *f*'s list of trigger words, we consider *s* to contain *f*. Some of these decisions are revoked at the later stages if no convincing role fillers are identified for *f* in *s*.

Stages 2 and 3. For role boundary detection, we first perform full constituency parsing of sentences using Charniak's (Charniak 2005) parser. We chose it over other parsers because it has state-of-the-art performance and is open-sourced which allowed us to modify it for online use. We then treat both remaining stages of SRL as classification tasks over the nodes of the parse tree.

We derive the following features for every node:

- Lemma of target word
- Phrase type (= Penn Treebank tag of node)
- Governing category (= parent node's tag; helps distinguish subjects from objects)
- Path from target to node
- Position relative to target (left/right)
- Passive/active voice of sentence. A sentence is considered passive if it contains an AUX↑VP↓VP↓VPN path.
- Lemma of node's head word. The head word is derived using widely adopted rules from Collins (1999).
- POS tag of node's head word.
- Verb subcategorization (= ordered list of children of VP immediately containing the node)

It has been shown that the choice of the classifier is not of critical importance; however, support vector machines (SVMs) are one of the most appropriate choices (Toutanova 2007; Carreras 2005). We use a linear SVM with ε=0.1, C=1/avg($\|x\|^2$) implemented in the svmlight toolset.

For stage 2 (role boundary detection) we use the above features and all of FrameNet's annotated data to classify each node as either *role* or *none*. We then discard all nodes which are classified as *none* with high confidence. The threshold was identified manually so that the pruning has about 95% recall and 55% precision. This significantly speeds up the role identification step and, perhaps even more importantly, greatly reduces class imbalance for the last step. In the role identification stage, we classify all the nodes remaining after the boundary detection stage into one of multiple classes: all the roles belonging to the frame and *none*. There is no clear consensus in the community on the best way to perform multi-class classification in this case, so we follow the recommendation from Hacioglu (2003) and use one-vs-all rather than pairwise classifiers or multi-class SVM.

When combining the votes, it is easy to satisfy the local constraints (each node should be assigned the class voted for with the highest confidence), but we should not neglect global constraints either (most importantly: a role appears only once in a frame, role fillers are strictly disjoint). We therefore employ a constrained greedy algorithm to assign roles. Votes for all nodes and all classes are sorted in descending order of confidence. They are then greedily assigned one by one; if an assignment would violate either of the two aforementioned global constraints, we discard the vote.

Additionally, based on observed algorithm bias towards nodes further from the root of the tree, we adjust the votes somewhat before sorting. Let us denote by $f(v,r)$ confidence of vote for role r on node v. If $f(v,r) > f(v, none)$ and, for some child node v' of v, it holds that $f(v',r) > f(v,r)$, then we set $f(v,r) := f(v',r)$.

Minor issues. To prepare training data, we map FrameNet's annotations (using word-level boundaries) onto parse tree nodes. In great majority of the cases, a perfect correspondence can be found; if, due to errors in parsing or due to a convoluted sentence structure, a perfect match does not exist, we map the role-filler annotation to the leftmost highest node in the tree which is completely contained the in annotation. We noticed that in English, this tends to preserve the semantic head of the role filler.

Akin to most of the existing work, we build a separate set of classifiers for every frame. This could be improved by taking into account that some roles (e.g. Place, Time) are shared across frames.

We limit ourselves to frames that describe actions, e.g. Drinking but not BiologicalState. There are several reasons for this: action frames are more informative, map to Cyc more cleanly and have better annotation coverage in FrameNet. Action frames were identified by having at least one verb trigger word and not more than 10 times as many non-verb trigger words. Of those, we discard frames with no annotated sentences. By hand inspection, we discarded further 20 frames deemed too generic or irrelevant (e.g. Undergoing with the definition "An Entity is affected by an Event."). We are left with approximately 550 frames. We also considered using roughly 100 additional frames where one of the roles generalizes to the generic Actor or Experiencer role, but decided against it for now as their mapping to Cyc is less straightforward (they mostly do not correspond to an #$Event; see section 3.1).

3 MAPPING FRAMENET TO CYC

As discussed in the introduction, our end goal is to obtain a semantic representation of input text. The SRL markup obtained using the method from the previous section, though, marks up syntactic constituents of the sentence. We thus still need to map the role fillers to an ontology. In general, this task is no easier than the one we started out with (mapping whole sentences), because role fillers can be whole relative clauses: for example, for frame Drinking, we can have the sentence [He DRINKER] [drank TARGET] [the strange stink emitting potion she had concocted for him before they left for the journey FLUID]. Mapping the Fluid role onto a set of ontological concepts is clearly no different from the original task. Luckily, it is reasonable to assume that the extra properties about the potion will be identified during analysis of other frames, e.g. Cooking: He drank [the strange stink emitting potion FOOD] [she COOK] had [concocted TARGET] [for him PURPOSE] [before they left for the journey TIME]. and Apperance[1]: He drank the strange [stink TARGET] emitting [potion PHENOMENON] she had concocted for him before they left for the journey.

Our problem therefore reduces to mapping only the semantic head of each role filler. The head is either a noun phrase (potion) or a verb phrase (meet in [He EXPERIENCER] [hoped TARGET(DESIRING)] [to meet her again EVENT].) or a noun phrase. The second case is easy to resolve: verb phrases are almost without exception targets of frames themselves; a verb-phrase head is therefore mapped simply to a whole frame (in the above example, SocialEvent). In the case of noun-phrase heads, we choose to simplify by discarding all adjectival information. This is motivated similarly to our limitation on a subset of frames: information conveyed by adjectives is in general less crucial and has poorer support in ontologies. What remains to be mapped is a very short noun phrase, typically consisting of a single word; in other words, we are left with the task of word sense disambiguation (WSD).

Before mapping the role filler, we of course have to choose an ontology. Ideally, this would be FrameNet as our frames and roles already come from it. However, FrameNet is not a general-purpose ontology; it does cover some entities (presented as frames that tend to require no roles), but for example has no satisfactory mapping for dog or seat. We therefore chose Cyc, an ontology created specifically for purposes like this and containing millions of concepts related to common knowledge.

The choice of Cyc (or any other ontology different from FrameNet) however introduces the necessity to map the frames and roles as well. This problem is known as ontology alignment.

We next describe our approach to both tasks.

[1] A somewhat unfortunately generic frame; taken, however, verbatim from FrameNet

3.1 Mapping Frames and Roles

Conceptually, it makes sense to first align the ontologies for two reasons. First, this is a task that only needs to be done once. Second, it offers support for WSD in that the ontology imposes selectional preferences and constraints on role fillers using its type system. This can aid in the role identification phase of SRL or at least be used immediately after it in a reranking postprocessing step. Our approach currently does not yet make use of this.

(Dis)similarities between the ontologies. Of the numerous concepts found in Cyc, of special interest to us are #$Event and #$BinaryRolePredicate. Specializations of the first are a natural counterpart of FrameNet's frames. Instances of the second are the counterpart of FrameNet's roles. They are connected by the #$rolesForEventType relation which specifies which roles apply to which events. In short, the structure of that part of Cyc is quite similar to that of FrameNet[2]. A majority of frames has a natural counterpart that is a specialization of the #$Event concept in Cyc. We currently discard the frames that do not; those fall in one of the following categories:

- Frame maps to more than one Cyc concept. For example, the frame Respond_to_proposal (with triggers reject, accept, refuse etc.) could map to Cyc's #$Refusing-CommunicationAct, #$Accepting-CommunicationAct, #$Rejecting-CommunicationAct and some others, but their only common generalization is #$CommunicationAct, which is too general. About 5% of frames are like this.
- Concept does not exist in Cyc. For example, Adjusting (triggers: adjust, tweak, calibrate, …). This does not necessarily mean the notion cannot be expressed in Cyc, but it would require a non-atomic expression. About 2% of frames fall into this category.
- About 2% of the frames map to relations rather than specializations of #$Event. For example, Evoking maps to the relation (#$evokes ARG1 ARG2) where ARG1 is an instance of #$Individual and ARG2 of #$FeelingAttribute.

With a moderate amount of additional work, frames from the last two categories could be accommodated as well, meaning that 95% of the frames we consider have a natural counterpart in Cyc. This supports our decision to use FrameNet for an intermediary representation of information.

It has to be noted, however, that not all mappings are perfect. In particular, we are sometimes forced to ignore certain subtleties in frame definitions. Consequently, several FrameNet frames might get mapped to the same Cyc concept. An extreme example of this is the #$Evaluating concept which is mapped to by Trying_out, Labeling,

[2] and it would be very reasonable to perform SRL directly using Cyc as the frame ontology, were it not for a complete lack of training data.

Regard, Judgment and Assessing. Another typical example of conflated frames are frame pairs of the form Cause_to_XYZ and XYZ. We map pairs like this to the same Cyc concept, but with different role mappings.

Semi-supervised mapping of frames. There are about 550 frames to be mapped and about 2000 roles. While the best accuracy would certainly be achieved by mapping by hand, this is prohibitively time-consuming. On the other hand, automatic approaches have few reliable features and no training data, so a completely unsupervised approach is also unrealistic. We opt for a semi-supervised scenario where an algorithm proposes several possible mappings and a human annotator chooses the best one among them.

When aligning ontologies, there are, broadly speaking, two types of features available: content-based, stemming from the properties of the nodes themselves (typically, glosses or sample instances), and structural. In our case, aiming at aligning the two ontologies structurally does not make sense as the two have different levels of granularity and coverage. We therefore make use only of the glosses and English denotation strings of entities in both ontologies.

When mapping frames, the trigger words provided with each frame prove to be much more valuable than the frame descriptions. Our method suggests for each frame all the concepts that have at least one of the trigger words of the frame listed as their English denotation. It also suggests all the common ancestors of these initially collected Cyc concepts in the generalization taxonomy: for example, the frame Inchoative_change_of_temperature is associated, among others, with trigger words chill, cool and heat. In Cyc, cool is not associated with any concept (English annotations are lacking), chill is associated with #$Chilling and heat is associated with #$HeatingProcess. One of their common ancestors is #$TemperatureChangingProcess, which is the right mapping for the frame in question.

To ease the annotator's job, the suggested Cyc concepts are ranked according to the number of FrameNet trigger words that map to them and their depth in Cyc taxonomy (more specific is better). However, the ranking did not prove essential as the number of suggestions is typically low. We therefore did not experiment with more complex ranking approaches based e.g. on similarities of glosses.

Automatic mapping of roles. Even with the semi-supervised approach, the time investment for mapping roles would be too large given their number. We therefore perform the mapping automatically, based on heuristics only. To increase accuracy, we only map the core roles[3] of each frame. This corresponds to roughly 80% of roles appearing in natural text. In Cyc, we do not have such information and therefore consider all roles; however, we discard those for which a more specific role (according to role hierarchy) is available as well.

To determine role similarity, we use the glosses and subject/object information. From glosses, a bag of words vector is constructed (with tf-idf weighting, Porter stemming and a stopword list). By subject/object information, we mean that the two most important roles tend strongly to be the subject and the object. For all Cyc roles, it is possible to infer (using role hierarchy) what the subject and the object are, if any. For FrameNet roles, a similar inference is sometimes possible (the hierarchy is much less principled and populated); when hierarchical info is unavailable, we heuristically assume that the first role to be listed is the subject with probability 0.7 and object with probability 0.3; and the other way around for the second role listed. For roles that have been identified as subjects or objects, this is added as an extra component to the sparse vector.

We define role similarity as the cosine between the two length-normalized vectors. To obtain the best global assignment, we create a bipartite graph of roles and weigh every edge connecting two roles r and r' with

$$w(r,r') := d(r,r')^{0.5}$$

where d is the cosine similarity between the feature vectors. We then use Hungarian method to find the maximum-weight assignment. The square root was introduced to further decrease the "greediness" of the method (propensity to choose the highest-scoring pair regardless of others). Another possible regularization is logarithmic (treating similarity scores as probabilities; the probability of the global assignment is then the product of pairwise probabilities, i.e. the sum of logarithms). We have also experimented with a few naïve greedy approaches, but found their performance to be worse.

In the above approach, we assume that no two roles from FrameNet map onto a single role in Cyc. It should be noted that this can be problematic. Especially for actions with "symmetric" roles, FrameNet assumes a somewhat confusing notation: for example, the frame Meeting contains roles Party_1, Party_2 and Parties. Some frame occurrences fill the first two roles and others fill only the third role – depending on the phrasing. In Cyc, all of these correspond to a single role (which may then have two distinct fillers).

3.2 Mapping Role-Fillers (WSD)

Identifying the head. For a role filler, we first identify its semantic head. This is different from the syntactic head used in the feature construction stage of SRL, so we derive a separate set of simple recursive rules. For NP nodes, descend into the rightmost noun-like child. For PP nodes that start with a preposition, descend into the child immediately following it. For S nodes, descend into the last verb (phrase). For VP nodes descend into the first verb (phrase). If no rule applies, stop.

Choosing among mappings. For mapping role fillers to Cyc, we use Cyc's built-in #$termStrings predicate which

[3] This is a FrameNet concept. Core roles are those that either have to be appear explicitly or their filler is implicitly understood from the context. A frame typically has two to four core roles.

connects concepts and English words. Often, a single word maps onto multiple concepts. At the task of WSD, simply mapping to the most common interpretation for the word will give an extremely strong baseline. Unlike WordNet, Cyc unfortunately has no "most common sense" information associated with each word. It does, however, have links from its concepts to WordNet. Although created semi-automatically and not of perfect quality or coverage, they allow us to rank all the Cyc concepts suggested by #$termStrings using commonness information from their WordNet counterparts. The highest ranking concept is then selected. If there are multiple highest-ranking concepts or if there is no WordNet information available due to absence of links, we give priority to the concepts first returned by the Cyc inference engine.
This is a very simple approach; we plan to later integrate a separate Cyc WSD engine currently being developed at our department.

4 RESULTS

SRL. To make our system more comparable with existing ones, we only measure performance on 10 frames, training on 300 annotated sentences. We achieve precision 56% and recall 61%. While we acknowledge that these results are lower than the state of the art (F1 in high seventies (Litkowski 2004)), there is also clear room for improvement. We expect our decision not to use out-of-the-box SRL packages to prove beneficial when we improve the pipeline as a whole and increase coupling between the SRL and the ontology alignment phase.
As described, we do not perform frame identification beyond trigger keyword matching, so we cannot comment on its performance.

Framenet-Cyc alignment. In this step, it makes little sense to compare ourselves with existing contributions to the field as the achievable performance depends highly on the actual ontologies we are trying to align.
For frame alignment (without the roles), we used a human annotator as described in section 3.1. There was only one annotator, so inter-annotator agreement has not been measured.
To evaluate the role alignment step, we manually inspected all 83 core roles in 25 randomly selected frames successfully mapped to Cyc. Accuracy is 35/83=42%; a perfect mapping could achieve at most 64/83=77% on this sample since for some frames, the corresponding Cyc concept is not associated with enough roles. We do have to note that mapping accuracy on the subject and object roles is higher, and because real-world sentences use these two roles more than others, the error rate introduced will be somewhat better than what the 42% above suggest.

Role filler alignment (WSD). Based on manual inspection of 50 role fillers, we estimate that the semantic head of the role filler is correctly identified in 78% of the cases. Mapping of role fillers to Cyc is correct in 60% of the cases in which the semantic head is identified correctly and thus in 48% of the cases overall. In this count, we ignore the pronouns he, she, her, him and his which are mapped to the generic #$Person concept or its gender specializations with hand-written rules.

Overall performance. It is very hard to estimate the recall of the complete pipeline or indeed even of SRL alone as there is no strict enough definition of what a frame is. If, however, for the sake of evaluation, we assume that FrameNet has perfect coverage, the recall of the pipeline at the frame level (i.e. frames successfully identified and mapped onto Cyc with at least one role) is about 65%. For intra-frame performance, refer to the previous paragraph.
As an illustrative example, we are attaching an excerpt from a newspaper article along with the automatically extracted facts. The text:

To understand and appreciate the Bush administration's policy regarding Israeli Prime Minister Sharon's disengagement plan, we must briefly reexamine the record. For three and a half years now, the administration's attitude toward the Israeli-Palestinian conflict/peace process has been characterized by high rhetoric and little action. On the one hand, President Bush is the first US leader to officially endorse the creation of a Palestinian state.

Facts from the first sentence:

```
(#$objectImproved #$Comprehending* #$OrganizationPolicy*)
(#$performedBy #$Comprehending* (ObjectDenotedByFn "we"))
(#$evaluationInput #$Evaluating* #$OrganizationPolicy*)
(#$performedBy #$ExercisingAuthoritativeControlOverSomething*
(ObjectDenotedByFn "we"))
(#$performedBy #$PurposefulAction* (ObjectDenotedByFn "Sharon"))
```

Facts from the second sentence:

```
(#$eventOccursAt #$DescribingSomething* #$Attitude*)
(#$senderOfInfo #$DescribingSomething* #$Action*)
(#$performedBy #$ExercisingAuthoritativeControlOverSomething*
(ObjectDenotedByFn "constitutes"))
```

Facts from the third sentence:

```
(#$performedBy #$Siding-SelectingSomething #$Bush*)
(#$doneBy ArrivingAtAPlace #$Bush)
(#$communicatorOfInfo #$Communicating #$Bush)
```

Some facts are very sensibly extracted (the first sentence does particularly well) while some of them are highly erroneous (e.g. most from the second sentence, or president Bush being mapped to #$Bush, the garden bush concept). A word on notation: With a star, we denote here "an instance of collection": for example, #$Dog is specified in Cyc to denote the collection of all dogs, so we use #$Dog* in the example above to denote a specific instance of #$Dog. In actual program output, this is denoted with multiple predicate statements and using #$isa. The notation (#$ObjectDenotedByFn "foo") means a concept Cyc does not know about, but is expressed in English as "foo".

5 RELATED WORK

SRL methods are well researched and numerous. Their basic design is unchanged since the first reported attempt at SRL (Jurafsky 2003). A basic preprocessing step is constituency parsing (although a few rare examples opt for chunking or other shallower methods (Punyakanok 2004). This gives rise

to most of the features; feature engineering was shown to be very important (Toutanova 2006). The problem is then typically divided into role detection and role identification steps; both are almost always performed using classic ML techniques. The best insight into SRL is offered by various challenges (Litkowski 2004; Carreras 2005; Ruppenhofer 2010).

The task of semantic fact extraction is much less researched. The better-known systems aim for high precision; this means that they only search for a limited number of relations and even within those do not focus on recall. TextRunner (Banko 2008) is is an example of such a system, though it does not completely meet our criteria in that the relations and entites it extracts are stil represented as textual strings. SOFIE (Suchanek 2009) is a recent system that performs ontology alignment as well. High-recall, general-domain oriented fact extraction has been attempted by Rusu (2009) by focusing on subject-verb-object triplets. The output is textual. Role filler alignment corresponds largely to the task known as Word Sense Disambiguation; refer to Navigli (2009) for a recent survey.

6 CONCLUSIONS AND FUTURE WORK

As demonstrated by the evaluation, automated fact extraction still has a long way to go. Because of the large gap between the textual and purely semantic representation, it is almost inevitable for approaches to employ long pipelines. While it is possible to achieve reasonable accuracy at each individual step, the pipeline length means a large number of errors accumulates. We believe this would remain a problematic factor even if we improved our individual methods – for which there is ample room. The solution is most likely in merging pipeline stages; in our approach, for example, the role labeling and Cyc mapping could be done in a more intertwined fashion, allowing the two to correct each other.

In the future, we would like to explore SRL based on simpler structural features, e.g. chunker output. There are several motivating factors for that. First, in the context of domain independence, full-parse features are problematic (Huang 2010, Croce 2010) because parsers are typically trained on the Penn Treebank (= annotated Wall Street Journal articles) and do not generalize well to other domains; SRL, in turn, shows high dependence on parser accuracy. Second, full constituency parsing is still quite slow and third, they are available for fewer languages. Last but not least, the simplification in features seems to affect performance by only about 2% (Surdeanu 2007).

Also related to SRL, we would like to explore ways of automatically increasing the amount of training data. We still see the lack of data as a major impediment; most papers and challenges on SRL limit themselves to only the few best-annotated frames.

As the primary motivation for extraction of semantic assertions is their further potential utility in text mining tasks, we plan to test their usefulness in this manner as well. In particular, because we chose an ontology with relatively large amounts of background knowledge and good inferencing capabilities, we would like to explore the value of facts inferred from the ontology.

7 ACKNOWLEDGMENTS

This work was supported by the Slovenian Research Agency and the IST Programme of the EC under PASCAL2 (IST-NoE-216886) and RENDER (ICT-257790-STREP).

References

Banko, M., Etzioni, O., Center, T.:: The tradeoffs between open and traditional relation extraction. In: Proceedings of ACL 08, pp. 28–36. Citeseer, (2008)

Carreras, X., Màrquez, L.:: Introduction to the CoNLL-2005 shared task: Semantic role labeling. In: Proceedings of the Ninth Conference on Computational Natural Language Learning, pp. 152–164. Association for Computational Linguistics, (2005)

Charniak, E., Johnson, M.:: Coarse-to-fine n-best parsing and MaxEnt discriminative reranking. In: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, pp. 173–180. Association for Computational Linguistics, (2005)

Collins, M.:: Head-driven statistical models for natural language parsing, PhD-thesis, University of Pennsylvania (1999)

Collins, M.:: Head-driven statistical models for natural language parsing. Computational linguistics 29, 589–637 (2003)

Croce, D., Giannone, C., Annesi, P., Basili, R.:: Towards Open-Domain Semantic Role Labeling. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pp. 237–246. Association for Computational Linguistics, (2010)

Hacioglu, K., Ward, W.:: Target word detection and semantic role chunking using support vector machines. Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology companion 25-27 (2003)

Huang, F., Yates, A.:: Open-domain semantic role labeling by modeling word spans. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pp. 968–978. Association for Computational Linguistics, (2010)

Litkowski, K.:: Senseval-3 task: Automatic labeling of semantic roles. In: Senseval-3: Third International Workshop on the, pp. 2-5. (2004)

Navigli, R.:: Word sense disambiguation: A survey. ACM Computing Surveys (CSUR) (2009)

Palmer, M., Gildea, D., Kingsbury, P.:: The proposition bank: An annotated corpus of semantic roles. Computational Linguistics 31, 71–106 (2005)

Punyakanok, V., Roth, D., Yih, W.-tau.:: The necessity of syntactic parsing for semantic role labeling. In: International Joint Conference on Artificial Intelligence, pp. 1117. Citeseer, (2005)

Punyakanok, V., Roth, D., Yih, W., Zimak, D.:: Semantic role labeling via integer linear programming inference. In: Proceedings of the 20th international conference on Computational Linguistics, pp. 1346–es. Association for Computational Linguistics, (2004)

Ruppenhofer, J. et al.:: FrameNet II: Extended Theory and Practice (2006)

Ruppenhofer, J. et al.:: Semeval-2010 task 10: Linking events and their participants in discourse. In: Proceedings of the 5th International Workshop on Semantic Evaluation, pp. 45–50. Association for Computational Linguistics, (2010)

Rusu, D., Fortuna, B., Grobelnik, M., Mladenic, D.:: Semantic Graphs Derived From Triplets With Application In Document Summarization. Informatica Journal (2009)

Suchanek, F.M., Sozio, M., Weikum, G.: SOFIE: a self-organizing framework for information extraction. In:Proc. of the 18th international conference on World wide web, pp. 631–640. ACM, (2009)

Surdeanu, M., Màrquez, L.: Combination strategies for semantic role labeling. Journal of Artificial Intelligence (2007)

Yih, S.W., Toutanova, K.:: Automatic semantic role labeling. In: Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Tutorial Abstracts on XX, pp. 309–310. Association for Computational Linguistics, (2006).

FIFTY WAYS TO DETECT A GHOSTWRITER

Katerina Zdravkova
Faculty of Computer Science and Engineering
University Ss. Cyril and Methodius in Skopje
Rudjer Boskovikj bb, 1000 Skopje, Macedonia
e-mail: katerina.zdravkova@finki.ukim.mk

ABSTRACT

Ghostwriting became students' most popular way to avoid writing of boring essays, or the best way to easily earn by writing on behalf of another student.

This paper presents several markers indicating a presence of potential ghostwriters. Proposed methodology suggests various inspection techniques, which do not prove anything in isolation. Whenever they are jointly implemented, they successfully cluster the essays, suggesting plausible absence, potential and almost certain presence of one or few ghostwriters. After the initial clustering, all the papers go through subtle linguistic check. In our sample, later approach discovered some unexpected phrases which confirmed the presence of the same ghostwriters not only in the current, but also in the previous generations.

1 INTRODUCTION

Massive storage technologies and search engines immensely provide and facilitate information access, and at the same time they enable a smooth and undoubted detection of most plagiarism sins. It is no longer a problem to discover students' naïve 'copy and paste' activities using search engines. When the original source is written in another language, Google Translate provides help for both, the students, and the teachers who easily capture the machine translated and usually unedited parts of essays. Even fairly translated essays are easily noticed, because students usually translate articles using an extremely professional writing style. Teachers can always use plagiarism detecting tools such as the famous iThenticate, Turnitin, or WriteCheck, (recommended by [1]) capable of comparing essays with the databases of stored texts.

In the recent years, writing essays, papers, and even theses has become a very popular and frequent activity. Magnificent article [2] reveals the presence of extremely well-paid professional writers such as Ed Dante, who "completed 12 graduated theses of 50 pages or more". Many ghostwriters work on their own, while others are organized by specialized agencies called essay or paper mills [3]. Chinese estimate "that university students spend up to half a billion yuan ($73 million) a year to have other people write their essays" [4]. Most of their works are still academically very inexperienced, and the presence of classical 'cut and paste' plagiarism is usually abundant. Unlike them, ghostwriters in America, Australia and Europe seem to be highly professional. They generate impeccable works with no evident plagiarism, and as aka El Dante claims in The Chronicle Review [2], they are all "based on specific instructions provided by cheating students". It seems that essay mills are on great demand in the academic world, and scholarly mercenaries daily finish many extremely ambitious tasks. Although not intended to support human writing, there are some authoring tools, such as GhostWriter [5, 6], which can facilitate the preparation of different contents. This case-based reasoning system can effectively support content authors suggesting them feature values.

If computer facilitated preparation of written products becomes reality, the opposite direction is not so prosperous. In spite of the presence of many plagiarism detecting tools, ghostwriting is rarely detected and almost impossible to prove. In the recent years it has become a lucrative business and young students or academicians willing to apply it are 'sprouting up like mushrooms after the rain'. As a consequence, procedural concerns grow, and one proposed solution in USA is to expand the federal rules to diminish its side effects [7]. Wherever the legislation is not prepared to handle with the problem, concerned faculties minimize the contribution of individual essays in the final grade [8].

Experienced teachers usually intuitively feel the cheat. Unfortunately, they have no means to prove it with indisputable certainty. Students always have an accurate and very rational excuse for all teachers' accusations. The only prove that the essay was not individually prepared is student's inability to tell what is written in it. But, very few teachers have the courage to find material evidence of the cheat, and time to personally enquire the student.

This paper presents several markers indicating the presence of potential ghostwriters who have prepared many essays on related topics for the same course over years. Second section is dedicated to the most obvious indicators derived from document properties and student activities in the learning management and storage system. Third sections presents the techniques connected with IP addresses through which students accessed the desired activity. The approach can be even more effective if the system has a track of all the previous accesses of all the students. After finishing the external inspection, the text itself becomes a target of a subtitle text and linguistic mining. These mining techniques are described in the fourth section.

Before presenting the final conclusion and further work, the results of the estimation of proposed approach are submitted. They are based on a sample of 185 short essays distributed into 5 groups of related topics. Although the correlation of proposed indicators is negligible, there are many exact matches between some of them. They confirmed the existence of several works done by another student, and suggest that there could be an anonymous ghostwriter. Before more reliable indicators are invented, the claim that a work was not done by the student are still a speculation.

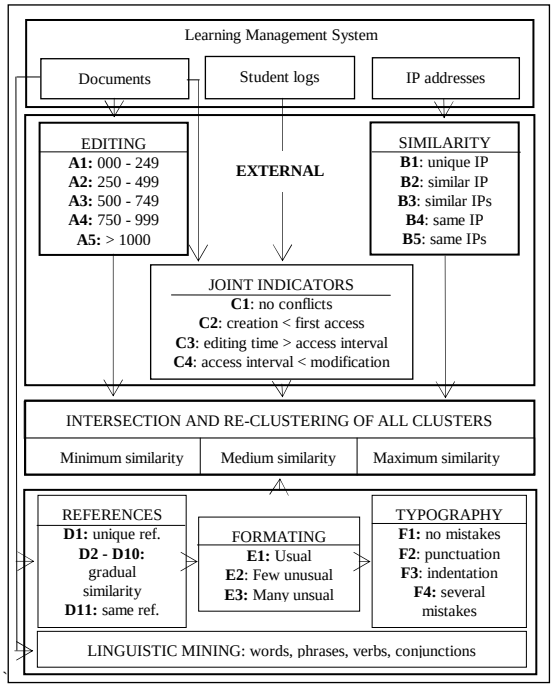


Figure 1: *System architecture*

2 EXTERNAL INDICATORS

External indicators are the information that can be extracted without looking into the contents of the essay. In order to be able to extract them, let's presume that:

- all the essays are stored and delivered as documents
- the course is maintained with a learning management system (or LMS) capable of reporting all the activities.

The document itself contains information such as: title of the paper, name of the first author, name of the user who last saved it, time of paper creation, revision number and total editing time. Only very naïve or extremely imprudent students deliver a document revealing initial creation in the past, no editing time, or a name of another colleague. But, they always have an acceptable excuse ("I used the template from last year.", "I use OpenOffice and saved it as Word.", "My computer is out of order, I went to my colleague".)

Traditional ghostwriters deliver the final essay without influencing the introductory activities prior to final upload of the essay. It seems that they deliver the essay as part of a mail message [9], which students simply copy into a newly opened document and deliver it with no editing time. More careful students use older documents or convert into pdf.

On the other hand, LMS reports offer information concerning the moment of first access of observed activity, how many times and when was it accessed, when was the document uploaded, how many times and how long has the student viewed the activity after final uploading of the essay, how many times and how long was the activity visited, and when was it last viewed. Using these times, time between first upload and first access and time between last upload and last access can also be calculated.

Having a long teacher's experience, the author of this paper can state that students who submit other's outcome usually access the definition of the task at most once, but afterwards regularly check it, sometimes upload it again, and eagerly wait for the final grade. Again, whenever someone is accused of uploading an essay 'borrowed' by another colleague, there are many excuses. The most frequent is: "We jointly prepared the essay".

There are at leats these joint indicators which can be derived from document properties and activity reports:

- the difference between time when the document with the essay was created and the first access of the definition of its topic,
- the difference between document total editing time and the difference between first uploading and first access,
- the difference between final uploading and document last modification.

They should all be positive. Surprisingly, we always discover few cases with negative differences. Predictably, students offer an excuse that they edited an earlier document rather than creating a new one. And, their excuse is very plausible.

All the external indicators are very useful to catch a student who delivers an essay prepared by a colleague from own generation, or from another student of previous generations. They can discover the presence of a ghostwriter who only delivers the final version of the essay. However, there are even more sophisticated ghostwriters. They take student's ID and password in their possession, and behave on student's behalf. ID fraudsters never make any mistake measurable with external indicators. Whenever they decide to commit an ID exchange, there is absolutely no proof that they exist. In order to catch them, additional techniques must be implemented.

3 INDICATORS BASED ON IP ADDRESSES

Most learning management systems are able to keep track of all users' IP addresses throughout all the courses. The first indicator is a comparison of all IP addresses participants use during the activities concerned with the essay. They are downloaded in a separate worksheet.

IP indicators are defined as follows:

- all the records are numbered consecutively according to the time of their access
- the whole worksheet is sorted according to IP address
- IP addresses during essay upload are separated in an individual set

Each IP address is labeled using following equations:

$$Label(IP_i) = n_i$$

where i is the consecutive index of IP address, and n_i is the number of students approaching to the activity from same IP address.

Faculty IP addresses are excluded from labeling, because all the students can access from any computer in the students' laboratories.

Particular attention is paid to IP addresses during upload. If a student uploaded the essay from an IP address with a label greater than 1, its IP value is doubled. This correction weakens the influence of joint work to further strict clustering.

After IP labeling, each student ID is also labeled

$$Label(ID_j) = \sum_{k=1}^{m_j} \frac{\ln(Label(IP_j))}{m_j}$$

where j is student's order in the course, and m_j is the number of IP addresses assigned to j^{th} student with $IP_j > 1$.

At the end, each student ID is included into one of five clusters.

$$Cluster(ID_j) = \begin{cases} 0{,}00 & 0 \le ID_j < M/5 \\ 0{,}25 & M/5 \le ID_j < 2M/5 \\ 0{,}50 & 2M/5 \le ID_j < 3M/5 \\ 0{,}75 & 3M/5 \le ID_j < 4M/5 \\ 1{,}00 & 4M/5 \le ID_j \le M \end{cases}$$

where M is the maximum average value of IP addresses labels assigned to all students. Students belonging to the first clusters are those who seem that prepare their essays individually.

Students belonging to the last cluster are checked most thoroughly. Whenever they belong to the same cluster in other course activities, they are also checked at other courses. The match is perfect. Furthermore, the access to same IP address is simultaneous. But, controversial IP addresses usually belong to student dormitories. Students usually insist on collaborative work, which is stimulated, rather than punished.

4 INTERNAL EXAMINATION

After opening the document itself, the exactness of the defined assignment topic with the prepared essay is checked. There are very seldom mistakes, but they always reveal deliberate swap of the topic enabling a fake or 'collaboration'.

4.1 References

It has been noticed that some very special references appear in several student essays. Therefore, references are considered a valuable ghostwriter indicators.

References are labeled using the same strategy as IP addresses. Before the reference labeling, all the references are subtracted from the texts and each reference is assigned to the student. They are sorted and labeled using exactly the same formulas presented in the left column of this page. Very popular sites, such as English Wikipedia, or popular aggregators of ICT news are excluded from labeling.

At the end, each student ID is again included into a corresponding reference cluster. As far as the number of references is usually very high, clusters can be more refined. We propose 11 clusters. Students who used very special references belong to the cluster with value 0, while those who used the same references as their colleagues belong to the cluster with value 1. The most suspicious in the light of ghostwriters are those students who do not belong to extreme clusters. Namely, the person preparing several essays on the same topic collects a limited number of references and carefully divides them into almost disjunctive sets. But, they have many things in common, such as the language of the original reference, news aggregators.

4.2 Formatting styles

It has been noticed that most essays are written using normal formatting style. However, unusual styles such as short_text, long_text, long_text + Arial, apple_style_span, apple-converted-text, or yellowfadeinnerspan appear in several essays. At the moment, we do not have an application capable of rearranging texts according to their styles. Therefore, we have manually distributed students into groups according to the most frequent styles. And again, those students who belong to same clusters were joined together. It can be a coincidence, but also an indicator of dishonest student behavior.

4.3 Typographic similarities

Recent trial about Facebook ownership includes several checks with techniques belonging to linguistic forensics [10]. The first two (apostrophes and suspension points) are typographic. Students usually make many typographic mistakes. They:

- forget to put a space after the punctuation,
- indent the line by adding several spaces,
- add a point after reference bracket, although teacher example excludes it

In absence of an application dealing with typographic mistakes and similarities, simple replacement with highlighted text is very useful. Essays are again divided into clusters according to the type of highlighted replacement.

Believe it or not, some students have already been united in several of these clusters. They had too many similarities, including the way of signing the paper, so it was evident that their effort was either joint, or done by few of them. However, these student cheatings are not as severe as the presence of an unwanted author who is gaining profit.

4.4 Linguistic similarities

The best way to catch a ghostwriter is to compare writing styles in all essays. One interesting approach is offered by Rong Zheng et al. who are dealing with authorship identification of online messages [11]. Apart from proposing their own framework, they also offer a comparison of previous studies in authorship identification.

The main reason to start chasing the ghostwriter was an essay with 20% identical titles. In all of them, the subtitle 'used sources' or 'literature' preceded the references.

All the essays were processed separately. The crucial elements were:

- frequency of the words and short phrases consisting of at most five words
- frequency of the most frequent verbs
- frequency of conjunctions

After examining these markers, it appeared that many students used the verbs: exist, create, and select. Some of them used either blessing or some religious phrases. Furthermore, the frequency of conjunctions of these students was higher than regularly. And, at the end, all of them were either with no editing time, or in pdf. The most interesting is the fact that all these students belonged to some of medium clusters.

5 EXPERIMENTAL RESULTS

The effectiveness of proposed approach was tested over a pool of 185 short student essays dealing with assistive technologies. They were first tested on plagiarism. Only three essays contained literal copies of texts found on the Internet. Another four essays contained unedited Google translations. They were excluded from further inspection.

Major external indicators were calculated and correlated mutually and with the essay grade and course final grade. There was a small correlation between editing time and essay final grade, very high correlation between number of views and views before upload. Most factors were close to zero. All of these proved nothing in particular.

Joint indicators derived from document properties and activity reports were much more sensitive to potential uploading of other's intellectual property. They located several students who knew nothing about the contents of the essay. At least one goal was accomplished, few cheaters were uncovered.

The best results were obtained using the clustering. As mentioned before, all the internal clusters were consisting of the same students. Students suspected of using ghostwriter services were always stuck together. This fact might be a proof that such a person existed, but he or she was not discovered.

In order to verify the cheat, typographic and linguistic check were performed over essays from previous generations. The arguments didn't exist few years ago, but their presence was noticed for the first time two years ago, first moderately, nowadays easily noticed.

6 CONCLUSION AND FURTHER WORK

This paper presented the attempt to uncover the dilemma, is there a ghostwriter among students. Almost fifty indicators were established to prove that such a person, or may be several of them exist. These indicators were very successful to catch students committing harmless fakes and cheats. They have also revealed that students are sometimes grouped together and jointly prepare their assignments. Whenever they are capable of presenting the contents of the essay, their fault or sin again remains unpunished.

All the indicators were sensitive to small student deficiencies. But, the professional outsourcer was never caught in the net. The crucial evidence of his or her presence were linguistic similarities.

We have already started the creation of a plagiarism tool intended to integrate student essays with search engines, Google Translate, and the pool of previous essays. The tool will be soon enlarged with ghostwriter detector. We do hope that it will discover the cheat and reduce it to the level of previous years.

References

[1] Plagiarism in a Digital Age: Voices from the Front Lines: What's Happening in High Schools Today? http://www.plagiarism.org/plag_webinar_high_schools.html

[2] P. Davis. The Ghowstwriter Behind Student Papers. Society for Scholarly Publishing, The Scholarly Kitchen, http://scholarlykitchen.sspnet.org/2010/11/18/the-ghostwriter-behind-student-papers/#comments

[3] Essay Mill, http://en.wikipedia.org/wiki/Essay_mill

[4] Medeiros, I., "Education in China: Ghostwriting hits frightening levels at universities", Thoughts on Design, Technology and Culture, http://designative.info/2010/03/26/education-in-china-ghostwriting-hits-frightening-levels-at-universities/

[5] D. Bridge, A. Waugh. Using experience on the read/write web: The GhostWriter System. Proc. of WebCBR, Reasoning from Experiences on the Web, Working Prorgamme at the 8th International Conference on Case Based Reasoning. 2009.

[6] D. Bridge, A. Waugh. An Evaluation of the GhostWriter System for Case-Based Content Suggestions. http://www.cs.ucc.ie/~dgb/papers/Waugh-Bridge-2010.pdf

[7] J. P. Justman. Capturing the Ghost: Expanding Federal Rule of Civil Procedure 11 to Solve Procedural Concerns with Ghostwriting. *Social Science Research Network. Minnesota Law Review*, Vol. 92, p. 1246, 2008

[8] I. Trajkovska. "I prepare seminar papers" http://www.vecer.com.mk/?ItemID=2CEA0A6FC904DB4B8E0664391C49478B. In Macedonian

[9] K. Zdravkova. Can Web 2.0 Reduce Plagiarism and Cheating. *The 8th International Conference for Informatics and Information Technology*. 2011.

[10] M. Liberman. High-stakes forensic linguistics. *Language Log*, http://languagelog.ldc.upenn.edu/nll/ ?p=3309

[11] R. Zheng, J. Li, H. Chen, Z. Huang. A Framework for Authorship Identification of Online Messages: Writing-Style Features and Classification Techniques. *Journal of the American Society tor Information Science and Technology*, 57(3):378–393. 2006

Zbornik 14. mednarodne multikonference

INFORMACIJSKA DRUŽBA – IS 2011

Proceedings of the 14th International Multiconference

INFORMATION SOCIETY – IS 2011

Sodelovanje, programska oprema in storitve v informacijski družbi

Collaboration, Software and Services in Information Society

Uredil / Edited by

Marjan Heričko

http://is.ijs.si

12. oktober 2011 / October 12th, 2011
Ljubljana, Slovenia

PREFACE

The transition to an information and knowledge society is being accompanied by a redefinition of business processes and a renewal of information solutions, networks and services. The main challenge remains unchanged – i.e. how to successfully manage and control the complexity of processes, software and services in order to improve collaboration between individuals, organizations and/or systems. A common focus of this year's contributions is on improving the understandability of existing and future software services, program code, networks, business processes, as well as socio-economic systems. A list of applied and/or proposed approaches and techniques include reverse engineering, modeling, layering, automatic transformations, simulations, prototyping, test-beds, context-sensitivity, interoperability frameworks, reference models, and also intersectionality theory.

We hope that these proceedings will be beneficial for your reference and that the information in this volume will be useful for further advancements in research and industry.

Our special thanks goes to the members of the Program Committee and additional reviewers as well as to the authors who have contributed to the success of this sub-conference.

prof. dr. Marjan Heričko
CSS 2011 – Collaboration, Software and Services in Information Society Conference Chairman

PROGRAMSKI ODBOR / PROGRAM COMITTEE

Dr. Marjan Heričko
University of Maribor, Faculty of Electrical Engineering and Computer Science

Dr. Ivan Rozman
University of Maribor, Faculty of Electrical Engineering and Computer Science

Dr. Kimmo Salmenjoki
Seinäjoki University of Applied Sciences

Dr. Lorna Uden
Staffordshire University, Faculty of Computing, Engineering and Technology

Dr. Gabriele Gianini
University of Milano, Faculty of Mathematical, Physical and Natural Sciences

Dr. Hannu Jaakkola
Tampere University of Technology Information Technology (Pori)

Dr. Mirjana Ivanović
Department of Mathematics and Informatics, Faculty of Science, University of Novi Sad

Dr. Zoltán Porkoláb
Eötvös Loránd University, Faculty of Informatics

Dr. Adel Taweel
King's College London, Division of Health and Social Care Research

Dr. Aleš Živkovič
University of Maribor, Faculty of Electrical Engineering and Computer Science

Dr. Gregor Polančič
University of Maribor, Faculty of Electrical Engineering and Computer Science

Dodatni recenzenti / Additional Reviewers

dr. Luka Pavlič, dr. Boštjan Šumak, Črt Gerlec
University of Maribor, Faculty of Electrical Engineering and Computer Science

V sodelovanju z / In collaboration with

UNDERSTANDING OLD ASSEMBLY CODE USING WSL

Doni Pracner, Zoran Budimac
Department of Mathematics and Informatics,
Faculty of Sciences, University of Novi Sad,
Trg Dositeja Obradovića 4, 21000 Novi Sad, Serbia
doni.pracner@dmi.rs, zjb@dmi.rs

ABSTRACT

Under the influence of time and changes in the environment, old code can become confusing even to the people that wrote it. This is especially true for legacy assembly libraries. They rarely have good (or any) documentation, the original high level code that was compiled into assembly is lost (or never existed) and often there were modifications directly in the libraries that make any original documents and/or code obsolete. This paper presents an attempt to make such code easier to understand and even automatically restructured.

For this purpose, we created a tool to translate assembly code into an existing high level language – WSL (*Wide Spectrum Language*), which was developed by Martin Ward. The results are then automatically transformed into higher level constructs, using the formal transformation system that is built into the language.

The main difference compared to earlier uses of WSL is that it tries to capture all of the aspects of the program in high level structures, without auxiliary files that are later used in translation to executable versions. This presents a potential for a better understanding of the transformed code and an option of running the programs directly in the interpreter, without translation to a third language.

The results achieved on a few test programs are promising: automated transformations show improvement in several software metrics. Maybe the most relevant is the custom weighted Structure metric which shows more than 30% improvement in the transformed code.

1 INTRODUCTION

One of the great problems of modern software engineering is the perceived ageing of software. Although an application correctly written 20 years ago should still work as designed, maybe the underlying system is no longer available, making the application useless, or on the other hand maybe the user now needs a different result due to changes in the "real" world. Parnas defined two main types of ageing: lack of movement and ignorant surgery (for instance quick fixes of small problems with no attention to the whole)[2].

Software Evolution is defined as the dynamic behaviour of programming systems as they are maintained and enhanced over their life times. Ward and Yang define the *life cycle of software* as a constant repetition of reengineering and its 3 phases: reverse engineering, functional restructuring and forward engineering[9]. Software evolution is therefore (in it's essence) repeated reengineering, and there is a need for tools which will help in this process.

This paper focuses on a tool for transforming low level assembly code into high level constructs, using *WSL (Wide Spectrum Language)*. Earlier uses of WSL created auxiliary files during the process (for instance memory mappings), which were later used in translation to lower levels. Our tool tries to capture all of the aspects of the program directly in high level structures. This presents a potential for a better understanding of the transformed code and an option of running the programs directly in the interpreter, without translation to a third language.

2 WSL AND FermaT

WSL (Wide Spectrum Language) is being developed by Martin Ward since 1989[4], with early versions of this system being written in LISP. A very important step was the addition *MetaWSL* which gave the user constructs to write programs that will be able to transform code (internally represented as abstract syntax trees) using formal transformations.

The current implementation is the *Fermat program transformation system*[1], and it is almost completely written in *MetaWSL*. This is then translated into a modified version of Scheme which enables efficient execution.

The main characteristics of the language is a strong mathematical core and the use of formal transformations. The *wide spectrum* in the name means that there are constructs in the language that can be used for a wide spectrum of applications in development: from abstract specifications to low level program code.

The system was successfully used in many projects migrating legacy assembly code to maintainable C/COBOL code[5, 6, 8].

One of the very useful features of this language that enabled this success is the existence of *Action systems*. This is a special structure in WSL which was specifically created to cope with unstructured jumps, which are very common in assembly code. It consists of a number of *actions* which can call each other:

```
ACTION start:
start == code block 1 END
name1 == code block 2 END
name2 == code block 3 END
...
...
ENDACTIONS
```

Once an action finishes it returns the control to the caller. Therefore an *action system* finishes when the start action finishes, or when a special, reserved action name "`Z`" is called which results in a momentary stop of the system.

3 OUR TRANSFORMATION PROCESS

Our transformation process consists of two basic steps: first we use our tool *Asm2wsl* to translate the assembly code to *WSL*, trying to capture all of the aspects of the original without much effort to optimise at this step. Second, we use *Trans.wsl* (a script written in *WSL*) for automated transformations on the translated code. There is always a possibility to apply manual transformations, either before or after the automated transformations.

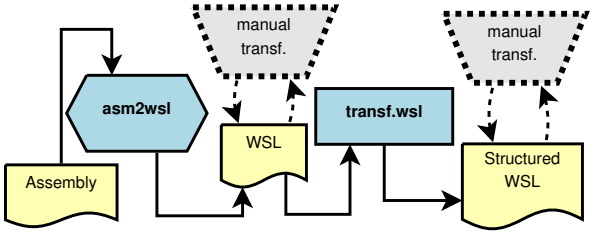


Figure 1: *Workflow diagram*

At it's base, this approach is similar to those in the past that used WSL. But the difference is that our main goal in the process is to get a high level version of the original program that will represent all the aspects of its functioning. The approach using WSL presented in a number of papers [5, 7, 8] creates additional files during the translation to WSL, which contain data about the variables and their mapping in the memory. These files are then used when the transformed and improved code is translated into (for example) *C* code and the appropriate pointer types are created.

Our approach should give a better understanding of the original code since we are looking at everything as high level structures, and it also enables us to run the translated programs directly in the WSL interpreter, without a need for an additional translator from WSL into another (semi) low level language. The downside is that assembler structures are often obscure, access data in different manners, and are therefore hard to understand and represent as high level, which limits the current version of the tools to a smaller subset of assembly code.

3.1 Translation

Asm2wsl is a tool that translates a subset of x86 assembly to *WSL*. To keep things simpler, at the moment it mostly presumes that we are working with a 80286. As newer processors were developed, they were mostly adding more registers, options to work with bigger words and more specialised commands, which are not of great importance to the concepts we are translating. The tool has been implemented in Java, making it platform independent and a good match to *FermaT*, which can also be run on a number of platforms. At its core, this is a line by line translator, with the focus on translating all aspects of the original code, without considering optimisation at this stage of the process. This generally results in programs that are much larger than the original assembly, but later on automatic transformations are able to reduce the size of the code. The same principle was successfully used in earlier translators which use WSL for transformations [6, 7, 8].

When run without parameters it gives a short usage help which includes the optional command line switches that influence the way the programs are handled and their default states. For instance the original comments from the assembly code are included as comments in the translated version, but this can be turned off.

An interesting option is to include the original assembly lines in comments next to their translated part, which can help understanding the workings of the code, as well as the translator itself. Being that many of the basic assembly programs just use processor registers and the memory without any output on the screen, another option of the translator is to add memory dumps at the end of the translated programs so that the final state of the program is more accessible.

Assembly commands work (more or less) directly with the processor. Being that high level languages don't do this, to capture all the aspects of these commands, we created a "virtual" processor. In it we have local variables to represent processor registers. Bits from the flag register are all defined as separate variables, which they practically are in the processor.

One of the simplest cases is translating the `mov` operation, which is in essence just assigning the value of one register to another. WSL also provides a construct for simultaneous assignments[9], which made possible translating an `xchg` command without an additional temporary variable.

```
mov ax, dx    ax := dx

xchg ax, dx   <ax := dx, dx := ax >
```

Being that we operate with variables that can be different sizes (in our case just 8 and 16 bits – one of the reasons for presuming a 80286 processor), an additional `overflow` variable is needed, which contextually tries to detect the size of the target variable. Based on the value the flag variables are set like they would be in the processor.

```
add dx, ax    overflow := 65536;
              dx := dx + ax;
              IF dx >= overflow THEN
                dx := dx MOD overflow;
                flag_o :=1; flag_c := 1;
              ELSE flag_o :=0; flag_c := 0;
              FI;
```

In a *x86* processor *Low* and *High* parts of 16 bit registers can be accessed independently (i.e. the lower and higher 8 bits). Being that our goal were high level structures, we wanted to exclude direct memory operations, so these were

implemented with additional operations that set the adequate parts of the register.

```
mov ah,n      t_e_m_p := n;
              ax := (ax MOD 256) + t_e_m_p * 256;

add al,12     overflow := 256;
              t_e_m_p := (ax MOD 255) + 12;
              IF t_e_m_p >= overflow THEN
                 t_e_m_p := t_e_m_p MOD overflow;
                 flag_o :=1; flag_c := 1;
              ELSE flag_o :=0; flag_c := 0; FI;
              ax := (ax DIV 256)*256 + t_e_m_p;
```

Labels in the original code are translated as *Action system* names (see Section 2). The whole system that we generate when we translate a "normal" assembly program is by nature *regular* (meaning that none of the calls ever return, that is, all of the actions just call other actions, and the system is finished with a `CALL Z`). Because of the special properties, these can be transformed easily into structured code.

Basic operations with arrays are also supported by the tool, with an automatic adjustment to the indexes, which is necessary since arrays start from 1 in *WSL*, and from 0 in assembly.

The processor's internal stack is implemented as a global list/array. The `pop` and `push` commands take and put elements on the start of this list directly. No additional checks (such as element size and compatibility, presence of elements on the stack) are performed, being that we presume to work with programs that worked correctly in their original form.

Macro structures are not translated at this stage, although there are some special names that are recognised and translated directly into *WSL* code to enable input/output operations. For instance `print_num x` and `print_str x` are directly translated to `PRINT(x)`.

The tool also has support for translating procedures from assembly. They are translated as nested Action Systems, so that local labels can be created, and it also enables us to return to the point of the original call once the procedure has finished it's work. The translated programs worked from the interpreter without modifications. Transformations were also successfull, despite the process resulting in action systems that are not *regular*. In the initial small tests, the procedures were simplified and then included in the main action system.

3.2 Transformations

The second part of the process consists of a small program written is WSL that goes through the abstract syntax tree of the translated program, and tries to apply some of the available transformations on adequate nodes. All of the transformations implemented in WSL need to have procedures that will test if they can be applied to the given node, so this part of the code is relatively simple. For example, a transformation that unrolls a `WHILE` loop will (among other things) first check if the given node is in fact a `WHILE` statement.

Some of the important transformations are collapsing of the action systems into endless loops with exits in the middle and the subsequent transform of those into `WHILE` loops. At the same time constants are propagated through the code and various redundant parts are removed.

3.3 Example

A nice example is a greatest common divisor program – a simple implementation of the Euclidean algorithm, presented in Figure 2.

```
model small
.code
            mov ax,12
            mov bx,8
compare:
            cmp ax,bx
            je theend
            ja greater
            sub bx,ax
            jmp compare
greater:
            sub ax,bx
            jmp compare
theend:
            nop
end
```

Figure 2: *GCD example assembly code*

Translation into WSL produces more than 60 lines out of the original 16 – an expected increase. Figure 3 shows this code after it was automatically cleared of redundant variables (such as the unused processor registers) – already reduceing it to around 30 lines.

```
VAR < flag_z := 0, flag_c := 0 >:
ACTIONS A_S_start:
A_S_start ==
  Ax := 12;
  bx := 8;
  CALL compare
END
compare ==
  IF ax = bx THEN flag_z := 1 ELSE flag_z := 0 FI;
  IF ax < bx THEN flag_c := 1 ELSE flag_c := 0 FI;
  IF flag_z = 1 THEN CALL theend FI;
  IF flag_z = 0 AND flag_c=0 THEN CALL greater FI;
  IF bx = ax THEN flag_z := 1 ELSE flag_z := 0 FI;
  IF bx < ax THEN flag_c := 1 ELSE flag_c := 0 FI;
  bx := bx - ax;
  CALL compare;
  CALL greater
END
greater ==
  IF ax = bx THEN flag_z := 1 ELSE flag_z := 0 FI;
  IF ax < bx THEN flag_c := 1 ELSE flag_c := 0 FI;
  ax := ax - bx;
  CALL compare;
  CALL theend
END
theend == CALL Z
END
ENDACTIONS
```

Figure 3: *GCD translated into WSL*

As explained earlier translation introduces a lot of flag variable testing. These can be removed quite efficiently with a "Flag Removal" transformation which reduces the size of the code considerably,

The next step is to apply transformations that will collapse the action system into a loop, and then to transform the end-

less loop with exits in the middle into a `while` loop. This gives a very clear and understandable high level implementation shown in Figure 4.

```
ax := 12;
bx := 8;
WHILE ax <> bx DO
  IF ax >= bx THEN ax := ax - bx
  ELSE bx := bx - ax FI
OD
```

Figure 4: *Transformed GCD example*

To see the results of automated transformations, several software metrics which are built into WSL were used – McCabe's cyclomatic complexity, number of statements, control/data flow, number of nodes and *Structure*, a custom weighted metric which should show a useful approximation of the complexity of the program. Only McCabe's CC showed higher numbers after the transformations, which is actually a result of the introduced higher level structures. All the other metrics (not suprisingly) showed improvements: *CFDF* 40%, *nodes* and *statements* around 20-30%, and (maybe the most relevant) *structure* showed more than 30% improvements.

4 CONCLUSIONS AND FUTURE WORK

This paper presents an approach to understanding and potentially restructuring old assembly code by converting it to a high level language. The distinction of this approach compared to others that use WSL is that all of the aspects of the program are translated into high level structures without additional files that are used in translation to a final language such as C or COBOL.

The basis of the work was an implementation of a tool for translating many basic functionalities of assembly languages, as well as a small program that uses formal transformations from MetaWSL. First results of automated transformations were promising, *structure* metric showed more than 30% improvement on the translated code.

Being that all of the results were made on rather basic programs, there is a lot of areas which could be improved in the future. The translator itself should be expanded with more support for a wider array of assembly commands. The transformation script also needs more work and to be expanded with more operations.

Finally the example code base should be enriched as much as possible to have a better insight into the quality of the translations. Some automated scripts that could check the outputs of the original assembly, the translated and the transformed versions of the programs would be useful as another step in the verification of the whole process.

The presented tool is currently in a closed prototype testing stage, but is expected to be published on-line under a free licence in the near future.

5 ACKNOWLEDGEMENTS

This work was supported by project "Intelligent techniques and their integration into wide-spectrum decision support" (no. 174029), of the Serbian Ministry of Science and Technological Development.

References

[1] M. Ladkau, F. Chen, S. Li, and S. Natelberg. FermaT Transformation Engine Tutorial. Technical report, Software Technology Research Laboratory, De Montfort University, Leicester, Oct 2006.

[2] D. L. Parnas. Software aging. In *ICSE 1994*, pages 279–287, 1994.

[3] D. Pracner and Z. Budimac. Restructuring assembly code using formal transformations. In *Proc. of Symposium on Computer Languages, Implementations and Tools – SCLIT 2011 (organized within ICNAAM 2011 conference)*, AIP proceedings, 2011. accepted.

[4] M. Ward. *Proving Program Refinements and Transformations*. PhD thesis, Oxford University, 1989.

[5] M. Ward. Assembler to C Migration using the FermaT Transformation System. In *IEEE International Conference on Software Maintenance (ICSM'99)*, pages 67–76. IEEE Computer Society Press, 1999.

[6] M. Ward. Reverse engineering from assembler to formal specifications via program transformations. In *7th Working Conference on Reverse Engineering, Brisbane, Queensland, Australia*. IEEE Computer Society, 2000.

[7] M. Ward. Pigs from Sausages? Reengineering from Assembler to C via FermaT Transformations. *Science of Computer Programming, Special Issue on Program Transformation*, 52/1-3:213–255, 2004.

[8] M. Ward, H. Zedan, and T. Hardcastle. Legacy assembler reengineering and migration. In *ICSM2004, The 20th IEEE International Conference on Software Maintenance*. IEEE Computer Society, 2004.

[9] H. Yang and M. Ward. *Succesful Evolution of Software Systems*. Artech House, 2003.

A COMPARISON OF BPMN NOTATION AND NATURAL LANGUAGE UNDERSTANDABILITY

Jernej Huber, Aleš Živkovič, Marjan Heričko
University of Maribor, Faculty of Electrical Engineering and Computer Science
Smetanova ulica 17, 2000 Maribor, Slovenia
Tel: +386 2 2207352
e-mail: jernej.huber@uni-mb.si

ABSTRACT

In this paper, we explored the understandability of BPMN notation as compared with the understandability of natural language. We conducted an experiment in which we examined the results of students who were completing two different types of questionnaires. Students from the first group analyzed BPMN diagrams, while others analyzed an unstructured text. The text and BPMN diagrams were from identical problem domains and both groups of students answered the same questions. We compared the results and conducted a statistical analysis using a t-test with independent samples. Because we were able to classify students according to their prior knowledge of BPMN notation, we also conducted a one-way ANOVA test and performed a detailed analysis of the collected data.

1 INTRODUCTION

Well-engineered business processes include the process performance measurement and influence the process performance in a desired direction by using these measurements to control the processes themselves. These measurements are essential in organizations that intend to reach a high level in the capacity of their processes. In other words, the business process measurement makes it possible for organizations to learn from the past in order to improve performance and achieve better predictability over time [1].
Quality considerations are of great importance when conducting a process modeling project. A variety of structural metrics for process models have been proposed, which are tailored to approximate usage characteristics, such as process model understandability. In [2] the authors verified how the understandability of BPMN models relates to these metrics by means of correlation and regression analysis and determined the threshold values to distinguish between different levels of process model quality.
On the other hand, there was also much research effort directed to the field of unstructured text readability analysis. Readability is what makes some texts easier to read than others. Readability metrics, such as the Flesch Kincaid Test, the Flesch-Kincaid Grade Level Test, SMOG, the Gunning Fog Index, the Coleman Liau Index and the Automated Readability Index, have been developed and tweaked for decades. Carrying out an analysis to determine readability is common in several industries, from medicine, business and the military to various legal professions [3].
In this paper, we have explored the understandability of BPMN notation [4] in comparison with natural language understandability. We therefore linked the fields of business processes and natural language understandability. In our review of existing literature, we did not find any similar articles that would link these fields of research.

2 EXPERIMENT STRUCTURE

The experiment was conducted in the form of a quiz on the Moodle learning system. Students were randomly assigned to one of two different quizzes. Students from the first group analyzing the BPMN diagrams, while the second group of students analyzed the official consolidated text of the Marriage and Family Relations Act (ZZZDR-UPB1) [5], on which basis we modeled the BPMN diagrams for the experiment. The questionnaires were the same in both groups and all students answered the same questions. The total length of text was 757 words or 4.107 characters without spaces. BPMN diagrams consisted of 1 process with 62 elements, 61 edges, 7 XOR forks and 1 subprocess with 47 elements, 53 edges and 26 XOR forks.
Each student was given a 15 minute time limit and was required to answer as many questions as possible. The problem domain of the experiment was the child adoption process, taken from ZZZDR-UPB1.
Each correct answer was evaluated with 1 point, and each incorrect one with 0 points. Multiple-choice answers were evaluated in a range from 0 and 1 points with 1/n step, where n is the number of possible choices. The experiment structure is shown in Figure 1.

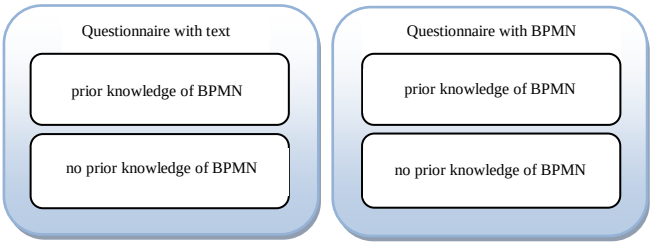


Figure 1: *The experiment structure.*

3 LIMITATIONS AND THREATS TO VALIDITY

To address one of the major threats to the validity of our research, namely the sample bias problem, we used random sampling and assigned students to one of the two quiz types, either BPMN or TEXT.

During the experiment, we employed a group that consisted of teaching assistants and young researchers at the Institute of Informatics. The group was intended to validate the BPMN diagrams and questionnaires, prepared for the experiment.

We also checked the impact of other factors, such as the time and duration of the experiment and found no statistical significance between groups with different factor values.

We think that threats to internal validity [6] were properly addressed and acted upon with the previously mentioned corrective actions.

The primary limitation to our research was the fact that our study population was limited to students of the FERI faculty. Therefore, we find it difficult to generalize the results to a wider environment. External validity is consequentially questionable [6].

4 CONDUCTED STATISTICS TESTS

The statistical tests were conducted to determine if statistical differences exist between different groups of students who have participated in the experiment. We conducted a t-test with independent samples. Because we were able to classify students according to their prior knowledge of BPMN notation, we also conducted a one-way ANOVA test.

4.1 T-test

An independent-samples t test was conducted to evaluate the following hypotheses; firstly, we declared the null hypothesis, which states that there is no difference between groups of students:

- H_{o1}: *There is no difference between groups of students who have completed the BPMN and TEXT quizzes, respectively* (D=0).

We also proposed two alternative hypotheses:

- H_{a1a}: *Students, who completed the BPMN quiz, achieved better results that students who completed the TEXT quiz* (D ≠0).
- H_{a1b}: *Students, who completed the TEXT quiz, achieved better results that students who completed the BPMN quiz* (D ≠0).

We met all the assumptions of the independent-samples t-test:

- The test variable is normally distributed in each of the two populations. We used the Q-Q plots to check the assumption.
- The variances of the normally distributed test variable for the populations are equal. We used Levene's test to check the assumption.
- The cases represent a random sample from the population and the scores on the test variable are independent of each other. We used a random assignment of students to each type of the quizzes to meet the assumption.

Levene's test of equality of variances between the two groups was not significant (F = 2.372, sig. = 0.125). Consequently, we can assume that population variances for the two groups are equal [7]. The t-test was significant, t(174) = 2.738, p = 0.007 and the results ran counter to the null hypothesis. We can accept the alternative hypothesis H_{a1b}, because of the students, who completed the TEXT quiz and attained better results. The SPSS tool that we used to conduct the statistical tests provides the results in the form of a two-sided t-test. Therefore, we have to additionally divide the resulting p value with two, so that p = 0.003 [8].

We expected that the quiz with BPMN notation would bring better results, but statistical findings opposed our assumptions. Either way, we cannot simply generalize our results to the hypothesis that the usage of BPMN notation reduces the understandability of business processes. We have to consider other factors, some of which we will give in Chapter 5.

Nevertheless, it would be prudent to carry out similar studies on a larger set of substantively different case studies.

4.2 One Way ANOVA

Figure 1 illustrates the structure of our experiment. When we divide participants according to their prior BPMN knowledge, we can compare the four groups of students. Students from the first group completed the TEXT quiz and had prior knowledge of BPMN notation. Similarly, the second group also completed the TEXT quiz, but they did not have prior knowledge of BPMN notation. Students from the third group completed the BPMN quiz and had prior knowledge of BPMN notation. Finally, the fourth group finished the BPMN quiz and did not have prior BPMN knowledge. A one-way analysis of variance was conducted to evaluate the relationship between the four groups in the experiment.

We set the following hypotheses:

- H_{o2}: *There is no statistically significant difference between groups of students with prior BPMN knowledge, who have completed the BPMN and TEXT quizzes, respectively* (D=0).
- H_{a2}: *There is a statistically significant difference between groups of students with prior BPMN knowledge, who have completed the BPMN and TEXT quizzes, respectively* (D≠0).
- H_{o3}: *There is no statistically significant difference between groups of students with no prior BPMN knowledge, who have completed the BPMN and TEXT quizzes, respectively* (D=0).
- H_{a3}: *There is a statistically significant difference between groups of students with no prior BPMN*

knowledge, who have completed the BPMN and TEXT quizzes, respectively (D≠0).

- H_{o4}: *There is no statistically significant difference between groups with and without prior BPMN knowledge, who have completed the BPMN quiz* (D=0).
- H_{a4}: *There is a statistically significant difference between groups with and without prior BPMN knowledge, who have completed the BPMN quiz* (D≠0).
- H_{o5}: *There is no statistically significant difference between groups with and without prior BPMN knowledge, who have completed the TEXT quiz* (D=0).
- H_{a5}: *There is a statistically significant difference between groups with and without prior BPMN knowledge, who have completed the TEXT quiz* (D ≠0).

We met all of the underlying assumptions of a one-way ANOVA test:

- The dependent variable is normally distributed for each of the populations. We used the Q-Q plots to check the assumption.
- The variances of the dependent variables are the same for all populations. We used Levene's test to check this assumption.
- The cases represent a random sample from the population and the scores on the test variable are independent of each other. We used a random assignment of students to each type of quizz to meet this assumption.

Levene's test for the equality of variances between the two groups was not significant (F = 2.372, sig. = 0.125). Therefore we can assume that population variances for all four groups are equal [9].

Figure 2 shows the mean values, standard deviations and size of our sample groups of students. The group name consists of the quiz type and information about prior knowledge of BPMN. For example, TEXT_YES represents the group of students who participated in the TEXT quiz and had prior knowledge of BPMN notation.

Group name	Mean	Std. Deviation	N
TEXT_YES	7.3305	1.11724	38
TEXT_NO	6.3129	1.63589	41
BPMN_YES	6.7449	1.58137	41
BPMN_NO	5.6151	1.78064	41
Total	6.4854	1.66256	161

Figure 2: *Mean values, std. deviations and sample sizes.*

The ANOVA was significant, F(3, 157) = 8.56, p = 0.0. Follow-up tests were conducted to evaluate pairwise differences among the means. We chose to assume equal variances among the four groups and conducted *post hoc* comparisons with the help of a Tukey HSD test. As shown in Figure 3, there were significant differences in the means between the following pairs of groups (differences are marked on Figure 3 with the symbol *):

- BPMN_YES and BPMN_NO
- TEXT_YES and TEXT_NO
- TEXT_YES and BPMN_NO

Figure 3 also shows 95% confidence intervals for the pairwise differences, as well as the means and standard deviations for each group.

					95% Conf. Interval	
(I) Group name	(J) Group name	Mean Diff. (I-J)	Std. Error	Sig.	Lower Bound	Upper Bound
TEXT_YES	TEXT_NO	1.0176*	.35036	.022	.1078	1.9274
	BPMN_YES	.5856	.35036	.342	-.3242	1.4955
	BPMN_NO	1.7154*	.35036	.000	.8056	2.6252
TEXT_NO	TEXT_YES	-1.0176*	.35036	.022	-1.9274	-.1078
	BPMN_YES	-.4320	.34365	.592	-1.3243	.4604
	BPMN_NO	.6978	.34365	.181	-.1946	1.5902
BPMN_YES	TEXT_YES	-.5856	.35036	.342	-1.4955	.3242
	TEXT_NO	.4320	.34365	.592	-.4604	1.3243
	BPMN_NO	1.1298*	.34365	.007	.2374	2.0221
BPMN_NO	TEXT_YES	-1.7154*	.35036	.000	-2.6252	-.8056
	TEXT_NO	-.6978	.34365	.181	-1.5902	.1946
	BPMN_YES	-1.1298*	.34365	.007	-2.0221	-.2374

Figure 3: *The results of post hoc pairwise comparisons.*

Based on these results, we cannot confirm or reject the null hypotheses H_{o2}, H_{o3} and alternative hypotheses H_{a2}, H_{a3}. On the other hand, we can reject the null hypotheses H_{o4}, H_{o5} and accept the alternative hypotheses H_{a4}, H_{a5}.

We also checked the statistical power of our experiment. On the basis of the resulting mean values and standard deviations of four groups, a G*Power tool [10] calculated a minimal required sample size of 35 students per group to have a 95 percent probability of not making a Type II error or a false negative decision and having a significant ANOVA with a p value smaller than 0.05.

We found it rather surprising that prior knowledge of BPMN notation affected the results of the TEXT quiz. On that basis, we can imply that learning BPMN notation has a positive impact on the understandability of unstructured text descriptions for business processes. In any case, additional studies with several different case studies are required.

The results showed that there is evidence of a statistically significant difference between the group of students with prior BPMN knowledge who completed the TEXT quiz and the group of students with no prior BPMN knowledge who completed the BPMN quiz. This additionally confirms the t-test results from Chapter 4.1, which showed that the students

were better at the TEXT quiz. The difference between the attained average values of these two groups was -1.7154, as we can see from Figure 3.
The differences within the BPMN quiz were expected. The group of students with prior knowledge of BPMN notation achieved better results than the group with no BPMN knowledge.
Figure 4 illustrates the statistically significant differences of the *post hoc* pairwise comparisons between the groups, which are marked with red-dashed lines.

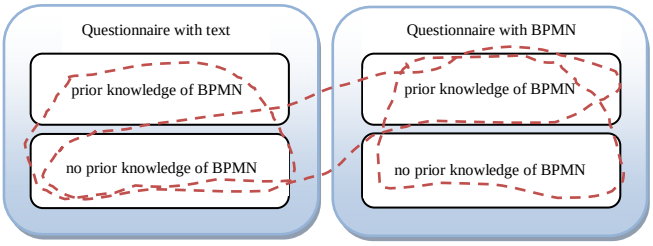


Figure 4: *Significant differences of pairwise comparison.*

We created a boxplot to show the distributions of the changes in dependent variables among the four groups of students who participated in the experiment. The boxplot is shown in Figure 5.

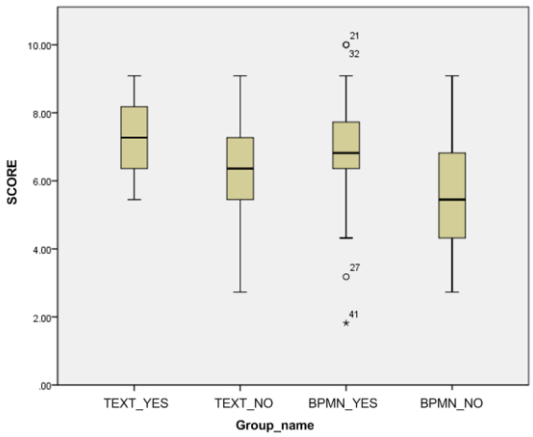


Figure 5: A *boxplot with distributions of the changes in dependent variables among the four groups of students.*

5 CONCLUSION

We expected that the BPMN quiz would garner better results among the students, but the results indicate otherwise. Nevertheless, we cannot simply deduce from the results that the usage of BPMN notation reduces the understandability of business processes when compared with the understandability of unstructured text. One of the biggest factors is surely the quality of the designed BPMN models. Another factor is the length of the unstructured text, which is also the basis for the BPMN diagrams. We cannot predict the results of much longer processes. We also do not know the results of the analysis of much more complex and interwoven processes, which combine several rules and laws. Also, the time needed for the same level of understanding of models and unstructured text can significantly affect the results.
Therefore, it would be prudent to conduct similar research on a much greater set of case studies from different problem domains.
For future research, we plan to more precisely link the fields of business process and natural language measurement. We will try to develop a metric to predict the structural complexity and understandability of associated process models on the basis of the understandability of the unstructured text, which is the source of the resulting BPMN models.

References

[1] L.S. González, F.G. Rubio, F.R. González, and M.P. Velthuis. *Measurement in business processes: a systematic review*. Business Process Management Journal. Vol. 16 (No. 1). pp. 114-134. 2010.

[2] L. Sánchez-González, F. García, J. Mendling, and F. Ruiz. *Quality Assessment of Business Process Models Based on Thresholds*. Proc. On the Move to Meaningful Internet Systems: OTM 2010. Vol. 6426. pp. 78-95. 2010.

[3] A. Mobasseri. *Improving Readability*. Computing Research Topics. Villanova University, Department of Computing Sciences, CSC 3990. http://www.csc.villanova.edu/~mdamian/csc3990/csrs2008/06-csrs2008-ArminMobasseri.pdf. Last visit in July of 2011.

[4] *Business Process Model and Notation*. Object Management Group. http://www.bpmn.org. Last visit in September of 2011.

[5] *Zakon o zakonski zvezi in družinskih razmerjih (uradno prečiščeno besedilo) (ZZZDR-UPB1)*. Register predpisov Slovenije. http://zakonodaja.gov.si/rpsi/r02/predpis_ZAKO4132.html. Last visit in July of 2011.

[6] W. Huitt et al. *Internal and External Validity*. Educational Psychology Interactive. http://www.edpsycinteractive.org/topics/intro/valdgn.html. Last visit in July of 2011.

[7] S. B. Green, N.J. Salkind. *Using SPSS for Windows and Macintosh, Fifth Edition*. Pearson & Prentice Hall, 2008

[8] How can I convert from a two-tailed to a one-tailed test? UCLA Academic Technology Services. http://www.ats.ucla.edu/stat/spss/faq/pvalue.html. Last visit in July of 2011.

[9] S. Archambault. *One-Way ANOVA*. Psychology Deparment, Wellesley College. http://www.wellesley.edu/Psychology/Psych205/anova.html. Last visit in July of 2011.

[10] F. Faul. *G*Power 3 - a statistical power analysis tool*. http://www.psycho.uni-duesseldorf.de/abteilungen/aap/gpower3. Last visit in September of 2011.

FEASIBILITY AND RATIONALE OF INTRODUCING LAYERS INTO BUSINESS PROCESS MODELS

Gregor Polančič[1], *Gregor Jošt*[1], *Jari Soini*[2]
[1] University of Maribor, Faculty of electrical engineering and computer science
Smetanova 17, 2000 Maribor, Slovenija
e-mail: {gregor.polancic, gregor.jost}@uni-mb.si
[2] Tampere University of Technology - Pori, P.O. Box 300, FIN-28101 Pori, Finland
e-mail: jari.o.soini@tut.fi

ABSTRACT

Business Process Model and Notation (BPMN) is the leader and de-facto standard in business process modeling, a sub-discipline of business process management. While BPMN's objectives are to be expressive, widely understood, interchangeable and executable, it has also became rich and complex. The objective of this article is to present an idea of lowering the complexity of BPMN by applying layers into BPMN based process models. As presented in the article, layers are the remaining complexity management approach, which has not been implemented in BPMN yet. The focal part of the article is a SWOT analysis of applying layers into business process models, as well as the future work and challenges which need to be solved in order to reach our research goal - increasing the objective and subjective understandability of BPMN based process models.

1 INTRODUCTION

Business process management (BPM) enables organizations to be more efficient, more effective, and more capable of change than a functionally focused, traditional hierarchical management approach [1]. A critical part of BPM is business process modeling. It is concerned with the representation of organizational processes, so that the current processes may be analyzed, executed and improved in the future.

An important criteria for business process modeling success is the ability to understand and exchange business process models within the company, between companies, with government institutions and customers. In addition, several business and technical roles commonly interact with a business process model (e.g. business analytics, software developer, quality managers and people who actually perform processes). The above criteria is in-line with the primary goal of Business Process Model and Notation (BPMN), the leader and the de-facto standard in business process modeling [1].

Besides the primary goal of BPMN, BPMN 2.0 focuses additionally on the executability and interoperability of business process models [2]. While this is a desirable BPM feature, it adds complexity to an already rich notation (BPMN 2.0 has over 110 elements). So, several empirical investigations of BPMN demonstrated that BPMN notation is complex, difficult to learn and hard to work with [3],[4], [5]. Figure **1** illustrates goals of BPMN in relation to its complexity.

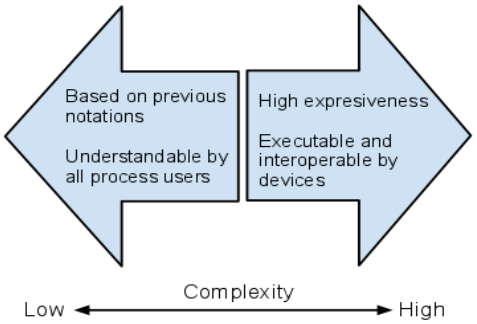


Figure 1: *Complexity conflict between different BPMN objectives*

To overcome above stated BPMN complexity challenges, we propose to lower the complexity considering following presumptions. First, we want to manage the complexity of BPMN models in a way that will be independent from the specification. And second, we want to focus primary on the "readers" of BPMN models, because more business users need to understand business processes than to model them.

2 MANAGING COMPLEXITY WITH LAYERS

The BPMN notation is perceived as complex and not easy to learn, even for business analysts. One of the main challenges for non-experienced modelers still lies in identifying the most suitable icon in a given moment, depending on the type of a specific event [4].

When we are addressing complex systems, we have to take into the consideration that complexity is hard to measure [6] and it is considered to be subjective [7]. For example, the number of elements and relations may not be enough to characterize complexity, because a large number of elements and relations can still present a simple system. Contrary, a few elements and relations can be more complex as a whole [8].

So, instead of measuring complexity it is more convenient to look for the signs of complexity, a method that is adapted from medicine (signs or symptoms help to confirm the diagnosis). Table **1** represents five signs of complexity [6] along with how they apply to BPMN and business process diagrams (BPD)

Table 1: *Five signs of complexity*

Sign of complexity	BPMN / BPD
Large number of components - a size has a direct influence of whether we perceive a system complex or not.	BPMN 2.0 has over 110 elements.
Large number of interconnections - even when we operate with a small amount of components, we can have many interconnections between them, making a system arguably complex.	In a BPD, BPMN elements need to be interrelated into a graph.
Many irregularities - we can have a large number of components and interconnections, yet they can still form a simple system. However, if there are many exceptions or irregularities in the system, we can perceive this as a sign of complexity.	Vertical BPMN domains, the meaning of non-standardized elements, the meaning of colors, etc.
A long description - to avoid complexity a methodological description of the system is advised. A shorter, systematic specification is favored over a long description of the system's properties.	BPMN 2.0 specification is 538 pages long.
A team of designers, implementers or maintainers of a system. Usually there are several roles involved in constructing and maintaining a system, so it should be simple enough for each individual to understand it.	Business process modeling involves different business and technical roles.

The source of complexity on a more specific level is mainly indicated through the [6]:

1) **Number of requirements.** While individual requirements might not be complex, a list of requirements adds their individually complexities, as well as complexity from their interactions.
2) **A specific requirement: maintaining high utilization.** This source of complexity represents the need for high performance or efficiency. However, when we strive for higher utilization, the complexity of the strategies for use, allocation and distribution increases.

There are several techniques of coping with complexity, which can be defined as [6]: modularity, abstraction, layering and hierarchy. Table **2** describes the purpose of each technique and how it can be applied to BPMN.

Table 2: *Complexity management techniques*

Approach	Purpose of the technique	Applying to BPMN
Modularity	Modularity is an approach for organizing complex products and processes efficiently by decomposing complex tasks into simpler portions so they can be managed independently and yet operate together as a whole [9].	Sub-processes, black-box pools.
Abstraction	Abstraction is the activity of finding similarities or common aspects while unimportant differences are left ambiguous, vague or undefined [10].	The three BPMN 2.0 abstraction levels: descriptive, analytic and executable.
		Abstract BPMN elements: complex gateway, multiple and parallel multiple events
Hierarchy	Hierarchy is a strict partially ordered set, which can be viewed as a graph. The fact that the graph represents a partial order implies that the graph is a directed acyclic graph [11].	Pool, lanes and sub-lanes can be hierarchically organized. In addition, each process can recursively include a sub-process.
Layering	A layer contains different modules, where they communicate with the peers within the same layer and the modules on the next higher and lower layers [6].	Not implemented yet.

In the context of large systems that require decomposition, the layer-based solution is extremely simple. The system is structured into an appropriate number of layers, which are placed on top of each other. The most common approach is to start at the lowest level of abstraction (layer 1), which represents the base of the system. The abstraction ladder is build by putting layer J on top of layer J-1 until the top level of functionality is reached [12]. As presented on figure below (Figure 2), layers might interact only with the neighbor layers.

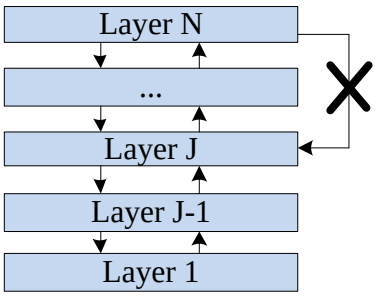


Figure 2: *Layering scheme*

3 APPLYING LAYERS INTO BUSINESS PROCESS DIAGRAMS

Considering the signs of BPMN and BPD complexity (Table 1) as well as the techniques for managing BPD complexity (Table 2) we found reasonable to investigate the feasibility and rationale of applying layers into BPMN based BPD.

An initial analysis of the proposed layered-based approach (hereinafter referred to as "LBA") demonstrates an opportunity to cope with the complexity of BPMN based BPD without influencing the BPMN specification in any way. This challenge is feasible, because LBA might be applied to BPD as presented below.

The proposed solution (LBA) is based on an idea of a basic BPMN layer (layer 1), which would represent the "core" BPMN elements" of BPD. So, layer 1 would represent the context and the main semantic of the BPD (e.g. the flow elements of the "positive" and the "most probable" process scenario with the corresponding connecting elements). In contrast to layer 1, every other layer would represent a specific subset of BPMN elements or a subset of BPD as following.

In case of BPMN elements, the dimension of "static layers" (layer S1, S2, Sn), which could be defined prior to a BPD, would be related to a specific subsets of BPMN elements. Several categories of BPMN elements are already defined in the BPMN standard [2]: flow objects, data, connecting objects, swimlanes and artifacts.

These categories are further decomposed to the level of specific graphical elements in a hierarchical way. A specific subset of semantically interrelated BPMN elements could represent a different aspect in form of a layer, thus enabling users to have faster and better overview of specific aspects of the model (Figure 3).

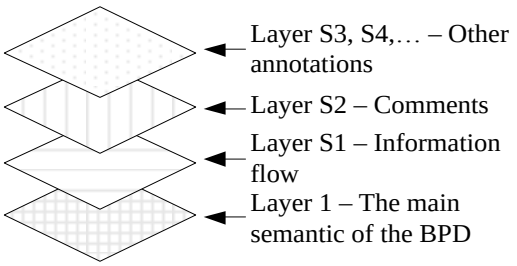


Figure 3: *Conceptual scheme of Type "S" Layer*

A similar approach was already represented in other domains, e.g. geographical maps. Using the technology, electronic maps have advanced the functionality of the maps, one of the latter are also layers. Layers in the electronic maps enable users to choose what kind of information they want to see displayed over the viewed place [13].
The second dimension of layers could be related to subsections of concrete BPD. This type of "dynamic layers" (Layer D1, D2, Dm) would be generated in the "run-time" and would be based on the actual focus within the BPD. Examples of these, dynamic layers are following (see also Figure 4):

1) **Alternative path layer.** This type of layer would represent elements relevant for a specific alternative path within the BPD.
2) **Exception layer.** This type of layer would represent elements, related to a specific exceptional flow.
3) **Black and white box layer.** This type of layer would represent the details (interior) of the focused pool. Other pools would be represented as black boxes.
4) **Data object layer.** This type of layer would focus on a specific data object and would present only flow elements, which interact with the focused data object.

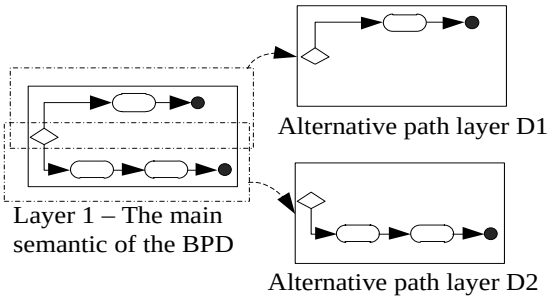


Figure 4: *Conceptual scheme of Type "D" Layer*

The proposed LBA is also in accordance with the research, done by La Rosa et al. in 2011. They introduced a pattern based approach for reducing BPMN model complexity. In their paper they identified patterns to reduce the perceived model complexity, thus to simplify the representation of the process model. They defined patterns as a capture of features to manage process model complexity, one of them being a graphical highlight, which is a visual mechanisms to emphasize certain aspects of a process model [14].

4 SWOT ANALYSIS OF APPLYING LAYERS INTO BUSINESS PROCESS MODELS

Table 3 represents an initial SWOT analysis in the field of LBA investigation.

Table 3: *SWOT analysis of LBA*

STRENGTHS	WEAKNESSES
▪ Smaller complexity of individual layers when compared to the complexity of the complete BPD. ▪ Complexity management approach with no influence on BPMN notation. ▪ Layers proved as successful in other domains (e.g. geographical maps, architecture, graphical design and anatomy).	▪ No research, which would investigate BPMN layers was found. ▪ BPMN specification does not define any layers. ▪ Applying layers in the phase of BPD design is difficult or even senseless. ▪ BPD represents a graph with many elements and connections; so dividing elements into layers might be difficult.

▪ LBA is extensible (new layers can be added). ▪ LBA does not influence other complexity management approaches.	▪ Existing BPMN tools do not support layers.
OPPORTUNITIES	**THREATS**
▪ Increased objective and subjective understandability of the models. ▪ Increased time-efficiency of reading the BPD. ▪ Faster learning curve of the BPDs. ▪ Incorporate the layer based approach into the BPMN specification. ▪ Increased acceptance of BPMN.	▪ Loss of the meaning while observing only a part (layer) of the BPD. ▪ Users would still prefer to read BPD without layers. ▪ Increased complexity of BPMN tools because of layer support. ▪ No BPMN vendor would be interested in implementing the proposed solution. ▪ Compatibility with tools that do not support layers.

5 FUTURE WORK AND CHALLENGES

As evident from the SWOT analysis (table 3), several challenges are related to LBA. The 1st challenge is to define LBA in a formal way. While BPD can be formally represented with a directed graph, we presume that graph as well as the set theory could be further investigated in order to apply them to BPD and LBA. The 2nd challenge is to define and to argue specific layers. We plan to investigate related notations and diagrams (e.g. UML diagrams, EPC diagrams, etc.) as well as other solutions that already incorporate layers. In addition we plan to analyze the BPMN standard in search for potential indicators that would help us define layers. Third, we need to investigate the efficiency and effectiveness of LBA. For that purpose we plan to conduct an empirical research, that would prove or reject benefits of LBA as well as the applicability of specific layers. If the results of the empirical research demonstrate the success of LBA, the 4rd challenge is to implement the solution. We plan to extend an open source solution (e.g. Oryx) with the LBA and share it with the community.

6 SUMMARY

This article presented an idea of lowering the complexity of BPMN based process models by introducing the concept of layers. The idea seems reasonable since layers represent the remaining complexity management approach which has not been implemented in BPMN yet. SWOT analysis demonstrated that layer based approach (LBA) offers several strengths and opportunities as well as threats and weaknesses. Several challenges still need to be addressed before we reach our final objective, to increase the understandability and efficiency of business process diagrams based on BPMN.

References

[1] M. Treat, "Business Process Management (BPM) - What is BPM Anyway?," 08-Dec-2006. [Online]. Available: http://www.bpminstitute.org/articles/article/article/what-is-bpm-anyway.html. [Accessed: 06-Sep-2011].

[2] OMG, "Business Process Model and Notation (BPMN) Version 2.0," 03-Jan-2011. [Online]. Available: http://www.omg.org/spec/BPMN/2.0/. [Accessed: 2011].

[3] G. Decker and F. Puhlmann, "Extending BPMN for Modeling Complex Choreographies," in *On the Move to Meaningful Internet Systems 2007: CoopIS, DOA, ODBASE, GADA, and IS*, vol. 4803, Springer Berlin / Heidelberg, 2007, pp. 24-40-40.

[4] H. F. Fernández, E. Palacios-González, V. García-Díaz, B. C. Pelayo G-Bustelo, O. Sanjuán Martínez, and J. M. Cueva Lovelle, "SBPMN — An easier business process modeling notation for business users," *Computer Standards & Interfaces*, vol. 32, no. 1-2, pp. 18-28, Jan. 2010.

[5] M. zur Muehlen and J. C. Recker, "How Much Language is Enough? Theoretical and Practical Use of the Business Process Modeling Notation," Berlin :, 2008.

[6] J. Saltzer and F. Kaashoek, *Principles of computer system design: An introduction*. Burlington, MA :: Morgan Kaufmann,, 2009.

[7] L. Biggiero, "Sources of Complexity in Human Systems," *Nonlinear Dynamics, Psychology, and Life Sciences*, vol. 5, no. 1, pp. 3-19, Jan. 2001.

[8] J. J. Padilla, A. Sousa-Poza, A. Tejada, and S. Kovacic, "Towards a Diagnostic Framework for Understanding Complex Situations," Old Dominion University, Norfolk.

[9] J. H. Mikkola, "Modularity assessment of product architecture: Implications forsubstitutability and interface management," presented at the DRUID's Nelson and Winter Conference, Aalborg, Denmark, 2001.

[10] I. Hassan and M. Mitchelmore, "The Role of Abstraction in Learning about Rates of Change," Pymble N.S.W., 2006.

[11] M. Gupte, P. Shankar, J. Li, S. Muthukrishnan, and L. Iftode, "Finding hierarchy in directed online social networks," 2011, p. 557.

[12] F. Buschmann, R. Meunier, H. Rohnert, P. Sommerlad, and M. Stal, *Pattern-Oriented Software Architecture Volume 1: A System of Patterns*. Chichester [u.a.]: Wiley, 1996.

[13] Google Inc., "Google Earth," 2011. [Online]. http://earth.google.com/support/bin/static.py?page=guide.cs&guide=22370&topic=22652&answer=148122. [Accessed: 25-Aug-2011].

[14] M. La Rosa, A. H. M. ter Hofstede, P. Wohed, H. A. Reijers, J. Mendling, and W. M. P. van der Aalst, "Managing Process Model Complexity via Concrete Syntax Modifications," *IEEE Transactions on Industrial Informatics*, vol. 7, no. 2, pp. 255-265, May 2011.

DEPLOYMENT OF NEW TECHNIQUES FOR SEARCHING, SORTING, PROCESSING AND PRESENTING DATA

Jari Soini, Pekka Sillberg, Petri Rantanen
Department of Software Engineering
Tampere University of Technology - Pori
P.O.Box 300, FIN-28101 Pori, Finland
Tel: +358 40 8262890; fax: +358 2 6272727
e-mail: jari.o.soini@tut.fi

ABSTRACT

As the amount and diversity of electronic data and information grow, ever more efficient data processing is required. The increase in the amount of information and the ensuing distribution of data processing capacity creates challenges for searching, sorting, processing and presenting data. More sophisticated search techniques such as associative and semantic search engines offer a partial solution, but diversified information distribution and the presentation in various user interfaces remains problematic. The ongoing two-year Adaptive Software Services (MOP) project will research software technologies that can enable context-based adaptive user interfaces and techniques to compose distributed software services will also be studied. This paper briefly presents the selected research themes and the preliminary results of the study.

1 INTRODUCTION

The increase in the amount of information and the ensuing distribution of data processing capacity creates challenges for searching, sorting, processing and presenting data. This commonly recognized situation can be considered the starting point for the research problem in this study. Moreover, the amount of electronic data and information and the growing diversity in types of information seen today require ever more efficient data processing. As a result, more sophisticated and intelligent search techniques, such as associative and semantic search engines have been developed. These offer a partial solution, but diversified information distribution and the presentation in various user interfaces remains problematic. Moreover, in the current IT age, it is very common that the required information is situated in scattered locations on different databases, registers and public data sources, so it must be possible to generate and combine technologies for retrieving the desired data On the other hand, the positioning of distributed services enabled by various types of players and servers is crucial, representing a useful development trend for ITC technology and for software service providers.

As a result of this situation, the research project Adaptive Software Services (Mukautuvat Ohjelmisto Palvelut - MOP in Finnish) was initiated at the beginning of 2011 by the Tampere University of Technology (TUT) [1] with a consortium consisting of three Finnish ICT companies and the City of Pori. The project also involves international research co-operation with selected research organizations. Universities from Slovenia (Maribor University), Germany (Christian Albrechts University) and Japan (Keio University SFC), and also the National Institute of Information and Communications Technology (NICT) pf Japan took part in the research, contributing their specialist knowledge of the particular research sectors studied in the MOP project. The general goal of this two-year research project is to research software technologies and also deploy new techniques that can be useful in solving the problems and challenges described above i.e. searching, sorting, processing and presenting data. The research problem is approached with the help of three separate, but interlinked, research topics – *semantic meta-data, dynamic user interface, and adapted software services*. The aim is to find solutions to the research problem by studying and combining the research results. The research topics and objectives are described briefly in Section 2, the preliminary research results in Section 3, and Section 4 summarizes the paper.

2 RESEARCH THEMES AND GOALS

The MOP project focuses on cross-platform solutions that provide software components and composed services that can be adapted in run-time for the preferred context or situation. The goal is to solve the problems and challenges faced by software developers when trying to create software for multiple platforms and devices that may not be natively compatible with each other while still providing similar user experience on all target devices. The technologies studied in this project are not limited to desktop or mobile clients, and the technologies can also be web-based, server-based or hybrid technology. During the project a prototype implementation of a system with the capability to adapt to the user's preferred context is planned using one or more of the technologies studied. A key element of this system is the source of the information, which in this case is to be

provided by Keio SFC, and it will be based on Kansei concept [2]. The prototype implementation is to be integrated with a consumer targeted search media service provided by the corporate partners of the project. Below, the three main research topics that will be used to approach the research problem in this project are described in brief.

2.1 Semantic Meta-data

The growth of the Internet and the information that is available electronically has created challenges for information processing and retrieval. This phenomenon has made it more difficult to find the desired information. Therefore when searching for information, one of the most important issues is recognizing the context of the information. The importance of context can be easily realized by running simple web searches – often the search terms or words match exactly, but the information content itself is not what was wanted. One solution to this problem is semantic computing, but often describing the meaning of words or context is by no means a trivial task. Simple keyword-based search services are being replaced by semantically associative search engines that depend heavily on metadata. This raises the question of how to gather or generate this metadata.

One of the main goals is of the MOP study and determine if - and how – Kansei-based methods and technologies [see e.g. 2, 3, 4] can be used to create usable metadata for information retrieval. Kansei-based systems try to provide a link between the user's preferred context and the data stored in local databases or in various information sources globally. The scope of research will be on the applicability of Kansei-based technology, and the aim is to find out how the concept and its functions [2] can be implemented in the field of information processing. This research includes both the initial creation of metadata – Kansei words [4] and matrixes [5][6] – and the question of how to adapt the metadata or the search algorithms that are used to match the user's requirements better. The general goal of the study is to prove the usability and applicability of Kansei in the context of information gathering (data mining), processing, and retrieval.

A key element of this system is the source of the information, which in this case is to be provided by Keio University (SFC), and based on Kansei. Research has already started on previous projects [7] conducted by TUT Pori Unit in co-operation with Keio University. The principles of Kansei Engineering [8] have proven successful in product design [9], but how well the same principles work in practice in the field of information technologies remains to be seen. However, Kansei seems a promising candidate for solving the challenges mentioned above.

2.2 Dynamic User Interface

One of the main focuses of the project is to study technologies that can enable context-based adaptive user interfaces and the project examines the distribution options for an interface that can be adapted according to the situation. Nowadays, fluent and impressive user interfaces (UI) have become more important, especially in mobile applications. This creates challenges for updating and modernizing UIs. It is also important to keep the software up-to-date as updates contain bug fixes, additional features, and security updates. A proper distribution model for the software and updates is also needed to ease the user's workload. But even then the end user may not care about – or want to – install and update software. As a whole, this can make software management and maintenance more difficult. So the research question in this case is that could there be a way to hide or even completely disable unwanted program features. In broader terms, could the UI be made adaptable and flexible enough to support the user's real needs?

On this basis, the selected research topic is to study whether we can have *adapter software* on the target device that has the capability to interpret a UI file. A suitable analogy for our adapter software with regard to UI files is like a WWW browser with regard to web pages. The main idea is that the adapter can update its UIs and extend functionalities by retrieving newer specification files from a system managing the UIs and their distribution. Overall, the goal is to study how to provide better updateability of individual software components and to optimize network usage while researching the possibility of allowing end user customization of user interfaces. This part of the project - Dynamic user interface - was scheduled for implementation first and the preliminary research results are described in more detail in Section 3.

2.3 Adapted Software Services

The third research topic relates closely to Service Oriented Architecture (SOA) [e.g.10]. It is hoped to exploit SOA in order to achieve more open and more easily integratable systems. When service oriented architecture is adapted to enhance the compatibility of the operational processes of several independent players, it is believed that significant benefits can be gained through the redistribution of functions, i.e. work. Service interoperability may demand coherence or harmonization services so as to achieve the common understanding and situational awareness required by the common process. The more complex the composed services in question are, the more likely it is that their implementation and the location of the runtime environment will be distributed between various players and service providers. However, the continual increase and fragmentation in the amount of information creates challenges for searching, sorting, processing and presenting data, as well as generating a need for more sophisticated search methods. These challenges will be met by the distribution of diverse information (protocols) and visualization (interfaces) in different devices offered by future associative and semantic search engines.

In this project, the research deals with extensive, multifaceted composed software services that can be adapted to each situation, which are based on distributed functionality as well as the positioning of different kinds of services on different levels of the architecture. Techniques to compose distributed software services will also be studied by means of a prototype system. The distribution options for the composed service that is formed and the interface that can be adapted according to the situation will be examined in this project from various viewpoints. A prototype implementation of a knowledge search service and an adaptive user interface will be created in the research project. From an economic point of view, the project will result in new knowledge about the practicability and distribution principles of software service composition and mobile technologies, which will enable all sizes of enterprises - including the SME sector - to impact global ecosystems.

3 RESEARCH RESULTS SO FAR – ADAPTIVE USER INTERFACE IN MOBILE APPLICATIONS

This section describes the preliminary results of the first part of the project, aimed at studying techniques and technologies that can be used in a dynamic user interface. The approach for this research topic is to use *adapter software* [11] for retrieving UI embedded with content and even for retrieving software updates. Earlier we used the analogy of adapter software relating to the UI files it uses as like a WWW browser with regard to web pages. The main difference between adapter software and a typical WWW browser is that the adapter software may use more specific and extended device capabilities than the WWW browser. The adapter software presented here uses declarative UI languages, such as XML User Interface Language (XUL) [12], Extensible Application Markup Language (XAML) [13], and Qt Meta-Object Language (QML) [14, 15]. Adapter software is to be created for each target platform but all different adapters may use the same UI/application description. It is up to the adapter software whether or not to follow the UI conventions (e.g. look-and-feel) of the target platform. This allows the software developer to create the application once but run it on multiple platforms.

The following two example cases illustrate the possibilities of adapter software. The first (Figure 1) concerns data binding and the second (Figure 2) runtime software updates. Figures 1 and 2 are screenshots from the development prototype. In both pictures an Integrated Development Environment (IDE) made with PHP Hypertext Preprocessor (PHP) running on Apache HTTP Server on a Linux server is shown on the left. The IDE in the picture does not present the final look and feel of the tool, but it gives a good impression of what can be done with the IDE. Currently the IDE is a simple way of creating and updating forms used on an end-user device, which in this case is an Android phone (on the right in Figures 1 and 2) that has adapter software made with Java. In the IDE – or developer view – the user can drag items from the list of Controls using the mouse and drop them freely anywhere in the preview window shown next to the Controls list. In this case three Labels, two Textboxes, one ComboBox and one Button (Submit) have been selected and the items have been named.

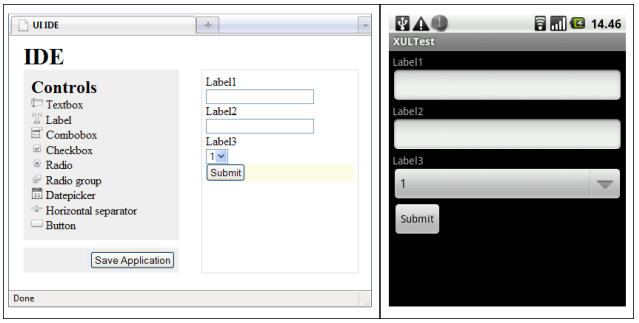


Figure 1: *Developer view (left) and user interface (right) [11]*

After the desired items have been chosen from the list of Controls, the Save Application button can be clicked to generate the actual code to be transferred to the client device. In this case XUL is used as the declarative language to represent the UI components, but for example the output format could be varied for any language using XSL transformations. This way the IDE could be used to create different code for different devices if so needed. As the declarative code is basically just text – no compilation to binary code is needed – it can be used for transfer to the client device using standard network protocols. The current prototype system uses a client-server model in which the client device checks for a new version of the user interface every time the user interface needs to be shown. In future versions, the functionality may be extended to provide push-functionality on the server side or to allow only the sending of updated elements of the user interface. Usually the amount of data that needs to be transferred is on a par with simple web pages (in this case 1-4 kilobytes), and because of their textual type they can be effectively compressed to save bandwidth.

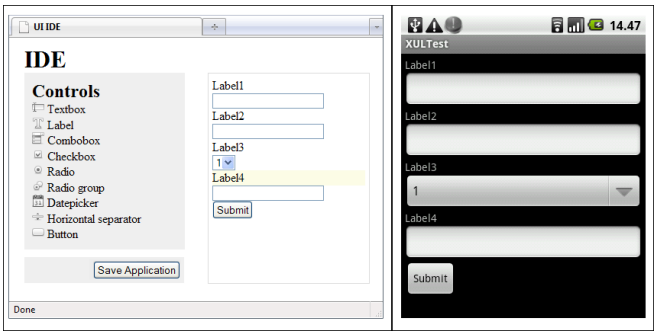


Figure 2: *Developer view (left) and user interface (right) [11]*

The difference between Figures 1 and 2 is that in the latter, a new Label and Textbox have been added to the preview window and the modifications have been saved for displaying on the device. The interface can be freely edited and when saved the changes will take effect on the device immediately. However, it should be noted that the changes cannot be downloaded on the client device unless a network connection is available. Naturally, in that case, the form cannot be submitted back to the server either so the user is unable to send back data using an old interface. This minimizes the need to worry about different software versions returning data in different formats. A more detailed description of the technologies behind the prototype above is presented in paper [11].

As observed during the study, the IDE and adapter software together would seem to provide a framework where applications may be managed with relatively little knowledge of software development. The ultimate goal of the framework studied here is to allow users to create applications, distribute applications to desired clients, use applications on any device with compatible adapter software, and to gather data. The research on this topic is ongoing, and will study how to add reporting applications that can be used to show the collected data to the users.

4 SUMMARY

This paper deals with a widely recognized phenomenon – information overload. Nowadays there is more information available electronically than ever before and therefore the challenge has arisen of how to find and compile the required and right kind of data from the mass of information located in various sources. The paper describes ongoing research activities as well as the preliminary results of the MOP research project, which aims to study software technologies and deploy new techniques that can be useful for solving this challenge. The project approaches the issue from three separate angles - *semantic meta-data*, *a dynamic user interface*, and *adapted software services*.

References

[1] Tampere University of Technology (TUT), Pori Unit, http://www.pori.tut.fi, Retrieved Aug. 4 2011.

[2] Y. Kiyoki, X. Chen. A Semantic Associative Computation Method for Automatic Decorative-Multimedia Creation with "Kansei" Information, *In Proc. APCCM '09*, Vol. 96, Australian Computer Society, Inc., ISBN 978-1-920682-77-4, 2009.

[3] S. Uemura, H. Arisawa, M. Arikawa, Y. Kiyoki. Digital Media Information Base, IEICE Transactions on Information and Systems, Vol. E82-D, no. 1, pp. 22-33, 1999.

[4] H. Kimoto. An Image Retrieval System Using Impressional Words and the Evaluation of the System, Special Issue on Computers and the HumanProcessing Society of Japan, Vol. 40, no. 3, pp. 886-898, 1999.

[5] Y. Kiyoki, T. Kitagawa, T. Hayama. A metadatabase system for semantic image search by a mathematical model of meaning, ACM SIGMOD Record, Vol. 23, no. 4, pp. 34-41, 1994.

[6] Y. Kiyoki, T. Kitagawa, K. Kurata. An adaptive learning mechanism for semantic associative search in databases and knowledge bases, Information modeling and knowledge bases VIII, IOS Press, pp. 345-360, 1998.

[7] H. Jaakkola, J. Soini, J. Leppäniemi. Seamless Solutions and Mobile Connectivity – Towards location independent and flexible system architectures. In Jaakkola H. (ed.), Selected Topics on Distributed Disaster Management: Towards Collaborative Knowledge Clusters. Tampere University Press, 2008.

[8] M. Nagamachi. *Kansei/Affective Engineering*, CRC Press, ISBN 978-1-4398-2133-6, 2010.

[9] The Institute of Technology at Linköping University – LiTH's Kansei Engineering Pages, http://www.kansei.eu, retrieved August 17, 2011.

[10] M. Papazoglou, D. Georgakopoulos. Service-Oriented Computing, Communications of the ACM, Vol. 46, no. 10, pp. 24-28, 2003.

[11] P. Sillberg, J. Raitaniemi, P. Rantanen, J. Soini, J. Leppäniemi. Flexibly managed user interfaces for mobile applications, *In Proc. IADIS 2011 Multi Conference on Computer Science and Information Systems*, Rome, Italy, 22-24 July, 2011.

[12] Mozilla Foundation. XML User Interface Language - MDC Doc Center, http://www.mozilla.org/projects/xul, Retrieved May 6 2011.

[13] Microsoft Corporation. Extensible Application Markup Language Overview, http://msdn.microsoft.com/en-us/library/ms752059.aspx, Retrieved May 6 2011.

[14] Nokia Corporation. Qt Meta-Object Language - Introduction to the QML language, http://doc.qt.nokia.com/4.7/qdeclarativeintroduction.html, Retrieved May 6 2011.

[15] Nokia Corporation. Qt Quick, http://qt.nokia.com/products/qt-quick, Retrieved May 6 2011.

TOWARDS DECISION SUPPORT SERVICES FOR COLLABORATIVE WORK

Jari Leppäniemi
Department of Software Engineering
Tampere University of Technology - Pori
P.O. Box 300, FIN-28101 Pori, Finland
Fax: +358 2 6272727
e-mail: jari.leppaniemi@tut.fi

ABSTRACT

This paper introduces the steps of a service design process for disasters and emergencies management in a service oriented ecosystem. The main concern of the users of these interoperable services is to have sufficient information for decision-making in different situations during the emergency management lifecycle. The basic concepts of the emergency domain, service design principles and the target system architecture are presented here.

1 INTRODUCTION

In a disaster, several kinds of participants are involved in the management of the situation [12], [22]. Participants typically represent state-based authorities (rescue, firefighting, policing, health and emergency, safety, etc) and nowadays very often also the public and several types of volunteer organizations. Also non-governmental organizations (NGOs), international coordination agents, ICT solution providers, EM and NGO professional and academic communities may be in crucial positions during the emergency management lifecycle [6]. An accurate and timely situational picture is a major problem in most cases.

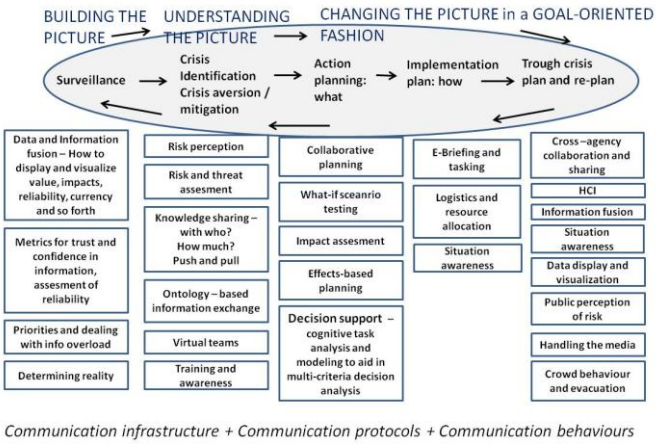


Figure 1: *The Emergency Management process. [3]*.

An example of the emergency management process was given by Carver and Turoff in [3]. They described the complexity of the environment and challenging context of the emergency management system. The emergency management process is illustrated in Figure 1. According to Carver and Turoff [3], the design of emergency management information systems should be based on human – computer interaction. A proposal for a checklist when comparing EM systems has been given by Grant [9]. Feng and Lee discussed next-generation service-oriented emergency management systems in [8].

There might also be other types of challenges besides technical - such as organizational and sociological issues. An example of a sociological-type challenge is the lack of common vocabulary and/or shared semantics between response organizations. The same may exist between organizations and the public [15].

The structure of this paper is as follows: The following section gives a short review of the concepts of the emergency domain. We introduce our approach to the information needs related to the different phases of the emergency management lifecycle and the type of emergency functions at different levels of the management dimension. Then we present the reference system architecture of a service oriented emergency management ecosystem and the proposed design process for the services. Finally, we state some topics for further research.

2 EMERGENCY MANAGEMENT LIFECYCLE AND ECOSYSTEM

The emergency management cycle is presented in Figure 2 [1], [16]. It divides the lifecycle of an emergency into four different phases (centermost ring). Before the main impact, there is usually time for assessment and planning in the *prevention and mitigation* phase and sometimes for starting preparations for what is expected and might be coming in the *preparation* phase.

The type of required information varies depending on the phase of the emergency management cycle. According to [23], the properties of a good emergency alerting system are: support for second languages, spontaneity, automated operation, locality, ubiquity and non-intrusiveness.

The good performance of an alerting system is a crucial feature. In [13] an example of the performance monitoring of

an emergency call-taking system is presented. They used a neural network (Self Organizing Map, SOM) based method to detect anomalies in emergency call data.

The spectrum of services and applications in use ranges from measurement devices capable of giving automated alarms to sophisticated and advanced anticipation applications using data mining and grid technologies etc. in order to forecast for example mudflows or weather.

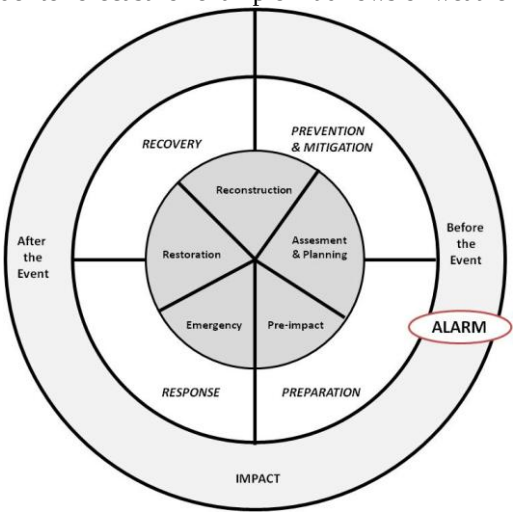


Figure 2: *The emergency management cycle (adapted from [1]; [16]).*

The lifecycle phase after the impact is *response* (life saving, protection of the environment, assets, and property). The next phase is *recovery*. During the recovery phase, the community returns back to normal operation [1]. The types of information systems and applications used in these lifecycle phases are thus very plentiful and the timeliness of the information that is required and available varies considerably. For example, the need for real-time information when planning possible actions for emergency training in the prevention and mitigation phase is much smaller than when rescuing people in a hectic response phase [1]. It is also important to differentiate between the types of information intended for authorities and the information directed to the public.

The type and function of the emergency task (i.e. rescue, police, health, etc.) and its location at functional level is also important. They place demands on the information required at the decision points of disaster management. An example of a service-based [7] system meant to support medical information exchange in real time is described by Hauenstein et al. in [10]. Cataldo and Rinaldi [4] gave another example of knowledge sharing using the Peer to Peer (P2P) approach and semantic web services [25]. Another example of the design of a distributed system for combining real-time data with predefined information sources in order to anticipate possible crisis or emergencies (WIPER) is described by Schoenharl et al. in [21].

Emergency functions could be classified in different levels. One example is presented in [16]. They divided emergency functions into three levels: *strategic*, *operational*, and *tactical*. At the strategic level, the functions focus on the top-level and long-term co-ordination of emergency management. At the tactical level, most of the functions and activities are performed in the response and recovery phases (search & rescue, firefighting, first aid and the longer term clean-up activities in the recovery phase). Management and decision support functions and activities for co-ordination, prioritization, and supporting tactical level functions are at operational level. At operational level, the quality and sufficient quantity of timely, accurate, and reliable information and data are crucial success factors [16]. Figure 3 presents an overview of the system architecture for a service-based [7], [17] and [18] emergency management knowledge ecosystem [14].

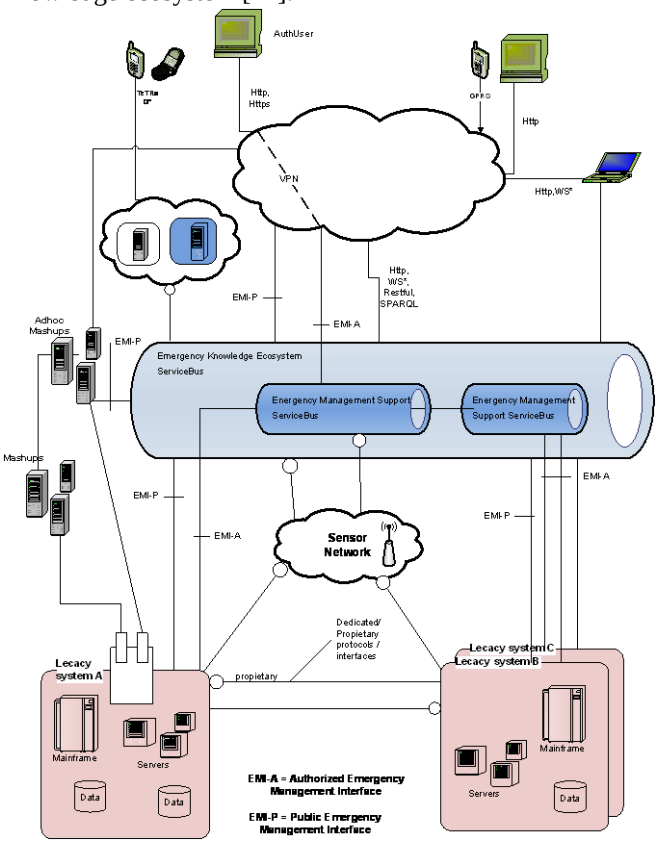


Figure 3**:** *Service oriented system architecture for Emergency Management Knowledge Ecosystem [14].*

To cite [14]: "The main interfaces are an authorized interface EMI–A for emergency management services and a public interface EMI–P for the general public and non-authorized users of the system(s). The concept and function of a service bus [7] is crucial for this ecosystem so its functionality should be fault-tolerant and distributable. The legacy systems (authorities, research organizations etc. located at the bottom of Figure 3) collaborate with a large service-based architecture via open standards [19] and [24]. They can preserve their proprietary interfaces and they can open new standard interfaces implemented by means of an

appropriate technology like WS –*, RESTful, http [11], [24] and [25]. Sporadic and temporary users can be connected as an information service provider or to a normal user via a public and open service interface like EMI–P in Figure 3. A more permanent connection to the ecosystem is established using either the common emergency service bus (EMI–P) or the more controlled interface for authorized users (EMI–A), which supports the coordination and supervision of shared business processes"[14].

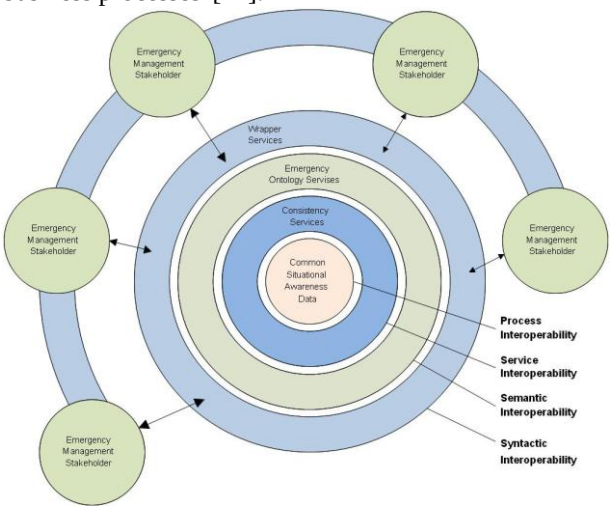


Figure 4: *Collaboration support for stakeholders of an emergency management ecosystem utilizing interoperability services [14]*

Figure 4 illustrates an interoperability (see also e.g. [2] and [5]) approach, which it is planned to achieve by introducing several sub-services to be used as part of a service composition. These sub-services could be called wrapper services for syntactic level interoperability, ontology services for semantic level interoperability, and consistency services for service level interoperability. Only after using these services will the various stakeholders - participating in the collaborative situation - be able to have a common situational awareness.

3 SERVICE DESIGN PROCESS FOR DECISION SUPPORT

In order to support the stakeholders of the service-based ecosystem and the decision makers, good situational awareness is needed. One challenge is to identify the critical information sources and requirements in dynamic situations. There are (at least) three different approaches to this problem:

1. Model the processes and identify the information flows in advance (e.g. a response plan).
2. Model the processes and identify the information flows afterwards (e.g. accident investigation report).
3. Identify and utilize the critical information in real time (i.e. during the crisis situation, e.g. ad hoc information provided by the public [20]).

The design process of the decision support services for a reference system architecture comprises several steps at different levels. These steps may include both domain-specific conceptual thinking and the evaluation of technology alternatives. The third approach (above) could be partly handled in the service-based ecosystem by lightweight development processes utilizing what are known as web 2.0 technologies - mashup platforms [20], RESTful web services [11], [24] and [25], etc. A draft proposal for an upper level knowledge management service design process utilizing the first two approaches above could be:

A. Identify (main) decision points:

1. Choose some known disaster type.
2. Identify the main stakeholders, their tasks, and concerns.
3. Identify the main decision situations and decision makers at different levels (strategic/operational/tactical) and in corresponding phases of the lifecycle (preparation, response, recovery, prevention and mitigation).
4. Analyze whether it is possible for decisions to be made also by others (persons / roles) and/or at a different moment of time – if the required information and decision-making power were available to them.
5. Document the options using B1 – B4.

B. For each decision point (A3, A4) try also to discover:

1. What data, information and knowledge are most essential for this situation?
2. Is it possible to support the analysts and decision makers merely by presenting *only* that information?
3. Do the currently used decision data and information need prioritizing and/or reduction?
4. Is it possible to enhance and/or accelerate the decision-making process by filtering, composing, or otherwise preprocessing the data before showing it to the analysts and decision makers at the tactical, operational, or strategic level?

The decision points can usually be identified by interviewing people and investigating guideline documents, current practices, plans and process models. It is extremely important that the processes are analyzed and modeled for collaboration from the viewpoint of interoperability. Earlier practices may have been sub-optimally managed due to tight organizational limits. The optimal solutions may need several iterations before the desired interoperability for enhanced collaboration and efficiency can be achieved. Modeling is affected by the technology solutions that are available and affordable. The same is true for improvements.

We are currently expanding our research related to service design process modeling to cover more formal methods (BPMN, EPC, etc.) in the ongoing "Adaptive Software Services" research project, (MOP in Finnish).

4 SUMMARY

The main concepts of the collaborative domain - emergency management - were described in this paper with a small review of related research. The paper introduces service-based reference system architecture and abstracted interoperability services for collaboration support. Finally we described our upper level service design approach to identifying crucial decision points and the optimized situational information to be used at these moments.

References

[1] D. Alexander. *Principles of Emergency Planning and Management*, US: Terra Publishing, Oxford University Press, 340 p.

[2] ATHENA Interoperability Framework, retrieved 29/02/11, http://www.modelbased.net/mdi/index.html and http://interop-vlab.eu/ei_public_deliverables/athena-deliverables.

[3] L. Carver and M. Turoff. "Human Computer Interaction: The Human and Computers as a Team in Emergency Management Information Systems", *Communications of ACM* 50 (3), pp. 33–38, 2007.

[4] A. Cataldo and A. Rinaldi. "Sharing Ontology in Complex Scenario using a Peer-To-Peer Approach", *International Journal of Computer Information Systems and Industrial Management Applications (IJCISIM)*, Vol. 1, pp. 92– 109, 2009.

[5] P. Chen. "Architecture-Based Interoperability Evaluation in Evolutions of Networked Enterprises," C. Bussler & al. (ed.), *BPM 2005 Workshops*, LNCS 3812, Springer-Verlag, Berlin, Heidelberg, pp. 293–304, 2006.

[6] EIIF(W3C) Incubator Group Report 8/ 2009, Retrieved 29/02/11, http://www.w3.org/2005/Incubator/eiif/XGR-EIIF-20090806/

[7] T. Erl. *Service-Oriented Architecture: Concepts, Technology, and Design*, Prentice Hall PTR, 760 p, 2005.

[8] Y-H. Feng and C.J. Lee. "Exploring Development of Service-Oriented Architecture for Next Generation Emergency Management System*". IEEE 24th International Conference on Advanced Information Networking and Applications Workshops*, pp. 557–561, 2010.

[9] T.J. Grant. "A Checklist for Comparing Emergency Management Information Systems". In *Proceedings of the 5th International Conference in Information Systems for Crisis Response And Management*, Washington DC, USA, pp. 752–763, 2008.

[10] L. Hauenstein, T. Gao, TW. Sze, D. Crawford, A. Alm and D. White. "A Cross-Functional Service-Oriented Architecture to Support Real-Time Information Exchange in Emergency Medical Response". In *Proceedings of the 28th IEEE EMBS Annual International Conference New York City*, USA, 3, 2006.

[11] IETF, the Internet Engineering Task Force, retrieved 14/02/11, http://www.ietf.org/.

[12] H. Jaakkola, J. Soini and J. Leppäniemi. "Seamless Solutions and Mobile Connectivity – Towards location independent and flexible system architectures". In *Publication 12, pp. 1–17, Tampere University Press*, 2008.

[13] P. Klement and V. Snášel. "SOM Neural Network – a Piece of Intelligence in Disasters Management", *International Journal of Computer Information Systems and Industrial Management Applications*, Vol. 2, pp. 243–251, 2010.

[14] J. Leppäniemi. "Domain Specific Service Oriented Reference Architecture. Case: Distributed Disasters and Emergency Knowledge Management", *International Journal of Computer Information Systems and Industrial Management Applications*. ISSN 2150-7988 Vol. 4, pp. 043–054.

[15] B.S. Manoj and A.H. Baker. "Communication challenges in emergency response", *Communications of ACM* Vol. 50 (3), pp. 51–53, 2007.

[16] OASIS EU-FP6 project, retrieved 15/02/11, http://www.oasis-fp6.org/

[17] OASIS, Reference Architecture for Service Oriented Architecture, retrieved 15/02/11, http://docs.oasis-open.org/soa-rm/soa-ra/v1.0/soa-ra-cd-02.pdf

[18] Reference Model for Service Oriented Architecture 1.0, retrieved 19/02/11, http://docs.oasis-open.org/soa-rm/v1.0/soa-rm.html.

[19] OPEN Group, retrieved 23/02/11, http://www.opengroup.org/projects/soa/.

[20] L. Palen, S.R. Hiltz and S.B. Liu. "Online forums supporting grassroots participation in emergency preparedness and response", *Communications of the ACM*, vol. 50 (3), pp. 54 – 58, 2007.

[21] T. Schoenharl, R. Bravo and G. Madey. "WIPER: Leveraging the Cell Phone Network for Emergency Response", *International Journal of Intelligent Control and Systems*, Vol. 11(4), 2006. http://www.nd.edu/~dddas/Papers/IJICS_WIPER2.pdf.

[22] J. Soini, J. Leppäniemi and H. Jaakkola. Towards Seamless Collaboration in Distributed Disaster Knowledge Management. Proceedings of the 11th International Multiconference INFORMATION SOCIETY – IS 2008, October 13th – 17th, 2008 Ljubljana, Slovenia.

[23] P. Verma and D. Verma. "Internet Emergency Alert System". In *Proceeding of Military Communications Conference*, Hawthorne, NY, USA, 2005.

[24] W3C, retrieved 23/02/11, http://www.w3.org/.

[25] WS-*, Web Service specifications, retrieved 24/02/11, http://en.wikipedia.org/wiki/WS-*.

ROUTES TO KNOWLEDGE SOCIETY INCLUSION: EXAMINATION OF THE N4C CASE

Maria Udén
Department of Business Administration, Society and Technology
Luleå University of Technology
S-97187 Luleå, Sweden
Tel: +46 (0)920-491 000
e-mail: maria.uden@ltu.se

ABSTRACT

In this paper the issue of routes towards knowledge society inclusion and the building of knowledge economies are examined. This is done via three different sub-cases identified within the EU FP7 project N4C (Networking for Communications Challenged Communities). Starting from the two separate theses that inclusion starts with either physical access or the social context, the conclusion is that inclusion is best understood as intersection of opportunities and obstacles that define each situation. One person or group may be successfully included in one instance but in other aspects stand outside the knowledge society

1 INTRODUCTION

"Networking for Communications Challenged Communi--ties: Architecture, Test beds and Innovative Alliances" was an EU FP7 STREP project that operated from May 2008 to April 2011. As indicated by the name, socio-economic concerns were included along with the focus on technical competitiveness. Both aspects guided the technical strategies. This has to do with the project's genesis, which is embedded in an analysis emphasising broader social inclusion in technology research and development as a condition for sustainable development [1] and, initiatives within a community of Sámi reindeer herders in northwest Sweden [2]. As complement to the technically oriented work plan, N4C included business model development and, social sciences in a broad sense such as Science and Technology Studies (STS). This paper addresses the issue of knowledge society inclusion, or the emergence of knowledge economies, based on high level analysis of the N4C process, in accordance primarily with an STS and sociological stance. Earlier works of Udén [3,4] and Udén and Wamala [5] that have not previously been presented together are here merged and further developed. Below, after setting the frames of the analysis and the case, selected sub-cases are discussed in the light of two different research strands addressing knowledge society inclusion.

2 POSSIBLE ROUTES TO INCLUSION

For inclusion of marginal groups and locations in the knowledge society, a crucial concern is the order of actions: what should be the first goal, the second, and so on. One type of investigation into this issue takes stance in the success-stories of the global economy. Pursuing this route, David and Foray [6] noted the following decisive characteristics for progress towards a knowledge economy;

1. *The acceleration of knowledge production* with the emergence of knowledge-based communities.
2. *The rise of intangible capital at macroeconomic level*, basically the human capital.
3. *Growing speed of innovation* including both formal and layman environments.
4. *The revolution in instruments of knowledge* where technologies for knowledge and information production and dissemination offer remote access, enormous quantities of information and, enhancement of creative interaction.

The topic can also be addressed via experiences in societies and nations that are engaged in processes towards fuller inclusion. Building on investigations along this route in Mongolia and similar settings in the African continent, Johnson et al [7] identify "the four pillars of the knowledge society";

1. *ICT and connectivity*: the backbone of a knowledge society is a well developed, well-maintained and affordable information infrastructure.
2. *Usable content*: affordable, available, timely, and relevant.
3. *Infrastructure and deliverability*: airports, railways, roads, trucks, warehouses and physical addresses of people.
4. *Human intellectual capability*: people who can create as well as utilize knowledge to ensure not only sustainability, but also prosperity.

Investigations at a more detailed level reveal that women's conditions are not obviously furthered in quests for knowledge society inclusion or building of knowledge economies. A report by Orbicom sets the scene stating that worldwide the capacity of women to engage in the knowledge society is under-developed and under-utilized.

Women are at risk of becoming increasingly marginalized and, much more than access to information technology is at stake. Women need access to new technologies but, also to education, entrepreneurship and employment opportunities, and the ability to participate fully in knowledge-based activities [8]. In terms of gender equality or the promotion of women's conditions there is no automatically emancipating power in the progress of the knowledge – or information – society [9].

3 INGOING PARAMETERS OF THE N4C PROCESS

The topic of knowledge society inclusion is suitable to assess in the light of N4C because of the project's inherent ambitions and perhaps even more, because of its genesis. Information can be gained that contributes to the formulation of problem areas and analytical frames. The project was characterized by a broadly endorsed good will to address the issue of access to information and communication technology (ICT). The drive was embedded in: 1. A theoretically founded identification of broader participation in technology development as a condition for sustainability [1] and 2. Inititiaves among semi nomadic reindeer herders in a remote location in northwest Sweden, including for better conditions for women in the herding business [3]. Reindeer husbandry exists around the Arctic, in particular the Eurasian continent. The reindeer herder community which was involved in N4C is semi nomadic and move their herds between locations as much as 400 km apart during the yearly cycle. The herders' access to ICT, such as cell phone coverage, is patchy. To several of the camps where the families during summer, not even radio broadcasts reach. At the other end of the access situation, in the winter homes, the herders may even be able to choose between different service providers and means of delivery for internet access, telephony, television, radio broadcast and so forth. It could be claimed that this should be enough. However, the husbandry industry is in no instance located to the permanent settings of the winter homes, why this access cannot be utilized as an integrated part of the industry routines. Furthermore, expectations increasingly fall upon the individual citizen, to handle errands with authorities as well as commercial transactions and other, via the Internet. In the same pace, the herders and their families face new challenges when not being able to see to their interests via the expected and often only available media.

The driving force behind the process leading to N4C was that not only can such circumstances be over won with a nomadic-friendly networking technology. The traditional life style, it was thought, can even be strengthened through new possibilities for business strategies, social networking, and development of such as better adapted schooling systems. It was also thought, that the technology has major potential for impact in wider areas including but not confined to remote and rural locations, and this was of course the ground for the application to the EU FP7. However, the focus here is on the above specified background and a set of results in the newly concluded project process. This provides a concrete material while the outcomes of the project impact at a more general level out of obvious reasons are yet to unfold.

4 CASE A: PHYSICAL ACCESS AS MAIN ISSUE

According to David and Foray [6] the distinction frequently drawn between information society 'haves' and 'have-nots' is overly simplified. The issue is not about a digital divide that can and should be overcome by providing universal access:

"…the more difficult and in a sense more fundamental problems are not simply those of providing greater technological access to information streams. Rather, they involve furnishing people with the cognitive capacities and intellectual frameworks than enable humans to interpret, select and utilise information in ways that augment their capabilities to control and enhance the material circumstances and qualities of their existence." (p. 10)

The N4C case cannot confirm such an unconditional assertion. It can be noted that nomadism and remote locations are shared factors of Mongolia and the reindeer herders who have been involved in the N4C process. Comparing the routes to knowledge society inclusion lay out by respectively David & Foray and Johnson et al, the latter better identify steps, and an order of steps, that mirror the challenges of the reindeer herders. First they place *ICT and connectivity*, next *Usable content* and as third in the list; *Infrastructure and deliverability*. Only as step four we see human capacity. It is not that David and Foray [6] do not see a role for ICT in the success stories of the knowledge society. To them this role revolves around "the capacity of the new information technologies to enable better integration of knowledge through helping bring down the cost of transporting it". Interestingly, they so deliver the explanation to why ICT is a matter also in the reindeer grazing areas. The issue is access where information and knowledge are generated and, where retrieving information can in real terms make a difference. In sum, the aim is better integration of knowledge in reindeer industry, and less cost for transporting it and this is a matter of tangible aspects of space, time and physical access. Wamala [9] finds that the problem of non-access is usually perceived as being user-based. Whether with reference to inequalities, or ability, users are constructed as problematic. This is based on their ethnicity, gender, regional location, or class formation. Instead, Wamala suggests intersectional analysis of access instead of the people who supposedly 'have' or do not 'have' it. Depending on the situation a person may be a user of ICT, or not.

5 CASE B: TEST BEDS AS AN OPPORTUNITY

Within the FP7 scope, the FIRE initiative has an aim to assemble a federation of test beds for the purposes of Internet experimentation and tests of new paradigms and products. One aim of N4C was that after the project

finished, at least one of the two test beds built for the purposes of the project would be presented as an offer to this federation. A three level strategy for governing models of sustainable N4C test beds of different types was developed for the project:

- Level 1: A research test bed platform on similar level as during the N4C FP7 project
- Level 2: A small scale test bed for present research partners and for a few new clients
- Level 3: A large scale federated test beds in collaboration with FIRE network

In the N4C offer to the FIRE federation, the Slovenian test bed is offered for level 3 [10]. The Slovenian test bed will be run by the N4C partner MEIS d.o.o., which is a company that already sells services that require a high level of technical and organizational preciseness in the customer relation. A commercially viable technical and management standard has been developed by this partner, for making a stable offer of the type expected for the integration with the European Commission plans for federation. This implies services where the technical and procedural specifications and performance is guaranteed by the test bed owner. The attractive features of the Swedish Lapland testing opportunity is the very remote location – the area is significantly more complicated to access, and the possibility to embed testing in the normal activities of the reindeer herders and their families during their stays in the summer camps. This possibility is not open to any actor, as the camps are located in a nature reserve (national park) where only accredited reindeer herders have the rights to reside. The Swedish test bed will continue rather as a level 1 activity, integrated in the activities of the N4C partner Tannak AB. This company is owned by two reindeer herders who were involved already in the first processes starting 2001 and which emanated in the N4C consortium and project. One former N4C partner travels to the site already a few months after the N4C end, as a step within the newly started larger, IP (integrated) FP7 project SAIL. To an extent the situation is thus very satisfactory. Considering the restrictions of activities including research in the national park, it is not evident how permanent or large scale testing routines would best be developed.

None the less, the Swedish test bed not being base for an offer to the FIRE federation points, to some complications in terms of knowledge society inclusion The possibility of opening a test bed integrated in the territory of the reindeer herders and with them as test pilots and providers of the test services emerged as an idea already five years prior to the N4C application was produced. In the early 2000's the test bed industry was favoured in the regional policies of the N4C Swedish Lapland test bed was situated. This brings back to the criteria for achieving a knowledge society and building a knowledge economy. As described in Case 1 above, the overall situation of the reindeer herders supports the appreciation of Johnson and co-writers [7] that is: physical access is step one to prepare for inclusion. In relation to the test bed development, the situation reverses. The decisive access is that to a certain type of organizational and technical expertise, ranked by David & Foray [6] as the first step towards inclusion.

6 CASE C: WOMEN AND INCLUSION

Considering the concerns commonly expressed regarding gender equality and the inclusion of women in knowledge society processes, it is noteworthy that women have from start been deeply involved in furthering the N4C process. Women have also gained employment and business opportunities through the project. By no terms has the project been 'gender equal' as in a 50/50 distribution. Yet, out of the project's eight work packages leader, three were women. This is close to forty per cent. Among the twelve partners, four had appointed women as lead for their N4C operations. In all, this makes female representation slightly above the mean levels for EU with regards to management positions and technical education. Eurostat figures for Tertiary education graduates show that about twenty per cent of the students in the Computing and Engineering sectors were women while women made almost sixty per cent of the total student corps at this level [11]. This difference is part of the issue of *Occupational and sectoral segregation*. The Directorate-General detects gender segregation also with respect to management positions:

"Despite the fact that more and more women are highly qualified and the labour market participation of women is on the increase, they are still largely outnumbered by men in positions of responsibility in politics and business, particularly at the top level. The number of female managers in the EU has remained stable over the last few years, averaging 30%, and figures are even lower in a majority of Member States." (p. 9).

In the N4C case, it can be noted that indigenous women from a traditional semi nomadic community participated in this effort of FP7 research and development, namely the founders of the partner Tannak AB. This may be interpreted as a significant win, in relation to goals of broadening the social basis for recruitment to science and technology research and development environments. However, the detailed level of the process offers further vital facts. The Tannak AB founders are reindeer herders. In terms of gender distribution and prestige this is a profession quite comparable to science and engineering. The profession has the highest status possible in the reindeer herder group. It renders unique rights to land and water in the grazing areas that no other Sámi or other Swedish citizens can achieve. Starting their company Tannak AB and furthering the process towards N4C, these women again stepped into a statistically male dominated, symbolically masculine environment of technical research and development. This situation emphasizes the danger of using too coarse categories in social classification.

Already before entering this particular process the women involved in initiating and leading N4C came from what in

their respective contexts are high status male dominated positions. But notably, being able to defend one's original position may be a substantial outcome. This implication emerges from putting N4C in a wider context. At least, if we accept that, speaking with David & Foray [6] being included in innovations systems is close to the sole means to survive and prosper in the highly competitive and globalized economies of to-day. This condition mirrors the concerns expressed by Huyer & Hafkin [8] that the capacity of women is under-utilized and as a result women are in danger of being marginalized.

7 CONCLUSIONS AND DISCUSSION

The three N4C sub-cases presented in this paper provide information that can contribute to the formulation of problem areas and analytical frames. The theme is if physical access to Information and Communication Technology or social readiness, is the first item for initiating the sought development, and generally the order in which different features should be addressed. The sub-cases raise the following sets of inquiry and suggest their own respective answers:

1. The semi nomadic reindeer herders of the case often enjoy good ICT access in their winter residencies, but not in the grazing territory. This affects the potentials for impact in their businesses, among other factors. This situation supports the hypothesis that physical access comes first.
2. The Slovenian N4C test bed was presented as an offer to the EU FIRE federation. Since long the ambition for the Swedish test bed has been to achieve such a result. These circumstances point towards the impact of education level and other aspects of social organization as decisive for the building of a knowledge economy.
3. The project has been run by both women and men in leading positions. The ratio is at about the same or slightly better than EU average figures. All leading women entered via high status male dominated sectors and professions.

The second and third items point to the importance of socio-economic preparations for gaining the capacity to make full use of knowledge society processes and keeping up one's competitiveness in the global economy. None the less, this does not change that regardless of for instance education level, if a woman or man who is a reindeer herder wishes to develop her or his company through ICT supported routines, it remains impossible without networks that support the conditions of herding. Eventually, neither the thesis that physical access nor that competence and social organization is decisive can be dismissed. Rather, drawing on the intersectionality theory developed by Wamala [9] the plethora of sub cases within N4C leads to thinking in terms of intersections of factors that occur with specific results in different settings even when the people or even groups involved are the same.

8 ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Community's Seventh Framework Programme (*FP7/2007-2011*) under *grant agreement* n° 223994.

References

[1] Udén, M. *Tekniskt sett av kvinnor*. (English title: Women technically speaking) Doctoral thesis 2000:05. Luleå: Luleå University of Technology. 2000.

[2] Lindberg, M; Udén, M. Women, reindeer herding and the Internet: An innovative process in northern Sweden. *Innovation: The European Journal of Social Science Research*, vol. 23, No. 2, pp. 169-177 2010.

[3] Udén, M. Reindeer and frozen stories. In Elovaara, P. et al (Edts) *Travelling Thoughtfulness – feminist technoscience stories*. Umeå: Umeå University pp. 193-208. 2010.

[4] Udén, M. Women, the Knowledge Society and ICT Access in the Reindeer Grazing Areas. Scheduled to be published in Krings, Bettina-Johanna (Edt) *Brain Gain or Brain Drain?* Berlin: sigma verlag, pp. 285-314. 2011

[5] Udén, M.; Wamala, C. *Horizontal project issues: Seminar conclusions* N4C Deliverable 2.4.2. http://www.n4c.eu/Download/n4c-ltu-080-D2.4.2-Seminar.pdf . 2011.

[6] David, A.; Foray, D. Economic Fundamentals of the Knowledge Society. SIEPR Discussion Paper Nr. 01-14. Version revised February 2002. Stanford Institute for Economic Policy Research, Stanford CA. http://ideas.repec.org/p/wop/stanec/02003.html. 2002.

[7] Johnson, C. A.; Ariunaa, L.; Britz, J. J. Constructing the Pillars of a Knowledge Society: The Challenge of Providing Access to ICTs in Rural Mongolia. In: Libri, 2005, vol. 55, pp. 216–224. 2005.

[8] Huyer, S.; Hafkin, N. Engendering the Knowledge Society: Measuring Women's Participation WIGSAT | Women, Knowledge, Technology www.wigsat.org Orbicom. 2007.

[9] Wamala, C. *Does IT count? Complexities between access to and use of Information Technologies among Uganda's farmers* Doctoral thesis. Luleå: Luleå University of Technology. http://pure.ltu.se/portal/files/5152377/Caroline_Wamala_Doc2010.pdf . 2010.

[10] Boznar, M. Z. et al. *Integration Plan for N4C Test Beds with other Future Internet Test Beds*. N4C Deliverable 2.3. http://www.n4c.eu/Download/n4c-wp2-043-D2-3-Offer-to-federation-10.pdf. 2011.

[11] Directorate-General. *Report on equality between women and men 2010*. DG for Employment, Social Affairs and Equal Opportunities Unit G1 Luxembourg: Office for Official Publications of the European Communities. http://ec.europa.eu/social/BlobServlet?docId=4613&langId=en . 2010

SMART VILLAGES VERSUS THE SMART CITIES CONCEPT FOR E-INCLUSION

Marija Zlata Božnar, Primož Mlakar and Boštjan Grašič
MEIS storitve za okolje d.o.o.
Mali Vrh pri Šmarju 78, SI–1293 Šmarje – Sap, Slovenia
Tel: +386 1 366 3226; fax: +386 1 366 3227
e–mail: marija.zlata.boznar@meis.si

ABSTRACT

One of the most important concepts in the present European development of ICT, are Smart Cities. Several 7th FP projects have developed new advanced products that will be available for citizens in urban areas. But we can identify a problem arising from only supporting this concept. Quite a large section of EU citizens do not live in an urban environment where everything is available in terms of ICT and e-Inclusion [1].
Outside the urban environment there are several areas where Broadband Internet connection is still a dream due to geographical, technical or economical reasons. The situation can be illustrated as an isolated Internet heaven being placed in the Internet desert.
People living in the "Internet desert" may quickly feel that they are becoming second-class citizens of the information society.
To overcome this problem, the Smart Villages concept is proposed in the paper. Recent developments in Delay and disruption tolerant networking (DTN) are presented [2,3]. DTN is one of the possible solutions for Smart Villages. Examples of good practices are discussed. In addition, scientific conferences that promote this type of technology are presented.

1 INTRODUCTION

One of the important concepts of the European policy is to ensure the uniform development of all regions and support the regions that are lagging behind in any sector. For this purpose, the cohesion funds as well as other funds, arising partly from the research programme of the Seventh Framework Programme, are aimed at achieving a uniform level of development.
This e–Inclusion is a very important concept for the information society in order that all citizens are involved in social processes which are to a large extent carried out via services available through the Internet. These services include everything from access to state authorities and their services via e–portals, to information and social networks that already determine to a large extent the social integration of individuals into society.
However, e–Inclusion is hampered in many ways which may also overlap and intensify.
The first obstacle is computer illiteracy, which is a specific problem of the older generation and partly also of the uneducated population.
Furthermore, e–Inclusion may also be a problem due to economic reasons, where poverty makes it impossible to access IT although it is practically at the door.
This article mainly focuses on problems regarding e–Inclusion arising from a geographical dimension where the problem is mainly caused by the population living in sparsely populated areas which are very isolated from urban centres.

1.1 Urban centres – the Smart Cities Concept is a driver for the development of the future Internet

Urban centres comprise a large part of the European population and are certainly an appropriate environment for the development of leading new technologies within the Smart Cities Concept which outlines and puts into practice new strategies for the future of Europe and the world in the area of the Internet and the information society. Within the European Seventh Framework Programme, there are several active projects, such as SmartSantander [4].

Figure 1: *Isolated Internet heaven being placed in the Internet desert*

Large urban centres are more or less followed by smaller centres which are also slowly turning from urban to rural centres. Rural centres are also trying to follow the

development and in such areas European funds are available for the development of Broadband Networks [5]. However, the lion's share of these funds goes mainly to centres of rural municipalities especially because the emphasis is put on the construction of networks based on fibre optic cables.

1.2 Non–urban areas – where to find them

These are the areas which do not belong to any of the categories listed in the previous section but are faced with issues on how to access the Internet.

First and foremost, these are areas that are sparsely populated and are geographically located far away even from rural areas and it is in fact inappropriate for them to expect a cable infrastructure. A low population density presents a good enough reason to discourage investments in cable networks not only for investors in the market but also for local municipalities whose mayors wish to achieve as wide response as possible with funds obtained at state competitions. However this may be achieved only by including the largest population possible.

We must not forget that the exclusion is also possible in places primarily due to economic reasons. We still have the suburbs of metropolitan areas and similar areas where people live so to speak on the verge of society and in poverty. Due to this situation, it is virtually impossible for them to access the Broadband Internet due to price.

From the perspective of European Union such areas can be found in poor slums, Lapland, South Balkan, and possible EU candidate countries from former Eastern Bloc (in former communist countries migrations from rural to urban areas was strongly encouraged which lead to highly populated and developed urban environment while rural areas became sparsly populated and lagged).

2 A POSSIBLE SOLUTION – THE SMART VILLAGES CONCEPT

If we are to prevent or at least mitigate the creation of marginalized groups regarding e–Inclusion and the accessibility of the population to the Broadband Internet, it is necessary to slightly modify the focus on the development of the Internet for the future at the EU–level.

We propose the development of a new umbrella concept, "Smart Villages" that was first presented at the FIA Budapest [6].

The Smart Villages Concept should be an umbrella concept for the projects with the aim of increasing the e–Inclusion of the non–urban area groups with new technologies.

Such new technologies will be useful to achieve uniform e–Inclusion in Europe, in potential EU–members, and of course, around the world where people only dream of access to the Internet.

These achievements of European research will also contribute significantly to the development of underdeveloped areas in other continents.

2.1 Obstacles with current technologies

Numerous technologies developed so far for accessing the Internet and thus ensuring e–Inclusion are also intended for the development of a wireless Internet infrastructure, however, its use is followed by the Smart Cities Concept and not by the items the Smart Villages Concept should provide.

Firstly, wireless networks based on mobile phones should be noted. These may be useless due to the too expensive construction of base stations in the areas where the population is insufficient for the consequential economic viability of the projects. Furthermore, there may be problems with protected areas like national parks where it is forbidden to construct transmitters on high towers.

Satellite telephony is usually too expensive for everyday needs and is used only in special cases like the rescuing, tracking expeditions in uninhabited areas etc.

Wireless WiMAX networks are much more suitable for rural areas yet the population density of the area we want to cover with such a network plays an important role in determining the economic justification for the construction.

Wi–Fi technology may cover nucleated villages; however, it is necessary to provide one link – a server with alternative ways of access to the broadband Internet.

2.2 Requirements for the Smart Villages Concept's technical solutions

Technological solutions which are suitable for Smart Villages shall of course be innovative. Most of all, in the final (non–research) stage they should be affordable as the purchasing power in remote and rural areas is generally lower. This fact is particularly important if we want to use these solutions for the underdeveloped countries in other continents.

In the early stage, these technological solutions should enable at least basic services relevant for e–Inclusion. It is desirable however, that further development is possible, that would bring these solutions to the solutions offered by the Smart Cities Concept.

The development of technical solutions should follow the concept which provide gender and social equality in the use of the technology. This means that it is desirable to include the future users already in the development stage as required by the Living Laboratories Concept [7].

3 WHAT DTN TECHNOLOGY CAN OFFER

Delay and disruption tolerant networking (DTN) was designed to be used instead of the "legacy" Internet in the areas where an Internet infrastructure is not available at all or in areas where it may be a cheap alternative to the existing high price solutions.

DTN is a novel architecture [8] where the Internet is available with possible significant time delays or in situations where significant disruptions should be overcome. Data are being delivered in packages that are carried by data mules. Data mules can be any kind of persons or objects in movement suitable for carrying data such as hikers carrying their net-books or smart phones, helicopters having mobile nodes on-board on their regular

flights for delivering and collecting goods in remote areas or any other human or device movement over longer distances.
At the current stage of development, DTN provides basic elements for a social e–Inclusion, for example e–mail, web cashing, podcast [9,10] as well as solutions for automated environmental data collection from remote areas where these activities are not being carried out by the in–situ method but to some extent only by remote sensing (remote sensing techniques).
The US space agency is testing DTN technology for Internet connection to the remote space [11].

3.1 Examples of good practice

Within the N4C project two test beds [12] were developed for testing the DTN infrastructure and several applications for basic e-Inclusion.
Two main solutions developed mainly by TCD and LTU in cooperation with Folly consulting Ltd., and NORUT [13] were tested in a Swedish test bed in Lapland where the nomadic Sami population are breeding reindeer during the summer period. The local population has started to use email and web cashing solutions for their everyday communication. Detailed result from Swedish test bed are described in paper [14]. Figure 2 presents solar and wind powered node during the winter time in Lapland.

Figure 2: *Setting up DTN node in Lapland (Picture from deliverable D8.4 [12])*

In Slovenia, an environmental DTN test bed was operating in a permanent mode to show the ability of such infrastructure to serve as a cheap and efficient way of collecting environmental data (meteorological for climate observations and radiological for monitoring disaster consequences). Detailed result from Slovenian test bed are described in N4C Wiki [15] Figure 3 presents setting up of an meteorological node in the Slovenian mountains. Remote areas and their inhabitants are also neglected in the terms of radiological early warning systems. Such safety availability issues are often neglected in e-Inclusion discussions, but in reality "after Fukushima", scenarios do also include remote and rural areas. Figure 4 shows radiological node mounted in a suitcase.

3.2 Applicability of DTN technologies for "after disaster" solutions

The DTN technologies as presented in the previous chapter are an excellent solution for setting up networks for communication within a population and for automatic monitoring data transfer in cases when the normal infrastructure has been severely damaged by a natural catastrophe or by a man-made major disaster such as a nuclear accident.
Within this framework, future developments of DTN solutions for Smart Villages can also be efficiently used after disasters.
From the scientific point of view, DTN test beds should be developed further to enhance the present premature solutions.

Figure 3: *Setting up an environmental DTN node in the Slovenian mountains*

Figure 4: *μ-GaRaMo DTN node for gamma dose rate radioactivity measurements*

Modules enabling DTN communications should become part of the operating systems for the Legacy Internet. In this way, the DTN network can be quite easily invoked in case of necessity after a disaster. It should be a final goal of developers that the technological solutions offering such new emergency features should be transparent from the user's perspective.

3.3 Extremecom

International conferences Extremecom are working for the exchange and promotion of DTN solutions for remote areas [16]. The first one was held in Lapland Sweden, the second one in the Himalayas and this year the third one in the Amazon.

4 CONCLUSIONS

Future Internet development should cover more than solutions for more or less urban areas, as their development is already under the umbrella of the Smart Cities concept.
The emphasis should also be on the Smart Villages development that will enable equal opportunities for e-Inclusion for people living in rural, remote or otherwise technologicaly forgotten areas.
DTN technology, which is at present at the start of its development, is one of the best promising solutions for the Smart Villages because it is high possibility that otherwise this areas will remain without Internet access at all for the next decade. Pros and cons about DTN technology are presented in paper [17] while additional effort to improve the technology is still being put in different projects (i.e. [18]).
In order to prevent Internet "secondary class citizens", DTN and other similar technologies should be developed further to offer interesting and attractive solutions for Future Smart Villages for Europe and the rest of the World.

5 ACKNOWLEDGEMENTS

The work was done under the contract:
SEVENTH FRAMEWORK PROGRAMME GRANT AGREEMENT No. 223994, "Networking for Communications Challenged Communities: Architecture, Test Beds and Innovative Alliances" N4C, Collaborative Project

References

[1] Europe's Information Society Thematic Portal. e-Inclusion. [Cited: 21.07.2011] http://ec.europa.eu/information_society/activities/einclusion/index_en.htm

[2] M. Z. Božnar. N4C non-scientific summary. *N4C home page.* [Cited: 07 21, 2011.] http://www.n4c.eu/N4C-info.php.

[3] N4C Consortium. *Main Project Page*. [Cited: 21.07.2011]http://www.n4c.eu.

[4] European Commission. *ICT Research in FP7*. [Cited: 21.07.2011]http://cordis.europa.eu/fp7/ict/fire/docs/fp7-factsheets/smartsantander_en.pdf.

[5] Ministrstvo za šolstvo, znanost in tehnologijo. *Javni razpisi.*[Cited:21.07.2011]http://www.mvzt.gov.si/si/o_ministrstvu/javne_objave/javni_razpisi/?tx_t3javnirazpis_pi1[show_single]=944.

[6] M. Božnar. Experimentation and user involvement. *FIA - Future Internet Assembly.* Budapest, Hungary 2011.

[7] Gurstein, Michael. Community Informatics: Enabling Communities with Information and Communications Technologies. s.l. : Idea Group, 2000.

[8] K. R. Fall and S. Farrell. DTN: An Architectural Retrospective. *IEEE Journal on Selected Areas in Communications 26.* 2008.

[9] O. R. Helgason, et al. Performance of Opportunistic Content Distribution under Different Levels of Cooperation. *European Wireless Conference* (EW). 2010.

[10] F. Legendre. PodNet - Mobile Distribution of User-generated Content. [Cited: 07 21, 2011.] http://podnet.ee.ethz.ch/.

[11] NASA. NASA Tests First Deep-Space Internet. [Cited: 21.07.2011]http://www.nasa.gov/topics/technology/features/internet-20081118.html.

[12] Romanowski, Krzysztof and Božnar, Marija Zlata. D8.4 Test Results - Documentation of test results from tests.[Cited:21.07.2011]http://www.n4c.eu/Download/8.4/n4c-wp8-d8.4-test-results-2-0.pdf.

[13] K. J. Grøttum, et al. N4C. *N4C Deliverables – Downloads.*[Cited:21.07.2011]http://www.n4c.eu/Download/n4c-wp3-D3-3_Prototypes_Final_v1-3.pdf.

[14] S. Farrell, A. McMahon, E. Meehan, S. Weber, K. Hartnett. Report on an Arctic summer DTN trial. *Wireless Networks, Volume 17, Number 5*, pp. 1127-1156. 2011

[15] N4C Wiki. Analysis of results from Slovenian permanent testbed. [Cited:06.09.2011] http://wiki.n4c.eu/wiki/index.php/Analysis_of_results_from_Slovenian_permanent_testbed

[16] ExtremeCom. ExtremeCom 2011. *ExtremeCom 2011 - 3rd Extreme Conference on Communication - The Amazon Expedition.* [Cited: 21.07.2011] http://www.extremecom.org/.

[17] M. Udén, C. Wamala. D2.4.2 Horizontal project issues. [Cited: 21.07.2011] http://www.n4c.eu/Download/n4c-ltu-080-D2.4.2-Seminar.pdf

[18] Delay Tolerant Networking Research Group. Home page. [Cited:06.09.2011] http://www.dtnrg.org/wiki

A TESTING PLATFORM SUPPORTING APPLICATION DEVELOPMENT FOR AD HOC AND CHALLENGED NETWORKS

Henryk Gierszal[1], *Witold Hołubowicz*[1], *Łukasz Kiedrowski*[2], *Krzysztof Romanowski*[2]
[1] Adam Mickiewicz University, ul. Umultowska 85, 61-614 Poznań, Poland
[2] ITTI, ul. Rubież 46, 61-612 Poznań, Poland
e-mail: Krzysztof.Romanowski@itti.com.pl

ABSTRACT

A network simulation software platform with focus on modelling ad hoc and challenged networks is presented. The platform is based on open source software components, including the *ns*-3 simulator, operating system virtualization software, and virtual private networking systems, adding specific network models, automation and visualization, and facilities for involving real hardware nodes in simulations. Applications in collaborative research projects concerning delay-tolerant networks and mobile ad hoc networks are shown.

1 INTRODUCTION

Developing practical distributed applications for services in communications challenged environments with no permanent network infrastructure requires a specific architectural approach – like the concept of delay tolerant networks (DTNs) [1] – as well as substantial amount of building and testing hardware nodes, transport media, network protocol stacks, and applications. Collaboration of multiple integration teams can be beneficial for thorough testing; however, access to target environments – which are often in remote or restricted areas, infeasible to be replicated – can be problematic. This calls for a testing platform that would simulate the characteristic behaviour of the target network, while at the same time facilitate examining the real hardware and software being developed.

This paper presents a software platform for supporting testing applications and protocol implementations for such challenged networks. The platform originated in the N4C collaborative research project [2], for testing and integrating software and hardware components developed by independent partner teams. The architecture and implementation of the platform are presented, followed by application examples and conclusions.

2 RELATED WORK

A number of open source as well as commercial network simulators have been developed, ranging from general packages like *ns-3* [3] to specialized software dedicated to DTNs, including DTNSim [4], DTNSim2[5], and ONE [6]. Support for DTNs, although not included in the *ns-3*, has been the subject of recent work (cf. [7]). That approach focuses generally on protocol research. Testing applications benefits from shifting the focus to observing practical protocol implementations on non-simulated nodes in a simulated network. This approach, followed in the N4C project, has been independently applied elsewhere (cf. *ns-3* usage in [8]). The platform developed for N4C is also being used in project TALOS [9].

3 ARCHITECTURE

The main components of the platform include the host operating system, the *ns-3* network simulator and a virtual machine engine both running on that operating system, and virtual machines run by the virtual machine engine. Each of the virtual machines runs a guest operating system and, on top of it, a networking software stack and services or applications that are subject to testing. The network connecting those virtualized nodes is modelled by the network simulator, with only minimum engagement of the host operating system's networking stack. The latter provides basic connectivity between each individual virtual machine and the simulator.

Connecting the virtual nodes by using the network simulator rather than the networking capabilities of the host operating system gives flexibility in configuring the simulated network. The host's networking stack could have been used to set up and tear down connections and shape network traffic, with some time scheduling. The network simulator, however, in addition to being able to provide the above functionality, makes it possible to model communication channel properties (delays, transmission rates, error and packet loss distributions), mobility patterns of wireless stations, networking device properties at the physical and link layers – all with detailed scheduling, logging, and packet tracing. If required, new models can be constructed and used in simulations.

In the presented solution, nodes that participate in simulated networks are not modelled in the *ns-3* simulator, although this is possible and could complement the setup made of virtual machines (e.g. to generate specific traffic). Instead,

virtualization of node hardware gives the possibility of installing a regular operating system and an almost native environment on which to install the tested software.

4 IMPLEMENTATION

In most of the tests a notebook with a dual-core and a PC with a quad-core processor were used, both of 64-bit x86 architecture with hardware virtualization support, and with 8GB RAM, hosting a 64-bit Fedora Linux distribution. The node types that can be virtualized in such a system are those based on Intel x86 or x86_64 architecture. The guest operating systems used in the tests were various Linux variants (mostly Ubuntu and OpenWrt), depending on the requirements of the applications.

Some of the nodes used in the N4C project were based on different architectures (ARM in smartphones, Broadcom or Atheros in WiFi routers) and could not be virtualized on the platform. Instead, they were integrated by physically connecting them into the system. The possibility of connecting a physical node to a simulation setup is important in itself, as it facilitates testing the complete real stack, including the hardware. In fact, the physical node can even be remotely connected over a VPN.

The virtualization system adopted for the platform is Oracle VirtualBox (VB) [10]. In addition to full virtualization, lightweight virtualization-like technology of Linux containers (LXC) [11] is also employed. Nodes hosted on LXCs use the operating system kernel of the host computer, with key resources (network interfaces, process namespaces, selected parts of file systems) isolated between the containers. As a result, an LXC node requires much less RAM and disk space than a full virtual node and can therefore be replicated in great numbers.

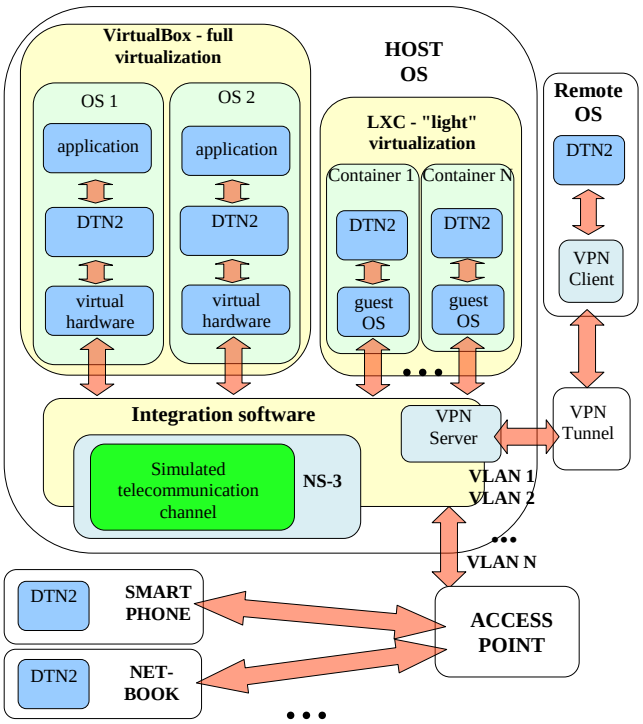


Figure 1: *Platform components*

The simulation setups for the *ns-3* simulator are written mainly in C++ and make use of components from the simulator library; some parts of the code use the Python language. There are also auxiliary subsystems for visualizing simulations and automating test setups (e.g. creating and cloning virtual machines and network connections).

Figure 1 presents the main components of the testing platform together with external (real) nodes. Figure 2 shows a screenshot from a simulation with 93 participating nodes, falling into four categories: stationary LXC nodes with positions generated prior to simulation, mobile LXC nodes with positions generated by the system at runtime according to a mobility model, stationary VB nodes with positions generated prior to simulation, and a stationary node mapping an external node on a remote computer connected over a VPN. The circles denote radio communication ranges. The line segments mark logical links between nodes that are currently in contact.

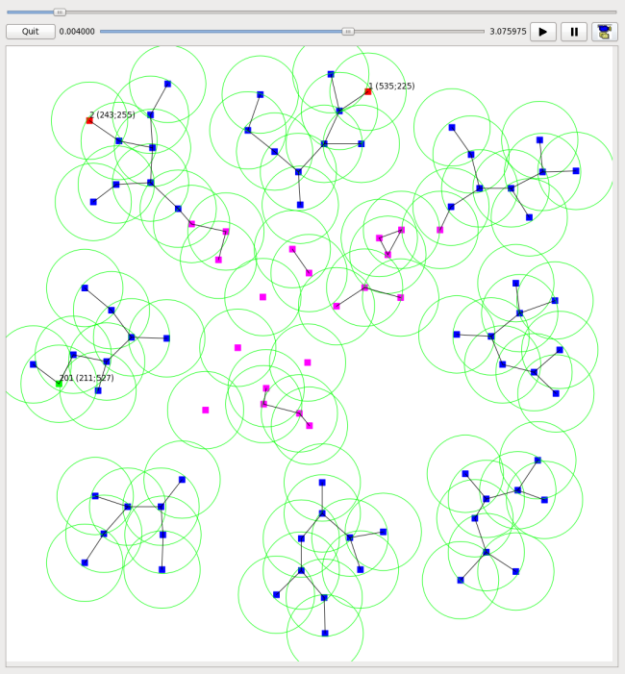


Figure 2: *Example 93-node simulation visualization*

5 APPLICATIONS

5.1 N4C

The N4C project involved designing and experimenting with network architecture, infrastructure and applications for communications challenged regions, and two test beds, one each in Sweden and Slovenia. The applications included animal tracking, meteorological and environmental data capture, Web caching and e-mail services, and Hiker's PDA (a suite of applications for hikers, hunters, herders, and rangers). The hardware nodes were developed on PCs, netbooks, dedicated routers, PDAs, and smartphones. The DTN protocol stack was based on two different

implementations: DTN2 from the DTN Research Group of the Internet Research Task Force and PRoPHET from the Luleå University of Technology. The physical data transport was provided by 'data mules' – mobile nodes carried by cars, helicopters, or hikers, using wireless communications (mostly WiFi) – or with special low-cost transport mode over USB pen-drives sent by snail mail (the SymbioNode [12]). This variety of hardware, software, and protocols provided many scenarios for simulation and testing. An example DTN with the various types of communication is shown in Fig. 3. Figure 5 presents a screenshot from a simulation involving four nodes: three virtual – hosted on VB guests – and one real – a Nokia N900 smartphone, with its screen contents shown via a remote display connection. The application being tested is the Hiker's PDA [13]; the displays show the status of node discovery as seen by each node.

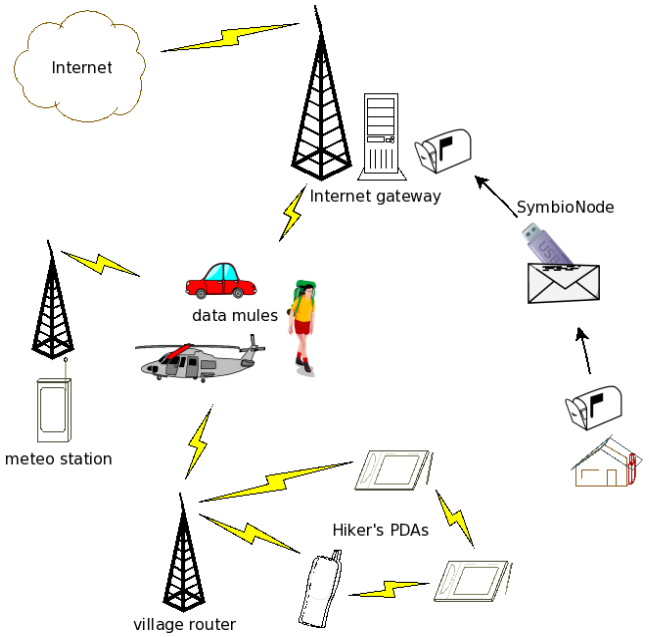


Figure 3: *DTN connections (N4C project)*

5.2 TALOS

The objective of the TALOS project is to design, implement and field-test a prototype of an adaptable and transportable border surveillance system. The TALOS system uses sensors intalled in UGVs (Unmanned Ground Vehicles), UAVs (Umanned Aircraft Vehicles), and static sensor towers along the border to detect people, vehicles, and hazardous substances which cross land borders. The testing platform was used to model a wireless mesh topology created among the UGVs. Testing the mesh network in the communication subsystem was done in simulation as well as emulation mode of the *ns-3*. Communication between the UUCC (Unmanned Units Command Centre) and the UGVs was evaluated in off-line tests in laboratory and in on-line demonstration in a demonstration area. The platform emulated the mesh topology created by a series of wireless hops between virtual UGVs that participated in the communication chain between the UUCC and the real UGV. Figure 4 presents a sample network topology.

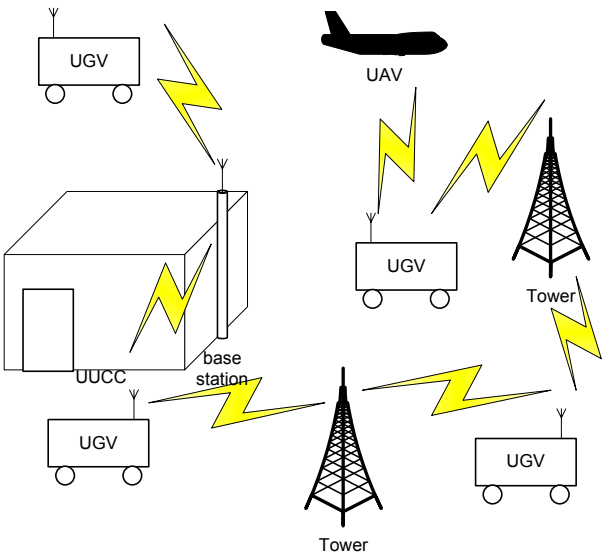


Figure 4: *MANET connections (TALOS project)*

6 CONCLUSION

With testbeds for experimenting with Internet services becoming large and connecting remote locations in different countries (cf. [14]), testing a large-scale application on a simulator before deployment can save later trouble of going after misconfigured devices in unaccessible remote regions, as well as limit personnel safety problems in vehicle demonstration areas. The platform has been found a useful tool for that.

The resource needs of the platform were observed to be modest. A four-core 2.8MHz system can accommodate hundreds of virtual nodes, with only slight increase of the memory resources used as the nodes are added. Generally, a simulation with 20 LXC nodes needs about 1.7GB of RAM; increasing the number of nodes to 80 takes up ca. 300MB more. Figure 6 shows dependence of the CPU usage on the number of simulated nodes for one of the more complex scenarios tested in N4C, with only LXC nodes. Memory and CPU usage place practical limitations on the platform in the range of hundreds of nodes.

Future work within the TALOS project will involve adding UAVs to simulation scenarios of MANET communication. Due to their higher speed as compared to UGVs, changes in protocol configuration will be needed. Beside TALOS, application of the platform to researching mesh networks in GSM environment is planned.

The code of the platform developed in the N4C project is available on the GPL2 licence at http://code.n4c.eu/code/simulation-tools.

7 ACKNOWLEDGEMENTS

This work was co-funded by the European Community Seventh Framework Programme under grants no. 223994 (N4C – Networking for Communications Challenged

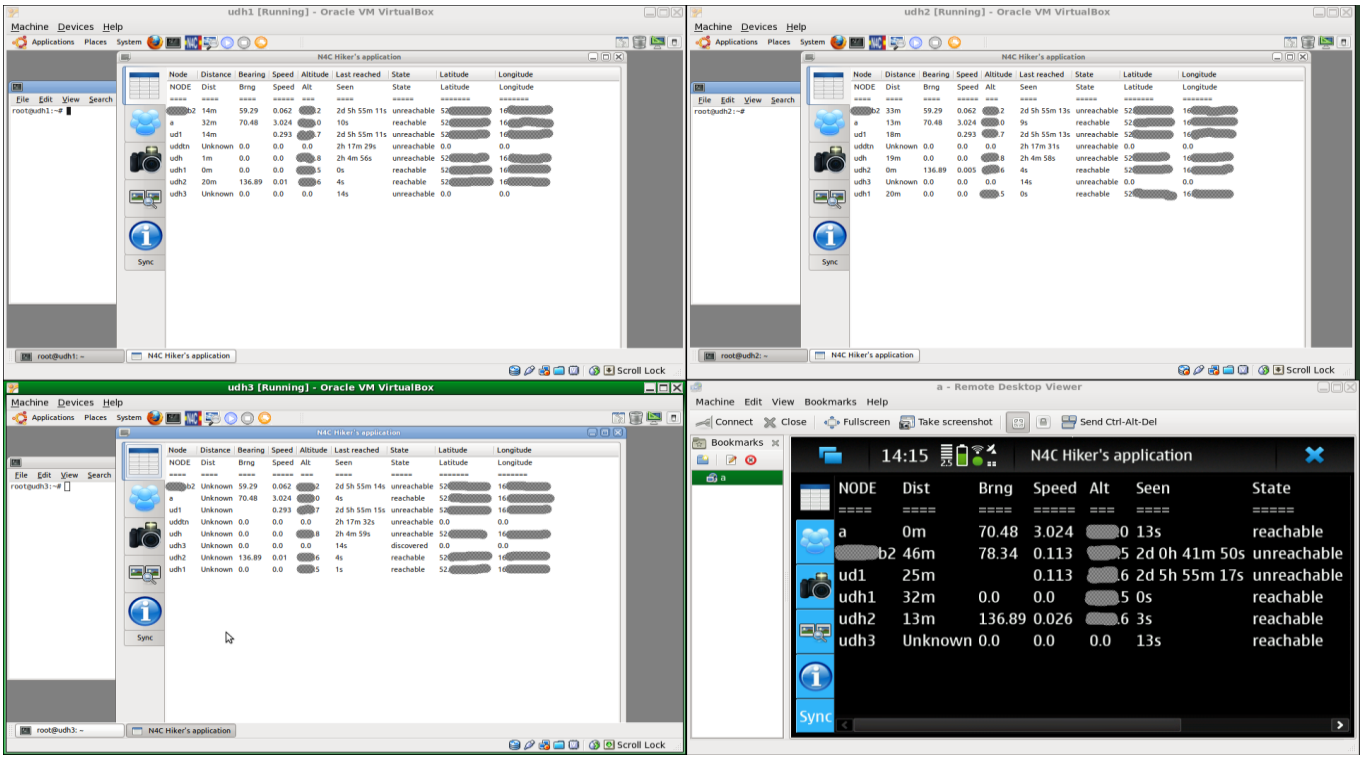


Figure 5: *Hiker's PDA application test (N4C project)*

Communities: Architecture, Test Beds and Innovative Alliances) and 218081 (TALOS – Transportable Adaptable Patrol for Land Border Surveillance).

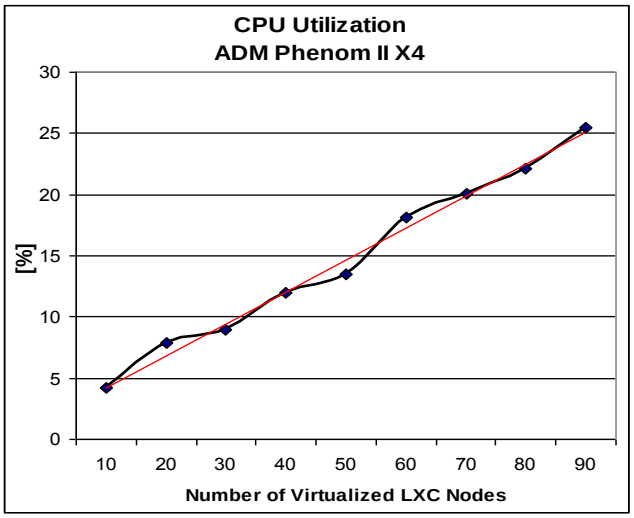


Figure 6: *CPU utilization for a complex test scenario*

References

[1] S. Farrell, V. Cahill: *Delay- and Disruption-Tolerant Networking.* Artech House, Boston, 2006.

[2] N4C project home page, http://www.n4c.eu

[3] NS-3 project home page, http://www.nsnam.org

[4] S. Jain, K. Fall, R. Patra: Routing in a delay tolerant network. *Proc. SIGCOMM '04*. ACM, New York, 2004.

[5] DTNSim2 project home page, http://watwire.uwaterloo.ca/DTN/sim

[6] A. Keränen, J. Ott, T. Kärkkäinen: The ONE Simulator for DTN Protocol Evaluation. *Proc. SIMUTools '09*. ICST, Brussels, 2009.

[7] F. Herbertsson: *Implementation of a Delay-Tolerant Routing Protocol in the Network Simulator NS-3.* LIU-IDA/LITH-EX-A--10/046—SE, Linköping University, 2010.

[8] A. Krifa, M. Mendonca, R. N. Bin Rais, C. Barakat, T. Turletti, K. Obraczka: Efficient Content Dissemination in Heterogeneous Networks Prone to Episodic Connectivity. *SIGCOMM '11*, Toronto, 2011.

[9] TALOS project home page, http://talos-border.eu

[10] VirtualBox home page, http://www.virtualbox.org

[11] LXC project home page, http://lxc.sourceforge.net

[12] S. Vrbinc, B. Grašič, M. Z. Božnar, P. Mlakar: SymbioNode Data Carrier in Delay and Disruption Tolerant Networking (DTN). *Collaboration, Software and Services in Information Society*, Ljubljana, 2010.

[13] K. J. Grøttum, S. Sjursen, B. Bergvall-Kåreborn, J. Näslund, E. Davies, B. Grašič, A. Lynch: *Prototypes. Final Version.* Norut, Luleå University of Technology, Folly Consulting, MEIS, http://www.n4c.eu/Download/n4c-wp3-D3-3_Prototypes_Final_v1-3.pdf, 2011.

[14] M. Z. Božnar, B. Grašič, S. Vrbinc, D. Popović, P. Mlakar, E. Davies: *Integration Plan for N4C Test Beds with other Future Internet Test Beds, including Offer to Federation by Slovenian Test Bed.* MEIS, Folly Consulting, http://www.n4c.eu/Download/n4c-wp2-043-D2-3-Offer-to-federation-10.pdf, 2011.

NOT-SO-INSTANT-MESSAGING SERVICE FOR THE DELAY TOLERANT NETWORKS

Samo Grasic
Luleå University of Technology
Sweden
e-mail: samo@ltu.se

ABSTRACT

Most of the research within the Delay Tolerant Network (DTN) area focuses on the networking layer and the routing of network traffic. In order to move the DTN research from the networking research area to the real-world usage the developments of services and applications that uses DTN are crucial. This paper presents the Not-So-Instant-Messaging (NSIM) service designed for DTN usage. The NSIM service allows users of the DTN to send and receive E-Mails, send out SMS and communicate efficiently with other users within the DTN region.

Keywords
DTN, Delay tolerant, Services, Not-So-Instant-Messaging, NSIM, Email

1 INTRODUCTION

The use of Delay Tolerant Network (DTN) [1] in sparsely populated rural areas without any available ICT infrastructure has proven to be a potential alternative [2][3] to expensive satellite communication systems. Usual networking scenarios in such deployments rely on the available public transportation means and the users' mobility as the main carriers of the network traffic. In this networking paradigm the user of the network service becomes a node and a part of this same network with its devices. The network user's mobility and his/her encounters with other DTN users are exploited for moving data through the network.

The DTN performance of this networking scenario relies heavily on the density of the nodes in deployed area. In order to provide the expected network traffic delivery rate and the maximum delivery delay in the DTN, a minimal critical density and mobility of the nodes in the deployed regions are required (e.g., in order to achieve an average delivery delay of less than a day from the isolated rural village, it makes sense that some sort of data carrier should reach this village at least once per day).

In order to attract more users who might contribute to the growth of this network, appealing and useful applications and services that can use the DTN are needed. The traditional internet Email service in DTN scenarios uses a client-server architecture, where the clients are located within the DTN region and the server is placed outside the DTN area. A drawback of this Email architecture is that every sent or received Email must first reach the server outside the DTN region, before it can be forwarded to its destination. As a consequence, an additional delay in the communication can be seen when users within the DTN area want to communicate with each other.

Another common problem that has been observed in the real-world DTN deployments [3] [7] is that the users cannot recall the Email addressed from the recipients. The recipient addresses are usually stored the in the Web-based Email clients or personal computers, something that complicates the use of Email services within the DTN region.

The name of the Not-So-Instant-Messaging service derives from the service used in one of the first real-world DTN deployment [7] that took place in Northernmost part of Sweden in 2006. This deployment was part of the Sami Network Connectivity project. At the same time the explosion of different kinds of Instant Messaging (IM) services (e.g., IRC, MSN, Skype) could be seen on the Internet.

The NSIM service was originally designed as an IM service for the DTN environment. The NSIM service addresses some of the listed problems above. This paper describes the evolution, architecture, implementation and evaluation of the NSIM service.

2 RELATED WORK

To push the DTN research field from the laboratory environment to the real-world use, real-world DTN testbeds are crucial. The DTN testbeds as described in [2], [5] and [6] has the potential to involve more end users and contribute to the growth of the DTN deployments. However, no mayor bursts of the real-world DTN users has been seen so far.

Discussing today's challenges of real-world DTN deployments in the [4], Lindgren et al. demonstrate the importance yet the current lack of so called “Killer Apps” available for the DTNs. The “Killer Apps” are applications and services for the DTN that facilitates the use of DTN. As an example of these services Lindgren et al. present tele-medicine service for developing regions, social networking services for the developing world and file sharing services.

They also point out the importance of user involvement in the design process of such applications.

A common way [7] of setting up the DTN Email service is to locate an Email server outside the DTN region. The users then communicate as clients through the DTN with the use of a proxy server that is located on the border of the DTN region. In their work, Hyyryläinen et al. [8] present various adaptations of the conventional Email system for the DTN environment. Besides the solution with the Email proxy server, they present a solution with an embedded mail user agent in every DTN node. In addition to the server-client Email distribution system, the embedded user agent supports the opportunistic peer-to-peer delivery within the DTN region.

3 SERVICE DESCRIPTION

The NSIM application was first developed as a very simple communication application that enabled people involved in SNC project [7] deployment to communicate within the DTN test field. This application became an immediate success, but the drawback was that it only allowed addressing of the network nodes and not the users themselves. Hence, certain users had to share certain nodes, something that hindered the users from attaining full integrity. Despite this flaw, the concept of and the need for the NSIM application on the field was still proven.

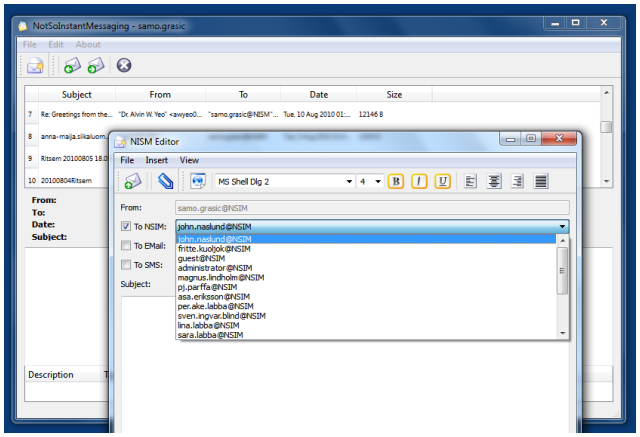


Figure 1: *A snapshot of the NSIM client application*

Within the N4C project, the NSIM application was redesigned and rewritten from scratch. To shorten the development time, parts of the graphical interface and email libraries were however borrowed from an open source project called Nuntius Leo Creator [9]. The code is based on the Nokia's QT cross-platform framework [10] and can be used on various operating system platforms. In the latest stage of the development, the NSIM service gained two new features: the conventional Email service and GSM SMS service.

In order to employ all the features of the NSIM service, two different applications are needed. The first one, called NSIM, is a client application that is similar to any email client application. NSIM needs to be installed on the computer of every user. A client application itself is enough if the users want to send messages within the DTN region. In order to provide the NSIM users with the Email services or GSM-SMS service, the NSIMGW application needs to be installed on the gateway that is connected to the Internet.

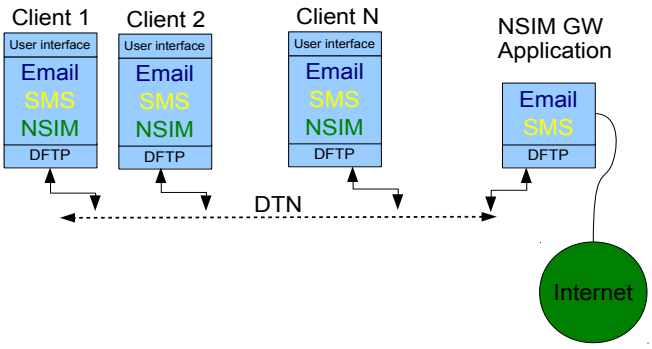


Figure 2: *The NSIM service architecture*

The NSIM application allows users to sign in with their own user name and password. When a new user is created for NSIM,the administrator needs to assign one or more DTN nodes on which the user will read his/her messages. This feature is important for the users who are mobile within the DTN region as well as those who do not use the same computer all the time. At the same time, it is important that any user is able to send a message from any machine within the DTN region by using his/her own user name account. The administration of the NSIM users accounts is then made possible from any DTN node within the field or Internet gateway that is connected to the Internet. This was a crucial feature in the last N4C summer test when most of the users were added to the NSIM system on the field.

Sending out a NSIM message, a list of other available NSIM recipients within the DTN area was available to the user from a drop-down list. Additional features such as reply and file attachments were also available from the user interface in the latest version of NSIM application.

4 EVALUATION

The NSIM service was tested and evaluated in the summer of 2010 using the N4C DTN testbed [3]. The testbed was located in the Padjelanta national park in the northernmost part of Sweden. The village, Staloluokta, where the DTN system was deployed, is mostly populated by the Sami reindeer herders and lies deeply embedded in the mountain area. Due to strict national park policies there is no ICT, power or road infrastructure available. Staloluokta can only be accessed by a four days hike or by one of the daily scheduled helicopter flights from the closest towns Ritsem or Stora Sjöfallet where power and ICT infrastructure are available. During the test periods the helicopters were equipped with the embedded DTN node and served as data carriers between the remote village and the helicopter bases that were located 60 kilometers from the testing area. The 2G Internet connectivity that was available in the helicopter bases provided the Internet connectivity for DTN border

nodes. The DTN gateway that was located in the office of Luleå Technical University. Due to a very limited GSM base cell capacity the 2G connection was frequently disrupted and most often unavailable during daytime.

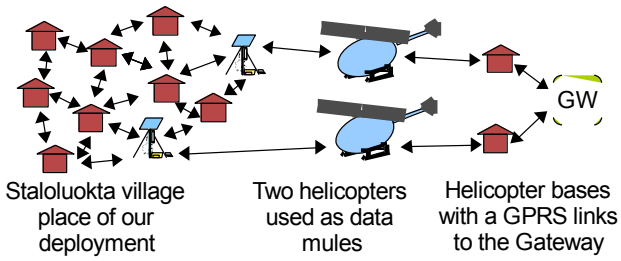


Figure 3: *The network topology of the N4C testbed in Sweden*

An approximate of 20 nodes was used in the deployed DTN region. More than 50 persons tried the application during the two months of testing period. The use of the NSIM services was recorded in the log files for the entire test period.

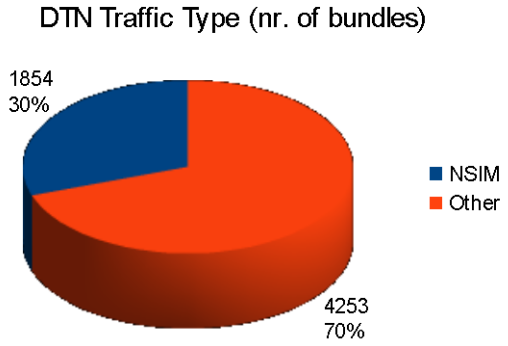


Figure 4: *Number NSIM bundles sent over the DTN*

5 RESULTS

Drawing on the analysis of DTN traffic, the collected log files from the test and the participatory observation on the test field, it was clear that the NSIM service was used extensively. In sum, the NSIM service generated around 1800 bundles, which represents almost one third of all the traffic sent over the DTN.

The provided NSIM service offered three different ways of usage. The most popular way of using the NSIM service was to send and receive Emails to and from the Internet. More than 1300 Email were transferred during the test period. One fifth of the messages that was transferred within the DTN region was generated by using the NSIM service. The least used NSIM feature was the SMS text messages. Around one tenths of the messages was sent out of the DTN region to the GSM network by using the SMS feature.

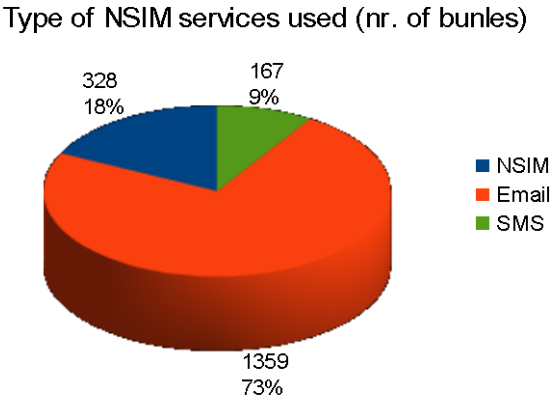


Figure 5: *Use type of the NSIM service*

In order to prove that the peer-to-peer message delivery mechanism that was used for the NSIM service offered a decreased delivery delay than the client-server delivery mechanism that was used in the conventional Email service an average delivery time per service type was analyzed.

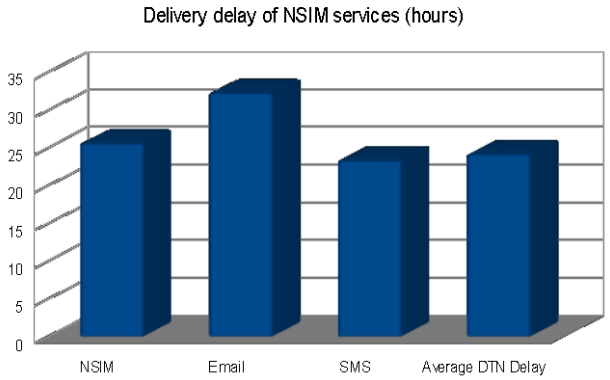


Figure 6: *Use type of the NSIM service*

As seen in *Figure 6*, the average delivery time of the NSIM messages is significantly smaller than the average time of the Email service and matched the total average DTN delivery delay of the whole DTN system. The delivery delay of the Email service is higher that the total DTN delivery delay, because every Email that was sent from the DTN region had to reach the Email server that was located outside the DTN region by using one of the helicopter data-mules, regardless weather the recipient was locate inside or outside the DTN region.

6 CONCLUSIONS

From the extensive use of NSIM service in the DTN testbed 2010 it can be concluded that there is a high potential for development of new user friendly DTN services. The involvement of users at an early stage of the service design is crucial for the success of the service. For example, the drop-down list of all available users within the DTN region was a

feature that was suggested from the users. This in turn engaged many users in their communication with each other.
Already, the availability of the NSIM service in the Staloluokta village has impacted positively on the communication among the inhabitants of Staloloukta. For instance, during the summer of 2010 the NSIM service was widely used to facilitate the technical support, and especially in cases when the user was located a couple of kilometers away. The impact could been also seen among the Sami reindeer herders who lived in the Staloluokta village. Instead of employing the traditional ways of communication (PMR walkies-talkies that did not provide privacy) the reindeer herders used NSIM service in the process of organizing the traditional calf marking event.
As seen in the results, the NSIM service provides a more efficient and comfortable communication within the DTN region as compared to the traditional Email service, especially since a reduced delivery time is crucial to any communication system.
The development of the NSIM service is however in the early stage and still needs to address some of the challenges in the DTN research field, for instance security, the lack of a common DTN API and interoperability.

7 ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Community's Seventh Framework Programme (*FP7/2007-2011*) under *grant agreement* n° 223994.

References

[1] V. Cerf et al., "Delay Tolerant Network Architecture", IETF, RFC 4838, April 2007.

[2] Anders Lindgren , Avri Doria , Jan Lindblom , Mattias Ek, Networking in the land of northern lights: two years of experiences from DTN system deployments, Proceedings of the 2008 ACM workshop on Wireless networks and systems for developing regions, September 19-19, 2008, San Francisco, California, USA

[3] Networking for Communications Challenged Communities (N4C) project website, http://www.n4c.eu

[4] Anders Lindgren and Pan Hui. 2009. The quest for a killer app for opportunistic and delay tolerant networks: (invited paper). In Proceedings of the 4th ACM workshop on Challenged networks (CHANTS '09). ACM, New York, NY, USA, 59-66.

[5] Delay and Disruption Tolerant Networking (DTN) test bed in Slovenia. GRAŠIČ, Boštjan, et al. Ljubljana : Evropska svetovalnica, 2010, Vols. International Conference InfoKomTeh 2010, 27th October 2010. COBISS.SI-ID 24104743.

[6] Aruna Balasubramanian , Brian Levine , Arun Venkataramani, DTN routing as a resource allocation problem, Proceedings of the 2007 conference on Applications, technologies, architectures, and protocols for computer communications, August 27-31, 2007, Kyoto, Japan

[7] Avri Doria, Maria Uden, and Durga Prasad Pandey, "Providing connectivity to the saami nomadic community," in Proc. 2nd Int. Conf. on Open Collaborative Design for Sustainable Innovation, 2002.

[8] Tuomo Hyyryläinen, Teemu Kärkkäinen, Cheng Luo, Valdas Jaspertas, Jouni Karvo, Jörg Ott: Opportunistic Email Distribution and Access in Challenged Heterogeneous Environments. Demo at the Second ACM SIGCOMM Workshop on Challenged Networks (CHANTS), Montreal, September 2007

[9] NuntiusLeo Project, http://nlcreator.sourceforge.net/nuntius_leo.html

[10] Nokia QT, http://qt.nokia.com/

SOFTWARE APPLICATIONS FOR ENVIRONMENTAL MEASUREMENTS USING DTN CONNECTIVITY

Boštjan Grašič, Sašo Vrbinc, Primož Mlakar, Marija Zlata Božnar
MEIS storitve za okolje d.o.o.
Mali Vrh pri Šmarju 78
SI-1293 Šmarje - Sap
Tel: +386 1 3663226; fax: +386 1 3663227
e-mail: saso.vrbinc@meis.si

ABSTRACT

In this paper software applications are presented that demonstrate usage of DTN (Delay Tolerant Network) technology for environmental purposes. DTN testbed [1] in Slovenia developed by MEIS and other members of N4C (Networking For Communication Challenged communities) project consortium was constructed using several environmental stations collecting meteorological and radiological data, as well as wild life photos on remote areas where no network infrastructure exists. Three examples are used to present how existing applications can be relatively easily adjusted to fully exploit DTN infrastructure for transferring of environmental data collected on remote areas to central database server.

1 INTRODUCTION

Governments that have responsibility for communication challenged remote areas have a pressing need for the observation of climatological data **(**including but not limited to basic meteorological observations). There are both local and global requirements for this information (from driving the everyday planning of outside work to feeding global models for assessing climate change and making daily weather prognoses).

Also the technically similar on-line measurements of gamma dose rate (radioactivity) are needed for local early warning system against potential pollution (such as from the nuclear reactor accidents at Chernobyl or in Japan).

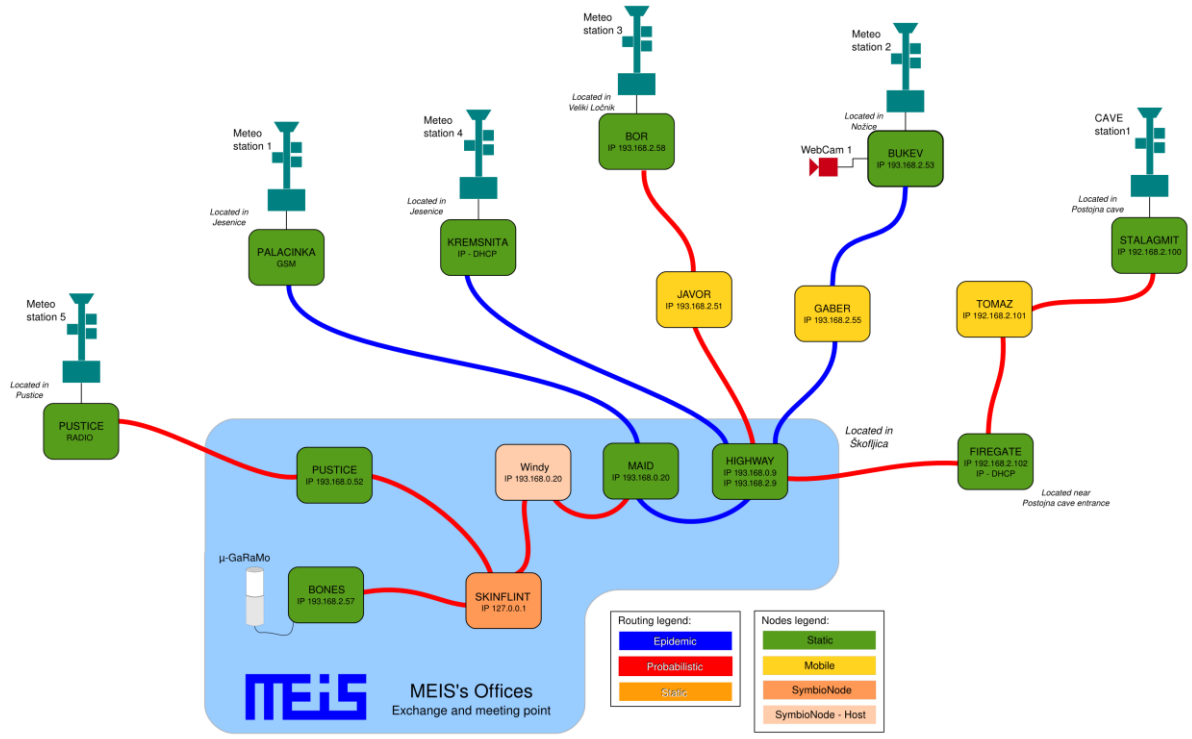


Figure 1: *Description of the winter tests 2011 in Slovenia*

These measurements are normally available for less remote areas of developed states but are often not available for remote areas. But such in-situ measurements are also needed for the research community to be able to evaluate past and present characteristics of natural and man-made pollution.
Therefore the aim of Slovene DTN test bed within N4C project (www.n4c.eu) was to develop a whole network (albeit on a small scale of up to 10 nodes) for this kind of meteorological and radiological measurements, based on DTN communication infrastructure and adapted to harsh environment.
In N4C we were developing Internet for the remote regions where it is not simple, or not cheap or, even, not feasible to have it in any of the conventional ways that town dwellers have come to expect. N4C is a research project that has successfully developed solutions for basic Internet access in such regions. The solutions are based on the novel approach of DTN – Delay– and Disruption– Tolerant Networking.
Based on DTN technology a dedicated topology was build in Slovenia for winter tests in 2011. The figure 1 depicts various environmental stations on remote locations that use DTN technology for data transfer to central database at MEIS. Data transfer between nodes was tested with regular data mules and SymbioNodes [2] in order to make comprehensive tests of developed technologies.

2 OPERATING SYSTEM OF EMBEDDED COMPUTER FOR METEOROLOGICAL STATION

As argumented on the OpenWRT development web site [3] we aimed the development according to the recommendations [4] from Klaas van Gend, were he addressed complex questions that needs to be answered in order to meet the requirements. These addressed questions are:

- ”How do I design a full-blown system based on customer requirements?
- What packages do I select, and from where do I get them?
- What package versions do I use and what are the dependencies?
- How do I glue all this together in a shippable product?
- What tools should I integrate for debugging, profiling and tracing and how do I provide them to my team - who have different host development environments?”

The result of addressing these questions has been the selection of OpenWrt which is described as a Linux distribution for embedded devices focused especially for the routers. The full description and manual of OpenWrt project is given at homepage http://www.openwrt.org where the short summary is the following: "Instead of trying to create a single, static firmware, OpenWrt provides a fully functional file system with package management. This frees you from the application selection and configuration provided by the vendor and allows you to customize the device through the use of packages to suit any application. For developer, OpenWrt is the framework to build an application without having to build a complete firmware around it; for users this means the ability for full customization, to use the device in ways never envisioned".
Installation of OpenWrt on the embedded devices is relatively simple when the binaries for the specific device are compiled. The binary must be flashed on the device using the boot-up manager that is already installed on the device. In our case the RedBoot boot-up manager was installed. The binary has been transferred to device using TFPT server that has been prepared especially for flashing the embedded devices. In the figures 2 and 3 the boot-up manager and boot process are presented.

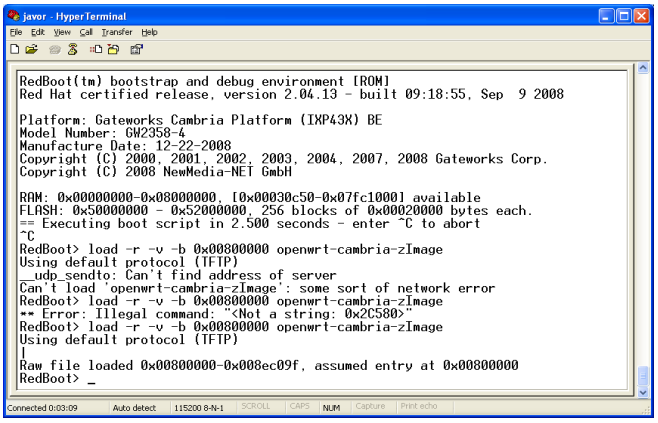


Figure 2: *Loading of binary image of OpenWrt into Cambria using the classic terminal window*

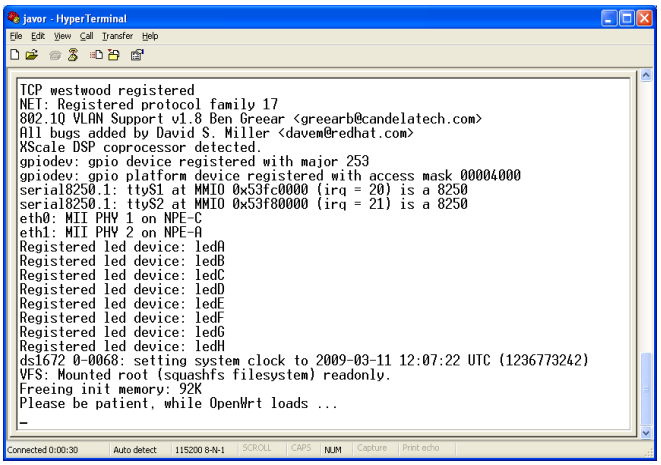


Figure 3: *Boot-up process of OpenWrt observed from the classic terminal window*

When installation of OpenWrt on the embedded device is finished the device must be configured. The configuration can be done in two ways: using the SSH terminal or using the graphical web interface LuCI. Following pictures show how the configuration of the device is performed using the web interface. Within this configuration the network properties (IP, DNS, Gateway of Ethernet and WiFi connection, routing, firewall) and common properties (name of the device, time server, time adjusting) of the device can be set, the system log files can be viewed, additional

software can be installed (different libraries, web server, ftp server, WebCam motion detection) and additional device drives can be installed (WiFi support, usb drivers, WebCam drivers, serial RS232 port drivers, compact flash drivers).

Figure 4: *Login window in LuCI web based configuration interface for OpenWrt*

3 METEOROLOGICAL DATA COLLECTION

MEIS software used in the environmental field was developed further to enable data collection and preparation of files to be transferred from Cambria DTN nodes.
The application SW to fulfill is called XMLJanko. The target meteorological data files format is defined in details in the chapter in deliverable D3.1 [5], in the chapter Meteorological station description. Example of file is presented figure 5. The measured data of each measuring environmental parameter is written in one line of the file. For example the line that starts with code P10 holds the measured data of air temperature where the first data is the average temperature, second is maximum temperature and so on.

P 0	11:00	10/03/09		00:00	01/01/00	
	34.8	4.8	12	00	43	11:00
	10/03/09		00	03	30	30
	30	1f				
P10	5.9	7.1	10:58	5.0	10:34	7.0
	0.7	0				
P11	3.6	3.7	10:30	3.5	10:35	3.6
	0.1	0				
P35	64.9	74.0	10:30	55.0	10:54	56.0
	6.1	0				
P43	696.0	820.9	10:59	232.5	10:39	806.7
	160.5	0				
P47	977.4	977.6	10:30	977.3	10:51	977.3
	0.1	0				
P100	0.7	32	2.3	336	10:46	0.0
	360	10:48	0.9	0.7	24	0.3
	0.5	0.4	0			

Figure 5: *Example of output file that contains measured meteorological data from XML Janko application*

XML Janko is configured using a XML format initial file. All operations are logged into special log files.

4 RADIOLOGICAL DATA COLLECTION

The program *mgm163* is designed to capture data from a variety of GM probes, which have serial communication. The program operates in the "preset time" mode: a probe is set to the desired averaging time and then data is being periodically collected according to the set averaging time. Format is compatible with the format P163 as developed by the J. Stefan Institute and described in details in deliverable D3.3[6].
The target radiological data files are generated according to the format that is defined for the meteorological data files. Example of target radiological file is presented on following picture. The measured data of each measuring environmental parameter is written in one line of the file. For example the line that starts with code P163 holds the measured data of gamma radiation where the first data is date of measurement, second is time of measurement and so on.

P 0	10:19	03/09/2009		09:13	05/08/03	
	25.2	3.00	12	00	41	10:00
	03/09/2009		03	03	30	30
	30	1f				
P163	03/09/09		10:19	03/09/09		10:00
	11	1	6.44E-08		5.3	A 2
	03/09/09		08:57	00	0.198	07:00
	1.464	0				

Figure 6: *Example of output file that contains measured radiological data from mgm163 application*

For the operation program needs a configuration file which is a simple ASCII text initial file named "ijsP163.ini". All operations are logged into dedicated log files.

5 WILDLIFE PICTURE COLLECTION SOFTWARE

For webcam picture collection open-source software name »motion« has been selected. It has been compiled and configured to work on OpenWrt operating system of the embedded router. Configuration has been set up to capture pictures when motion has been detected. Captured pictures had been automatic passed to the DTN using special scripts and sent to central nodes located at MEIS offices (maid.si.n4c.eu). Furthermore, data are later automatically sent to LTU's main gateway (gw-lime.dtn.n4c.eu) over DTN. [7] Transferred data have been automatic distributed to web server to be accessible over legacy internet on following addresses:

- MEIS webpage: *http://193.77.212.133/n4c-data/index.html*.
- LTU (*Luleå* University of Technology – N4C project leader) webpage: *http://dtn.n4c.eu/*

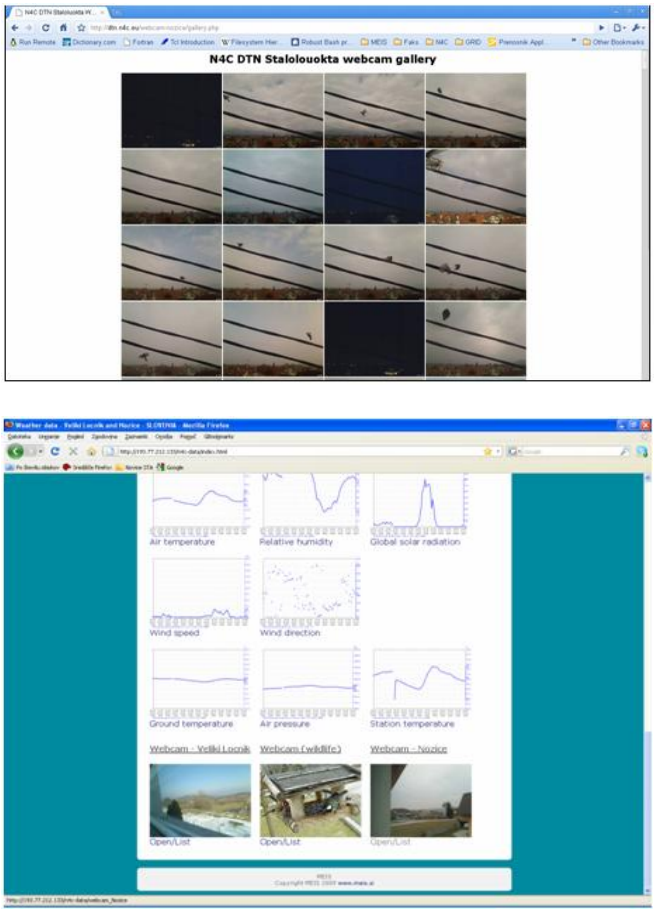


Figure 7: *Pictures from webcam nodes accessible on legacy internet (top: LTU's gateway, bottom: MEIS's gateway).*

6 EVALUATION

From the beginning of winter testing in 2009 on-going tests have been started that have continued until the end of the project, through the last winter test period in 2011. For this period QA/QC reports have been made to evaluate the environmental measuring system. These QA/QC reports are attached to the milestone reports from summer and winter tests: M8.1, M8.2-MEIS, M8.3-MEIS, M8.4-MEIS, M8.5-MEIS and M8.6-MEIS.[8] In the reports statistics of the measured environmental parameters are given where statistical parameters such as average, minimum and maximum measured values are given. Also the statistics of successfully transferred data and the quality of data is made. Additionally also all measured environmental parameters (air temperature, relative humidity, air pressure, precipitation, solar radiation, radioactivity, etc.) are presented on time scale graphs where also any delays in data delivery are shown. For the wind measurements also wind roses are drawn for each month of measurements.

7 CONCLUSION

DTN test bed in Slovenia developed within N4C research project successfully demonstrated the usefulness of DTN technology for environmental applications. It has been presented how existing applications can be relativly easily adjusted to fully exploit DTN infrastructure. Demonstration has been made using three different environmental applications: transfering of environmental data (meteorological, radiological and webcam pictures) collected on remote areas to central database at MEIS. For the near future extension of the testbed in Slovenia we are developing an application for collection of air quality data.

8 ACKNOWLEDGEMENTS

The work was done under the contract:
SEVENTH FRAMEWORK PROGRAMME GRANT AGREEMENT No 223994, "Networking for Communications Challenged Communities: Architecture, Test Beds and Innovative Alliances" N4C, Collaborative Project

References

[1] GRAŠIČ, Boštjan, et al. Delay and Disruption Tolerant Networking (DTN) test bed in Slovenia. Ljubljana : Evropska svetovalnica, 2010, Vols. International Conference InfoKomTeh 2010, 27th October 2010. COBISS.SI-ID 24104743.

[2] Vrbinc, Sašo, et al. SymbioNode data carrier in delay and disruption tolerant networking (DTN). Ljubljana : Institut Jožef Stefan, 2010, Vols. Zbornik 13. mednarodne multikonference Informacijska družba - IS 2010, 11.-15. oktober 2010 : zvezek A : volume A. COBISS.SI-ID 24047399.

[3] OpenWrt project. https://openwrt.org/. Accessible on date 28.07.2011.

[4] Gend, Klaas Van. Addressing the Top 5 Pains in Linux System Build and Design. http://www.soccentral.com/results.asp?EntryID=27956. Accessible on date 17.07.2011.

[5] Grøttum, Karl Johan and Sjursen, Sigurd. Functional Specification. 2009. http://www.n4c.eu/Download/N4C-WP3-2-fs-1-5.pdf

[6] Grøttum, Karl Johan et. al. D3.3 Prototypes. http://www.n4c.eu/Download.php. Accessible on date 16.07.2011.

[7] GRAŠIČ, Samo. M8.7 Winter test report (Swedish test field). Lulea University of Technology. http://wiki.n4c.eu/wiki/images/2/2d/M8.7_N4C-WP8.7-1.2.pdf

[8] Consortium N4C. N4C Project Wiki- Deliverables and Milestone reports. http://wiki.n4c.eu/wiki/index.php/Milestone_reports. Accessible on date 29.07.2011.

INTRODUCING ENRICHED CONCRETE SYNTAX TREE

Gordana Rakić, Zoran Budimac
Dept. of Mathematics and Informatics, Faculty of Sciences, University of Novi Sad
Trg Dositeja Obradovića 4, 21000 Novi Sad, Serbia
+381 21 4852877, +381 21 458888
goca@dmi.uns.ac.rs, zjb@dmi.uns.ac.rs

ABSTRACT

In our earlier research [9] an area of consistent and systematic application of software metrics was explored. Strong dependency of applicability of software metrics on input programming language was recognized as one of the main weaknesses in this field. Introducing enriched Concrete Syntax Tree (eCST) for internal and intermediate representation of the source code resulted with step forward over this weakness. In this paper we explain innovation made by introducing eCST and provide idea for broader applicability of eCST in some other fields of software engineering.

1 INTRODUCTION

We introduce the new type of Syntax Trees to be used as intermediate representation of the source code. We call this type of tree "enriched Concrete Syntax Tree (eCST), because it is enriched by so-called "universal" nodes. Universal nodes are common for all programming languages and add additional value to these trees. In this way the tree becomes broadly usable for different algorithms of software metrics and independent on input programming language. Besides original application field in building software metrics tools and other static analyzers, eCSTs can be further manipulated and transformed and thus applied in other fields as well (e.g. software testing and source code translation). In this way eCST can be used for numerous purposes related to development, maintenance and analysis of software systems. The major benefit of usage of eCST is its universality and independency on programming language.

2 BACKGROUND

Syntax trees are usually secondary product of compiler and parser generators. These generators usually have embedded mechanisms to generate syntax trees as internal structures. Additionally, these mechanisms can be extended with mechanism for enrichment of syntax trees with additional information about language or input source code. This opportunity is our key instrument.

Parser generators take the language grammar as its input and return parser for that language as an output. This grammar is provided in some default form (e.g. EBNF – Extended Backus-Naur form) or in some generator-specific syntax. In this project parser generator is used to generate scanner and parser with embedded rules and functions for generating trees and for managing the content of its nodes. In fact, we use it to generate parser that will produce eCST including insertion of universal nodes.

The main idea is to add corresponding universal node as a parent of sub-tree that represents specific element in a source code (e.g. COMPILATION_UNIT, FUNCTION_CALL, BRANCH_STATEMENT, LOOP_STATEMENT, etc.). Key characteristic of these nodes is that these are equivalent in all programming languages.

3 RELATED WORK

Syntax trees, abstract or concrete, are broadly used in numerous fields of software engineering. Abstract Syntax Tree (AST) is used as representation of source code and model.

Baxter [1] and Ducasse [2] use abstract syntax trees for representation of the source code for duplicated code analysis. Those trees have some additional features designed for easier implementation of the algorithm for comparison. Koschke et al. [5] propose similar but fresh idea for code clone detection using abstract syntax suffix trees

In [3] the role of AST as representation of model in Model Driven Engineering is described. ASTs were also used for monitoring of changes (Neamtiu et al. [6]) in the change analysis of code written in programming language C.

Even if the construction of AST is language independent; the content of these trees is always strongly related to language syntax. That can be clearly concluded from all papers related to usage of AST referred in this article.

In [9] a detailed motivation for initiating the research in proposed direction is described. It is related to problems in application of software metrics and early work on introducing eCST in that particular field. In [9, 10] is described eCST, its construction and its role in development of SMIILE (Software Metrics - Independent of Input LanguagE) tool, as well.

Additionally, we can propose [7] as an introduction to automatic building of syntax trees by generated language parser. It also provides mechanism for adding universal nodes into tree that is to be generated.

4 INTRODUCING eCST

Related research shows that there is no fully consistent support for software development and maintenance. All tools used for these purposes have some limitations, e.g. limited programming language support, weak and inconsistent usage of metrics and/or testing techniques, etc.
In the field of software evolution, which enforces techniques such are advising, recommending and automating of refactoring and reengineering, solutions based on a common intermediate structure could be a key supporting element. This support could be based on metrics, testing and deeper static analysis. Development of such support would introduce new values into software engineering field. For all offered reasons, proposed universal tree could be an appropriate internal representation applicable toward all stated goals. Universality of internal structure is important for meeting consistency in all fields.
By realization of this idea key benefit could be made from language independency of eCST and its universality and broad applicability.

4.1 Motivation

Motivation for introducing eCST as a new intermediate representation of the source code lays in intention to fulfil gaps in field of systematic application of software metrics by improving characteristics of software metric tools. One of the important weaknesses of available metric tools is the lack of support for calculation of metric values independently on input programming language.
Originally, Concrete Syntax Tree (CST) is used for representation of a source code. CST represents concrete source code elements attached to corresponding construction in language syntax. Although this tree is quite rich, it is still unaware of sophisticated details about meaning of syntax elements and their role in certain problems (e.g. Algorithms for calculation of software metrics). We enrich CST by adding universal nodes to mark elements to become recognizable independently on input programming language.
To illustrate the technique to achieve independency of programming language we provide a simple example. It illustrates problems in calculation of Cyclomatic Complexity (CC) metric by predicate counting method.
The simple loop statement written in Modula-2 is stated as follows:

```
REPEAT
        …Some statements…
UNTIL (i > j);
```

The equivalent loop in Java would look like:

```
do{
        …Some statements…
}while (i <= j);
```

Although given statements have different syntax they express the same functionality: “some statements” in the code will be repeated until parameter “i” becomes greater then the parameter “j”. Beside the different syntax, condition for leaving the loop is oppositely stated. First condition express what condition should be fulfilled to leave the loop, while the second one states condition to continue looping.
Simplified syntax trees representing given statements are illustrated by Figure 1.

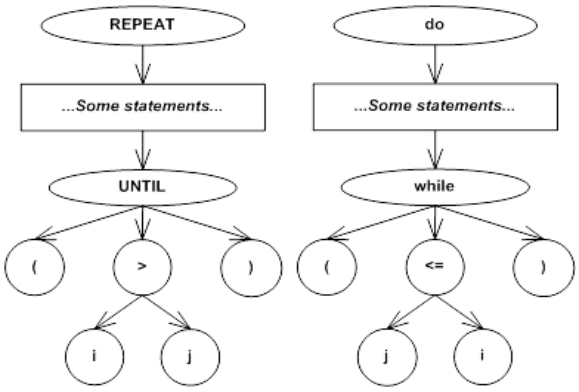


Figure 1: *Simplified syntax trees for REAPEAT-UNTIL (left) and do-while (right) statements*

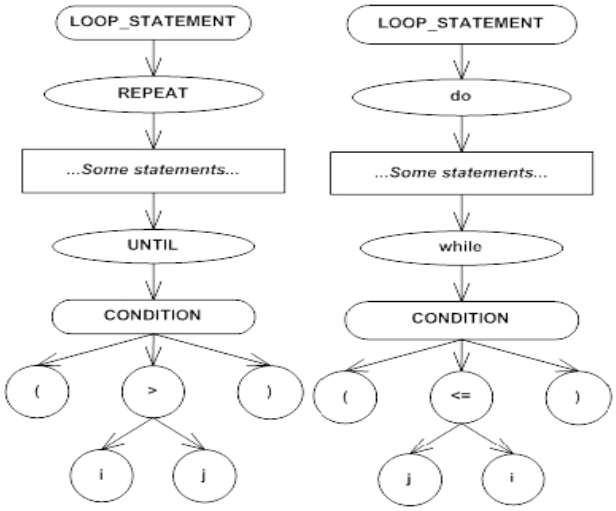


Figure 2: *eCSTs for REAPEAT-UNTIL (left) and do-while (right) statements*

For the implementation of CC algorithm we should recognize “REPEAT” and “while” as loops and to increment the current CC value by 1. It is clear that by using CST to represent source code we would need two implementations or at least two conditions to recognize these loops in the tree. By adding universal nodes “LOOP_STATEMENT” as parent of sub-trees that represent these two segments of source code we meet our goal by only one condition in implementation of the CC algorithm. Also we add universal node CONDITION to mark condition for leaving the loop repetition (Figure 2).
Additional enrichments for some other purposes could be including information about logical value that condition should have to leave the loop.
By adding all of needed universal nodes we implemented algorithms for CC metric independently on programming language. All we need is language grammar to modify and generate appropriate parser that is used for generating eCST.

New prototype of SMIILE tool that use eCST in metric calculation is described in [10]. This related paper describes language independent implementation of CC software metric based on universal nodes. It concentrates on CC as characteristic example for demonstration of usefulness of eCST in the direction of language independency of described tool. The paper provide table of used universal nodes in this prototype and provide way of usage in case of three characteristic languages – object-oriented Java, procedural Modula-2 and legacy COBOL.

4.2 Possible broader applicability

eCST was originally used in the development of language independent software metrics tool (SMIILE) [9]. Current prototype is implemented to support two software metrics and three languages.

However eCST has a limitation - it represents only separate compilation units. By translating all compilation units we get a set of autonomous trees. For the implementation of e.g. design metrics these trees should be connected by creating directed graph.

Idea for connecting compilation units is based on information about function (procedure, method, etc) calls contained in other functions. These calls could be placed either in the same or in some other compilation units.

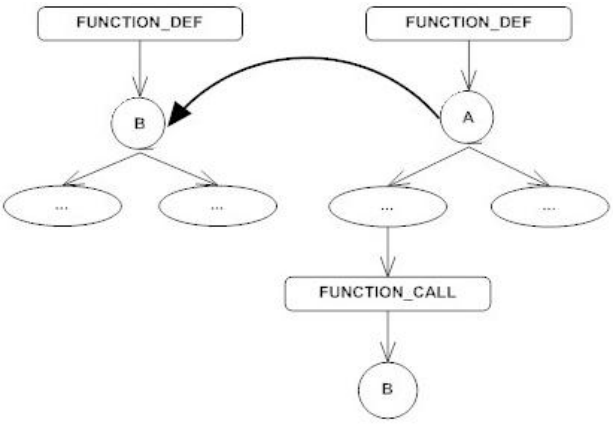


Figure 3: *Connecting of compilation units into call graph*

If function A contains call of function B than directed branch would lead from node representing function B to node that represent function A (Figure 3). Universal nodes (FUNCTION_DECL, FUNCTION_DEF and FUNCTION_CALL) would be used to locate the fragment of source code that contains function declaration, definition or call respectively. Information about function is placed in sub-tree of corresponding universal node.

Generated graph is a specific call graph. Maybe we can use even complex network (e.g. the one in [12]), but by creating the network by connecting eCSTs it would become language independent.

Additional possibility is to transform eCST to language independent enriched Control Flow Graph (eCFG), by inserting branches that represent possible execution paths through program (Figure 4).

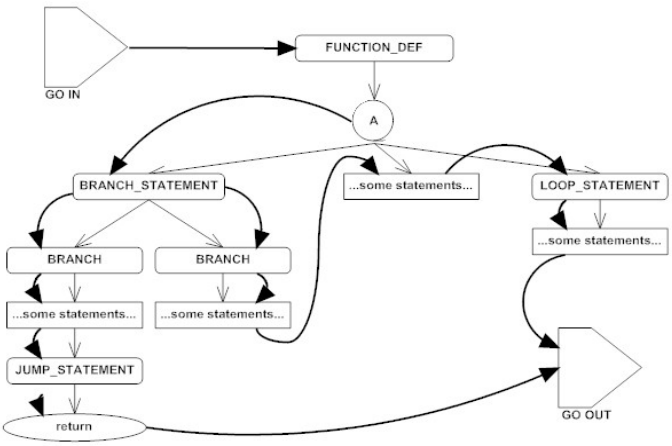


Figure 4: *Simplified view of eCST to eCFG transformation*

Generated eCFG could be used in software testing [4] (e.g. for development of automatic test case generator), in dead and duplicated code discovering, code-clone analysis [1],[2],[5], etc., but also as a basis for connecting compilation units instead of original eCST. In this case language independent call graph would be created by connecting eCFG components that represent compilation units.

Furthermore, we can notice that eCST could be used for automatic source code translation between programming languages. If we again consider given example we can conclude that given statement could be translated from Java to Modula-2 or from Modula-2 to Java. For automatic translation by using eCST we would have reflection table with rules for translation. In this example we should have rule about:

- How to translate the loop
- How to translate the condition
- How to translate inside statements

In this concept for translation we will not get perfectly written source code but by defining proper rules for translation we could manage to get correct source code. This limitation could be corrected by several cycles of code transformation [8].

SMIILE tool which is based on eCST for short-term goal had language independency, but as long term objective we stated smart software metrics tool that would recommend to developers how to improve their source code. It is planed to develop input language independent metric based advising system which would communicate with its user not only about metric values, but by concrete advices for corrections and refactoring of the source code based on calculated software metrics. For this purpose metric values should be stored in data storage. This storage could be well organized XML file system as primarily was proposed by SMIILE team, but also external software metrics repository could be used. Integration of SMIILE prototype with particular software metrics repository described in [11] is basis for further work in this direction.

Opportunity for improvement refactoring process gives additional value to described potential application of eCST

in code translation because needed after-translation refactoring could be automatically suggested or even applied.

The tool that integrates all described functionalities, including ones planed for SMIILE tool, would provide important features for consistent development of heterogonous software systems consisting of different components, implemented in different programming languages.

Furthermore SMIILE tool has possibility of keeping history of source code and corresponding software metrics. For keeping history of the source code eCST is stored to XML file created according to structure of eCST. By adding code-change analysis to the planed it would become important support in software reengineering process [6].

5 CONCLUSION

In this paper we introduce eCST and propose its usage in source code and model representation in development of universal tool to support different software engineering techniques and processes.

Idea for introducing eCST is supported by example of successful development of the prototype of language independent software metrics tool.

As this paper provide still fresh idea, it is clear that there exist numerous open questions and further work in proposed directions is planned.

ACKNOWLEDGMENTS

The authors acknowledge the support of this work by the Serbian Ministry of Education and Science through project "Intelligent Techniques and Their Integration into Wide-Spectrum Decision Support," no. OI174023.

References

[1] Baxter I.D, Yahin A, Moura L, Sant'Anna M, Bier L, Clone Detection Using Abstract Syntax Trees, *Proceedings of International Conference on Software Maintenance*, 1998. pp. 368-377

[2] Ducasse S., Rieger M., Demeyer S.,1999, A Language Independent Approach for Detecting Duplicated Code, *Proceedings. IEEE International Conference on Software Maintenance (ICSM '99)*, pp 109-118

[3] Fischer G., Lusiardi J., Wolff von Gudenberg J., Abstract Syntax Trees – and their Role in Model Driven Software Development, *In Proceedings of International Conference on Software Engineering Advances(ICSEA 2007)*, 2007

[4] Guangmei Z., Rui C., Xiaowei L., Congying L. The Automatic Generation of Basis Set of Path for Path Testing, In the *Proceedings of the 14th Asian Test Symposium (ATS '05),* 2005

[5] Koschke R, Falke R, Frenzel P, Clone Detection Using Abstract Syntax Suffix Trees, *Proceedings of the 13th Working Conference on Reverse Engineering (WCRE'06)*, 2006

[6] Neamtiu I., Foster J.S, Hicks M. Understanding Source Code Evolution Using Abstract Syntax Tree Matching, In *Proceeding of the International Conference on Software Engineering 2005, international workshop on Mining software repositories*, ISBN:1-59593-123-6, pp 1–5

[7] Parr T., *The Definitive ANTLR Reference - Building Domain-Specific Languages*, The Pragmatic Bookshelf, USA, 2007, ISBN: 0-9787392-5-6

[8] Pracner D., Budimac Z, Restructuring Assembly Code Using Formal Transformations, In *Proceedings Of International Conference of Numerical Analysis and Applied Mathematics ICNAAM2011, Symposium on Computer Languages, Implementations and Tools (SCLIT), September 19-25, 2011, Greece (in print)*

[9] Rakic G., Budimac Z., Problems In Systematic Application Of Software Metrics And Possible Solution, In *Proceedings Of The 5th International Conference on Information Technology ICIT 2011*, Jordan

[10] Rakic G., Budimac Z., SMIILE Prototype, In *Proceedings Of International Conference of Numerical Analysis and Applied Mathematics ICNAAM2011, Symposium on Computer Languages, Implementations and Tools (SCLIT), September 19-25, 2011, Greece (in print)*

[11] Rakic G., Gerlec Č., Novak J., Budimac Z., XML-Based Integration of the SMIILE Tool Prototype and Software Metrics Repository, In *Proceedings Of International Conference of Numerical Analysis and Applied Mathematics ICNAAM2011, Symposium on Computer Languages, Implementations and Tools (SCLIT), September 19-25, 2011, Greece (in print)*

[12] Savić M., Ivanović M., Radovanović M., Characteristics of Class Collaboration Networks, In *Large Java Software Projects, Information Technology and Control Journal*, Vol.40, No.1, 2011, pp. 48-58.

Zbornik 14. mednarodne multikonference

INFORMACIJSKA DRUŽBA – IS 2011

Proceedings of the 14th International Multiconference

INFORMATION SOCIETY – IS 2011

Kognitivne znanosti

Cognitive Sciences

Uredili / Edited by

Urban Kordeš, Matjaž Gams, Olga Markič

http://is.ijs.si

10.-11. oktober 2011 / October 10th -11th, 2011
Ljubljana, Slovenia

PREDGOVOR

Letos že osemnajstič nadaljujemo tradicijo vsakoletnih srečanj kognitivnih znanstvenikov v okviru Slovenskega društva za kognitivne znanosti in že trinajstič pod okriljem multikonference »Informacijska družba«.

Po Wikipediji so kognitivne znanosti »interdisciplinarne znanosti uma in miselnih procesov. Raziskujejo, kaj je in kako deluje kognicija. Pomemben del kognitivnih raziskav je študij procesiranja informacij v kognitivnih procesih (percepcija, jezik, spomin, razmišljanje, čustva), kako je razmišljanje povezano z obnašanjem pri ljudeh, živalih in računalnikih. Med vedami so psihologija, umetna inteligenca, filozofija, nevroznanost, lingvistika, antropologija, sociologija in vzgoja.«

Kognitivna znanost torej sestoji iz širokega spektra znanstvenih disciplin s skupnim ciljem, da bi pojasnile delovanje možganov in nevronskega sistema ter poiskale povezavo z duševnimi procesi in vedenjem.

To je dokaj novo in izrazito interdisciplinarno znanstveno področje, ki sodi v zadnjih deset letih med najbolj razvijajoče se znanstvene panoge. Od svojih začetkov v petdesetih letih dvajsetega stoletja in je doživela številne paradigmatske spremembe. Združuje različna raziskovalna področja, med katerimi so temeljne discipline nevroznanost, antropologija, umetna inteligenca, jezikoslovje, filozofija in psihologija. V začetnih letih razcveta kognitivne znanosti sta imeli vodilno vlogo umetna inteligenca in psihologija, kasneje – s pojavom metod za neinvazivno slikovno raziskovanje možganov – pa je prevzela vodstvo nevroznanost. V zadnjem času pa kognitivna znanost že dovolj dozorela, da med njene sestavne dele štejemo tudi področja aplikacije bazičnih spoznanj (kot so na primer pedagoške vede).

Kognitivne znanosti so nesporno v vzponu, kar med drugim dokazuje čedalje večje število dogodkov, npr. konferenc, pa tudi razpisi 7OP. Evropski okvirni program (FP) sestavljajo najbolj napredne usmeritve evropske znanosti in razvoja, pretežno v smeri informacijske družbe, npr. raznih naprednih programskih storitev. S 7OP pa se je prvič med programi pojavila tudi beseda »kognitivno« in to pri »kognitivni robotiki«. Gre za en spekter kognitivnih znanosti, za posnemanje človeških kognitivnih lastnosti in storitev na robotih.

Prav tako nesporno napredujejo spoznanja in dosežki kognitivnih raziskovalcev. V letu 2010 sta se zgodila dva izjemno pomembna kognitivna dogodka, ki jih je javnost v veliki meri povsem prezrla. Prva znanstvena objava je dokazala, da nevroni ne delujejo kot stikala, ampak kot računalniki (Stanford 2010). Besedilo navajamo v izvirniku: »: … the brain's overall complexity is almost beyond belief, said Smith. "One synapse, by itself, is more like a microprocessor —with both memory-storage and information-processing elements — than a mere on/off switch. In fact, one synapse may contain on the order of 1,000 molecular-scale switches. A single human brain has more switches than all the computers and routers and Internet connections on Earth," he said.«

Druga izjemna objava razlaga algoritem razmišljanja v možganih: »…neuroscientists uncovered strong evidence that neurons also communicate with each other through weak electric fields. The study, published in the journal Nature Neuroscience, by Dr. Costas Anastassiou (Caltech), explains how every time an electrical impulse races down the branch of a neuron, a tiny electric field surrounds that cell ..«

Vsekakor pa vsa pisana zgodovina te vede jasno kaže, da je problem raziskovanja in razumevanja človeških kognitivnih procesov tako zapleten, odprt in pomemben, da je kvalitetne rezultate moč pričakovati le v usklajenem interdisciplinarnem povezovanju različnih področij ter različnih vidikov opazovanja. Posamezne discipline lahko dajo delne odgovore, ki pa morajo biti združeni na višjem nivoju. Eno od temeljnih vprašanj kognitivnih znanosti je torej, kako integrirati raznolike pristope, saj se vsaka od disciplin problemov loteva s svojega zornega kota in uporablja svoj strokovni jezik in svoje metode. Da bi lahko prišlo do tovrstne integracije, se je najprej potrebno med sabo spoznati in razumeti (ter spoštovati) medsebojne razlike. Tudi letos nam je uspelo zbrati širok spekter raziskovalcev, ki se ukvarjajo z vprašanji, aktualnimi za kognitivno znanost. V letošnjem naboru prispevkov, ki so predstavljeni v pričujočem zborniku, lahko najdemo konceptualno-fiozofske razprave, poročila o kliničnih študijah, psihološke raziskave, sociološko-fenomenološke študije, poročila o pedagoško-edukacijskih raziskavah in še mnoge druge kamenčke iz bogatega mozaika kognitivnih znanosti.

Uredniki zbornika upamo, da se bo za vsakega bralca našel kak prispevek, ki ga bo inspiriral in zvabil izven ozkih okvirov lastnega disciplinarnega področja.

Urban Kordeš, Matjaž Gams, Olga Markič

KAKO IGRAJO MOŽGANI IZVAJALCA GLASBE

Viktor Avbelj
Odsek za komunikacijske sisteme
Institut Jožef Stefan
Jamova 39, 1000 Ljubljana, Slovenija
Tel: +386 1 4773824; fax: +386 1 4773111
e-pošta: viktor.avbelj@ijs.si

ABSTRAKT

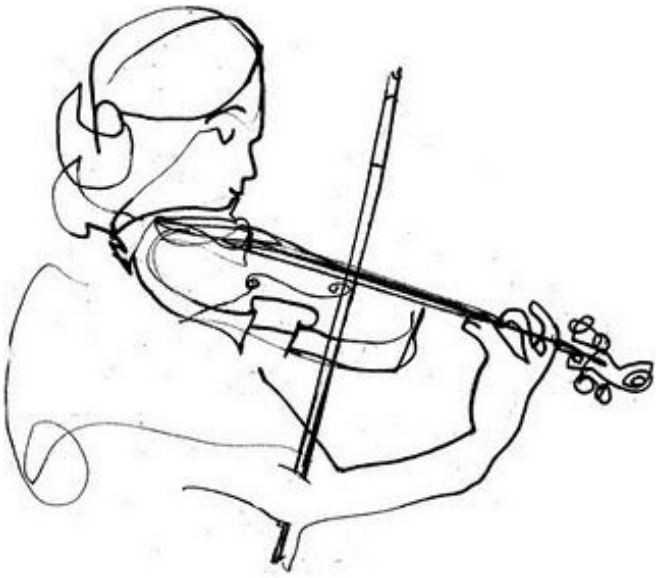
Kadar človek igra instrument, poje ali žvižga, poleg premikanja delov svojega telesa posluša še zvoke. Pri petju, žvižganju in igranju nekaterih instrumentov, kot so na primer godala, mora izvajalec poleg ostalega poskrbeti še za natančno višino tonov. V dolgotrajnem procesu vadbe izvajalec pripravi svoj organizem do ravni, ko poteka uravnavanje višine tonov v polju nezavednega. Možgani izvajajo drobne premike mišic, ki so povezani z rezultatom analize zvoka, ki prihaja v uho izvajalca. Gre za povratno zanko, ki se oblikuje v izvajalcu, poslušalci pa so deležni le zvoka, ki ga oddaja instrument oziroma slišijo glas, ki ga oddaja izvajalec. Tudi zato je izvajanje glasbe drugačno doživetje, kot je le njeno poslušanje. V prispevku se dotaknemo vprašanja, kako hitro poteka dogajanje v omenjeni povratni zanki med izvajanjem glasbe v polju nezavednega in kaj se v izbranem primeru dogaja v zavesti izvajalca.

1 UVOD

"Imaš dober posluh, lahko boš igral violino", so rekli otoku na sprejemnem izpitu v glasbeni šoli, potem ko je zapel pesmico, uganil nekaj glasbenih intervalov in ponovil ritmične vzorce s ploskanjem. Tistim, ki so imeli "slabši posluh" pa so svetovali igranje klavirja ali kaj drugega. In v čem je razlika med instrumenti? Pri violini mora violinist med igranjem s prsti skrbeti za natančno višino tonov (slika 1), medtem ko so višine tonov pri klavirju že v naprej določene - za vsako tipko jih je določil uglaševalec klavirja. Podobno kot violinist, mora višino glasu med petjem kontrolirati tudi pevec, le da pri tem uporablja glasilke.

Nedavno so v dveh raziskavah s pomočjo EEG in evociranih potencialov ugotovili, da se možgani ljudi "brez glasbenega posluha" odzivajo na posebnosti v predvajani glasbi, vendar se to dogaja ne da bi se teh posebnosti zavedali [1, 2]. S pomočjo slik, dobljenih z magnetno resonanco, so v tretji raziskavi ugotavljali anatomske posebnosti v zgradbi takih možgan [3].

Tudi brez kompleksnih merilnih naprav in postopkov je mogoče dobiti vpogled v dinamiko procesov, ki se odvijajo med kontrolo višine glasu, ki jo izvaja poskusna oseba. To je mogoče, ker se v povratni zanki nahaja sicer za meritve težko dostopen živčni sistem poskusne osebe, a tudi instrument, ki oddaja zvok in je zato za meritve idealno dostopen. Česa se zaveda poskusna oseba ob tem pa jo je mogoče povprašati.



Slika 1: *Violinistka med igranjem posluša zvok in uravnava višino tonov z lego prstov leve roke na ubiralki violine. (slika je iz vira /talknologytips.blogspot.com/)*

V nadaljevanju opišemo eksperiment, kjer smo namesto igranja na instrument izbrali žvižganje in predstavimo rezultate.

2 EKSPERIMENT

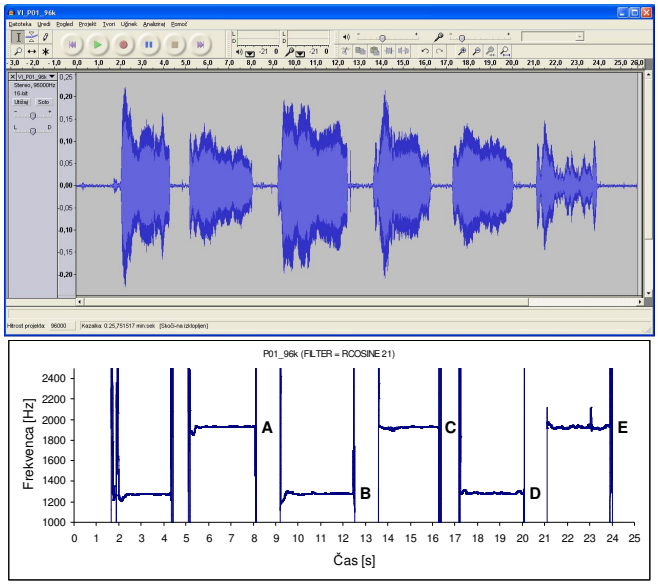
2.1 Opis

Osebi z osnovno glasbeno izobrazbo (štirje razredi violine) je bila v poskusu zastavljena preprosta naloga, da večkrat zapored zažvižga osnovni ton in kvinto navzgor ter pri tem posamezen ton zadrži nekaj sekund. Oseba si je lahko sama izbrala višino osnovnega tona in tudi število ponovitev. Po izvajanju oseba z besedami opiše, česa se je zavedala med izvajanjem naloge v zvezi z višinami tonov.

Zvok je bil posnet z diferencialnim mikrofonom (Logitech), ki zmanjša jakost motečih zvokov iz okolice. Mikrofon je bil nameščen ob strani ust, tako da oseba med žvižganjem ni pihala vanj. Signal z mikrofona je bil digitaliziran s frekvenco 96 kHz z zunanjo zvočno kartico Creative Labs, model SB1090, povezano z osebnim računalnikom preko USB priključka.

2.2 Rezultati

Oseba je izvedla poskus s tremi ponovitvami. Po vsakem tonu je vdihnila. Iz posnetka smo določili časovni potek frekvence tonov (slika 2).



Slika 2: *Zvočni posnetek poskusa s 6 toni (zgoraj) in časovni potek frekvenc, ki so bile dobljene z analizo tega zvočnega posnetka (spodaj, horizontalni deli grafa). S črkami A-E so označeni komentarji osebe in tudi izvajani toni.*

Oseba je po koncu izvajanja opisala, česa se je zavedala med izvajanjem in povedala, da posamezne misli, ki so se pojavile med izvajanjem naloge, niso bile v obliki stavkov (v nadaljevanju A-E), ampak so se pojavile kot fragmenti, ki bi jih lahko opisali z besedami "previsoko", "izboljšaj", "nepričakovano", "kompenzirano", "dobro", "ne gre več".

A Prva kvinta navzgor je previsoka - izboljšaj.
B Drugi osnovni ton je bil nepričakovano visok, a ker to zmanjša prejšnjo napako, ga vzamem kot dobrega in nadaljujem s kvinto navzgor od tu dalje.
C Dobro
D Dobro
E Pojavil se je nepričakovan šum in nestabilnost, raje končam.

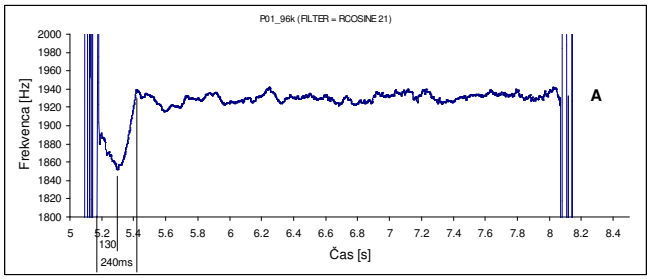
Frekvenco tona smo izračunali kot povprečje trenutnih frekvenc v drugi sekund izvajanja posameznega tona (Tabela 2). Za tonom, ki je označen s črko A, se je oseba zavedala, da je kvinta previsoka, medtem ko pri intervalu B-C zavedanja o netočnosti ni bilo. Oseba je kljub zavedanju, da je prva kvinta previsoka, vztrajala pri začetni višini tona (interval posnetka od 5. do 8. sekunde).

Tabela 1: *Povprečne frekvence izvajanih tonov, odstopanje kvint* in valovitost frekvence v drugi sekundi izvajanja vsakega tona.*

Ton	Frekvenca [Hz]	Odstopanje kvinte [Hz, (cent)]	Valovitost st. dev. [Hz]
prvi	1274.5		4.8
A	1930.7	19 (17.1)	4.2
B	1279.5		5.0
C	1926.9	7.7 (6.9)	5.8
D	1283.8		4.3
E	1928.1	2.4 (2.1)	8.2

* Odstopanje kvint je izraženo v Hz in v centih. Kvinto smo vzeli v pitagorejski uglasitvi [4], torej je razmerje frekvenc 2:3. Interval poltona je 100 centov.

Eksperiment je bil narejen z namenom, da bi lahko opazovali dinamično dogajanje, ki poteka v povratni zanki izvajalca (možgani, ustnice in jezik, zvok, uho, možgani). Dinamični del je viden v začetku izvajanja tonov. Potek frekvence ob izvajanju prve kvinte, pred znakom A na sliki 2, si podrobneje oglejmo na sliki 3.



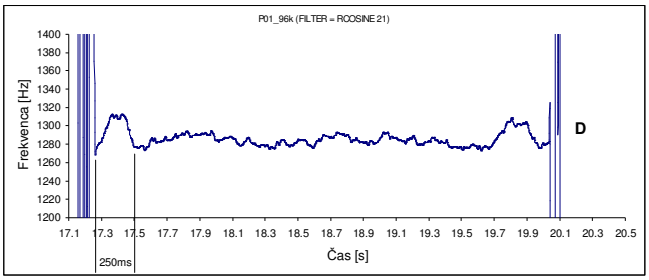
Slika 3: *Dinamična nastavitev frekvence pri kvinti A. Prehodni pojav izzveni v času 240 ms.*

V času 130 ms po začetku izvajanja tona, se je frekvenca nižala, čeprav je bila v začetku že nižja od končne. V naslednjih 110 ms je frekvenca hitro narasla do končne vrednosti in se tam skokovito zaustavila. Tudi v drugih primerih se je prehodni pojav zaključil že v prvih 300 ms. Zanimiv je primer D, kjer je oseba že v začetku izvajala natančno frekvenco, pa se je prehodni pojav vseeno zgodil in se zaključil v 250 ms (slika 4). Oseba je izvajanje tega tona komentirala kot "dobro".

2.3 Diskusija

Iz navedenih primerov izhaja, da se oseba začetnega prilagajanja frekvence ne zaveda in zato ne more zavestno posegati vanj. Nastavitev frekvence se izvaja avtonomno v povratni zanki, ki jo tvorijo možgani, ustnice in jezik, uho, možgani. Zavedanje o pojavih časovno ne sledi hitri

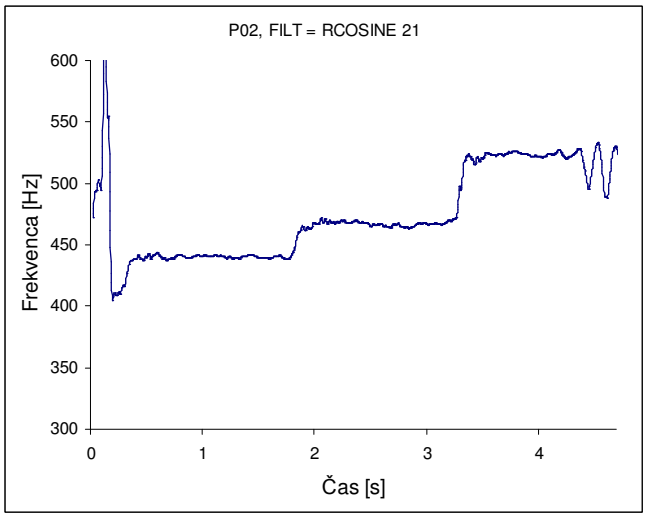
dinamiki dogajanja. Ko se oseba zave nenatančnosti izvajanja to uporabi pri izvajanju naslednjih tonov.



Slika 4: *Začetni del nastavitve frekvence pri osnovnem tonu z oznako D izzveni v času 250 ms.*

Ocenili smo premik jezika, ki je potreben za valovitost frekvence po začetnem prehodnem pojavu na sliki 3 in dobili velikostni razred 0,1 mm. S takšno natančnostjo morajo možgani nastavljati lego jezika pri žvižganju, da nas nenatančnost frekvence ne moti.

V našem eksperimentu je "instrument", ki ga "igra" poskusna oseba, kar del te osebe. Enako velja pri petju. Na sliki 5 je prikazan potek frekvence v času, ko je pevka zapela tri zaporedne tone in pri tretjem prešla v tremolo (valoviti del na koncu grafa) - intervala sta polton in cel ton. Prehodni pojav po spremembi tona ni izrazit.



Slika 5: *Frekvenca pri petju treh zaporednih tonov.*

Na podoben način bi bilo mogoče opazovati igranje pravih instrumentov, kjer mora izvajalec sam skrbeti za višino tonov (npr. godala). Pričakujemo, da bi prehodni pojav tu trajal dalj časa kot pri žvižganju, saj gre poleg drugega tudi za večje razdalje mišic od možgan.

3 ZAKLJUČEK

Vsak otrok ima pravico do glasbene vzgoje, kjer lahko razvije svoj glasbeni posluh ne glede na prirojene sposobnosti [5]. S tem je nekaterim prihranjeno zmerjanje da "fuša", kadar želi kaj zapeti.

Zaključimo lahko, da izvajalec in instrument predstavljata ubrano celoto, s katero nam izvajalec preko instrumenta daje svoj zavedni in nezavedni del sebe. V enostavnih primerih je pojave mogoče meriti, pri drugih nam ostane vse, kar je najlepšega v glasbi.

Reference

[1] Peretz I, Brattico E, Järvenpää M, Tervaniemi M. The amusic brain: in tune, out of key, and unaware. Brain. 2009 May;132(Pt 5):1277-86.

[2] Braun A, McArdle J, Jones J, Nechaev V, Zalewski C, Brewer C, Drayna D. Tune deafness: processing melodic errors outside of conscious awareness as reflected by components of the auditory ERP. PLoS One. 2008 Jun 11;3(6):e2349.

[3] Hyde KL, Zatorre RJ, Griffiths TD, Lerch JP, Peretz I. Morphometry of the amusic brain: a two-site study. Brain. 2006 Oct;129(Pt 10):2562-70.

[4] Vencelj M. Glasba in matematika - 1. del, tonski sistemi. Presek. 2002/2003;30(5);267-72.

[5] Internet: http://aiem-willems.org/monde/index.php?page=bioWillems&lang=si&code=

DOES BRAIN VOLUME REDUCTION AFTER MILD HYPOXIA IN NEAR-TERM BABIES AFFECT QUALITY OF LIFE IN ADOLESCENCE?

Tina Bregant[1], Milan Radoš[2], Lana Wasung[2], Vesna Zadnik[3], Metka Derganc[4], Ivica Kostović[2] , David Neubauer[1]

1) Department of Pediatric Neurology, University Children's Hospital, University Medical Centre Ljubljana, Ljubljana, Slovenia
2) Neuron Policlinics, Croatian Institute for Brain research - HIIM, Medical University Zagreb, Croatia.
3) Department of Epidemiology, Institute of Oncology, University Medical Centre, Ljubljana, Slovenia
4) Department of Pediatric Surgery and Intensive Care, University Medical Centre, Ljubljana, Slovenia

Contact: e-mail:tina.bregantdrmed@gmail.com

ABSTRACT

We studied the cohort of adolescents, who experienced mild to moderate hypoxic-ischaemic encephalopathy in neonatal period. We asessed their health status, performed the EEG and MRI, including the DTI and fMRI, which enabled to study the regional volume differences. We compared the cohort to a healthy, gender and age matched group of medical students.

Volume loss in specific brain regions was observed, hippocampus and right temporal lobe being most affected. Despite health issues, self-reported health status was better in the cohort of hypoxia-ischaemia than in a healthy group. Response shift was integrated into explanation of this phenomenon.

1 INTRODUCTION

Improved medical treatment increases the number of adolescents and young adults growing up with chronic diseases of childhood or consequences of acute disease. Hypoxic-ischemic encephalopathy (HIE) during the perinatal period is a commonly recognized cause of severe, long-term neurological deficits in children [1] . Neonatal encephalopathy occurs with the incidence of 1–6/1000 [2]. Among newborns affected by perinatal brain injury 20-50% die during the newborn period, and 25–60% of the survivors suffer from permanent neurodevelopmental handicap, including cerebral palsy, seizures, mental retardation, and learning disabilities [2-4].

Following a hypoxic-ischaemic insult, neuronal death occurs in two major phases, the primary neuronal cell loss at the time of the insult and the delayed neuronal cell loss [5]. Neuroimaging studies have shown, that brain injury following a hypoxic-ischaemic event evolves over days, if not weeks [6]. In the human infant, the severity of the secondary brain injury is correlated with adverse neurodevelopmental outcome.

Long-term changes in brain structures as well as in brain functioning may be observed. Hypoxia in a critical developmental period may affect brain development and maturation, affecting the outcome more than we previously thought. In determining neurological outcome the effects of neonatal brain injury on subsequent brain development are probably more important than simple tissue loss.

Neurological outcome does affect quality of life. However, one's self evaluation and self-perception may differ grossly from the clinical and objective changes in health status. Incorporating the model of response shift we can better understand how quality of life is affected by changes in health status, as well as by by medical or psychosocial interventions.

2 METHODS

2.1 Subjects

We have contacted 31 Slovenian families who had an infant, born near term between 1988-1990, and admitted to the PICU, University Medical Centre Ljubljana, due to mild to moderate HIE. Final cohort was of 16 adolescents who volunteered and did not receive any payment. 16 healthy, term children, age and gender matched, at the time of examination students of medicine, were recruited from the local community.

2.2 Measurements

The adolescents were neurologically examined and interviewed. Health-related quality of life was self-assessed by using SF-36v2 questionnaire, Rosenberg self-esteem inventory and Unwholesome behaviour questionnaire. We have used a multi-purpose, short-form health survey SF-36, second version, which is internationally accepted. The survey yields an 8-scale profile of functional health and well-being scores as well as psychometrically-based physical and mental health summary measures. It enables comparison of

relative burden of diseases among different populations and is also suited for older adolescents. We adressed the common causes of morbidity in adolescents by using the unwholesome behaviours scale which include smoking, alcohol drinking, overeating, watching television, playing computer games, and use of drugs. EEG after sleep deprivation was peformed.

2.3 Imaging

The adolescents were scanned with 3T Magnetom Trio Tim, Siemens, head coil 12 channels, regular sequences and sagittal 3D magnetization-prepared rapid acquisition gradient echo (MPRAGE) sequence with voxel size 1mm x 1mm x 1 mm. MPRAGE sequence consisted of TR=2300 ms; TE=3 ms; flip angle 9 °; matrix : 256 x 256.

Volumetric analysis was performed with the interactive software package DISPLAY developed at the Brain Imaging Center of the Montreal Neurological Institute. This program allows simultaneous viewing and segmentation of volumes in coronal, sagittal and horizontal orientations.

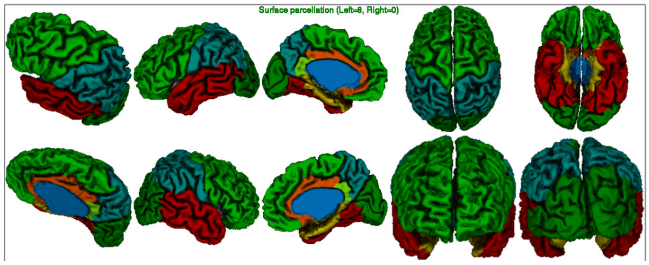


Figure 1: *Surface parcellation which enables to identify and measure selected brain structures. Parcellation is shown on 21 years old male brain who had neonatal HIE.*

3 RESULTS

3.1 Health status

16 young adults, 7 girls (43,8%) and 9 (56,3%) boys of average age 21,69 years (SD=0,87), born at the mean gestational age of 35,75 weeks (SD=3,80) and mean birthweight of 2644g (SD=815),were selected among the Slovenian infants who were admitted to the PICU due to mild (68,7%) to moderate (31,3%) HIE.

In a healthy students group only one (1) has inhallatory allergies. Health problems in the HIE group were: in three (3) myopia, in two (2) strabismus, where one (1) had been surgically corrected, in one (1) bruxizem, in two (2) epilepsy, where one (1) has a cortical dysplasia which is probably not linked to mild HIE, in one (1) diplegia (ambulatory with crutches), in one (1) backache with hip pain (tendinitis), in two (2) tension headache, two (2) had several episodes of bronchiolitis in childhood and inhalatory allergies, one (1) had food and one (1) inhalatory allergy. Sports activity is the same in both groups.

Adolescents with HIE had attained a diverse educational level: one (1) has finished elementary school (1), nine (9) have finished secondary school and six (6) have finished the baccallaureate programme which enables the study at the faculty. Four (4) continue the education at the college; six (6) continue at the university.

Despite the fact that adolescents with HIE reported higher additional morbidity they claimed to have a good quality of life. Health-related quality of life is rated high: for adolescents with HIE M=81,9 (SD=11,2) and for healthy adolescents M=75,3 (SD=11,5); p=0,112. Mental health – summary was reported statistically significantly higher in adolescents with HIE. Mental health components, such as vitality, social functioning, role emotional and defined mental health were rated higher.

Adolescents with HIE did not distinguish from healthy adolescents in self-esteem (p=0,68) and unwholesome behaviours, except for over-eating (p=0,01) Comparison of unwholesome behaviours in adolescents with HIE and healthy adolescents, showed no statistical difference in smoking, alcohol consumption, television watching, computer games, and use of drugs.

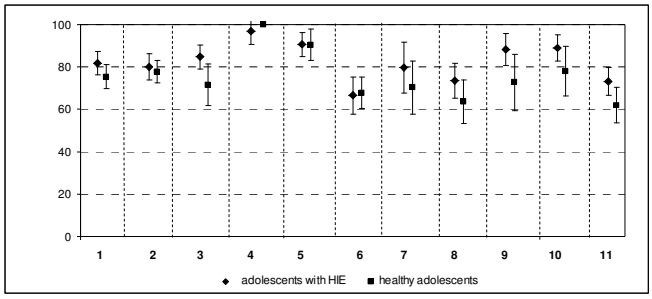


Figure 2: *Comparison of quality of life in adolescents with HIE and healthy adolescents, where columns represent: 1- quality of life; 2 - physical health (summary); 3 – mental health (summary); 4 - physical functioning; 5 - role-physical, 6 – general health; 7 - bodily pain; 8 – vitality; 9 - social functioning; 10 - role-emotional; and 11 – mental health. P value calculated with the t-test for quality of life P=0,112; physical health (summary) P= 0,61; physical functioning P=0,99; general health P=0,818; bodily pain P=0,301; vitality P=1,46; mental health P=0,046. P value calculated with the Mann-Whitney U test for mental health (summary) P=0,026; role-physical P=0,806; social functioning P=0,061; role-emotional P=0,19. Cronbach alpha coefficients for all adolescents ranged from 0,61 to 0,83; adolescents with HIE gave more reliable answers.*

3.2 Imaging

Despite the general well being of young adults MRI findings decline from the normal MRI in terms of white matter changes – predominantly changes that resemble non-cystic periventricular leucomalacia (PVL), lateral ventricles enlargements and thinned corpus callosum.

Imaging showed regionally specific significant reductions in adolescents who had HIE when compared to healthy adolescents. Statistically significant differences were more observed in males. Statistically significant changes were observed in gyrification index of gray matter for both hemispheres (p=0.052 for left and p=0.018 for right hemisphere; p=0.026 for right hemisphere in males). Statistically significant difference was found in lobe areas for whole group (p=0.032 right temporal, p=0.040 right hippocampus); in males (p=0.036 left temporal; p=0.043 left parietal; p=0.057 left frontal and p=0,047 for all left lobes, p=0.011 right temporal, p=0.045 right hippocampus, p=0.029 total right lobe area); right lobe areas for females (p=0.045 right parietal, p=0.029 right cingulum). Right lobe thickness was significantly different (p=0.031 for others in whole group, p=0.072 in other areas in males); including right isthmus cinguli (p=0.038). Right lobe volume was statistically significantly smaller in males (p=0.042 right temporal, p=0.057 right isthmus cinguli, p=0.062 right hippocampus);) and in both sexes (p=0,028 right temporal, p=0.083 right hippocampus). Total brain volume was smaller in males (p=0.021).

From the data we can see the preponderance of right hemisphere involvement with male subjects showing more statistical differences in specific regions, esp. temporal and hippocampal region. We also observe a near normal MRI findings in some cases of mild HIE. In these cases thinner corpus callosum and larger pineal gland cyst is observed.

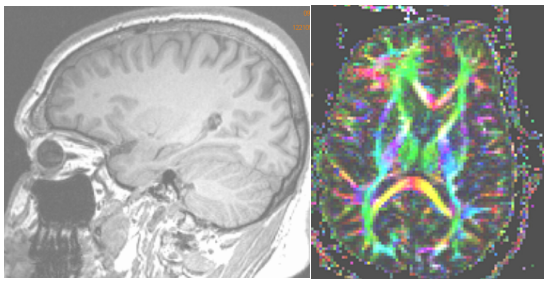


Figure 3: *Difusion tensor imaging (DTI) with 3T Magnetom Trio Tim, Siemens, regular sequences and sequence MPRAGE, voxel size 1mm x 1mm x 1 mm, 30 directional FA in a 21-year old boy, who had neonatal HIE. He has bothsided hippocampal atrophy with globally atrophic white matter.*

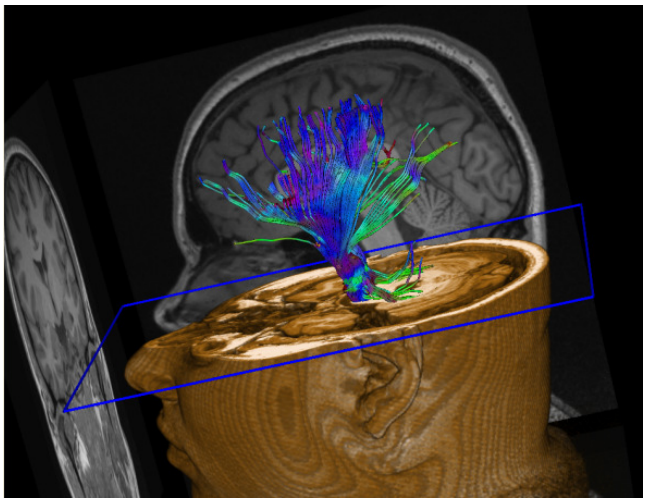


Figure 4: *Difusion tensor imaging (DTI) with 3T Magnetom Trio Tim, Siemens, regular sequences and sequence MPRAGE, voxel size 1mm x 1mm x 1 mm, 30 directional FA in a 20-year old boy, who had neonatal HIE. Now he has no health concerns, he studies at Faculty of forestry and is a recipient of gifted-children bursary. His volumetric data are comparable to other healthy adolescents except for a thinner corpus callosum and a large pineal cyst.*

4 DISCUSSION

Hypoxic brain injury is age-dependent and brain region-specific [7]. Our findings suggest that even mild perinatal hypoxia-ischaemia causes permanent structural changes in the brain that persist into adulthood. Rapid neurobiological development in late gestation and early infancy is disturbed by HIE leading to regional specific deficits, which we have observed in our studied group. From the data we can see the preponderance of right hemisphere involvement with male subjects showing more statistical differences in specific regions, especially in the temporal and hippocampal region. Hippocampal lobe volume is affected bilaterally. Volume loss is sugestive of neuronal necrosis selective to the area of high energy demand. This suggests hippocampus being very active and also highly vulnerable during the late gestation-early infancy. In our group the border-line statistically significant reductions were observed also in fronto-parietal lobe areas for both hemispheres. It has been shown, that near-term neural injury involves the parasagittal cortex, the dorsal horn of the hippocampus, and the cerebellar neocortex after insults including pure ischaemia [8] and prolonged partial asphyxia [9-10]. In both adults and children, lesions in these areas are seen after systemic hypotension [11]. Statistically significant changes were more observed in volume changes in the male HIE group, indicating that male brain is more vulnerable to perinatal mild-moderate hypoxia. The presence of large pineal gland cyst is suggestive of endogeneous mechanism of brain repair in near-term brain [12].

In the interviews, we expected at least some memory problems to be reported, since hippocampus is significantly affected. Only the boy with epilepsy (1) reported significant memory problems. On detailed questioning, three (3) reported some memory problems. Despite the medical issues the group with HIE has, they do not report them as difficult to manage. Their quality of life does not deteriorate significantly with the condition. It is known, that patients, confronted with a life-threatening illness or chronic condition are faced with the necessity to accommodate to their illness. An important mediator of adaptation is »response shift«, which involves changing internal standards, values, and conceptualization of quality of life [13]. Response shift can be well incorporated into other existing theories: adaptation theory, discrepancy theory, stress-coping theory. In adolescents with HIE as well as in interviews with the parents the most striking difference was observed in their internal standards, values and conceptualization, when compared to the medical student's group (Fig. 5).

Within the studied group, health care was well coordinated and interdisciplinary oriented. The families of adolescents with HIE promoted the adolescents towards adult independence. They claimed the role of education and employment of utmost importance. The level of education was adapted to individual capabilities, which in some cases were relatively low. Schooling was structured, so that children were directed to the secondary schools and only when successful, redirected to the baccalaureate programme. The whole family, including the siblings, were involved in care for the children with HIE. Studies show that parental support contributes significantlly to emotional adjustment in adolescents preventing them from emotional disturbances that are liable to become chronic [14]. Mental health was indeed better in the group of HIE. Two major predictors of subjective quality of life in adults are known to be self-esteem and a sense of primary control, which is more used in younger children [15]. Self-esteem in HIE group was higher, but statistically not significantlly different from healthy students. Probably self-esteem in HIE group was high enough to contribute to the sense of good quality of life. We think that adequate family functioning with a school support and well coordinated health care enpowered the adolescents with HIE to feel confident about themselves. We hope that society will support these young adults and their independence beyond the adolescence by offering them an appropriate employment.

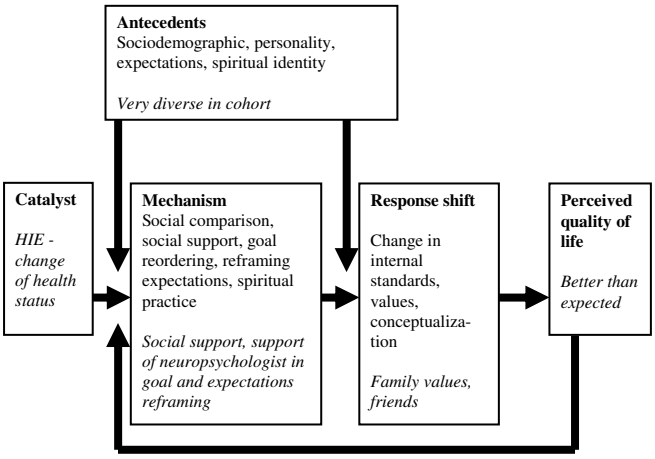


Figure 5: *A model of response shift and quality of life in adolescents with HIE.*

5 CONCLUSIONS

Our data suggest that mild to moderate perinatal hypoxia-ischaemia can be associated with structural volume loss observed in the early adulthood. Decrements in brain volume associated with mild to moderate hypoxia-ischaemia are localized to hippocampus and are more observed in males.

Despite the objective changes the adolescents with perinatal hypoxia have, their quality of life is very good. Their mental health is rated even higher than in healthy students. We think, that response shift from higher to lower academic aspiration, orientation towards family instead of career values, vawing goodbye to to the imaginatory "ideal" child nearly 20-years ago and the help of family support, well-coordinated health care, adequate school system and neuropsychological backup, they developed compensatory strategies, which enabled them to have a good quality of life. We can hope that their compensatory strategies do not become insufficient when they reach adulthood but enable them a high quality of life instead.

References

[1] Vannucci RC. Hypoxic-ischaemic encephalopathy. *Am J Perinatol* 2000; 17(3): 113-20.

[2] Volpe JJ. *Neurology of the newborn*. 4th ed. Chicago: Saunders; 2001.

[3] Inder TE, Volpe JJ. Mechanisms of perinatal brain injury. *Semin Neonatol* 2000; 5(1): 3-16.

[4] Mercuri E, Guzzetta A, Haataja L, Cowan F, Rutherford M, Counsell S, et al. Neonatal Neurological Examination in Infants with Hypoxic Ischaemic Encephalopathy: Correlation with MRI findings. *Neuropediatrics* 1999; 30: 83-9.

[5] Inder TE, Warfield SK, Wang H, Huppi PS, Volpe JJ. Abnormal Cerebral Structure Is Present at Term in Premature Infants. *Pediatrics* 2005; 115 (2): 286 -294. doi: 10.1542/peds.2004-0326.

[6] Volpe JJ. Brain injury in premature infants: a complex amalgam of destructive and developmental disturbances. *Lancet Neurol* 2009; 8 (1): 110-24. doi: 10.1016/S1474-4422(08)70294-1.
[7] McQuillenPS, Ferriero DM. Selective vulnerability in the developing central nervous system. *Pediatr Neurol* 2004; 30: 227–235.
[8] Mallard EC, Gunn AJ, Williams CE, Johnston BM, Gluckman PD. Transient umbilical cord occlusion causes hippocampal damage in the fetal sheep. *Am J Obstet Gynecol* 1992; 167(5): 1423–30.
[9] de Haan HH, van Reempts JL, Vles JS, de Haan J, Hasaart TH. Effects of asphyxia on the fetal lamb brain. *Am J Obstet Gynecol* 1993; 169(6): 1493–501.
[10] Ikeda T, Murata Y, Quilligan EJ, Choi BH, Parer JT, Doi S, et al. Physiologic and histologic changes in near-term fetal lambs exposed to asphyxia by partial umbilical cord occlusion. *Am J Obstet Gynecol* 1998; 178(1 Pt 1): 24–32.
[11] Torvik A. The pathogenesis of watershed infarcts in the brain. Stroke 1984; 15(2): 221–3.
[12] Bregant T, Radoš M, Derganc M, Neubauer D, Kostović I. Pineal cysts - a benign consequence of mild hypoxia in a near-term brain. *Neuroendocrinology letters 2011*; in print.
[13] Sprangers MAG, Schwartz CE. Integrating response shift into health-related quality of life research: a theoretical model. *Social Science and Medicine* 1999; 48: 1507-15.
[14] Lipschitz-Elhawi R, Itzhaky H. The contribution of internal and external resources to emotional adjustement: a comparison of at-risk and normative adolescents. *Child and Adolescent Social Work Journal* 2008; 25 (5): 385-96.
[15] Marriage K, Cummins RA. Subjective Quality of life and Self-Esteem in Children: The role of primary and secondary control in coping with everyday stress. Social *Indicators research 2004;* 66 (1, 2): 107-22.

CASE STUDY IN INNOVATION DIFFUSION: AIRCRAFT ELECTRICAL SYSTEM

Jelena Ćosić, Josip Stepanić
Department of Quality
Faculty of Mechanical Engineering and Naval Architecture – University of Zagreb
I. Lučića 1, 10000 Zagreb, Croatia
Tel: +385 1 6168 592; fax: +385 1 6168466
e-mails: jelena.cosic@fsb.hr, josip.j.stepanic@fsb.hr

ABSTRACT

Civil aviation technology includes two tendencies. One tendency is to unify optimal, proven methods and techniques in construction, manufacturing and maintenance related to aircrafts and aviation in general. The other tendency is to broaden, virtually without limits imposed within the aviation industry, set of existing solutions exploited.
This paper presents an analysis of these two tendencies, and context within which they evolve.
We state a conjecture about the possible limits, i.e. the envelope of exploitation of diverse construction, manufacturing and maintenance methods and techniques.

1 INTRODUCTION

On the most general level, aviation enables people to conduct their tasks using air as a medium. That medium may serve as a transport route, place for collecting information, strategic position enabling some, usually larger groups of people to gather advantage in power in comparison to other groups, etc.
Such tasks have been successfully conducted during the last century. When projected onto the level of aircraft technology, they are transformed into the following guidelines, rather stringent and perpetually imposed onto all parts and systems utilised:
(i) minimize power consumption,
(ii) minimize mass,
(iii) maximize reliability.
These guideliens are not mutually independent, as e.g. lowering the mass of some part automatically lowers the power which the engines need to develop in order to keep the whole aircraft in the air. Furthermore, they are not mutually congruent, since e.g. reliability is augmented using multiplication of equipment and utilizing more than one working principle of some method. That, in turn, is contrary to other two guidelines.
Moreover, reliability enhancement, or preservation in otherwise modified aircraft, is further developed into several prescriptions: use modular architecture of aircraft systems, minimize number of moving parts, etc.
Listed guidelines, and consequently prescribed approaches, have brought about during the last century, which contains most of the aircraft flying, two solidly founded, spontaneously developed, constantly present tendencies. Let us call them convergind and diverging tendency.
Converging tendency means that aircraft's producers and users try to utilize some solutions which have been proven as contemporary optimal. Either by their own development, or by adopting otherwise existing solutions, that tendency is realized as diffusion of best practices. Overall, a consequence is that best practice gradually suppresses other practices, thus in some segment of equipment and systems, realized solutions converge toward the contemporary optimal one.
Diverging tendency means that some realization of a system or its parts, gradually develops in number of variants in order to become better aligned with specific requirements imposed on the aircraft systems and performances. Requirements are caused e.g. by customers, by efforts to mutually align different aircraft's systems as good as possible.
Listed notions, i.e. "best practices", "tendencies", as well as meny realizations which are consequences of the three guidelines, are simply innovations. Diverse processes related to them, such as introduction, adoption, etc. are covered by the notion of innovation diffusion.
Innovation diffusion is a process having rather generally unique dynamics.
In this article we set the foundation for analysis of capabilities of some system, aircraft in particular, regarding the different amount of realization of some innovations within hts systems. In particular, our goal is to observe the dynamics of aggregated quantities, generalized potentials like entropy and free energy, in the capabilities of aircraft. In that sense, this paper is continuation of previous work on that subject [1-3].
In the second section we briefly state main characteristics of innovation diffusion. In the third section aircraft electric system is considered in some details. Main steps in its development during the last century are listed, and are interpreted from the point of view of innovation diffusion. Last section concludes that case study and provides the readers with the projections of further work.

2 ELEMENTS OF INNOVATION DIFFUSION

Innovation diffusion is a process of adopting an innovation by entities of a some system [4]. It is rather sociological process, since the technical development of innovation must be in a considerable part finished in order that the innovation becomes interesting to the entities. These entities can be individuals in some social system, firms in a regional or global economy [5], etc. In this paper, by entities we consider aircraft producers.

According to E. Rogers, there are five main stages of entities' adoption of innovations:

(i) knowledge,
(ii) persuasion,
(iii) decision,
(iv) implementation and
(v) confirmation.

In particular, in the context of aviation technologies, **knowledge** implies a set of known methods, techniques, principles, experiments, testing. Some parts could have been known in scientific or other communities for decades before they become utilized within the aviation industry. An example of the former approach is the concept *Fly-by-wire*, while an example of the later are optical gyroscopes exploiting Sagnac's effect.

The **persuasion** means that facts within the knowledge ar actively collected and that entities are interested in a particular set of facts forming knowledge. Naturally, entities could initiate applied research & development processes in order to enlargen understanding of some principle, or in order to develop some method or technique. The example of such an approach is *Fly-by-Wire* concept.

It is interesting to note that in this case, persuasion precedes knowledge.

Stage of **decision** is the very stage in which an innovation is adopted or rejected. It is usually interrelated with analysis of previously formed knowledge. However, some decision stages may induce further collecting of knowledge, e.g. through comparison of innovation performance in diverse conditions, or through comparison of performances of several innovations in the same conditions.

Implementation is a stage in which an innovation becomes part of the entity's regular processes. The percentage in which that innovation is used may differ considerably among diverse systems. Similarly to previous stages, in this stage knowledge is broadened. However, that broadening occurs spontaneously, through observations of regular processes with adopted innovation.

Finally, innovation diffusion ends with the stage of **confirmation**. That, functionally marks the end of adopted innovation as an innovation, and transforms it into an established practice. It is expected then, that gradually that practice becomes more and more ubiquituous and that its further interference with innovations occurs only if its disadvantages arise.

Innovation diffusion by the aircraft producers are somewhat different than innovation diffusion by individuals. The reasons are that set of operations is more restricted in the aviation industry than in life of individuals, and furthermore that stages like persuasion and decision are conducted using different criteria.

If number of entities adopting some innovation are considered, than time dependence of that adoption follows the logistic curve. Thus in the early phase of innovation diffusion, rather small percentage of entities, the more innovative ones, adopt an innovation. That is followed with increase in percentage of innovation adopters. That forms the **diverging tendency**, as more and more adopters rise the number of variants of an innovation.

However, when most of the entities adopted an innovation, the percentage of its adoption by the rather small part of further possible adopters generally implies that knowledge about innovation collected in implementation and confirmation stages of early adopters contributed to further profiling of characteristics of related innovations. In that sense, spontaneously, the hierarchy of related innovations forms pointing to one or a small subset of related innovations which are further adopted. That contributes to **converging tendency**.

Overall, constantly there are innovations which contribute to diverging and converging tendencies, regarding different aircrafts' systems.

3 CASE STUDY: AIRCRAFT ELECTRICAL SYSTEMS

Modern aircrafts are similar to powerplants regarding their generation of electricity, and to villages regarding their use of electricity. However, during previous decades, intensity of utilization of electrical energy in aircrafts have been developing in many steps.

Historically, first aircrafts did not use electricity [6]. That was folloed with the aircrafts utilizing electricity in modest amount, e.g. for radio-navigation, lights, some instruments etc. An important step for the use of electricity in aircrafts was, in fact, start of using of hydraulics (for landing gear, flaps, etc.) as the hydraulics required well pressurized system. The commonest pumps for setting the hydraulic system became the electric motor driven pumps.

Thus, constantly the number of components, for which the electric energy was needed, arose [7]. Furthermore, the very electric network developed, both because of more and more components and because of the redundancy realized as multiplication of systems. The fundamental characteristic of the electric energy, the electromotive voltage and character of the current also developed, hence also broadened in number of realizations. While first aircrafts with electric power utilized direct current, later the aircrafts usually utilized several types of electric energy: direct current of low and higher voltage, single-phase alternating current and triple-phase alternating current of frequency 400 Hz and lower and higher voltages, etc.

It is, however, interesting to note that such diverting tendency stil persists with introduction, or at least experimenting with different voltages and frequencies of the electric energy utilized.

On the other hand, contemporary, most of the newly introduced aircrafts utilize completely new, and rather different electric current: the current of variable frequency. Since it has rather useful characteristics regarding the introductory stated tendencies, main aircraft producers have exploited that concept in their new products such as Airbus A 380 or Boeing 787.
Overal, importance of the electric energy for aircrafts have been formulated within the rather novel concept of the *More-Electric-Aircraft*, along with its limit, the *All--Electric-Aircraft*.

4 CONCLUSIONS AND PERSPECTIVES

This paper develops the approach to aircrafts as systems in which dynamics of innovation diffusion can be observed, and related to the overall aircraft capabilities.
This provides setting for application of previously defined system-level, i.e. aggregated, potentials as indicators of the overall aircraft capabilities.

References

[1] J. Stepanic, H. Stefancic, M.S. Zebec and K. Perackovic: Approach to a Quantitative Description of Social Systems Based on Thermodynamic Formalism", Entropy **2**(3), 98-105, 2000,
[2] J. Stepanić, G. Sabol, M.S. Žebec: Describing social systems using social free energy and social entropy. Kybernetes 34(**5**-**6**), 857-868, 2005.
[3] K. Frenken and L. Leydesdorff: Scaling Trajectories in Civil Aircraft (1913-1997).
[4] E.M. Rogers: Diffusion of Innovations. Free Press, New York, 1983.
[5] A.D. Zimm: Derivation of a Logistic Equation for Organizations, and its Expansion into a Competitive Organizations Simulation. Computational & Mathematical Organization Theory **11**, 37–57, 2005.
[6] S. Ransom and R. Fairclough: English Electric Aircraft and Their Predecessors. Putnam, London, 1987.
[7] I. Moir and A. Seabridge: Aircraft Systems. 3rd ed. Willey, 2008.

UGOTAVLJANJE ASERTIVNOSTI MED PROFESORJI SREDNJE ŠOLE JESENICE

Nevenka Dolenc
E-pošta: nevenka.osredkar@gmail.com
Srednja šola Jesenice

IZVLEČEK

Z asertivnostjo opisujemo način ravnanja z ljudmi, ki udejanjanje lastnih ciljev vodi na odprt in pošten način. V prispevku opisujem ugotovitve asertivnosti v kolektivu profesorjev Srednje šole Jesenice. Rezultati anketnega vprašalnika nakazujejo, da je 33 anketiranih profesorjev v povprečju ravno nad mejo (45) še sprejemljive oziroma primerne asertivnosti, saj rezultat maksimalno število doseženih točk 68 (starost 54 let), kot tudi minimalno število točk 35 (starost 50 let) ankete sta si prislužili dve profesorici.

1. UVOD

Namen prispevka je bil odkriti asertivnost kolegov v službi. Glavni razlog za izbor anketnega vprašalnika (Poličnik, str. 16) o tej problematiki je, da že nekaj časa opažam slabo komunikacijo v kolektivu, ki se predvsem odraža v nezadovoljstvu do nadrejenih.

2. TEORETIČNA IZHODIŠČA

Na delovnem mestu prevladuje poslovno komuniciranje, kjer si izmenjujemo informacije, mnenja, rešujemo probleme, raziskujemo… Poslovno komuniciranje poteka interno med sodelavci in nadrejenimi ter eksterno (sodelovanje s starši). Z vidika uporabe simbolov med sodelavci ločimo besedno in nebesedno komuniciranje (Nose, K. Prevodnik, str. 24).

Poznamo dve vrsti tipičnih komunikacijskih vedenj, ki predstavljata obe skrajnosti. Agresivnost pomeni, da se vedno postavimo zase in zgolj za svoje pravice (ljudi napadamo, kaznujemo, se radi prepiramo…). Drugo vedenje je izrazito pasivno, saj pogosto ne izrazimo svojih čustev, prepričanj in mnenj. Zato potrebujemo še bolj jasno izoblikovan tretji način komuniciranja, ki nam bo pomagal pri reševanju zapletenih situacij (Možina in drugi, str. 422).

Z asertivnostjo opisujemo način ravnanja z ljudmi, ki udejanjanje lastnih ciljev vodi na odprt in pošten način (Poličnik, str. 24; Možina in drugi, str. 423). Pri asertivnosti se zavedamo, da (Možina in drugi, str. 423):

- Imamo potrebe. Prav tako jih imajo tudi drugi.
- Imamo pravice. Prav tako jih imajo tudi drugi.
- Lahko nekaj prispevamo. Prav tako lahko prispevajo tudi drugi.

Asertivna oseba jasno izraža svoja čustva, želje in potrebe. Išče rešitve, ki so sprejemljive za obe strani. Deli svoje informacije, mnenja in občutke s sogovorniki. Ko doseže dogovor, se dobro počuti. Je aktivni poslušalec. Jasno pove svoja stališča in odločitve ter se odločno sooča s sogovornikom (Poličnik, str. 28; Možina in drugi, str. 424).

Koristi asertivnega vedenja (Možina in drugi, str. 424):

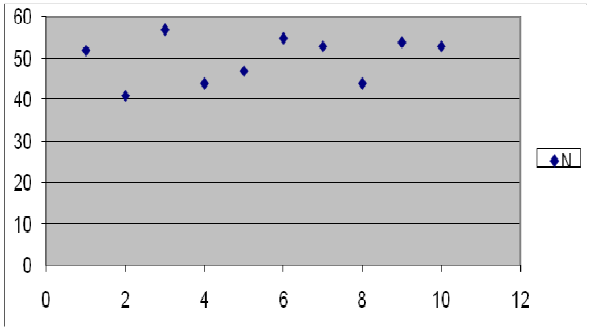
- Ljudje so bolj zadovoljni sami s seboj.
- Z asertivnostjo povečujemo svojo učinkovitost in lažje dosežemo cilje.
- Bolje obvladujemo konflikte in druge težke situacije.
- Občutno lahko zmanjšamo stopnjo stresa na delovnem mestu.

3. PREDSTAVITEV IZBRANIH PODATKOV O PRIMERU IZ PRAKSE

Anketni vprašalnik "Preverite svojo asertivnost" je vključeval 14 trditev, katere so anketiranci ocenjevali od 1 do 5. Zato so bili možni rezultati od 14 do 70. Rezultat anket pod 45 opozarja na to, da bi bilo dobro izboljšati vedenje, saj bi bilo v vzpostavljanju in vzdrževanju medsebojnih odnosov veliko lažje (Poličnik, str. 16).

V anketo je bilo vključenih 33 (54,1 %) od 61 zaposlenih profesorjev na Srednji šoli Jesenice. Sodelovalo je 10 profesorjev moškega spola, v povprečju starih 43,7 let (SD = ± 10,8) in 23 profesoric, v povprečju starih 42,7 (SD = ± 9,1).

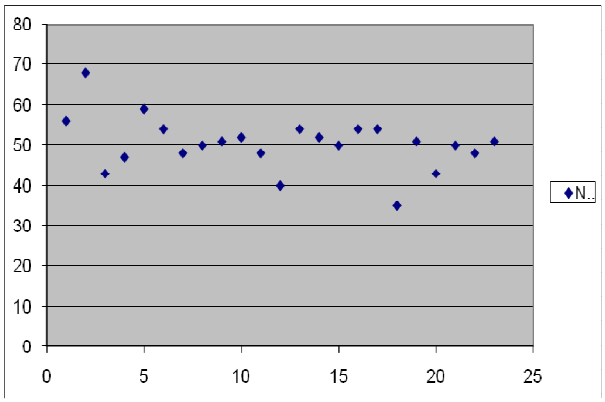
Graf 1: Grafični prikaz asertivnosti profesorjev
Povprečno Legenda: N – stopnja asertivnosti posameznega profesorja



Število točk asertivnosti pri moških je bilo 50 (SD = ± 5,5), maksimalno 57 in minimalno doseženo število točk 41.

Graf 2: Grafični prikaz asertivnosti profesoric

Legenda: N –stopnja aertivnosti posamezne profesorice



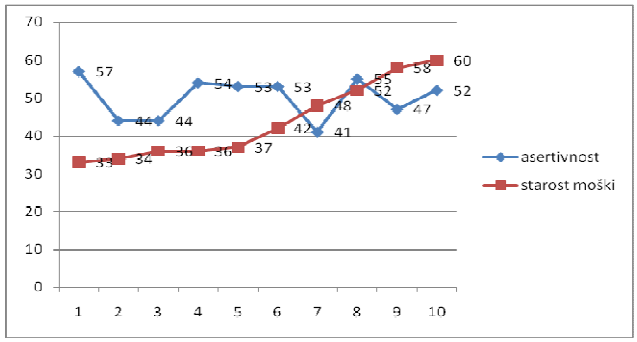
Povprečno število točk pri ženskah je bilo 50,3 (SD = ± 6,6), maksimalno 68 in minimalno doseženo število točk 35.

4. ANALIZA IN INTERPRETACIJA IZBRANIH PODATKOV – SOOČENJE S TEORIJO

Rezultati anketnega vprašalnika nakazujejo, da je 33 anketiranih profesorjev v povprečju ravno nad mejo (45) še sprejemljive oziroma primerne asertivnosti, saj rezultat nakazuje naslednje: povprečje dobljenih točk iz ankete je 49,8 (SD = ± 6,3). Tako maksimalno število doseženih točk 68 (starost 54 let), kot tudi minimalno število točk 35 (starost 50 let) ankete sta si prislužili dve profesorici.

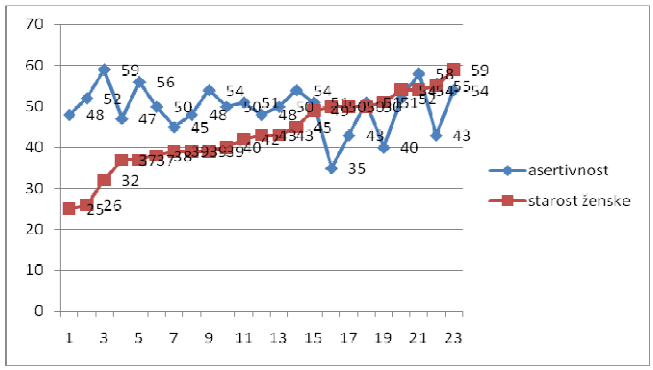
Kljub temu že izračun standardnega odklona (SD) nakazuje, da je nekaj profesorjev manj asertivnih. 4 profesorice (17,4 %) od 23-ih imajo rezultat pod 45 točk in sicer je njihovo povprečje seštevka točk 41. Zanimivo je tudi to, da so vse štiri anketirane profesorice stare 50 let in več. Pri profesorjih je slika izračuna popolnoma drugačna. Od desetih anketiranih trije profesorji (30 %) ne dosegajo meje 45-ih točk, ki je mejna vrednost.

Graf 3: Primerjava starosti in asertivnosti profesorjev



Grafikon nakazuje, da so mlajši profesorji bolj asertivni kot starejši profesorji. Očitno imajo mlajši še “upanje”, da se z zdravo komunikacijo lahko več doseže.

Grafikon 4: Primerjava starosti in asertivnosti profesoric



Prav tako pri profesoricah grafikon nakazuje, da so mlajše profesorice bolj asertivne kot starejše.

5. SKLEPI IN PRIPOROČILA

Tudi sama na splošno občutim med kolegi več asertivnosti kot negodovanja. Rezultati kažejo, da je večina profesorjev asertivnih. Zanimivo bi bilo izpolnjevanje anketnega vprašalnika za posameznega profesorja s strani nevtralne osebe in brez njihove vednosti.

Menim, da se lahko še veliko naredi pri izboljšanju kvalitete komunikacije v kolektivu. Kot prvo predlagam ponovno uvedbo glavnega odmora, ki je bil zaradi tople malice dijakov ukinjen. Profesorji bi tako med sabo več komunicirali. Na splošno kolektiv pričakuje konkretnejše in jasnejše cilje ter načrte s strani vodstva, kar pa trenutno ni ravno jasno definirano. Zanimivo bi bilo pridobiti mnenje aktiva o asertivnosti vodstva. Naslednja ideja, ki se mi je porodila je, da bi na temo dobra komunikacija organizirali seminar na šoli ali izven šolskega okoliša. Skrajni čas pa je, da se šola prične vključevati tudi v različne projekte, zaradi

katerih bi imeli možnost izoblikovati (oziroma najti) dobre time ter s tem izboljševati komunikacijo med kolegi. V slednjem primeru bi bili primorani dodatno komunicirati in usklajevati svoje ideje in načrte. Hkrati bi bil to tudi doprinos celotni šoli v smislu njenega dobička.

Morda še primer iz prakse: Definicija o asertivnosti govori, da asertiven človek išče rešitve, ki so sprejemljive za obe strani ter deli svoje informacije, mnenja in občutke s sogovorniki. Ravno iz tega vidika smo imeli na naši šoli zanimivo izkušnjo, ko so se spred dobrimi tremi leti spreminjali programi v srednjih strokovnih šolah. Takrat je vsaka šola sama lahko določila določen fond ur iz odprtega kurikuluma, kako bo te ure razdelila med predmete. Na šoli imamo pet različnih programov in večina usmeritev se je dogovarjala tako v prid dijakom kot tudi v prid kolektiva. Na eni usmeritvi je bila komunikacija med kolegi precej na nizkem nivoju, saj je dejansko želela pobrati prav vse ure iz odprtega kurikuluma stroka. Ravnali so sebi v prid ne glede na dolgoročne posledice tehnoloških viškov med profesorji. Žal so zmagali. V tem primeru menim, da bi zdrava pamet in komunikacija morali prevladati.

6. SEZNAM UPORABLJENE LITERATURE

1. Košmrlj Prevodnik, T. in Nose, S. 2007. Tajniško poslovanje. Ljubljana, Tehniška založba Slovenija, d.d. 24-25.
2. Možina, S. in drugi 2004. Poslovno komuniciranje (Evropske razsežnosti). Maribor, Obzorja, založništvo in izobraževanje. 422 – 435.
3. Poličnik, V. 2011. Modul ljudje v organizaciji (gradiva za vaje). Kranj. 16-18
4. Poličnik, V. 2011. Ljudje v organizaciji (interno gradivo). Kranj. 28

SAMOPODOBA IN FUNKCIJA METAKOGNITIVNEGA ZAVEDANJA V TERAPEVTSKIH INTERVENCAH PRI OSEBAH Z MOTNJAMI V RAZPOLOŽENJU

Barbara Dolenc, Lilijana Šprah
Družbenomedicinski inštitut, Znanstvenoraziskovalni center SAZU
Novi trg 2, Ljubljana
tel: 01 470 64 39; faks: 01 426 14 93
e-pošta: dolenc.barbara@gmail.com

POVZETEK

Samopodobo definiramo kot množico odnosov, ki jih posameznik zavestno ali nezavedno vzpostavlja do samega sebe. Raziskave kažejo, da se pri pacientih z motnjami v razpoloženju pojavljajo deficiti v samopodobi in samospoštovanju, ki so prisotni v akutni fazi bolezni, kot tudi v remisiji, ko simptomov bolezni ni opaziti. Popačenost predstav o samih sebi naj bi pri osebah z motnjami v razpoloženju bila posledica pomanjkanja metakognitivnega zavedanja, to je sprejemanja tega, da so misli, sklepi in prepričanja notranji dogodki in procesi in ne odrazi objektivne realnosti. Prav to naj bi tako bilo tudi predmet terapevtskih intervencij, kjer skušamo spremeniti posameznikove izkrivljene notranje predstave o samem sebi, s čimer posledično vplivamo tudi na posameznikovo zaznavanje sebe. Ker se bolezenska simptomatika in posameznikova samopodoba prepletajo s kvaliteto življenja, je tovrstna terapevtska intervencija za posameznika izrednega pomena.

1. OPREDELITEV ZAZNAVANJA SEBE: SAMOPODOBA IN SAMOSPOŠTOVANJE

Samopodoba je množica odnosov, ki jih posameznik zavestno ali nezavedno vzpostavlja do samega sebe (Guillon, Crocq in Bailey, 2003; Kobal Grum in Ucman, 2003), leži pa tudi v ozadju posameznikovega prepričanja o lastni kompetentnosti in možnosti uspeha. S takšno organizirano celoto pojmovanj, stališč, sposobnosti in doživljanj posameznik uravnava in usmerja svoje ravnanje ter povezuje svoj vrednostni sistem z vrednostnim sistemom ožjega in širšega družbenega okolja.

Ko govorimo o samopodobi, pa ne moremo mimo njenega najpomembnejšega korelata, to je samospoštovanja. Le-tega opredeljujemo kot pozitivno ali negativno stališče do sebe (Kobal Grum in Ucman, 2003). Na samopodobo in s tem tudi na samospoštovanje pomembno vpliva mnenje, ki ga ima človek o sebi, torej določene izkušnje s samim seboj, vplivajo pa tudi informacije, ki jih o sebi pridobi od drugih, in njegova ocena, kaj drugi menijo o njem. Ta ocena je subjektivna in je precej odvisna od posameznikove življenjske naravnanosti.

2. ZAZNAVANJE SEBE PRI OSEBAH Z MOTNJAMI V RAZPOLOŽENJU

Med motnje razpoloženja po Klasifikaciji Ameriškega psihiatričnega združenja DSM-IV (American Psychiatric Association, 1994) in Svetovne zdravstvene organizacije ICD-10 (World Health Organization, 1992) spadajo depresivne motnje ter bipolarna motnja. Pri obeh boleznih se poleg nihanja razpoloženja motnja odraža tudi v mišljenju in vedenjskih vzorcih posameznika (Benazzi, 2007).

Eden od simptomov velike depresivne motnje je tudi negativna pristranost pri procesiranju informacij o samem sebi, torej nizka stopnja samozavesti in samopodobe (Fennell, 2004; Serretti, Olgiati in Colombo, 2005). Raziskave kažejo, da je le-ta pogosta tudi pri bipolarni motnji, tako med depresivnimi, kot tudi med hipomaničnimi epizodami (Scott in Pope, 2003). Pri pacientih z motnjami v razpoloženju je prisoten deficit v samopodobi in samospoštovanju (Daskalopoulou in dr., 2002; Guillon in dr., 2003; Pardoen, Bauwens, Tracy, Martin, Mendlewicz, 1993; Serretti in dr., 1999, 2005), pri čemer med osebami z veliko depresijo in bipolarno motnjo ne prihaja do razlik v konceptu zaznavanja sebe (Serretti in dr., 2005). Poleg tega pri obeh skupinah pacientov raziskovalci poročajo tudi o nihanju samopodobe v fazi remisije (Blairy in dr., 2004; Serretti in dr., 1999; Shapira in dr., 1999), kar nakazuje, da ta primanjkljaj v samopodobi in samozavesti predstavlja stabilno lastnost afektivnih motenj, ki lahko vpliva tudi na klinični potek in rezultat same bolezni (Daskalopoulou in dr., 2002; Johnson, Meyer, Winett in Small, 2000; Serretti in dr., 2005).

Johnson s sodelavci (2000) nadalje poroča, da je nizka stopnja samopodobe pomembno povezana tudi z resnostjo razpoloženjske motnje. Te ugotovitve nekateri povezujejo z mediacijskim modelom, kjer socialna opora vpliva na razvoj simptomov skozi čas (Johnson in dr., 2000). Posamezniki, ki imajo občutek, da jih njihovi prijatelji pozitivno ocenjujejo, bodo tako z večjo verjetnostjo poročali o boljši samopodobi, ki se povezuje z blažjo simptomatiko.

Novejše študije bolnikov z bipolarno motnjo v remisiji počasi spreminjajo pogled na samo bolezen. Vedno več raziskav namreč kaže, da ti pacienti kažejo pomembne fluktuacije razpoloženja tudi v evtimični fazi bolezni, ko ni opaziti niti depresivnih niti maničnih simptomov (Henry in

dr., 2008; Knowles, Tai, Jones, Highfield, Morriss in Bentall, 2007). Tovrstna afektivna nestabilnost bi lahko bila tudi posledica nestabilne samopodobe in samospoštovanja, torej zaznavanja samega sebe (Pavlova, Uher, Dennington, Wright in Donaldson, 2011), čeprav ostaja odnos med afektivno nestabilnostjo, ki vztraja skozi vse faze bipolarne motnje, in velikimi afektivnimi epizodami nejasen (Pavlova in dr., 2011). Poleg nestabilne samopodobe pri pacientih z bipolarno motnjo razpoloženja (Knowles in dr., 2007; Pavlova in dr., 2011; van der Gucht, Morriss, Lancaster, Kinderman in Bentall, 2009), pa se nihanje samozaznavanja pojavlja tudi pri pacientih z depresivno motnjo (Knowles in dr., 2007, Pavlova in dr., 2011) v primerjavi z zdravimi posamezniki. Bolniki z motnjami v razpoloženju tudi v fazi remisije reagirajo na relativno majhne z uspehom povezane dogodke z velikimi spremembami v samopodobi (Pavlova in dr., 2011), ta reaktivnost in nihanje samopodobe pa se povezuje tudi z bolj kroničnim pojavljanjem bolezni in ponovnem pojavu simptomatike (Butler, Hokanson in Flynn, 1994). Zaradi tega nekateri raziskovalci menijo, da tovrstna odzivnost samopodobe predstavlja faktor tveganja za nadaljnji razvoj afektivnih epizod, tako pri depresiji, kot tudi pri bipolarni motnji razpoloženja (Pavlova in dr., 2011). Posamezniki z bipolarno motnjo lahko reagirajo na manjše uspehe in neuspehe s povečanim oziroma zmanjšanim razpoloženjem in vedenjem, kar vodi do motenj v socialnih in bioloških ritmih, kar pa še dodatno pripeva k razvoju velikih afektivnih epizod (Mansell, Morrison, Reid, Lowens in Tai, 2007).

Študije kažejo, da kljub medikamentoznemu zdravljenju obstaja visok procent bolnikov z bipolarno motnjo in depresijo, pri katerih se bolezen kronično ponavlja (Solomon, Keitner, Miller, Shea in Keller, 1995), zaradi česar je zelo pomembno tudi vzdrževanje pozitivne samopodobe. Le-ta namreč pacientom pomaga pri spopadanju z negativnimi vplivi bolezni na njihovo življenje in tako poveča kvaliteto njihovega življenja (Hayward, Wong, Bright in Lam, 2002). Poleg tega je nizka samopodoba povezana z višjim tveganjem za kronični potek depresije (Ezquiaga, Garcia, Pallares in Bravo, 1999) in s samomorilnostjo (Daskalopoulou in dr., 2002). Vse to je razlog, da je nihanje samopodobe skupaj s samospoštovanjem pogosto tudi primarna tarča terapevtskih intervencij afektivnih motenj (Pavlova in dr., 2011).

3. SAMOPODOBA KOT PREDMET TERAPEVTSKIH INTERVENCIJ

Čeprav genetski dejavniki ključno sooblikujejo razvoj bipolarne motnje (Johnson in dr., 2000), igrajo pomembno vlogo pri napovedi poteka bolezni tudi psihosocialno okolje (Johnson in dr., 2000), življenjski dogodki (Hammen, Ellicott in Gitlin, 1992; Johnson in Miller, 1997; Johnson in Robers, 1995) in metakognitivno zavedanje. Metakognitivno zavedanje je definirano kot sprejemanje tega, da so misli, sklepi in prepričanja notranji dogodki in procesi in ne odrazi objektivne realnosti (Fennell, 2004), zato je ta vidik pomemben tudi v terapevtskih intervencijah. Prav pri razpoloženjskih motnjah se namreč pogosto dogaja, da pacienti svoja notranja prepričanja, četudi popačena, razumejo kot objektivna dejstva (Beck, 2005). Teasdale, Segal in Williams (1995) so predlagali, da bi lahko bil ključni proces pri uspešnosti zdravljenja afektivnih motenj prav stopnja, do katere pacienta priskrbi z alternativnimi perspektivami o izkušnjah in vztrajnosti simptomatike (metakognitivno zavedanje). Metakognitivno zavedanje oziroma prepoznavanje kognicij kot notranjih dogodkov in procesov sicer ni nujno posledica učinkovite kognitivne terapije, je pa zagotovo prvi pogoj za uspešno zdravljenje bolezni (Fennell, 2004). Veliko preteklih raziskav se je osredotočalo na način, kako interpretacije, ki jih ljudje povezujejo s simptomi, ki jih občutijo, lahko služijo kot vzdrževanje ali celo povečanje psiholoških težav (Fennell, 2004), hkrati pa so številni raziskovalci iskali tudi ustrezne načine, kako pomagati ljudem dvomiti in testirati te interpretacije ter oblikovati verodostojne alternative, ki dovoljujejo pacientom, da se odzivajo na njihove lastne misli, emocije in telesne občutke na nov in boljši način (Fennell, 2004).

Beck (1967, v Fennell, 2004) je že zgodaj v svojih raziskavah kognicije pri osebah z depresijo opazil, da se pri teh pacientih pogosto pojavlja nizko samospoštovanje, nerealno razvrednotenje samega sebe, poveličevanje svojih napak in neuspehov, zmanjšanje vrednotenja svojih kvalitet in uspehov, ekstrapolacija iz majhnih napak v splošno ničvrednost, samokritičnost in občutja krivde in podobno. Vse te simptome je Beck (1967, v Fennell, 2004) identificiral kot posledico sistematične pristranosti proti sebi, karakteristične za kognicijo depresivnih posameznikov. Ta pristranost se dotika vseh treh elementov negativne kognitivne triade, ki so vključeni v vsebino depresivnega mišljenja (Beck, Rush, Shaw in Emery, 1979, v Fennell, 2004), in sicer so to negativne distorzije o sebi, izkušnje, ki vključujejo podcenjevanje sebe, interpretacijo specifičnega neuspeha kot splošnega, negativne primerjave z drugimi, in podobno, ter distorzije o prihodnosti oziroma mnenje, da trenutnih primanjkljajev ni mogoče spremeniti. Vendar pa imajo v kontekstu depresije te negativne predstave o samem sebi tri različne funkcije (Fennell, 1998, v Fennell, 2004). Kot prvo so lahko primarno produkt depresivnega razpoloženja in izginejo ob prvem dvigu le-tega. Kot drugo so lahko to odzivi na samo depresijo (t.i. depresija zaradi depresije, Teasdale, 1983, v Fennell, 2004). Učinkovito zdravljenje depresije bo v tem primeru zelo verjetno povrnilo občutek samovrednosti. Tretja možnost pa je, da so negativna prepričanja o samem sebi odraz širše kognitivne strukture (negativna prepričanja), ki so neodvisna od depresivnega razpoloženja in prispevajo k sami občutljivosti na depresijo kot tako (Hollon in dr., 1996, v Fennell, 2004). V tem primeru je zdravljenje težko, saj lahko zgolj zdravljenje depresije pusti nizko samopodobo nedotaknjeno, zaradi česar posameznik ostane občutljiv na prihodnje težave, četudi se samo razpoloženje izboljša (Fennell, 2004). Vse te tri funkcije pa so lahko tudi tesno prepletene, saj lahko oseba z dolgotrajno nizko samopodobo postane visoko samokritična ko se depresija še poglobi.

Določene življenjske situacije lahko reaktivirajo konstelacijo misli, čustev, telesnih občutij in vedenja, ki so del depresije, vključno s situacijami, ki vodijo do zmanjšanja samozavesti in samopodobe. To vodi do zoženja polja zavesti v smislu preferiranja procesiranja negativnega materiala, kar privede do povečanega fokusa na negativne aspekte sebe in lastnih izkušenj (Fennell, 2004). Na ta način pride do samovzdrževane sekvence izkušnje, ki jo interpretiramo kot neuspeh, kar vodi do krivde in občutka splošne ničvrednosti in samoponiževanja. Ker na te defekte oseba gleda kot na intrinzične dele sebe, tako ne daje možnosti spremembi, zato se v shemo vključi še brezup.

Fennell in Teasdale (1982, v Fennell, 2004) sta se spraševala, če bi izkušnje predolge in ponavljajoče se depresije lahko spremenile lastno percepcijo o sebi, saj bi lahko depresivni deficiti postali aspekti posameznikovega »pravega selfa« in ne kot začasni odkloni od »normalnega selfa«. Vendar pa avtorja nista pomislila, da bi bilo možno tudi ravno obratno – ponavljajoča ali kronična depresija je lahko tudi produkt stalnih negativnih mišljenj o samem sebi. Sprememba tovrstnih mišljenj in prepričanj je bila vključena v zdravljenje depresije šele kasneje, ko se je s pomočjo kognitivne terapije začelo zdraviti tudi kronične in globlje oblike depresivne motnje (Moore in Garland, 2003).

Mehanizem, skozi katerega samopodoba moderira povezavo med življenjskimi dogodki in depresijo, lahko razumemo skozi medosebnostno in kognitivno perspektivo. Brown, Bifulco, Veiel in Andrews (1990) so se osredotočili na povezanost med slabo samopodobo in majhno socialno mrežo ter odnos pokazali skozi analizo poti. Nizka samopodoba namreč vodi do zmanjšane socialne opore, kar se v kriznih časih kaže kot povečano tveganje za depresijo. V nasprotju s tem pa večina literature samopodobo konceptualizira iz kognitivne perspektive kot moderator razpoloženja. Občutki krivde in notranje, stabilne in splošne atribucije ob negativnih življenjskih dogodkih naj bi vodile do povečanega notranjega fokusa, ruminacije in slabšega pogleda nase, kar posledično povzroči depresijo (Beck, 1987; v Miller, Warner, Wickramaratne in Weissman, 1999; Nolen-Hoeksema, 1991).

Ena izmed najbolj pogostih psihoterapevtskih intervencij, ki se dotika prav popačenih notranjih prepričanj, je tako kognitivna psihoterapija (Beck, 2005). Kognitivna terapija torej temelji na predpostavki, da naše misli in interpretacije dogodkov vplivajo na občutke, telesne znake in vedenje. Ko se disfunkcionalna bazična prepričanja v določenih okoliščinah aktivirajo, sprožijo val negativnih avtomatskih misli, kar je povezano z neprijetnimi občutki. Z negativnimi avtomatskimi mislimi posameznik izkrivljeno interpretira sedanje dogodke, predvideva o prihodnosti in ponovno negativno interpretira dogodke preteklosti, vse to pa vodi v nastanek ostalih znakov depresije. Ti znaki pa nato povratno krepijo negativne avtomatske misli in tako simptomi praktično poglabljajo samega sebe (Vidmar Vengust, 2004). To pomeni, da pri depresivnih ljudeh nastopi problem začaranega kroga, saj nastanek depresivnih simptomov posameznik ponovno negativno interpretira, kar še poglablja depresijo. Negativne misli torej predstavljajo določen sprožilec, vzpodbujevalec in vzdrževalec drugih simptomov.

Terapija pri depresivnih je tako aktivna, direktivna, časovno omejena in strukturirana, bolnika pa spodbuja k razvoju in oblikovanju veščin samopomoči (Vidmar Vengust, 2004). Temelji na kognitivnem modelu emocionalnih motenj. Orientirana je na problem in osredotočena na tista področja, ki vzdržujejo težave. Depresivnega posameznika terapija uči, da lastne misli in prepričanja postavlja pod vprašaj in jih v življenju preverja. Neposredni cilj terapevtskega procesa je zmanjševanje simptomov depresije, na daljši rok pa je cilj predvsem oblikovanje strategije za reševanje problemov, pa tudi preprečevanje morebitnih prihodnjih epizod depresije (Vidmar Vengust, 2004).

4. ZAKLJUČKI

O motnjah razpoloženja se v današnjem času vedno več govori, saj je depresivna motnja najpogostejša motnja povsod po svetu (Kores Plesničar, 2006). S spreminjanjem posameznikovih popačenih notranjih predstav o samem sebi lahko posledično vplivamo na simptomatiko razpoloženjskih motenj, kar vodi tudi do boljše samopodobe ter pretrga začaran krog kognitivnih distorzij. Znano je, da je kvaliteta življenja tako bolnikov z depresijo, kot tudi tistih z bipolarno motnjo pomembno nižja v primerjavi s kvaliteto življenja zdravih posameznikov (Angermeyer, Holzinger, Matschinger, Stengler-Wenzke, 2002; Michalak, Yatham in Lam, 2005), zato je tovrstna terapevtska intervencija izrednega pomena.

5. LITERATURA

[1.] American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders, 4th edition*. Washington: American Psychiatric Association.

[2.] Angermeyer, M.C., Holzinger, A., Matschinger, H. in Sengler-Wenzke, K. (2002). Depression and quality of life: results of a follow-up study. *International Journal of Social Psychiatry, 48(3)*, 189-99.

[3.] Beck, J. (2005). *Cognitive Therapy for Challenging Problems*. New York: The Guilford Press.

[4.] Benazzi, F. (2007). Bipolar disorder – focus on bipolar II disorder and mixed depression. *The Lancet, 369 (9565)*, 935 – 945.

[5.] Blairy, S., Linotte, S., Souery, D., Papadimitriou, G., Dikeos, D., Lerer, B., Kaneva, R., Milanova, V., Serretti, A., Macciardi, F. in Mendlewicz, J. (2004). Social adjustment and self-esteem of bipolar patients: a multicentric study. *Journal of Affective Disorders, 79(1-3)*, 97-103.

[6.] Brown, G.W., Bifulco, A., Veiel, H.O. in Andrews, B. (1990). Self-esteem and depression: II Social correlates of self-esteem. *Social Psychiatry and Psychiatric Epidemiology, 25(5)*, 222-234.

[7.] Butler, A.C., Hokanson, J.E. in Flynn, H.A. (1994). A comparison of self-esteem lability and low trait self-esteem as vulnerability factors for depression. *Journal of Personality and Social Psychology, 66(1)*, 166-177.

[8.] Daskalopoulou, E.G., Dikeos, D.G., Papadimitriou, G.N., Souery, D., Blairy, S., Massat, I., Mendlewicz, J. in Stefanis, C.N. (2002). Self-esteem, social adjustment and suicidality in affective disorders. *European Psychiatry, 17(5)*, 265-271.

[9.] Ezquiaga, E., Garcia, A., Pallares, T. in Bravo, M. (1999). Psychosocial predictors of outcome in major depression: a prospective 12-month study. *Journal of Affective Disorders, 52(1-3)*, 209-216.

[10.] Fennell, M. J. V. (2004). Depression, low self-esteem and minfulness. *Behaviour Research and Therapy, 42(9)*, 1053-1067.

[11.] Guillon, M.S., Crocq, M.-A. in Bailey, P.E. (2003). The relationship between self-esteem and psychiatric disorders in adolescents. *European Psychiatry, 18(2)*, 59-62.
[12.] Hammen, C., Ellicott, A. in Gitlin, M. (1992). Stressors and sociotropy/autonomy: A longitudinal study of their relationship to the course of bipolar disorder. *Cognitive Therapy, 16(4)*, 409-418.
[13.] Hayward, P., Wong, G., Bright, J.A. in Lam, D. (2002). Stigma and self-esteem in manic depression: an exploratory study. *Journal of Affective Disorders, 69(1-3)*, 61-67.
[14.] Henry, C., Van den, B.D., Bellivier, F., Roy, I., Swendsen, J., M'bailara, K., Siever, L.J. in Leboyer, M. (2008). Affective lability and affect intensity as core dimensions of bipolar diosrders during euthymic period. *Psychiatry Research, 159(1-2)*, 1-6.
[15.] Johnson, S. in Miller, I. (1997). Negative life events and recovery from episodes of bipolar disorder. *Journal of Abnormal Psychology, 106(3)*, 449-457.
[16.] Johnson, S. in Roberts, J.R. (1995). Life events and bipolar disorder: implications from biological theories. *Psychological Bulletin, 117(3)*, 434-449
[17.] Johnson, S.L., Meyer, B., Winett, C. in Small, J. (2000). Social support and self-esteem predict changes in bipolar depression but not mania. *Journal of Affective Disorders, 58(1)*, 79-86.
[18.] Knowles, R., Tai, S., Jones, S.H., Highfield, J., Morriss, R. in Bentall, R.P. (2007). Stability of self-esteem in bipolar disorder: comparisons among remitted bipolar patients, remitted unipolar patients and healthy controls. *Bipolar Disorder, 9(5)*, 490-495.
[19.] Kobal Grum, D. in Ucman S. (2003). Bivanje samopodobe v teoriji. V D. Kobal Grum (ur.), *Bivanje samopodobe* (str. 19-69). Ljubljana: I2 družba za založništvo, izobraževanje in raziskovanje d.o.o.
[20.] Kores Plesničar, B. (2006). Epidemiologija, etiologija, klinična slika in diagnostika depresije. *Farmakološki vestnik, 57(4)*, 241-244.
[21.] Mansell, W., Morrison, A.P., Reid, G., Lowens, I. in Tai, S. (2007). The intepretation of, and responses to, changes in internal states: an integrative cognitive model of mood swings and bipolar disorders. *Behavioural and Cognitive Psychotherpy, 35(5)*, 515-539.
[22.] Michalak, E.E., Yatham, L.N. in Lam, R.W. (2005). Quality of life in bipolar disorder: A review of the literature. *Health and Quality of Life Outcomes, 3(72)*, 117-134.
[23.] Miller, L., Warner, V., Wickramaratne, P. in Weissman, M. (1999). Self-Esteem and Depression: ten year follow-up of mothers and offspring. *Journal of Affective Disorders, 52(1-3)*, 41-49.
[24.] Moore, R.G. in Garland, A. (2003). *Cognitive therapy for chronic and persistent depression*. Chichester: Wiley.
[25.] Nolen-Hoeksema, S. (1991). Responses to depression and theis effects on the duration of depressive episodes. *Journal of Abnormal Psychology, 100(4)*, 569-582.
[26.] Pardoen, D., Bauwens, F., Tracy, A., Martin, F. in Mendlewicz, J. (1993). Self esteem in recovered bipolar and unipolar outpatients. *British Journal of Psychiatry, 163*, 755-762.
[27.] Pavlova, B., Uher, R., Dennington, L., Wright, K. in Donaldson, C. (2011). Reactivity of affect and self-esteem during remission in bipolar affective disorder: An experimental investigation. *Journal of Affective Disorders, Članek v tisku*.
[28.] Scott, J. in Pope, M. (2003). Cognitive styles in individuals with bipolar disorders. *Psychological Medicine, 33*, 1081-1088.
[29.] Serretti, A., Cavallini, M.C., Macciardi, F., Namia, C., Franchini, L., Souery, D., Lipp, O., Bauwens, F., Smeraldi, E. in Mendlewicz, J. (1999). Social adjustment and self-esteem in remitted patients with mood disorders. *European Psychiatry, 14(3)*, 137-142.
[30.] Serretti, A., Olgiati, P. in Colombo, C. (2005). Components of self-esteem in affective patients and non-psychiatric controls. *Journal of Affective Disorders, 88(1)*, 93-98.
[31.] Shapira, B., Zislin, J., Gelfin, Y., Osher, Y., Gorfine, M., Souery, D., Mendlewicz, J. in Lerer, B. (1999). Social adjustment and self-esteem in remitted patients with unipolar and bipolar affective disorder: a case-control study. *Comprehensive Psychiatry, 40(1)*, 24-30.
[32.] Solomon, D.A., Keitner, G.I., Miller, I.W., Shea, M.T. in Keller, M.B. (1995). Course of illness and maintenance treatments for patients with bipolar disorder. *Journal of Clinical Psychiatry, 56(1)*, 5-13.
[33.] Teasdale, J.D., Segal, Z.V. in Williams, J.M.G. (1995). How does cognitive therapy prevent relapse, and why should attentional control (mindfulness) training help?. *Behaviour Research and Therapy, 33(1)*, 225-239.
[34.] van der Gucht, E., Morriss, R., Lancaster, G., Kinderman, P. in Bentall, R.P. (2009). Psychological processes in bipolar affective disorder: negative cognitive style and reward processing. *British Journal of Psychiatry, 194(2)*, 146-151.
[35.] Vidmar Vengust, M. (2004). Kognitivno vedenjski pristop k depresivni motnji. V N. Anić (ur.), *Prispevki iz vedenjsko-kognitivne terapije*, str. 211 – 225. Ljubljana: Društvo za vedenjsko in kognitivno terapijo Slovenije.
[36.] World Health Organization (1992). *The ICD-10 classification of mental and behavioral disorder: Diagnostic criteria for research*. Geneva: World Health Organization.

KOGNITIVNA (R)EVOLUCIJA, INTELIGENCA IN ORODJA

Matjaž Gams
Odsek za inteligentne sisteme, Institut Jožef Stefan
Jamova 39, 1000 Ljubljana, Slovenija
Tel: +386 1 4773900; fax: +386 61 1251038
e-mail: matjaz.gams@ijs.si, http://ai.ijs.si/mezi/matjaz.html

Povzetek: Predstavljen je avtorjev antropološki in evolucijsko navdahnjen pogled na razvoj človeških kognitivnih sposobnosti v preteklosti in kako se bodo razvijale vnaprej.

1 UVOD

Kako se je razvijal človeški um v preteklosti in kaj se mu bo zgodilo v prihodnosti? V tem prispevku bomo predstavili dogajanja v preteklosti predvsem skozi razvoj praljudi, nato pa skozi razvoj zadnja desetletja, ko je očiten izreden vpliv IKT na delovanje ljudi tako fizično kot kognitivno.

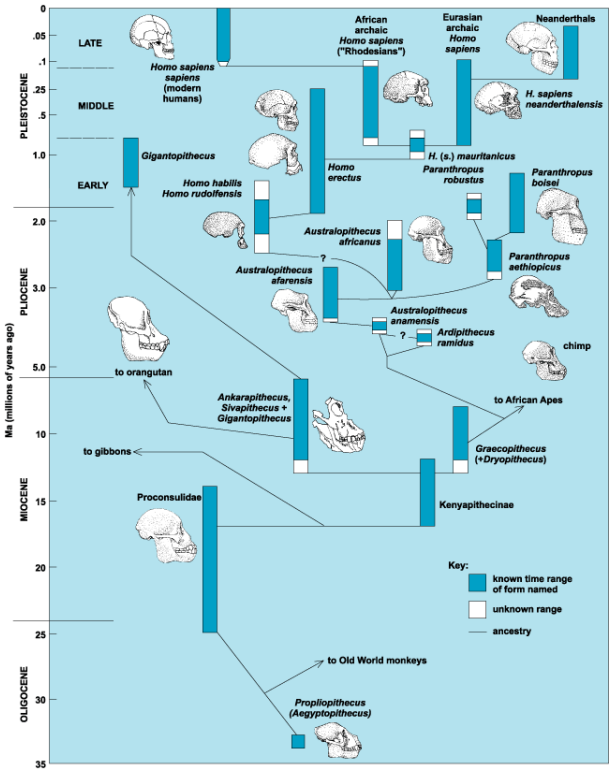
Za razliko od primerjav med sposobnostmi ljudi in računalnikov oz. Turingovih strojev (Gams 2003) je analiza narejena bolj v smislu sinteze (Brooks 2008; Gams 2007), ali miselne-simbolne ali celo fizične. V tem prispevku prav tako ne analiziramo sistemov umetne inteligence oz. inteligentnih sistemov (AI Magazin 2008) oz. lastnosti možganov, predvsem mnogoterosti (Gams 2001).

V prvem delu prispevka se ukvarjamo z razvojem praljudi (Geary 2004); Gross 2010).

V drugem delu prispevka nas zanima relacija med umom-kognicijo in IKT, kot konceptualna, tj. npr. internet, ali pa kot fizična, tj. dejanski sklop možganov in računalnikov.

2 PRALJUDJE

National Geographic (2011) med svoje najpomembnejše teme običajno izbira evolucijo in razvoj ljudi. Tudi v avgustovski številki leta 2011 navajajo najnovejša odkritja. Na osnovi 2 milijona let starih fosilov avstralopitekov z nekaj človeškimi lastnostmi so predložili hipotezo, da so našli prednike ljudi. Prevladujoča hipoteza danes je, da so avstralopiteki slepa veja v razvoju ljudi, da so izumrli, medtem ko smo se ljudje razvili iz drugih prednikov.



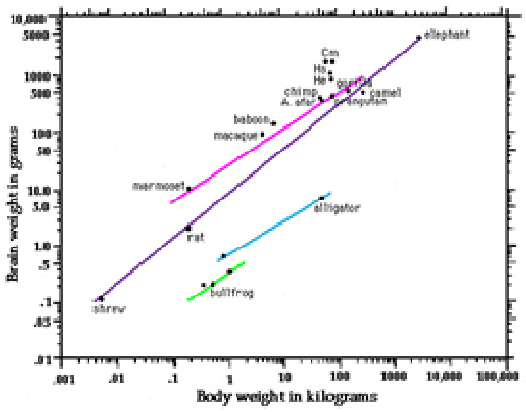
Slika 1: Razvoj ljudi je bil očitno zelo zapleten. Zakaj?

Če pogledamo sliko 1, vidimo izredno razvejanost razvoja praljudi, ki je opisana tudi v National geographicu. Nikjer ne vidimo preprostega linearnega razvoja, kjer bi predniki, podobni npr. šimpanzu, polagoma postajali čedalje bolj človeški, s čedalje več človeškimi lastnostmi. Tako naj bi hodili čedalje bolj vzravnano, naj bi imeli čedalje bolj razvito roko in čedalje večje možgane. Zgodovina naših prednikov pa je prepredena – eni imajo večje možgane in drugi bolj spretno prijemalno roko, ki omogoča boljše prijemanje s palcem in kazalcem, česar opice še danes ne zmorejo. Zakaj je bil razvoj tako razvejan in ne premočrten? Tudi če privzamemo antropološki princip, bi moral biti razvoj bolj ali manj linearen.
Zdi se, da poznamo vsaj delen odgovor. Možgani so namreč veliki potrošniki energije, saj jo danes porabijo približno petino in so zelo občutljivi na poškodbe, hrano itd.

(Wikipedia). Danes večina ljudi nima velikih težav z zadostno količino energije, praljudje pa so imeli bistveno drugačne razmere, kjer je za časa lakote sposobnost preživetja z malo potrošnjo kalorij igrala ključno vlogo. Torej: sposobnejši in energijsko bolj potrošni možgani ali bolj preprosti in energijsko bolj varčni?

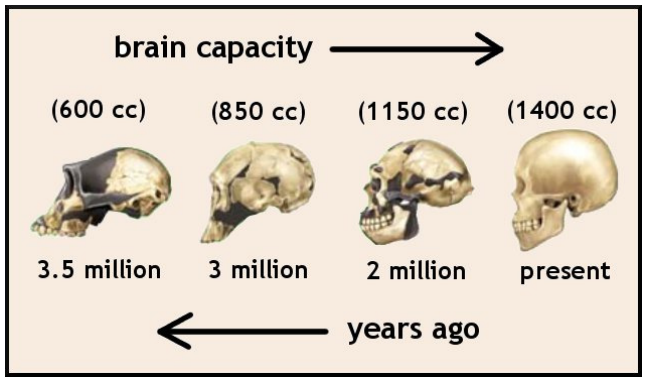
Podobno dilemo srečamo še danes v marsikateri napravi, npr. pri sesalnikih Roomba. To so prvi roboti, katerih uporaba se je štela v milijonih. Pri robotskih sesalnikih je osnovna dilema: ali naj presesajo sobo bolj ali manj naključno, ali pa naj se naučijo prostora in potem posesajo sistematično enkrat ali dvakrat (drugič pravokotno)? Izkaže se, da ni bistvene razlike v kvaliteti, kvečjemu v času sesanja in ceni. Za večjo ceno dobimo sposobnejši robotski sesalnik, ki nekaj hitreje posesa. Ker pa je cena na trgu zelo pomembna, se večina uporabnikov odloča za preprosto verzijo in ne za bolj inteligentno.

Evolucijsko se zdi, da večja inteligenca, ki najbrž odgovarja velikosti možganov vsaj pri sorodnih vrstah, prinese več, kot pa bitje izgubi za to. Če pogledamo razmerje med težo telesa in maso možganov, vidimo, da ljudje izstopamo. Na diagramu na sliki 2 sta obe osi logaritemski, saj slon ni le nekaj malega težji kot človek, pa tudi možganov nima samo malo težjih. Pomembne pa so črte, ki kažejo razvoj možganov od preprostih opic do ljudi na zgornji levi črti. Ljudje odstopamo celo od te črte.

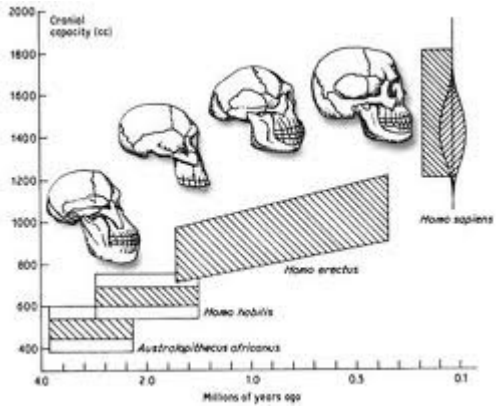


Slika 2: Razmerje med težo možganov in težo telesa se prikaže na logaritemskih skalah kot dokaj linearne črte pri sorodnih živalskih vrstah.
Ne smemo pozabiti, da sta obe osi logaritemski, torej je pravo razmerje v resnici dokaj linearno. Kaže se tudi, da je inteligenca znotraj sorodnih vrst dokaj sorazmerna teži možganov oz. razmerju med težo možganov in težo telesa, saj so npr. preproste opice vse nižje kot ljudem podobne.

To odvisnost lahko prikažemo tudi na sliki 3, kjer so navedeni najbolj značilni predniki praljudi in volumen njihovih možganov v časovni odvisnosti.



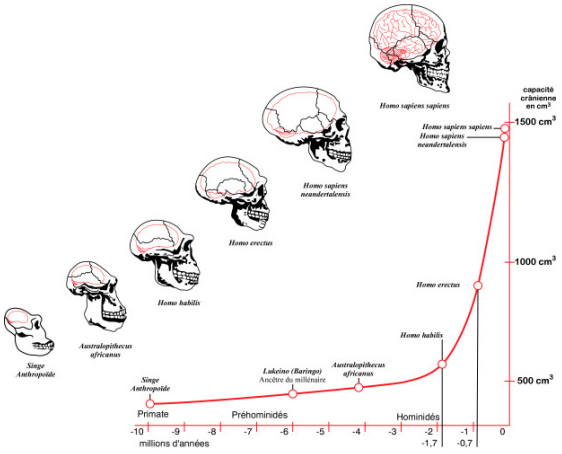
Slika 3: Linearen prikaz razvoja velikosti možganov pri prednikih ljudi.



Slika 4: Podrobnejši prikaz rasti velikosti človeških možganov po vrstah praljudi.

Tudi na sliki 4, ki nekoliko bolj podrobno prikazuje razvoj velikosti človeških možganov po vrstah ljudi, se zdi, da je v globalu gledano šlo za bolj ali manj linearen razvoj. Če torej pozabimo, da se je evolucija lovila sem in tja in da je v nekih konkretnih razmerah občasno manj inteligentna verzija praljudi bolj uspešno preživela pomanjkanja, pa lahko nekako vseeno posplošimo, da je šlo za linearen razvoj? Odgovor je negativen, kar nam kaže slika 5.

Namesto linearne rasti se je počasna rast okoli 2 milijonov let nazaj pospešila in začela hitro rasti.



Slika 5: Novejša spoznanja o hitrosti razvoja volumna možganov kažejo, da razvoj ni bil linearen.

Po Wikipediji se se prva orodja pojavila okoli 2.5 do 2 milijoni nazaj. Ocena je problematična, saj tudi šimpanzi uporabljajo nekaj orodij, od paličic za lovljenje termitov do kamnov in debelih vej za razbijanje orehov in vej za bezanje srednje velikih živali iz drevesnih duplin. Tudi nekatere vrane in druge ptice uporabljajo paličice za podaljšanje dosega v duplinah. Šimpanzi tudi obdelajo orodje, npr. odstranijo listje s paličic. Vendar so bila orodja že skoraj 2.5 milijona let nazaj izdelana z obdelavo kamnov s kamni. Bolj komplicirana orodja kot kamnite sekire so se pojavile nekaj manj kot dva milijona let nazaj.

Pojav orodij torej sovpada s hitro rastjo možganov, zato lahko naredimo za ta prispevek najpomembnejši sklep oz. hipotezo:

Razvoj naprednih orodij sovpada z razvojem uma-možganov.

Morda je izjema pojav pravega jezika, ki se je zgodil približno 500.000 let nazaj (manj kvaliteten jezik se je pojavil že več kot 2 milijona let nazaj in ga avstralopiteki še niso osvojili), vendar tudi orodja sama zase lepo sovpadajo z razvojem človeškega uma. Poleg tega lahko orodja oz. izdelke lepo zgodovinsko sledimo.
Iz zgodovine orodij je razvidno, da se je veliki skok naprej zgodil okoli 40 do 50.000 let nazaj. Takrat so se pojavili tudi prvi izdelani okraski itd.

Sledil je razvoj v bakreni dobi itd. do razvoja IKT, ki ga bomo posebej analizirali.

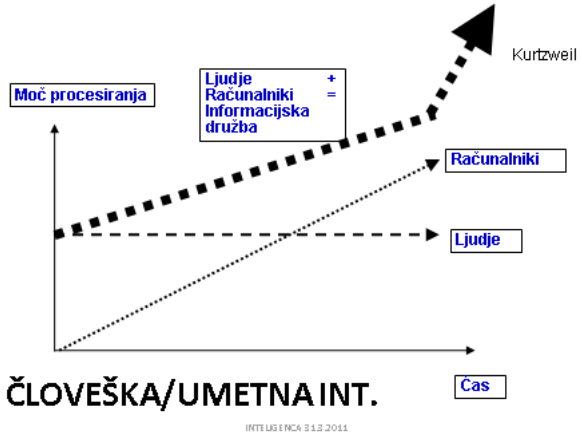
3 MODERNI LJUDJE IN IKT

Računalniki so se pojavili sredi prejšnjega stoletja in so ključno prispevali k razvoju IKT, ki ga opažamo v raznih oblikah od fizičnih – računalnika in mobitela, do konceptualnih – interneta.

Če so na človeške možgane močno pozitivno vplivali že klasična orodja in stroji, potem lahko predstavimo naslednjo hipotezo:

Z razvojem IKT so se človeške kognitivne sposobnosti radikalno povečale.

Osnovne relacije so prikazane na sliki 6.



Slika 6: Razvoj IKT je že in bo še bolj vplival na hitro rast človeškega uma – kognicije.

Na sliki 6 je predstavljenih nekaj zakonitosti, npr. rast računalniških kapacitet po Moorovem zakonu (spodnja poševna premica). Čas je na tej skali linearen na osi x in se začne nekje okoli 1940, moč procesiranja pa je na osi y in je logaritmična. Rast povprečne teže človeških možganov ali osnovnih miselnih sposobnosti se v zadnjih 100 letih praktično niso spremenile, zato je to predstavljeno z vodoravno črto.

Pač pa je sposobnost človeškega razmišljanja oz. reševanja problemov predstavljena z debelo vrhnjo črto. Sovpada s Flynnovim efektom (Wikipedia). Po svoje je preprosto razumljiv – če imate na voljo internet, potem lahko vrsto problemov rešite bistveno bolj ugodno, kot če ste brez njega. Človeške miselne sposobnosti torej rastejo tako, kot če bi sešteli računalniške in človeške fizično pogojene.

Za prihodnost nekateri strokovnjaki kot Kurzweil (2000) predvidevajo bistveno hitrejšo rast – skok na desni strani vrhnje daljice na sliki 6. Ena od možnosti bi bila, da se človeški možgani obogatijo z neposredno povezavo z računalniki. To kombinirano bitje človek-računalnik bi še bolj uspešno uporabljalo oba svetova. Je pa to le ena izmed futurističnih napovedi.

4 ZAKLJUČEK

Če teorija o pozitivni povezavi med človeškimi orodji-stroji in kognitivnimi sposobnostmi-umom velja, potem so pred nami res zlati časi. Razvoj se bo pospešil in postajali bomo čedalje pametnejši in sposobnejši.

To seveda postavlja tudi praktična vprašanja – npr. ali je smotrno na izpitu zahtevati, da študentje ne smejo uporabljati interneta, ko pa v vsakdanjem delu in življenju dosegajo signifikantno boljše rezultate? Ali ni bolj smiselno deliti izpite na dve vrsti – na take, ki zahtevajo znanje v fizičnih možganih in na take, kjer je normalno uporabljati oboje, možgane in računalnike?

Obstajajo tudi zahtevnejše dileme – bodo novi kiborgi-ljudje res »samo« še ljudje?

Zahvala:
Zahvaljujemo se ARRS za financiranje raziskovalne programske skupine Umetna inteligenca in inteligentni sistemi.

Literatura:

AI Magazin (2008). Achieveing Human-Level Intelligence through Integrated Systems and Research, AI Magazine, vol. 27, no. 2, summer 2006.

Broooks, M. (2008). I, Rodnez Brooks, Am a Robot, IEEE Spectrum Online, http://www.spectrum.ieee.org/jun08/6307

Gams, M. (2001). Weak intelligence : through the principle and paradox of multiple knowledge, (Advances in computation, vol. 6). Huntington: Nova Science, XIX, 245 str., graf. prikazi, ilustr. ISBN 1-56072-898-1.

Gams, M. (2003). Are minds stronger than Turing machines?. V: ZAJC, Baldomir (ur.). Zbornik ERK

Gams, M. (2009). V postelji z robotom : umetna inteligenca leta 2009. Delo (Ljubl.), 12. sept. 2009, l. 51, št. 211, str. 36.

Geary, D.C. (2004). The Origin of Mind: Evolution of Brain, Cognition, and General Intelligence, ISBN: 978-1-59147-181-3

Gross R.D. (2010). Psychology: The Science of Mind and Behaviour, Hodder Arnold, ISBN 9781444108316

Kurzweil, R. (2000), The Age of Spiritual Machines: How We Will Live, Work, and Think in the New Age of Intelligent Machines, Penguin books 2000.

National Geographic (2010), avgust.

Wikipedia (2010). http://www.wikipedia.org/

SENZORIČNA METROLOGIJA, VEDA O MERJENJIH Z OSEBAMI

Gregor Geršak, Janko Drnovšek
Univerza v Ljubljani
Fakulteta za elektrotehniko
Tržaška 25, 1000 Ljubljana, Slovenija
e-mail: gregor.gersak@fe.uni-lj.si

ABSTRACT

This paper describes sensory metrology, science of measuring, which is not based solely on measurement of conventional physical and chemical properties, but includes also quantities which are (currently still) without reference values, without standards, scales and units. Sensory metrology deals with psychological, sociological, economic parameters, such as pain, fear, joy, anxiety, smell, comfort, poverty, social capital ...
Sensory metrology is a highly multidisciplinary science, with scientists sharing measuring methods and measurement instrumentation to achieve the main objective – a complete measurement result with measuring error and uncertainty.
Measuring anxiety as one of the uncertainty contributions within sensory metrology is defined and described.

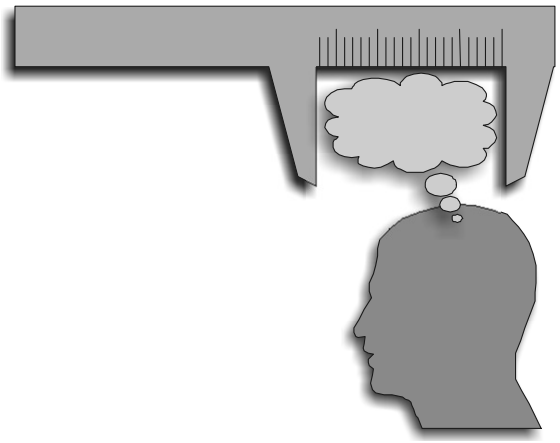
Prispevek opisuje in obravnava najmlajše področje meroslovja, vede o merjenju, senzorično meroslovje (ang. soft metrology, sensory metrology, perception-based measurements). Senzorično meroslovje je veda o merjenju, ki ni osnovano samo na merjenju klasičnih fizikalno-kemičnih veličin, kot so tlak, sila, električni tok, molska masa, ampak vsebuje tudi veličine, ki so (še) brez referenčnih vrednosti, brez etalonov, brez skal in brez enot. Senzorično meroslovje se ukvarja z merjenji psiholoških, sociološkiih, ekonomskih parametrov, kot so bolečina, strah, veselje, tesnoba, smrad, ugodje, revščina, socialni kapital... Senzorično meroslovje je izrazito multidisciplinarna znanost, saj v njem sodelujejo fiziki, psihologi, sociologi, inženirji, zdravniki, kemiki, arhitekti, ekonomisti in drugi, ki si izmenjujejo merilne metode in merilno instrumentacijo za dosego cilja – popolnega merilnega rezultata.
V klasičnem meroslovju ali tradicionalni merilni tehniki se človek običajno pojavi zgolj kot upravljavec merilnega sistema - merilec. Njegov vpliv je večinoma omejen na spretnost pri ravnanju z instrumenti, s katerimi se opravljajo meritve, in interpretaciji rezultatov meritev, na primer pri določanju merilne negotovosti, povezane z omejenim znanjem o meritvi.

Senzorično meroslovje pa se ukvarja z merjenji, odvisnimi od človeške zaznave in / ali tolmačenja. To vključuje zaznane (ang. perceived) lastnosti proizvodov in storitev, kot so kakovost ali zaželenost, in družbene parametre, kot sta varnost in dobro počutje. Največji problem senzoričnega meroslovja je doseganje soglasja o tem, kako lahko generična meroslovna vprašanja, terminologijo in merilne metode, merilne negotovosti, odločanja in presoja vplivov uporabljamo pri merjenju z osebami. Meroslovje na tem področju je kljub velikem potencialu za inovacije relativno slabo razvito, a sega preko tradicionalnega fiziološkega meroslovja in se ukvarja tako z merjenjem vseh človeških čutov, kot tudi duševnimi, vedenjskimi in kognitivnimi procesi. To je zlasti pomembno na področjih, kjer so ljudje in njihove zaznave pomemben sestavni del sistemov, npr. kjer sta zdravje in varnost človeka ogrožena.

1 UVOD

Senzorično meroslovje se že v veliki meri uporablja na različnih znanstveno raziskovalnih področjih, od ugotavljanja kvalitete oljčnega olja, pečenega jagenjčka, merjenja hrustljavosti ocvrtega krompirčka, merjenja občutka svežine zaradi pitja obarvanih izotoničnih pijač, merjenja zaznane kakovosti zraka, merjenja občutkov pri zapiranju avtomobilskih vrat, merjenja stopnje pomirjenja zaradi zvočne kulise, merjenja zadovoljnosti kupcev ob hladilnih omarah v velikih supermarketih, merjenja čustvene vzburjenosti med igranjem računalniških iger do ugotavljanja stopnje treme, bolečine ali psihološkega stresa.
Tako je na primer pri določanju zgornjih meja akustičnih signalov, ki so jim izpostavljeni ljudje, treba poleg nivojev zvoka, ki lahko fizično poškodujejo bobnič, upoštevati tudi zaznano raven akustičnih motenj. Sodni primeri, ki se ukvarjajo z okoljsko akustiko, pogosto poudarjajo, da zvok povzroča tudi psihološke motnje pri človeku (npr. zmanjšana sposobnost učenja pri pouku) in ne samo fiziološke učinke [1]. Podoben primer je vidno zaznavanje. Zgolj fizikalni opis človeškega očesa z zakoni optike ne more v celoti razložiti učinke, kot so obrat slike (slika na mrežnici se ne zazna kot na glavo obrnjena), vid starejših ljudi je zaznan kot boljši, kot v resnici je, ali pri bolnikih z umetnimi mrežnicami, ki vidijo jasne slike, čeprav z omejenim številom optičnih detektorjev.

Kompleksne veličine, kot so bolečina, strah, tesnoba, skozi vrsto fizioloških, živčnih, psihofizičnih, duševnih in vedenjskih odzivov obsegajo celoten razpon senzoričnega merjenja, npr. fizično uporabljeno mehansko silo.



Slika 1: *Senzorično meroslovje v svojo metodologijo vključuje človeka z njegovimi čutili in kognitivnimi sposobnostmi.*

Merjenje z osebami lahko v splošnem razdelimo na dve veliki skupini; "merjenje človeka" in "človek kot merilni instrument" [2]. Na ta način človek vstopi v merilni sistem v dveh, zelo različnih kontekstih.

2 MERJENJE ČLOVEKA

Na področju "merjenja človeka« je predmet meritve človek. Meritve v senzoričnem meroslovju segajo od merjenja dogajanja v človeku med dejavnostmi (npr. merjenje fMRI signala ali EEG signala med branjem, dinamično opazovanje aktivnosti človeških možganov, vzdraženosti človeške kože ali grimas na obrazu pri miselnih dejavnostih, igranju računalniških igric, poslušanju glasbe) do zaznavanju govorice telesa. Uporabljajo se tudi računalniški vid, termografija, EKG, EMG, MEG, EDA, fMRI in druge naprave za merjenje fizioloških parametrov, kot je elektrodermalna aktivnost, dihanje, gibanje zenice, razpoznavanje vzorcev, itd.
Vendar fiziologija človeka leži na stičišču psihofizičnih, duševnih in vedenjskih funkcij organizma, pri čemer je vse bolj jasno, da obstoji tudi pomembno prepletanje vseh treh. Psihofizični, duševni in vedenjski procesi vplivajo na človekov odziv na različne dražljaje (na primer pri bolečini).
Posledično narašča tudi potreba po spremembi in nadgrajevanju klasičnega fiziološkega meroslovja. Na vse več reguliranih aktivnih področjih tradicionalni fiziološko osnovani predpisi vedno pogosteje vključujejo človeške dejavnike, na primer kognitivne sposobnosti človeka, kar kaže na dejstvo, da je človeški odziv odvisen tudi od psihofizičnih, duševnih in vedenjskih procesov v kritičnem primeru.
Da se to doseže, je treba najprej opredeliti ustrezno primerne merjene veličine in nato zagotoviti, kadar je to mogoče, njihovo sledljivost na ustrezne enote SI sistema. V nekaterih primerih je treba razviti tudi smernice za oceno merilne negotovosti, na primer kadar so fiziološke veličine iz različnih merilnih lestvic (na primer ordinalnega tipa), ali če so meritve kvalitativne in/ali multivariatne.

2.1 Najbolj uporabljani psihofiziološki parametri

Električna prevodnost kože
Električna prevodnost kože (PGR – psihogalvanski refleks, SC - skin conductance, EDA - electrodermal activity, GSR – galvanic skin response) je ena najbolj splošno uporabljenih metod za opazovanje aktivnosti človeške kože in meri spremembo električnih lastnosti kože v obliki večje aktivnosti znojnic. Klinično se uporablja v psihopatologiji, dermatologiji, nevrologiji za diagnoze in ovrednotenja terapij [3, 4, 5]. Prevodnost kože se običajno meri med prstoma ene roke.

Temperatura kože
Ugotovljeno je bilo, da je, drugače kot pri dolgotrajnih termoregulacijskih spremembah v prekrvavitvi kože zaradi hlajenja (vazokonstrikcija) in gretja (vazodiletacija), sprememba pretoka krvi zaradi psihološkega dražljaja kratkotrajna. V topli osebi, ki ima tipično temperaturo prstov (33±2) °C, tako na primer simpatični stimulus povzroči vazokonstrikcijo v neporaščenem delu kože [6]. Temperatura se sicer običajno meri na konici prsta s pomočjo hitroodzivnega termometra.

Srčni utrip in krvni tlak
Vrednost srčnega utripa in sprememba srčnega utripa sta parametra, ki sta korelirana s psihološkim stanjem opazovane osebe [7, 8, 9]. Eden osnovnih fizioloških parametrov, ki se močneje spreminjajo zaradi sprememb v psihološkem stanju opazovane osebe, je arterijski krvni tlak [10, 11]. Krvni tlak v psihofizioloških eksperimentih lahko merimo diskretno s pomočjo oscilometrije ali avskultatornih merjenj, zaradi hitrih sprememb pa je primernejše zvezno merjenje, predvsem neinvazivno [12].

Ostale meritve
Merjenje *respiratorne frekvence* je ena osnovnih psihofizioloških metod [13, 14]. Telo oddaja toploto v okolico na več načinov – v obliki konvekcije toplote skozi kožo v kontaktu z zrakom, prevajanjem toplote skozi oblačila, izparevanjem znoja na koži, izparevanjem izdihane vlage in toplotnega sevanja. Temperaturi soroden parameter je tudi *toplotni tok*, ki je definiran kot razmerje toplotne konvekcije v skupni notranji energiji, ki se odda v okolico [15]. Med ostale neinvazivne meritve uvrščamo različne tehnike, kot je recimo *fotopletizmografija*, in različne oblike *opazovanja možgan* [16]. Aktivnost možganov lahko

opazujemo z metodami, ki merijo električni tok, ki se generira zaradi proženja nevronov, ali pa merijo magnetno polje, ki ga ta tok generira. Prvi tip metode je na primer elektroencefalografija (EEG) ali EPS, drugi pa njena magnetna različica magnetoencefalografija (MEG). Metode glede na parameter zaznavanja delimo na tiste, ki neposredno merijo elektromagnetno aktivnost sproženih nevronov, in tiste, ki posredno merijo aktivnost sproženih nevronov preko pretoka krvi. Po načinu delovanja delimo metode na pasivne in aktivne. Aktivne metode vplivajo na možgane in šele nato lahko izmerijo njihovo aktivnost (stimulacija z magnetnim poljem pri fMRI in MRI, injeciranje radioaktivnih organskih molekul pri PET), medtem ko pasivne le zaznavajo elektromagnetno aktivnost sproženih nevronov.

3 ČLOVEK KOT MERILNI INSTRUMENT

Na področju »Človek kot merilni instrument« se namesto merilnih instrumentov in opreme za merjenje predmetov uporabljajo človeška čutila. Primeri senzoričnega merjenja so recimo merjenje kvalitete zraka v prostoru (v smislu dober zrak, slab zrak), merjenje občutka ugodja zaradi dotikanja različnih tekstur površin, merjenje naravnosti (to je podobnosti z naravnimi) materialov, merjenje pomirjevalnega učinka glasbe…

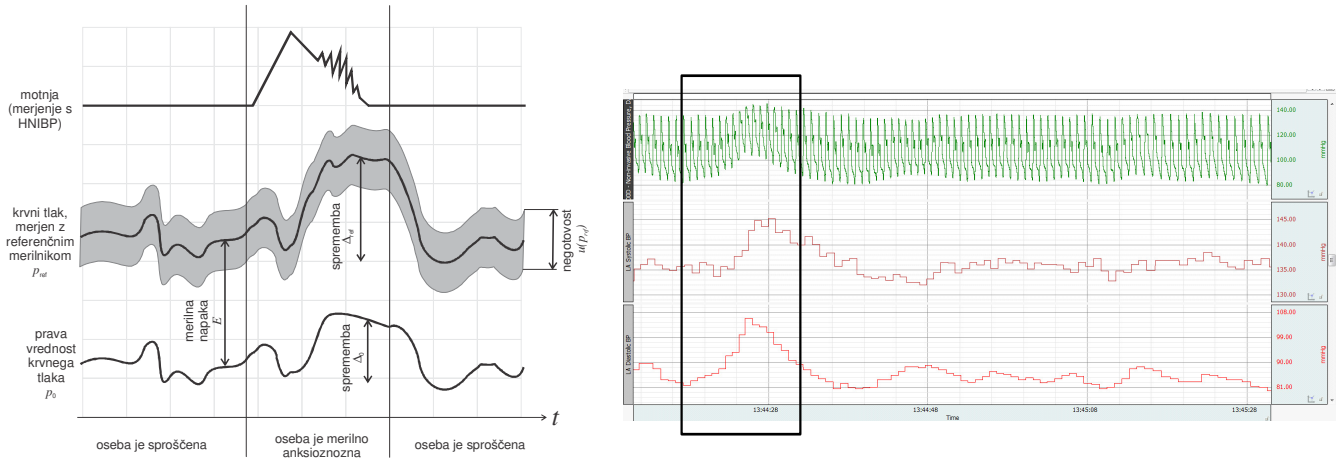
3.1 Merilna anksioznost kot prispevek negotovosti meritve

Če obravnavamo človeka kot merilni instrument, lahko v splošnem na rezultat meritve (in negotovost meritve) vpliva več dejavnikov. Poleg merilnih napak in negotovosti uporabljenih pomožnih merilnih instrumentov, merilnih metod, okolice merjenja lahko vpliva tudi človek, (uporabljen kot merilni instrument). Izmerjene vrednosti človeka kot merilnega instrumenta so odvisne od psihološkega stanja človeka, kar lahko opišemo s pojmom merilnan aksioznost.

Anksioznost je stanje psihične napetosti. Različne oblike anksioznosti so tesnoba, trema, napetost, živčnost, strah in zaskrbljenost, ki se razlikujejo tudi v intenzivnosti.

Merilno anksioznost posledično definiramo kot stanje psihične napetosti opazovane osebe zaradi izvajanja meritev, oziroma sodelovanja v merilnem procesu, zaradi česar se spremeni merilni rezultat.

Merjenje kot motnja, ki lahko povzroča psihološki odziv opazovane osebe, lahko vpliva na fiziologijo osebe v več nivojih (slika 2). Motnja lahko vpliva na referenčni zvezni merilnik, s katerim merimo fiziološko stanje osebe, na dejansko fiziologijo osebe ali pa na kombinacijo obeh. Ker se merilna anksioznost lahko izrazi v vseh nivojih, je potrebno eksperiment sestaviti tako, da rezultati lahko diferencirajo med različnimi nivoji.



Slika 2: *Levo - shematični prikaz vpliva merjenja s pomočjo naprave za domače merjenje krvnega tlaka (HNIBP, zgoraj) na časovni potek fiziološkega parametra krvni tlak (srednja krivulja). Spodnja krivulja prikazuje naravno fiziološko gibanje srednjega arterijskega tlaka. Desno - vpliv motnje (merjenja z HNIBP, črni pravokotnik) med opazovanjem sproščene osebe. Prikazana je časovna odvisnost zvezno merjenega krvnega tlaka sproščene osebe (zgoraj arterijski tlak, na sredini sistolični in spodaj diastolični tlak).*

4 ZAKLJUČEK

Klasično meroslovje (ang. hard metrology) se v z zadnjih letih zaradi potreb industrije in moderne človeške družbe vedno bolj usmerja k individualizaciji in personalizaciji merilnih metod in merilnih instrumentov. T.i. senzorična metrologija postaja vedno pomembnejša znanost, ki inherentno vključuje interdisciplinarnost in multidisciplinarnost široke skupine znanstvenih ved, od naravoslovnih, tehničnih do družboslovnih in humanističnih.

V zadnjih letih se industrija Evropske unije vedno bolj poslužuje senzoričnih merjenj, saj jo poleg običajnih fizikalno-kemičnih lastnosti izdelkov zelo zanima tudi subjektivna zaznava le-teh v očeh človeka (kupca). Tako na področju senzoričnega meroslovja delujejo veliki proizvajalci, kot so avtomobilska industrija, proizvajalci

pohištva, vezanih plošč in iverk, izdelovalci specifičnih vonjav (ang. brand specfic odour), predelovalci hrane, oblikovalci zvoka (ang. sound design), ipd.

Prispevek obravnava tudi enega pomembnejših prispevkov k merilni negotovosti kateregakoli fiziološkega parametra, ki je odvisen od avtonomnega živčnega sistema opazovane osebe. Merilna anksioznost - psihična napetost, živčnost oziroma utesnjenost med, pred in po merjenju z instrumentom lahko precej spremeni merilni rezultat.

Literatura

[1] Measurements With Persons: Theory, Methods and Implementation Areas, Taylor&Francis Psychology Press, 2011

[2] L. R. Pendrill, R. Emardson, B. Berglund, M. Gröning, A. Höglund, A. Cancedda, Gabriele Quinti, F. Crenna, G. B. Rossi, J. Drnovšek, G. Geršak, G. van der Heijden, K. Kallinen, and N. Ravaja, Measurement with Persons: A European Network, Measure, vol. 5, no. 2, June 2010

[3] Z Poh, NC Swenson, RW Picard, A Wearable Sensor fo Unobtrusive, Lon-Term Assessment of Electrodermal Actovity, IEEE Trans. Biomed. Eng. vol.57, no. 5, May 2010.

[4] H. Pazderka-Robinson, J. W. Morrison, P. Flor-Henry, »Electrodermal dissociation of chronic fatigue and depression: evidence for distinct physiological mechanisms", International Journal of Psychophysiology, vol. 53, pp.171– 182, 2004.

[5] DC Fowles, MJ Christie, R Edelberg, WW Grings, DT Lykken, PH Venables, Publication Recommendations for Electrodermal Measurements, Psychophysiology, vol 18, no. 3, 1981.

[6] A Kistler, C Mariauzouls, K von Berlepsch, Fingertip temperature as an indicator of sypathetic resposes, International Journal of Psychophysiology 29 , 35-41, 1998.

[7] N Hjortskov, D Rissen, AK Blangsted, N Fallentin, U Lundberg, K Søgaard, The effect of mental stress on heart rate variability and blood pressure during computer work, Eur J Appl Physiol, vol 92, 84–89, 2009

[8] R McCraty, M Atkinson, WA Tiller, G Rein, AD Watkins, The effects of emotions on short-term power spectrum analysis of heart rate variability, American Journal of Cardiology, vol 76, 1995.

[9] Heart rate variability, Standards of measurement, physiological interpretation, and clinical use, European Heart Journal, vol 17, 354–381, 1996.

[10] LC Becker, CJ Pepine, R Bonsall, JD Cohen, AD Goldberg, C Coghlan, PH Stone, S Forman, G Knatterud, DS Sheps, PG Kaufmann, Left Ventricular, Peripheral Vascular, and Neurohumoral Responses to Mental Stress in Normal Middle-Aged Men and Women, Circulation, vol 94, 2768-2777, 1996.

[11] P Fauvel, C Cerutti, P Quelin, M. Laville, M. P. Gustin, C. Z. Paultre, M. Ducher, Mental Stress–Induced Increase in Blood Pressure Is Not Related to Baroreflex Sensitivity in Middle-Aged Healthy Men, Hypertension, vol 35, 887-891, 2000.

[12] J Ogorevc, A Podlesek, G Geršak, J Drnovšek, The effect of mental stress on psychophysiological parameters, 2011 IEEE International Symposium on Medical Measurements and Applications (MeMeA 2011), Proceedings, Bari, Italy, 642-645, 2011.

[13] I Homma, Y Masaoka, Breathing rhythms and emotions, Exp Physiol 93.9, 1011–1021, 2008.

[14] D. Novak, J. Ziherl, A. Olensek, M. Milavec, J. Podobnik, M. Mihelj, M. Munih, "Psychophysiological Responses to Robotic Rehabilitation Tasks in Stroke", IEEE Transactions on neural systems and rehabilitation engineering, vol. 18, no. 4, pp. 351 - 361, 2010.

[15] CB Liden, M Wolowicz, J Stivoric, A Teller, C Kasabach, S Vishnubhatla, R Pelletier, J Farringdon, S Boehmke, Characterization and Implications of the Sensors Incorporated into the SenseWear Armband for Energy Expenditure and Activity Detection, BodyMedia Inc., 2002

[16] J Allen, Photoplethysmography and its application in clinical physiological measurement, Physiol. Meas. 28, 2007

NAD IN ONKRAJ SAMOURESNIČUJOČE SE PREROKBE SKOZI RAZISKOVANJE MOŽNOSTI OBJEKTIVISTIČNE TER KONSTRUKTIVISTIČNE EPISTEMOLOŠKE POZICIJE

Simona Gomboc
Oddelek za biopsihologijo
Fakulteta za matematiko, naravoslovje in tehnične vede
Univerza na Primorskem
Muzejski trg 2, 6000 Koper, Slovenija
e-mail: simona.gomboc@upr.si

POVZETEK

Prispevek na osnovi predstavitve pojava samouresničujoče se prerokbe naprej predstavlja možnosti zmanjševanja pogostosti pojavljanja le-tega skozi prevzemanje objektivistične epistemološke pozicije, nato pa ponudi še razmišljanje o pojavu skozi konstruktivistično epistemološko pozicijo.

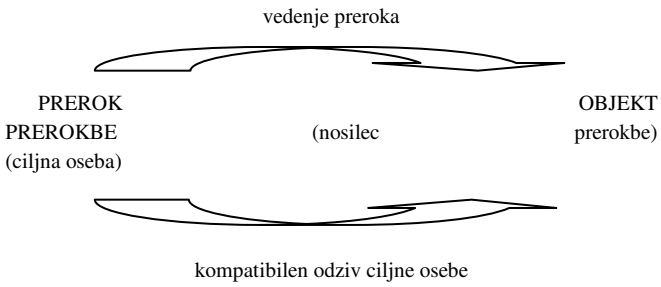
1 PREDSTAVITEV POJAVA SAMOURESNIČUJOČE SE PREROKBE

Koncept samouresničujoče se prerokbe je leta 1948 prvi predstavil sociolog Robert K. Merton, da bi opisal situacijo, v kateri pričakovanja ene osebe (ali skupine) o drugi osebi (ali skupini) vodijo do uresničitve teh pričakovanj. Pojav je opredelil na sledeč način: »Samouresničujoča se prerokba je v začetku *napačna* definicija situacije, ki izzove novo vedenje, katero nadalje povzroči, da se primarno napačna ocenitev situacije uresniči. Navidezna veljavnost samouresničujoče se prerokbe ohranja prevlado napake, kajti prerok bo dejanski izhod dogodkov videl kot dokaz, da je imel že od začetka prav.«[1]. Mertonu je pojav samouresničujoče se prerokbe služil za pojasnjevanje dogajanja v družbi. Kot bazično predpostavko za razumevanje dogajanja med ljudmi je privzel Thomasov teorem, ki pravi, da v kolikor posameznik označi situacijo kot resnično, je ta resnična v vseh svojih posledicah. Dodatno je Merton opozoril tudi na implikacije teorema, da se namreč posamezniki ne odzivajo le na objektivne značilnosti situacije, ampak, še pomembneje, na pomene, ki jih pripišejo objektivni situaciji. Ti pomeni potem določajo njihovo vedenje.

V psihologijo je pojav samouresničujoče se prerokbe uvedel Robert Rosenthal, ki ga je prvi začel preučevati eksperimentalno. Potem ko je potrdil obstoj vpliva eksperimentatorjev na (rezultate) udeležence(v) raziskav, je z Jacobsenovo [2] izvedel danes klasično Pygmalionovo študijo, kjer sta preverjala vpliv učiteljevih pričakovanj na inteligenčni razvoj učencev. Pojav samouresničujoče se prerokbe je z odzivi, ki jih je spodbudila Pygmalionova raziskava, postal pomembno raziskovalno področje znotraj socialne psihologije. Prvotna raziskovalna vprašanja so se dotikala obstoja pojava ter velikosti njegovega učinka, danes pa se raziskovalci ukvarjajo s poskusi razlage mehanizmov fenomena, iščejo pogoje, ki vplivajo na njeno pojavljanje in velikost njenega učinka ter preverjajo, v katerih okoljih je ta pojav moč zaslediti [3].

2 MOŽNOSTI ZMANJŠANJA POGOSTOSTI POJAVLJANJA SAMOURESNIČUJOČE SE PREROKBE KOT JIH OMOGOČA OBJEKTIVISTIČNA POZICIJA

Objektivizem (tudi pozitivizem, realizem) vidi opazovalca sveta ločenega od sveta »tam zunaj«, ki ga ta opazuje in predpostavlja, da je neposredno zaznavanje tega sveta strani opazovalca možno [4]. Z objektivistične pozicije neodvisnega opazovalca je pojav samouresničujoče se prerokbe viden na način, kot ga prikazuje spodnja slika:



Slika 7: Samouresničujoča prerokba z vidika zunanjega opazovalca.

Predstavitev mehanizma v pojava nam naprej omogoča razmislek o možnostih zmanjšanja pogostosti pojavljanja samouresničujoče se prerokbe. Mnenja raziskovalcev o pozitivnih in negativnih vidikih tega pojava so sicer različna.

Medtem ko možnosti, ki jih samouresničujoča se prerokba ponuja, nekatere navdušujejo: »edine še preostale meje za široko praktično uporabo samouresničujoče se prerokbe so meje naše domišljije« [5], so drugi pesimistični: »ni vsako življenje tisto, kar naredimo iz njega. Življenja nekaterih so to, kar drugi ljudje naredijo iz njih.« [6]. Najti pa je tudi manj skrajna mnenja: »Znanje o tem, da imajo pričakovanja lahko samouresničujoče se posledice, je tako opozorilno sporočilo, kot tudi sporočilo upanja. Opozarja, da lahko negativna pričakovanja o osebi povzročijo, da ta ne razvije potencialov, ki jih ima. Je pa tudi sporočilo upanja, saj pozitivna pričakovanja pomembnih oseb – staršev, učiteljev, nadrejenih – lahko pomagajo doseči uspehe, ki so bili prej le sanjani.« [3]. Čeprav pozitivni vidiki samouresničujoče se prerokbe obstajajo, so njene možne negativne posledice resne in zato vredne preučitve možnosti zmanjšanja pogostosti njenega pojavljanja [7].

2.1 Vidik ciljne osebe samouresničujoče se prerokbe

Poglejmo najprej možnosti za zmanjšanje pojavljanja samoizpolnjujoče se prerokbe, ki jih ima na voljo ciljna oseba. V zadnji zgornji navedbi lahko opazimo, da je ciljna oseba dojeta kot pasivni objekt, prepuščen vplivu pričakovanj drugih. Takšno pojmovanje ciljne osebe vsaj implicitno sprejema večina raziskovalcev, kljub temu pa so bile opravljene tudi raziskave, ki so pokazale na aktivno vlogo ciljne osebe pri uresničitvi samouresničujoče se prerokbe. Že nekaj let pred Mertonovo prvo opredelitvijo pojava [1], je Lecky [8] poročal o težnji ciljne osebe, da uskladi informacije z obstoječim konceptom sebe. Swann [8] je nadalje oblikoval teorijo o samopreverjanju (angl. *»self-verification«),* ki pravi, da ljudje nismo nagnjeni le k zaznavanju informacij, ki potrjujejo naš že oblikovani koncept sebe, ampak tudi aktivno delujemo v smeri, ki poveča možnost, da bodo informacije z njim konsistentne. Ugotovljeno je bilo, na primer, da sta tako zapomnitev kot priklic informacij, ki so skladne s konceptom sebe, boljša ter da so informacije, ki so skladne s konceptom sebe videne kot bolj verodostojne in dojete kot tiste, ki bolje odražajo dejanske sposobnosti ciljne osebe od informacij, ki s konceptom sebe niso skladne [9]. Te ugotovitve odpirajo možnost, da ciljne osebe pogosto ohranijo obstoječi koncept sebe na način, da pričakovanja preroka zaznavajo v skladu z lastnim konceptom sebe in jih v nekaterih primerih v skladu s konceptom sebe tudi spremenijo. Kot sklepa Madon s sodelavci: »ciljne osebe samouresničujoče se prerokbe so le redko tako voljne, da bi preprosto postale to, kar drugi od njih pričakujejo.« [8].

Pri ciljni osebi je težnja po ohranjanju koncepta sebe torej pogosto tista, ki pomembno vpliva na uresničitev samouresničujoče se prerokbe. Nadalje so raziskave tudi nakazale, da do samouresničujoče se prerokbe redkeje prihaja v primerih, ko ima ciljna oseba nizko potrebo po socialni potrditvi [10] ter visoko stopnjo samospoštovanja [11], vendar ostaja proces samopreverjanja razumljen kot tisti, ki poteka vzporedno s procesom samouresničujoče se prerokbe in skupaj z njim vodi do končnega učinka medsebojnega delovanja preroka in ciljne osebe [8].

Raziskave o vplivu prerokovih pričakovanj na ciljno osebo so bile vezane predvsem na spremembe na kognitivni ravni (konceptu sebe) ciljne osebe [12], čeprav teoretično obstaja tudi možnost, da ciljna oseba sicer spremeni koncept sebe pod vplivom prerokovih pričakovanj in iz njih izhajajočega vedenja, a *hkrati* na ravni vedenja ne da kompatibilnega odziva, ki bi imel za posledico uresničitev prerokbe. V kolikor, na primer, sin zazna, da oče od njega pričakuje, da se bo v določeni socialni situaciji vedel negotovo, se sicer lahko začne dojemati kot sramežljiva oseba (pride do spremembe koncepta sebe), vendar še vedno obstaja možnost, da njegovo vedenje ne bo takšno. Možnost za nepotrjujoči odziv se poveča, v kolikor se sin zaveda, da so očetova pričakovanja napačna, v tem primeru pa se zmanjša tudi možnost, da bo sin sploh spremenil koncept sebe. Premislek nas torej pripelje do sklepa, da je za učinek napačnih pričakovanj preroka na ciljno osebo pomembno zavedanje ciljne osebe o prerokovih pričakovanjih in predvsem o napačnosti le-teh.

2.2 Vidik preroka samouresničujoče se prerokbe

Če se zdaj prestavimo v pozicijo preroka samouresničujoče se prerokbe, vidimo, da ima tudi on možnost vplivanja na zmanjšanje pogostosti pojavljanja samouresničujoče se prerokbe, in sicer – tako kot ciljna oseba – na ravni vedenja (sprememba vedenja ali vzdržnost pri vedenju) ter na ravni mišljenja (sprememba pričakovanj).

Ker je vedenje preroka tisto, ki izzove kompatibilen odziv s strani ciljne osebe, lahko njegovo vedenje, ki je namenoma v nasprotju z njegovimi lastnimi pričakovanji, prepreči pojav samouresničujoče se prerokbe. Kot pravi Franzoi [6], ko daje napotke, kako ravnati z ljudmi, ki se do nas vedejo odklonilno: »Naslednjič, ko boste v interakciji s temi ljudmi, pustite ob strani svoja negativna pričakovanja glede njihovega obnašanja do vas in se, nasprotno, do njih vedite, kot da bi bili vaši najboljši prijatelji. Na podlagi ugotovitev [...] je zelo verjetno, da boste s tem, ko jih boste sami začeli dojemati drugače, tudi pri njih povzročili novo definiranje socialne resničnosti.« V kolikor je vedenje preroka drugačno od vedenja, ki bi sledilo iz njegovih pričakovanj, ne pride do samouresničujoče se prerokbe, ampak do tako imenovane samoizničujoče se prerokbe (angl. *»self-defeating prophecy«*). Pri tem pojavu pričakovanja preroka povzročijo posledice, ki so ravno nasprotne pričakovanim. Na primer, zaradi pričakovanja prometnega zamaška na določenem odseku ceste se vozniki usmerijo na druge poti in tako povzročijo, da do pričakovanega prometnega zamaška ne pride. Prerok lahko prepreči pojav samouresničujoče se prerokbe tudi s tem, da se vedenja, ki bi izviralo iz pričakovanj, vzdrži. Oseba, ki, na primer, pričakuje prepir s sosedom zaradi določenega parkirnega mesta, lahko ne

parkira na tem mestu. V kolikor vedenja, ki bi izviralo iz pričakovanj ni, le-to ne more izzvati kompatibilnega odziva ciljne osebe.

Na zmanjšanje oziroma preprečitev pojavljanja samouresničujoče se prerokbe pa prerok lahko vpliva tudi s spremembo svojih napačnih pričakovanj do ciljne osebe. Sprememba napačnih pričakovanj je sicer možna le v kolikor se prerok njihove napačnosti zaveda. Zavedanje je tudi sicer ključno pri zmanjševanju pojavljanja samouresničujoče se prerokbe – raziskave so pokazale, da prihaja do tega pojava redkeje pri posameznikih, ki dosegajo visoke rezultate na na lestvicah samo-opazovanja in socialne prepoznave [13] – in je skupno vsem načinom zmanjševanja pogostosti pojavljanja samouresničujoče se prerokbe, tako s strani preroka kot ciljne osebe. Pomembno je zavedanje celotnega pojava, kar vključuje zavedanje tako napačnosti pričakovanj kot tudi moči pričakovanj in načina, kako ta povzročijo svoje uresničenje.

3 MOŽNOSTI, KI JIH PONUJA KONSTRUKTIVISTIČNA EPISTEMOLOŠKA POZICIJA

Konstruktivistična epistemologija gradi na predpostavki, da je védenje konstrukt samega objekta védenja [14], pri čemer poskuša izgraditi koherentno epistemologijo brez vključitve predpostavke o objektivni resničnosti. Dualizem med »svetom tam zunaj« in našimi predstavami tega sveta s pozicije konstruktivizma več ni mogoč, ampak postaneta oba nivoja, torej stvarnost (»svet tam zunaj«) na eni strani ter epistemološki področji dejanskosti ter resničnosti (»svet, kot se nam kaže«) komplementarna in soodvisna. Predstave sveta ne morejo obstajati brez predpostavke o obstoju sveta, medtem pa je svet omejen s spoznavnimi možnostmi v domeni področij dejanskosti ter resničnosti. Kot pravi Kordeš [15]: »Našo izkušnjo sveta konstruira "zunanji svet" in nasprotno: »zunanji svet« je konstruiran v skladu z našim doživljenjem sveta. Obe strani rekurzije sta soodvisni, obe sta hkrati generator in generirano... Vsako stran lahko opazujemo posamično, lahko pa ju opazujemo kot medsebojno odvisna dela celotnega cikla spoznavanja.« »Svet tam zunaj« in naše predstave sveta so neločljivo povezane, pri čemer ju veže konstitutivni proces – spoznavna krožnost.

V kolikor prevzamemo konstruktivistično pozicijo in svet vidimo kot porajajoč se skozi proces spoznavanja le-tega, nam to omogoči drugačno razumevanje samouresničujoče se prerokbe, kot pojava, kjer so napačna pričakovanja tista, ki se uresničijo. Namreč, je sploh mogoče govoriti o napačnosti pričakovanj, če pa se ta uresničijo? Logičen sklep je negativen.

Tovrstno razumevanje tako znanstvenika, ki zagovarja naravoslovni model, kot vsakdanjega posameznika, ki zdravorazumsko dojema svet kot ločen od njega, oropa možnosti objektivnega spoznanja in s tem možnosti predvidevanja. V kolikor zunanjega sveta ni mogoče spoznati, tudi ni mogoče opredeliti pravilnosti oziroma napačnosti naših predstav o njem. V kolikor ne vemo, kakšen svet *je*, tudi ne moremo napovedati, kako se bo obnašal v prihodnje, kar izključi možnost znajdevanja v njem. Kajti predvidljivost sveta je tisto, kar omogoči upravljanje, spreminjanje in kontrolo ter s tem možnost obvladovanja, obvladovanje pa pomeni temeljno varnost pri znajdenju v okolju. Že Rosenthal in Jacobsenova sta zasledila, da ljudje bolj težimo k predvidljivosti kot k resnici. Prvo izberemo tudi na račun slednje ali pa slednjo vsaj lažje sprejmemo zaradi prve: »Neprijaznost resničnosti je lažje prenesti, v kolikor je pričakovana.« [2]. Ko razmišljata o funkcionalnosti tega, da so nam presenečenja neljuba, ponujata sklep, da smo ljudje zainteresirani »za doseganje čim večjega razumevanje sveta in preko tega razumevanja za doseganje večje kontrole nad resničnostjo. Ta predpostavka kot pomembna področja osvetljuje pomen magije, religije pa tudi znanosti. Prav znanost ključuje verjetnostne opredelitve določenih možnosti, s katerimi se pojavi presenečenj zreducirajo na minimum. Boljše razumevanje resničnosti nam omogoča večjo moč kontrole resničnosti pa tudi prilagoditve resničnosti ter zmanjševanje možnosti presenečenj. Če je bilo človeštvo zaradi možnosti napovedi bolj sposobnosti preživeti, /.../ potem je razumljivo, da so ljudje zainteresirani za doseganje čim večje pravilnosti svojih napovedi.«

Motivacija za temeljno varnost v okolju in s tem strukturiranje okolja (kar pomeni predvsem oblikovanje socialne resničnosti) torej vodi v oblikovanje pričakovanj, in še pomembneje, v nenehno prisoten interes, da so ta pričakovanja *pravilna*. Ko konstruktivistična pozicija odvzame možnost določiti pričakovanjem njihovo pravilnost ali napačnost, odvzame tudi možnost pravilnega napovedovanja prihodnjih dogodkov in s tem možnost nadzora nad okoljem. Vendar le z objektivistične pozicije, ki predvideva obstoj »sveta tam zunaj«, glede na katerega lahko določimo pravilnost oziroma napačnost pričakovanj. Konstruktivistična pozicija namreč namesto pravilnosti zaznav sveta, na kateri gradi občutek varnosti objektivizem, ponudi možnost drugačne varnosti: oblikovanje stabilne interakcije s svetom.

Konstruktivizem namreč namesto točne slike zunanjega sveta ponuja možnost znajdenja v »nespoznavnem« svetu, ki postane stabilen in s tem varen skozi nenehno interakcijo: ni zunanji svet tisto, kar spoznavamo, ampak je zunanji svet tisti partner, s katerim ustvarjamo celoto bivanja. Namesto linearnega spoznavanja »sveta tam zunaj« konstruktivizem ponudi krožno vzročnost, strah pred tovrstno vzročnostjo, kot pravi Kordeš [16] pa je opaziti že od Aristotela naprej. »Zdi se, da bi z opustitvijo vztrajanja pri linearnem pogledu zamajali trdnost, kredibilnost in celo samo možnost znanstvenega diskurza. Pojavil bi se zloglasni začarani krog,

ki hodi z roko v roki s paradoksom in nesmislom.« (prav tam, s. 128). Vendar, »zloglasni krog« postane s konstruktivistično pozicijo »čarobni ali kreativni krog«, kot ga imenujeta von Foester [16] in Varela [17], kar jasno demonstrira pojav samouresničujoče se prerokbe. Pojav namreč predstavlja primer, ko se, iz objektivistične pozicije, napačna pričakovanja, izmikajo preverjanju, saj, kot pravi Merton, »navidezna veljavnost samouresničujoče se prerokbe ohranja *prevlado napake*, kajti prerok bo dejanski izhod dogodkov videl kot dokaz, da je imel že od začetka prav« [1] in tako demonstrirajo našo nezmožnost neposrednega preverjanja pravilnosti ali napačnosti pričakovanj. Konstruktivistična pozicija vzročno krožnost, ki objektivno preverjanje onemogoča, zaobjame in posamezniku namesto možnosti preverjanja in predvidevanja ponudi možnost ustvarjanja in odgovornost za ustvarjeno. Medtem ko v kontekstu samouresničujoče se prerokbe to pomeni, da postane prerok odgovoren za uresničitev prerokbe, to širšem kontekstu pomeni, da posamezniki prevzamemo odgovornost za svet, v katerem kot simbolna bitja bivamo; torej za našo socialno resničnost.

4 ZAKLJUČEK

Samouresničujoča se prerokba predstavlja pojav, skozi katerega lahko jasno primerjamo možnosti, ki jih odpirata obe epistemološki poziji. Pri tem namen prispevka ni bil nakazovanje pravilnosti ene na račun kazanja šibkosti druge, ampak sta obe, skozi perspktivo konstruktivizma 2. reda, videni kot dopolnjujoči se.

Viri:

[1] R. Merton. The Self-Fulfilling Prophecy. *The Antioch Review,* 8 (2). 193-210.1948.

[2] R. Rosenthal, L. Jacobsen. *Pygmalion im Unterricht: Lehrerwartungen und intelligenzentwicklung der schüler.* Beltz verlag. Weinheim, Basel. 1973.

[3] M. J. Harris, M. J. The self-fulfilling prophecy. V R. F. Baumeister in K. D. Vohs (ur.), *Encyclopedia of Social Psychology*, 828-830. Sage Publicatins. Los Angeles, London, New Delhi, Singapore. 2007.

[4] U. Kordeš. Raziskovanje v konstruktivizmu? Kairos, 4 (1-2). 99-113. 2010.

[5] O. R. Davidson, D. Eden,D. Remedial self-fulfilling prophecy: Two field experiments to prevent golem effects among disadvantaged women. *Journal of Applied Psychology,* 85 (3), 386-398. 2000.

[6] S. L. Franzoi. *Social psychology.* McGraw – Hill. Boston. 2006.

[7] R. Rosenthal. Covert commucation in classrooms, clinics, courtrooms, and cubicles. *American Psychologist,* 57 (11). 839-849. 2002.

[8] S. Madon, A. Smith, L. Jussim, D. W. Russel, J. Eccles, P. Palumbo, M. Walkiewitz. Am I as you see me or do you see me as I am?: Self-fullfilling prophecies and self-verification. *Personality and Social Psychology Bulletin,* 27, 214-1224. 2001.

[9] W. B. Swann. Identity negotiation: Where two roads meet. *Journal of Personality and Social Psychology,* 53,1038-1051. 1987.

[10] M. J. Harris, R. Rosenthal. Counselar and client personality as determinants of counselor expectancy effects. *Journal of Personality and Social psychology,* 50 (2), 362-369. 1986.

[11] L. Jussim. Self-fulfilling prophecies: A theoretical and integrative review. *Psychological Review,* 93 (4), 429-445. 1986.

[12] J. M. Harris, R. Rosenthal, S. E. Snodgrass, The effects of teacher expectations, gender, and behavior on pupil academic performance and self-concept. *Journal of Educational Research,* 79 (3), 173-179. 1986.

[13] M. J. Harris, R. Rosenthal. Mediation of interpersonal expectancy effects: 31 meta-analyses. *Psychological Bulletin,* 97 (3), pp. 363-386. 1985.

[14] A. Ule. Dosegljivost resnice. Filozofska fakulteta. Ljubljana. 2004.

[15] U. Kordeš. *Konstrukcije stvarnosti v distribuiranih epistemskih sistemih.* Magistrsko delo. Filozofska fakulteta. Ljubljana. 2000.

[16] U. Kordeš. Od resnice k zaupanju. Studia Humanitatis. Ljubljana. 2004.

[17] F. Varela. Two principles of self-organization. V H. Ulrich in G. J. B. Probst (ur.), *Self-Organisation and Management of Social Systems Insights, Promises, Doubts and Questions, 25-32.* Springer-Verlag. Berlin. 1984.

O NAČELIH ZAVESTI

Fatih Hodžić
Organizator informacijskih sistemov
Robova 8, 3000 Celje, Slovenija
Tel: +386 59 013084; GSM: +386 31 380707
e-mail: fatih_hodzic@t-2.net

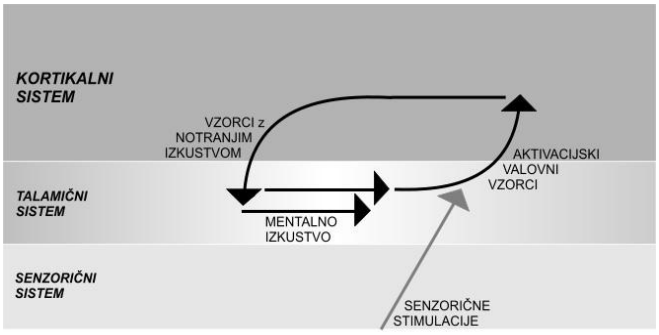
IZVLEČEK

Izhodišča: Izhajajoč iz študij o aktivnostih v diencefalonu, retikularni formaciji in delovnem prostoru talamusa, je pomembno izpostaviti vlogo slednjega kot pacemaker-ja pri ustvarjanju talamičnih aktivnosti[1] in aktivacijskih vzorcev[5]. Pojav valovanj, sinhronih oscilacij in ritmičnih aktivnosti v retikalno-talamičnem sistemu, povzroča večino kortikalnih aktivacij[16]. Poleg tega pa se v povsem drugačni luči dojema vloga mielinskih proteinov v kognitivnem modelu zavestnih procesov.

Zaključki: Izmenjevanje energijskih stanj z valovanji, ustvarjenimi v retikularno-talamičnem aktivacijskem sistemu – TAS in sklenjenim v talamičnem procesnem sistemu – TPS, poteka v aktivacijskih cikli. Zaradi nenehne kavzalnosti v valovanju vpošteva se retardiran čas, ki povezuje spremembe energijskih stanj v talamusu, na začetku in koncu vsakega aktivacijskega cikla. Epilog kortikalnih aktivacij in procesiranja – v času trajanja aktivacijskega cikla - je multiplikacija akcijskih potencialov v asociativne akcijske potenciale in pojav elektromagnetne interference[10,12]. Aktivacijski vzorci iz talamičnega aktivacijskega sistema se sestavljajo z valovanjem, ki izvira pretežno iz mielinskih proteinov. Le ti so v ovojnicah ali plaščih, na izrastkih nevroglijalnih celic, katere obdajejo nevrite[15]. Elektromagnetni valovi iz stimuliranih mielinskih proteinov spremenijo (po končani interferenci) dele valovnih sekvenc, ali fragmentov v aktivacijskem valovnem vzorcu. Skupni imenovalec vseh interferenc v kortiko-kortikalnih zankah je sestavljanje elektromagnetnih valov z aktivacijskimi vzorci, oziroma rekurzivni vnos notranjega izkustva (ang. *inner experience*) v akcijske in asociativne akcijske potenciale. Čeprav se v raziskavah zaznavajo in definirajo kot gama možganski valovi[14], so notranja izkustva valovni vzorci v energiji valovanja, oziroma množica podatkov na katerih se izvajajo motorične, senzorične in psihične funkcije. Oblike krepitev (ali slabitev) valovnih vzorcev s konstruktivno (ali destruktivno) interferenco mogoče je razumeti kot ‘motnje’ v sinhronih oscilacijah, vendar so te iste motnje v valovnih sekvencah aktivacijskih vzorcev notranja izkustva in tisto kar je ‘berljivo’ v procesnih centrih mentalnega izkustva in zavesti.

1 UVOD

Talamus predstavljajo dve ovalni simetrični strukturi v diencefalonu nad možganskim deblom, med možgansko skorjo (korteksom) in mezencefalonom. Talamična jedra – gre za približno 15 dorzalnih, ventralnih, retikularnih jeder - so funkcionalne enote, sekcije ki imajo obsežen aferentni dotok signalov iz senzornih sistemov kot tudi signalov, aktivacijskih vzorcev iz kortikalnih regij. Vsi signali iz senzoričnih vlaken (z izjemo nekaj olfaktivnih) v talamusu 'prestopijo' na tirnice aktivacijskih valovnih vzorcev kateri bodo uvedli senzorne signale v regije možganske skorje, ter tam procesirane in obogatene z notranjim izkustvom prinesli nazaj v talamični procesni sistem.
Dražljaji (signali iz senzornega sistema) se usmerjajo v kortikalne regije z interferenčnimi tokovi, sestavljenimi skupaj z aktivacijskimi valovnimi vzorci, razpršenimi po vlaknih nevronskih in nevroglijalnih sklopov. Med trajanjem aktivacijskega ciklusa (slika 1), valovni vzorci so v inerciji, interferenčno se krepijo (ali slabijo) in refinirajo z notranjim izkustvom. Po distribuciji v procesnim centrima TPS in pretvorbi notranjega v mentalno izkustvo, preostanek valovanja postane spet, delno ali v celoti, del ali sekvenca aktivacijskega valovnega vzorca, ritmičnega signala v novem aktivacijskem ciklusu.



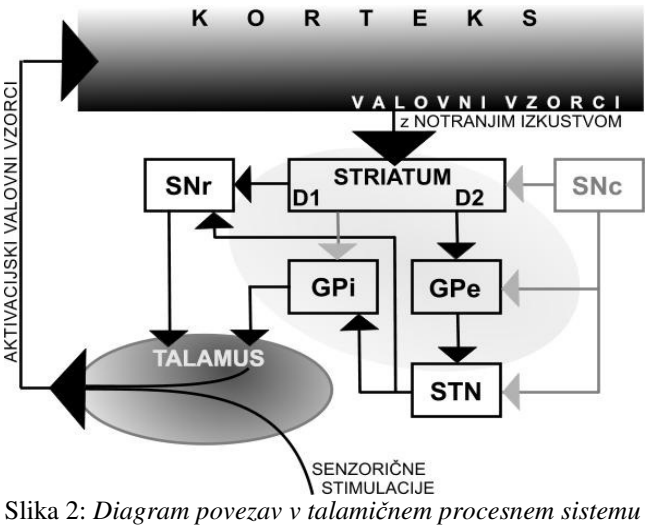
Slika 1: *Shema aktivacijskega ciklusa.*

2 ERTAS

Koncept opisanega kognitivno-nevroznanstvenega modela, ki se nanaša na aktivnosti, na stimulacije in dražljaje, aktivacijske valovne vzorce in njihovo procesiranje (multiplikacijo akcijskih potencialov v asociativne akcijske

potenciale in interferenco elektromagnetnih valov), temelji na razširjenem retikularno-talamičnem aktivacijskem sistemu ERTAS – ang. *extended reticular-thalamic activating system* [11]. V funkcionalnem smislu, gre za primer interakcije stimulacij (dražljajev iz senzornega sistema), aktivacijskih valovnih vzorcev (ritmičnih signalov iz talamičnega aktivacijskega sistema) in notranjega izkustva na mestih sprejemanja. Gre tudi za transmisije in medsebojne odnose znotraj več anatomskih struktur, ki so tesno povezane z vedenjskimi in kongnitivnimi vsebinami ter zavestjo. V tem procesnem delovnem prostoru so zajete, skupaj s talamusom in njegovimi jedri, tudi retikularne formacije možganskega debla in mezencefalona (srednjih možganov), bazalni gangliji (corpus callosum) in limbični sistem. Sem spada tudi razpršen projekcijski sistem, katerega vlakna so razvejana po vseh delih korteksa in kortiko-kortikalnimi povezavami na perifernih področjih možganov. Vse te strukture skupaj tvorijo ERTAS.
Večina kortikalnih aktivacij[16] izzvana je z aktivacijskimi vzorci, ki izvirajo iz talamičnega aktivacijskega sistema. ERTAS se lahko tolmači kot globalni delovni prostor psihičnih funkcij interaktivno povezan (zaradi prilivov in odlivov) s senzorno-motorični sistemi in kortikalnimi regijami. S tega vidika, ERTAS ima ključno vlogo v procesih, pogojenih z zavestjo[2].

Bazalni gangliji skupaj s talamusom predstavljajo jedro talamičnega procesnega sistema, sistema sekcij psihičnih funkcij in modulacije zavesti.
Signali iz korteksa, sprejeti v TPS, v bazalne ganglije in usmerjeni v talamus (slika 2), so podvrženi glutamatergičnimi (glutamat), GABAergičnimi (gama-aminomaslena kislina - GABA) in dopaminergičnimi (dopamin) transmisijami.

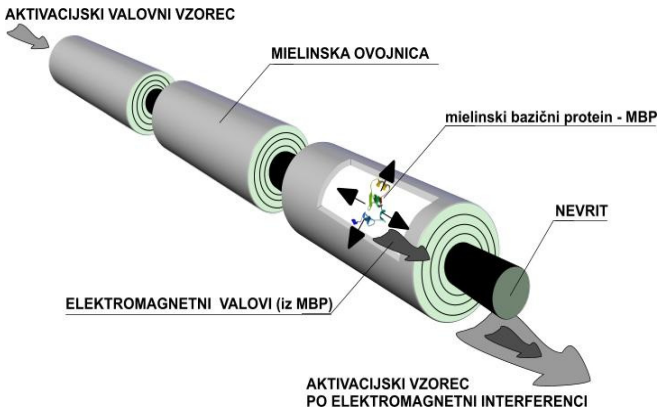


Slika 2: *Diagram povezav v talamičnem procesnem sistemu*

Sprejemanje aktivacijskih vzorcev v TPS poteka skozi vhodna jedra striatuma ekscitatornimi glutamatergičnimi transmisijami, v talamus pa izključno z GABAergičnimi transmisijami. GABAergične in kolinergične transmisije iz striatuma in drugih jeder bazalnih ganglij (izhodnih jeder: substantia nigra pars reticulata - SNr in notranji globus pallidus - GPi) usmerjene proti talamusu, so podvržene dopaminskimi transmisijami vezanimi na D2 in D3 dopaminske receptorje[3]. Več raziskav o distribuciji dopaminskih receptorjev razkriva relativno visoko gostoto v jedrih talamusa in njihovo heterogeno porazdelitev[13]. Zaradi takšne gostote D2/D3 receptorjev je razumljivo, da se večina ciljnih, prominentnih projekcij v talamusu spodbuja in ureja z dopaminom.

3.1 VALOVANJA IZ MIELINSKIH PROTEINOV

Več kot 50% celic v osrednjem živčevju CNS – ang. Central nervous system, so specifične nevroglije: astrociti, oligodendrociti v CNS ter Schwannove celice v perifernem živčevju PNS - ang. Peripheral nervous system[17]. V mielinskih ovojnicah prevladujejo plasti lipidov (do 70%), znotraj katerih se nahajajo mielinski proteini (do 30%) različnih struktur in stadiumov[6,7]. Vloga lipidov in obnašanje proteinov v mielinih so primer popolnosti v funkcionalnem smislu. Neizogibnost mielinskih proteinov v funkcionalnih shemah CNS temelji na elektrokemijskih lastnostih in valovanjih iz proteinov s katerimi se tudi identificirajo notranja izkustva.
Razen zagotavljanja idealnih pogojev za dejavnosti proteinov, lipidi v ovojnicami varujejo in ohranjajo proteine v njihovi izvirni molekularni strukturi. Lipidi ne obvarujejo proteine od valovanj iz okolja, lahko pa preprečijo da se te, zelo občutljive molekule ne bi – zaradi zunanjih vplivov - strukturno spreminjale, ali stopile v stik s sosednjimi proteini. V prostoru okrog mielinskih proteinov - zaradi električnega naboja in valovanj iz akcijskih potencialov - nastajajo električna in magnetna polja, katera proteini zaznavajo in se na njih ustrezno odzovejo z nihanjem in električnim nabojem. Na ta način spodbujeni in iritirani proteini postanejo vir elektromagnetnih valov (slika 3), čigar valovne vrednosti (amplitude, frekvenca vala) so značilne in enoznačne za vsako od teh molekul.

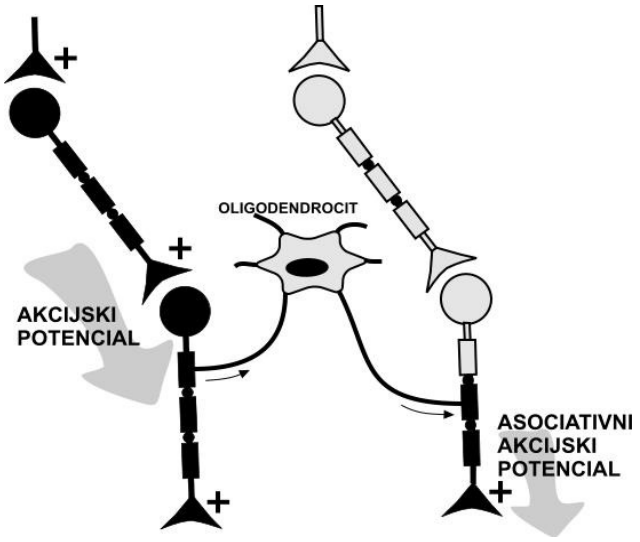


Slika 3: *Primer elektromagnetne interference na nevritu, po aktivaciji nevrona*

Elektromagnetna interferencija je način refiniranja valov iz mielinskih proteinov z valovanjem (akcijskimi potenciali) v aktivacijskih valovnih vzorcih. Njihova električna polja se kombinirajo, združujejo v nekakšno rezultatno polje po načelu superpozicije. Konačno stanje je nova ali spremenjena valovna sekvenca v aktivacijskem valovnem vzorcu. Ta razlika se lahko identificira kot 'kvant' notranjega izkustva ki se kasneje, v kombinaciji z drugimi notranjimi izkustvi, prepoznava v zavesti. V kvantni mehaniki velja po Heisenbergovem načelu nedoločenosti, da je istočasno določanje lege in gibalne količine delca s poljubno natančnostjo nemogoče in nezanesljivo. Tudi določanje lege in gibalne količine 'vzorca' notranjega izkustva, pridobljenega v valovnem vzorčenju, je nezanesljivo.

3.2 AKCIJSKI IN ASOCIATIVNI AKCIJSKI POTENCIALI

Električni naboj mielinskih proteinov se ohranja in prenaša iz mielinske ovojnice, po telesu oligodendrocita, njegovem citoskeletu (strukturami mikrotubul in mikrofilamentov) do drugih izrastkov z mielinskimi ovojnicami. Na enem oligodendrocitu je lahko tudi do 50 mielinskih ovojnic ki obdajejo nevrite več drugih nevronov, v funkcionalno isti ali v sosednjih kortikalnih regijah[4]. Čez sinaptične kanale (prostor med mielinskim plaščem in nevritom), se električni naboj prenese neposredno v nevrit, v telo nevrona in nadaljuje svojo pot vzdolž nevritov kot asociativni (podvojeni) akcijski potencial (slika 4). Ranvierovi zažetki, mesta na nevritu med sosednjima mielinskima ovojnicama omogočajo, da se asociativni akcijski potenciali – od mesta na nevritu kjer vstopijo v telo nevrona - usmerjajo proti njihovim končnicam.



Slika 4: *Primer multiplikacije akcijskega potenciala v asociativni akcijski potencial (po oligodendrocitu).*

Ob sprožitvi asociativnega akcijskega potenciala, ionski kanali Na+ in K+ koncentrirani na Ranvierovimi zažetki, ustvarijo prehoden preobrat membranskega potenciala (depolarizacijo)[9]. Zaradi prepusne polarizacije v Ranvierovih zažetkih, se asociativni akcijski potenciali prenašajo vedno vzdolž nevrita. Schwannove celice na živčnih vlaknih v PNS so v enaki funkciji kot mielinske ovojnice in Ranvierovi zažetki v CNS. Schwannove celice vzdržujejo, neposredno usmerjajo (spodbujajo ali zavirajo) in v elektromagnetnoj interferenci ojačajo signale (dražljaje) pri njihovem prenosu.

4 RESONANCA IN DOPAMINERGIČNO DUŠENJE

Ko se 'psihična' energija, kot neko stanje v procesnih centrih - v jedrih in lamelah talamusa - pretvarja izmenično v kemično in toplotno energijo ter spet v električno energijo, postajajo talamične sekcije prizorišča doživljajev, prizorišča svetlobe in zvoka, razpoznavanja vonjav, občutkov strahu ali zadovoljstva, sanj, domišljije, misli, spominov…

Resonanca je sistemski pojav v možganih. Pojavi se pri pretvorbi dražljajev (po percepciji) že v senzornem sistemu in pri aktivaciji valovnih vzorcev v talamičnem aktivacijskem sistemu. Pojavi se pri sprožanju asociativnih akcijskih potencialov ter elektromagnetne interference v kortikalnih regijah, pri distribuciji valovnih vzorcev iz korteksa v jedra bazalnih ganglijev, ter ob sklenitvi vsakega aktivacijskega ciklusa v sekcijah talamusa. S slikanjem možganov z magnetno resonanco (MR) ali z tehniko funkcionalne magnetne resonance je mogoče snemati aktivnosti v 'žariščih' kortikalnih regij[8], kjer se izvajajo procesi multiplikacije signala, pojav asociativnih akcijskih potencialov in elektromagnetnih interferenc valov. Z prisotnostjo nevrotransmiterjev, kot sta dopamin ali acetilkolin, se ohranjajo razmere, v katerih se aktivacijski vzorci transmitirajo in transformirajo. Dopamin je izraziti stimulator in modulator resonance pa tudi – na račun svoje presnove - njen izvršitelj v fazi do končnega dušenja.

V izmeničnem pretvarjanju energije, pri resonanci do njenega popolnega ali delnega dušenja, se pojavljajo z električno energijo tudi kemična in toplotna energija. Mogoče je sklepati, da bi učinek vsake od teh oblik energije pustil trajne posledice v celični infrastrukturi sekcija talamusa ali kjerkoli drugje v CNS. Okrepljeno valovanje (električnega naboja in toplote) bi ali požgalo, ali prekomerno segrelo tkiva v neposredni bližini. Poleg tega da spodbuja resonanco, dopamin je sposoben pretvorjeno toplotno energijo enostavno 'absorbirati' in obdržati stanje v okolju v mejah dopustnega. Vsaka resonanca znotraj sekcij talamusa je dovolj, da se navzven zlije in ovekoveči eno notranje izkustvo več.

Do končnega dušenja, se dopamin razgradi (v homovanilično kislino) in izloči iz sistema. Nizke koncentracije homovanilične kisline v urinu zanesljivo kažejo na motnje in pomanjkanje dopamina v CNS.

5 PSIHIČNA ENERGIJA V ENAČBI

V aktivacijskem ciklusu se računajo razlike v količini energije: stanja na začetku v TAS in na koncu v TPS. Energija v TAS je stanje z aktivacijskimi valovnimi vzorci, kateri so v interakciji in v interferenci s signali iz senzornega sistema. Energija v TPS je stanje s prilivi valovnih vzorcev iz kortikalnih regij, vzrokovano z odlivi aktivacijskih vzorcev iz TAS. Ob aktivacijah so poti valovnih vzorcev v TPS (skozi bazalne ganglije do talamusa) ali ekscitatorne (spodbujevalne), ali inhibitorne (zaviralne).
Vpoštevajoč retardirani čas, je prevladujoča količina 'psihične' energije (E_P) - tista ki se v resonancah spremeni v doživljaje, dejansko razlika med energijami vsakega aktivacijskega ciklusa v TPS (E_{TPS}) in v TAS (E_{TAS}):

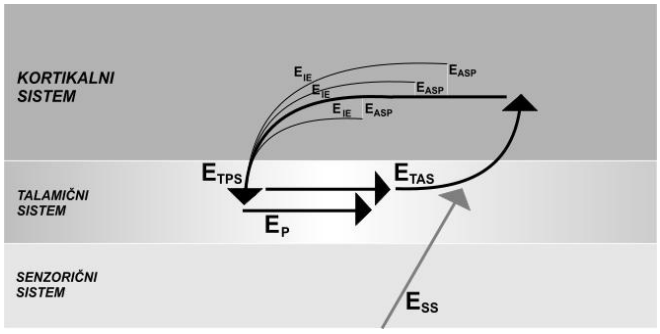
$$\Delta E_P = E_{TPS} - E_{TAS}$$

V stanju, ki bi ustrezalo najtežji kvantitativni motnji zavesti (stanju kome), E_{TPS} bi bila enaka E_{TAS} in bi $E_P = 0$. Skupna energija v TPS je:

$$E_{TPS} = E_{TAS} + E_{SS} + \Sigma E_{ASP} + \Sigma E_{IE}$$

E_{SS} pomeni energijo dražljajev, signalov iz senzornega sistema, ΣE_{ASP} vsoto energij asocijativnih akcijskih potencialov in ΣE_{IE} vsoto energij valovnih vzorcev z notranjimi izkustvi (slika 5). Tako kot se ves čas E_{TAS} vzorči in spreminja sprejemanjem E_{SS} ter E_{IE}, tako se tudi E_{ASP} spreminja sprejemanjem E_{IE}. Količina psihične energije v resonanci s popolnim dušenjem ali v resonanci z delnim dušenjem je:

$$E_P = \Delta E_{TAS} + \Delta E_{SS} + \Delta E_{ASP} + \Delta E_{IE}$$



Slika 5: *Shema energijskih stanj v aktivacijskem ciklusu.*

S pretvorbom 'psihične' energije se valovni vzorci notranjega izkustva transformirajo v mentalno izkustvo. Po resonanci, se del energije asociativnih akcijskih potencialov, dražljajev in aktivacijskih valovnih vzocev z vzorci notranjega izkustva prenese v E_{TAS}, v energijo aktivacijskih valovnih vzorcev naslednjega aktivacijskega ciklusa.
Glede na to, da je dušenje v resonancah ponavadi nepopolno (delno), del E_{SS} ali E_{IE} ostane integriran v aktivacijskem valovnem vzocu naslednjega aktivacijskega ciklusa. Na ta način se – v novem ciklusu - obnavljajo ali pridobivajo novi asociativni akcijski potenciali in valovni vzorci z notranjimi izkustvi, kateri so v zavesti kognitivna nadgradnja izkustvu iz prejšnjega, ali prejšnjih aktivacijskih ciklusov.

6 ZAKLJUČEK

Govor in uporaba jezika, kot obliki verbalne komunikacije, izražanje misli in občutkov…, so samo refleksije iz zaporedja kombinacij, katerih spremenljivke so valovni vzorci notranjega izkustva. Nepredstavljivo veliko je teh spremenljivk, ki v vsakem trenutku in nenehno v ciklusih preplavljajo procesne centre zavesti.

Govoriti o zavesti kot izključno človeški prvini je le pogojno pravilno. Nihanja ali razlike med gibalnimi količinami proteinov in drugih delcev v TPS, oziroma v delovnem prostoru talamusa se odražajo skozi različna stanja ali nivoje zavesti. Med pomembnejše dejavnike teh razlik štejejo zmogljivost nevronskih mrež, stopnja mielinizacije, kvaliteta transmisij in kvantiteta nevrotransmiterjev. Izkustva, pridobljena v kognitivnem razvoju, s katerimi se človek poslužuje v življenskem vsakdanu omogočajo vzdrževati in ohraniti raven zavesti. Razlika v energiji aktivacijskega ciklusa, med TPS in TAS (z ali brez E_{SS}) je določljiva tudi pri drugih organizmih, zlasti pri višjih sesalcih, vendar so te razlike sorazmerne številu asociativnih akcijskih potencialov in posledično tudi valovnih vzorcev z notranjim izkustvom. Navadno prevladujejo vzorci, ki predhodijo nagonskem vedenju ali tisti ki so posledica stimulacij in zunanjih dražljajev. Med njima so tudi 'moteči' vzorci iz kortikalnih aktivnosti, ki bi ustrezali kategorijam notranjega izkustva. Število teh vzorcev z notranjim izkustvom je zanemarljivo, ampak ni nepomembno. Dejstvo je, da možgani višjih sesalcev omogočajo in zagotavljajo podporo tudi različnim aspektom kognicije in vedenja, s katerimi bi se ti organizmi ne samo zavedali sveta okrog sebe, ampak tudi - in pod določenimi pogoji - zavedali sebe v tem svetu.

V zvezi s talamusom in talamičnim aktivacijskim sistemom, ne bo odveč še ena zanimiva podrobnost. Poškodbe različnih anatomskih ali funkcijskih enot v CNS povzročajo bolezni z nevropsihičnimi motnjami. Kakršnakoli poškodba talamusa vodi do najhujše motnje zavesti in stanja kome.

Literatura

[1] **Alkire** M.T., Haier R.J., Fallon J.H., Barker S.J. (1996) *PET imaging of conscious and unconscious verbal memory*, Journal of Consciousness Studies, Volume 3, Numbers 5-6, pp. 448-462.

[2] **Baars** B.J. (1988) *A Cognitive Theory of Consciousness,* Cambridge University Press, Cambridge.

[3] **Barto**, A. G. (1995) Adaptive Critics and the Basal Ganglia. In *Models of Information Processing in the Basal Ganglia*, Cambridge, MA: MIT Press.

[4] **Baumann** N., Pham-Dinh D. (2001) *Biology of Oligodendrocyte and Myelin in the Mammalian Central Nervous System,* Physiological

Reviews, Vol. 81, No.2, Copyright ©2001 by the American Physiological Society, April, pp. 871-927.

[5] **Cacioppo J.,** Tassinary L., Berntson G.(2000) *Handbook of psychophysiology*, Cambridge University Press, NY.

[6] Pedraza L, Fidler L, Staugaitis SM, **Colman** D.R. (1997) *The Active Transport of Myelin Basic Protein into the Nucleus Suggests a Regulatory Role in Myelination*, Neuron. Apr. 18(4), pp. 579-89.

[7] Brophy PJ, Boccaccio GL, **Colman** D.R. (1993) *The distribution of myelin basic protein mRNAs within myelinating oligodendrocytes,* Trends Neurosci. Dec; 16(12), pp. 515-21.

[8] **Coyle** S., Ward T., Markham C. (2003) *Brain–computer interfaces: a review,* Interdisciplinary Science Reviews, Maney Publishing, Vol. 28, Number 2, June, pp. 112 – 118.

[9] **Hubel** D.H. (1988) *Eye, brain and vision*, Scientific American Library (Distributed by W.H. Freeman), New York.

[10] **McFadden** J. (2006) *The CEMI Field Theory: Seven Clues to the Nature of Consciousness,* In Tuszynski J. *The Emerging Physics of Consciousness*. Berlin: Springer. pp. 385–404.

[11] **Newman** J. (1997) *Putting Together the Puzzle*, Part-I, *Towards a General Theory of Neural Correlation of Consciousness* Part-II, Journal of consciousness Studies, Volume 4, Numbers 1-2, pp. 46-66, 101-121.

[12] **Pockett** S. (2000) *The Nature of Consciousness*, Writers Club Press, Lincoln, NE.

[13] **Rieck** R.W., Ansari M. S. , Whetsell W.O., Deutch A.Y. & Kessler R.M. (2003) *Distribution of Dopamine D_2-Like Receptors in the Human Thalamus: Autoradiographic and PET Studies Neuropsychopharmacology,* Advance online publication, Vol.29, Nov. pp. 362–372.

[14] **Rubik** B. (2011) *Neurofeedback-Enhanced Gamma Brainwaves from the Prefrontal Cortical Region of Meditators and Non-Meditators and Associated Subjective Experiences*, The Journal of Alternative and Complementary Medicine, 17(2): 109-115.

[15] **Suzuki** M., Raisman G. (1994) *Multifocal pattern of postnatal development of the macroglial framework of the rat fimbria*, Glia, Vol. 12, Issue 4, pp. 294–308.

[16] **Traub** R. D., Contreras D., Cunningham M.O., Murray H., LeBeau F.E.N., Roopun A., Bibbig A., Wilent W.B., Higley M.J. and Whittington M.A. (2005). *Single-Column Thalamocortical Network Model Exhibiting Gamma Oscillations, Sleep Spindles, and Epileptogenic Bursts*. Neurophysiol. 93: 2194-2232.

[17] **Ullian** E.M., Christopherson K.S., Barres B.A. (2004) *Role for glia in synaptogenesis«*, Special edition: Glial Control of Synaptic Function, Vol. 47, Issue 3, pp. 209–216.

VPLIV ZAVEDANJA METAKOGNITIVNIH PROCESOV PRI SAMOREGULACIJSKEM UČENJU NA DOŽIVLJANJE UČENJA PRI GLASBENI VZGOJI

Ada Holcar
Zavod Republike Slovenije za šolstvo
Predmetna skupina za glasbo
Parmova 33, 1000 Ljubljana, Slovenija
Tel: +386 1 23 63 120; fax: +386 1 30 05 199
e-mail: ada.holcar@zrss.si

IZVLEČEK

V pričujoči raziskavi sem preučevala, kako učenci doživljajo učenje, preverjanje, ocenjevanje in samoregulacijo pri glasbeni vzgoji v osmem razredu osnovne šole. V tem prispevku so predstavljeni rezultati na vprašanje, kakšen vpliv ima zavedanje metakognitivnih procesov pri samoregulacijskem učenju na doživljanje učenja. Izsledki raziskave so pokazali, da so učenci, ki so zavestno spremljali in uravnavali svoje znanje in učne procese, v učenju videli večji smisel, v krajšem času dosegli boljše glasbene dosežke ter prišli do spoznanja, da je samoregulacija področje, ki ga bodo lahko prenesli in uporabili tudi v svojem življenju.

1 UVOD

Ugotovitve raziskav s področja samoregulacije kažejo, da uporaba kognitivnih in čustvenih vidikov samoregulacije pripelje učence do boljših učnih rezultatov. Večina raziskav narejenih na področju glasbe se nanaša na preučevanje vadenja inštrumentov, Fritz-Smolejeva (2006) pa je preučevala vpliv samoregulativnega učenja pri nauku o glasbi. V raziskavi je preučevala, na kakšen način so povezani samoregulativni procesi in različni dejavniki učenja (spol, starost, splošne sposobnosti, zaznavno vedenje staršev, zaznavno učiteljevo vedenje, učni uspeh), v kakšni zvezi so kognitivno-metakognitivni in motivacijski vidiki samoregulacije ter kateri vidiki najbolje napovedujejo učni uspeh pri nauku o glasbi (Fritz-Smolej, 2006).

Raziskave, ki se nanašajo na področje vadenja inštrumentov, lahko razdelimo v dve skupini: prva se nanaša na preučevanje strategij vadenja pri otrocih (Mc Pherson in Renwick, 2001), druga pa na raziskovanje motivacijskih in kognitivno-metakognitivnih procesov pri izkušenih glasbenikih (Sloboda in Davidsonova, 1996).

Za poglobljen pristop k raziskovanju vpliva zavedanja metakognitivnih procesov na doživljanje učenja sem se odločila, ker sem v prvem akcijskem ciklu učence spodbujala k samoregulaciji učenja (učenci so postavljali svoje učne cilje, spremljali, regulirali in kontrolirali lastni učni proces), pa se je kljub temu v drugi fokusni skupini pokazalo, da učenci metakognitivnih procesov pri samoregulacijskem učenju niso ozavestili. Ob vprašanju ali so dosegli večji napredek in boljše glasbene dosežke, ker so razvili samoregulacijske spretnosti, so učenci ostali brez besed in se kljub dodatnemu pojasnjevanju in vprašanjem v pogovor o samoregulaciji, niso vključili. Zato sem samoregulacijo v drugem akcijskem ciklu vodila bolj strukturirano.

2 METODA

Kot temeljni raziskovalni pristop sem uporabila metodo akcijskega raziskovanja. Načrtovana in izpeljana sta bila dva akcijska cikla, v katerih so bili učenci deležni poučevanja s poudarkom na razvijanju samoregulacijskega učenja in samoocenjevanja. Fazi akcijskega raziskovanja je sledila analiza rezultatov v obliki multiple študije primera. Ker je šlo za izrazito kvalitativen problem, je bil način dela hermenevtičen, kar pomeni prilagajanje raziskovalnih vprašanj in raziskovalnih pristopov glede na dinamiko dela z učenci. Med in na koncu akcijske raziskave so bili učenci spodbujeni k oblikovanju refleksij, ki so jih zapisali v razvojne delovne zvezke. Na začetku in po vsakem akcijskem ciklu je bil z učenci izveden skupinski intervju v fokusnih skupinah. Ker ni šlo za statistično relevanten, kvantitativen vzorec, je bilo idealno akcijsko raziskovanje, saj so me zanimali načini za optimizacijo modela poučevanja.

2.1 Udeleženci

Vzorec je bil namenski. Pri akcijskem raziskovanju in pri preverjanju uspešnosti modela sem sodelovala v vlogi učiteljice glasbene vzgoje, v 8. razredu osnovne šole v mednarodnem oddelku, v katerega je bilo vpisanih 12 učencev.

2.2 Pripomočki za analizo

Analiza razvojnih delovnih zvezkov, dnevnika in transkripta skupinskih intervjujev v fokusnih skupinah.

Skozi leto so učenci uporabljali razvojni delovni zvezek, ki jih je spodbujal k samostojnemu uravnavanju lastnega učenja, izmenjavi povratnih informacij in samoocenjevanju. Učencem je omogočal razvoj kritičnosti, spremljanje procesa nastajanja glasbenih dosežkov in zapis refleksij. Po vsaki uri sem pisala raziskovalni dnevnik, ki je vseboval različne zabeležke in razmišljanja.

2.3 Postopek

Študija je potekala v dveh fazah – v akcijskem in komparativnem delu.

Akcijski del

Pilotni preizkus modela je bil spremljan z akcijskim raziskovanjem. Akcijsko raziskovanje je omogočilo proučevanje in reševanje problemov, ki sem jih identificirala. Vsak od načrtovanih ciklov naj bi trajal eno polletje, v praksi pa se je izkazalo, da sem za prvi cikel potrebovala en mesec več časa. Na začetku sem definirala akcijske korake, ki so temeljili na proučevanih teoretičnih izhodiščih. Po koncu prvega cikla je bila izvedena evalvacija. Vključevala je kvalitativno analizo zgoraj omenjenih pripomočkov. Pri analizi sem uporabila metodi kvalitativnega kodiranja in klasteringa. Nato so bili določeni akcijski ukrepi za naslednji cikel. V prvem ciklu so bili učenci deležni poučevanja s poudarkom na razvijanju samoregulacijskega učenja, v drugem ciklu pa na razvijanju samoregulacijskega učenja in samoocenjevanja. Raziskovalnim vprašanjem sem na koncu prvega akcijskega cikla dodala vprašanje: »Kašen vpliv ima zavedanje meatkognitivnih procesov pri samoregulacijskem učenju na doživljanje učenja pri glasbeni vzgoji?«. Izveden je bil nov cikel. Na koncu je bila izvedena nova evalvacija. Na osnovi tega je nastala (multipla) študija primera s poskusom izdelave utemeljene teorije.

Komparativna faza

Zaključku akcijske faze je sledila komparativna faza. V tej fazi je bila izvedena primerjava analiz pedagoškega procesa in izidov ciklov. Dinamiki dela obeh ciklov sta bili analizirani vzdolž časovne premice. Izvedeno je bilo opazovanje kvalitativnih razlik med procesom v obeh ciklih s poudarkom na iskanju odgovorov na raziskovalna vprašanja.

3 OPREDELITEV SAMOREGULATIVNEGA UČENJA

Samoregulacijsko učenje je definirano kot način učenja, kjer si učenec sam postavi svoje učne cilje ter nato skuša spremljati, regulirati in kontrolirati lastni učni proces ob upoštevanju lastnih značilnosti in značilnosti okolja (Paris in Paris, 2001; Schunk, 2005). Samoregulirajoči učenci razumejo svoje učenje kot sistematični proces, ki se odvija pod njihovo kontrolo, kot proces, v katerem so metakognitivno, motivacijsko in vedenjsko aktivno udeleženi (Zimmerman, 1994).

Samoregulacija učenja poteka kot krožni proces s tremi fazami (Zimmerman, 1998, str. 3) – predhodno razmišljanje, izvedba in zavestna kontrola in samorefleksija. Faza predhodnega razmišljanja vključuje procese in prepričanja, ki se pojavljajo pred učenjem in predstavljajo pripravo na učenje. Faza izvedbe vključuje procese, ki potekajo med učenjem in pomagajo učencu, da se osredotoči na učenje. Faza samorefleksije vključuje presojanje procesov in rezultatov učenja ter na koncu krožnega procesa vseh treh faz zopet vpliva na prvo fazo – fazo priprave na učenje.

Najvidnejša in razlikovalna lastnost samoregulacijskega učenja je proaktiven pogled na učenje kot proces, usmerjen k posamezniku samemu, v katerem pretvori svoje umske sposobnosti v akademske spretnosti, ki so povezane z nalogami. Učenci se samoregulirajo do tiste mere, ki jim omogoča, da so metakognitivno, motivacijsko in vedenjsko aktivni udeleženci v lastnem učnem procesu. Ti učenci sami ustvarjajo misli, občutke in dejanja, da bi dosegli svoje učne cilje (Zimmerman, 2001). Osnovni elementi proaktivnosti obsegajo zastavljanje lastnih ciljev, vztrajnost in prilagodljivost. Čeprav morda izgleda samoregulacijsko učenje primerno še posebej za izkušenjsko učenje, e-učenje, učenje z odkrivanjem, pa je izredno pomembno tudi pri tradicionalnih oblikah učenja, v vseh učnih aktivnostih v šolskem kontekstu.

4 OPREDELITEV MEATKOGNICIJE

Metakognicija predstavlja kontrolne strukture višjega reda, ki omogočajo posamezniku razumevanje in regulacijo lastne miselne aktivnosti (Demetriou in Efklide, 1989). Flavell je metakognicijo kot znanje in razmišljanje o kognitivnih pojavih opredelil že leta 1979, Kluwe (1982) pa je opredelil dve temeljni značilnosti, ki določata metakognitivne aktivnosti:

- misleči subjekt ima znanje o svojem lastnem mišljenju ter o drugih ljudeh in
- misleči subjekt lahko spremlja in uravnava tok svojega mišljenja.

Večina avtorjev, ki se ukvarja z raziskovanjem metakognicije pojem opredeljuje kot "/.../ védenje o posameznikovi kogniciji; procese; kognitivna in čustvena stanja; sposobnost, da zavestno in namerno spremljamo in uravnavamo svoje znanje, procese, kognitivna in čustvena stanja" (Hacker 1998, str. 11).

V metakognicijo sodijo aktivnosti, ki nam omogočajo, da se uspešno učimo in rešujemo probleme, kot so: načrtovanje svojega dela, nadzorovanje razumevanja gradiva, evalvacija napredka v učenju, izbira strategije reševanja specifičnega problema, poznavanje lastnih spominskih zmožnosti, ocena časa, potrebnega, da se naučimo določeno gradivo, poznavanje, katera učna strategija je učinkovita pri učenju

določene snovi in katera ne, uporaba učinkovite strategije za priklic predhodno zapomnjene informacije ipd. (Tancig, 2004).

Raziskave v novejšem času se v veliki meri posvečajo vprašanju, kakšen natančno je odnos med različnimi komponentami samoregulacije učenja (npr. Sperling, Horward, Stanley in DeBois, 2004). Dosedanje raziskave so ugotovile, da sta metakognitivni komponenti – védenje o kogniciji in regulacija kognicije – med seboj pozitivno povezani. Več znanja vodi v boljšo kontrolo, boljša kontrola pa v konstrukcijo novega metakognitivnega znanja.

Odprto je tudi vprašanje o vlogi metakognicije v modelu samoregulacije. Na osnovi preteklih raziskav se je oblikoval konsenz, da učenci, ki imajo več metakognitivnega znanja in regulirajo učni proces na specifičnem področju, uporabljajo učinkovitejše kognitivne strategije in so boljši pri reševanju problemov (Boekaerts, 1997). Po drugi strani pa Pokay in Blumenfeld (1990) ugotavljata, da je v obdobju spoznavanja novega področja uspeh pri nalogah pozitivno povezan s specifičnimi strategijami, nepovezan s splošnimi kognitivnimi strategijami in celo negativno povezan z uporabo metakognitivnih strategij. Večji del miselnih kapacitet zaseda učenje strategije in zato ostaja le malo prostora za spremljanje tega učenja. Winne (1995) je ugotovil, da metakognitivno spremljanje učenja v zgodnih fazah pridobivanja spretnosti škoduje učenju, zlasti pri manj sposobnih učencih z malo predhodnega znanja. Slika je precej drugačna, ko se učenec že nekaj časa ukvarja z nekim področjem – takrat je uspeh pri nalogah s tega področja pozitivno povezan z uporabo metakognitivnih strategij. Zdi se, da mora novinec oz. slab učenec najprej osvojiti specifične strategije na nekem področju, šele nato pa lahko učinkoviti uporabi metakognitivne strategije (Pokay in Blumenfeld, 1990). To se je pokazalo tudi v moji raziskavi, saj so se manj sposobni učenci pri uporabi metakognitivnih strategij načrtovanja, spremljanja in uravnavanja učenja v prvem učnem sklopu včasih spraševali o smiselnosti samoregulacije učenja. V drugem učnem sklopu, ko so že bili seznanjeni z glasbenim procesom ustvarjanja, tovrstnih vprašanj ni bilo.

5 UGOTOVITVE

V tretji fokusni skupini so učenci v pogovoru o samoregulacijskem učenju, med seboj primerjali učni proces v prvem in drugem akcijskem ciklu. Izpostavili so pomen lastnih spoznanj, do katerih so prišli v obeh ciklih in so bila po njihovem mnenju ključnega pomena za ozaveščanje metakognitivnih procesov pri učenju. Najprej so izpostavili pomen izkušnje iz prvega akcijskega cikla, ko so ustvarjanju glasbenih vsebin lahko posvetili toliko časa, kot se jim je zdelo, da ga potrebujejo in nato spregovorili še o drugem ciklu, v katerem je bil ustvarjalni proces časovno natančneje opredeljen. Učenci so v drugem učnem sklopu dobili učni list z natančno opredeljenimi glasbenimi dejavnostmi, s pomočjo katerega so samostojno načrtovali posamezne korake v procesu ustvarjanja glasbenih vsebin. Raziskava je pokazala, da so učenci v drugem akcijskem ciklu, v krajšem času dosegli bolj kvalitetne glasbene dosežke. Samoregulacijo učnega procesa so doživljali kot področje, ki osmišlja učenje, omogoča doseganje boljših glasbenih dosežkov in jo je mogoče prenesti v svoje življenje. Spregovorili so o pomenu načrtovanja in organiziranja svojega dela, prevzemanja odgovornosti za svoje učenje ter vrednotenja le-tega. O tem govorijo njihovi citati:

Menim, da je bilo dobro, da smo v prvem učnem sklopu imeli toliko časa, kot se nam je zdelo, da ga potrebujemo za ustvarjanje svoje skladbe. To nas je privedlo do spoznanja, da je načrtovanje pomembno. Za nas je zelo pomembno, da sami odkrijemo stvari in pridemo do svojih spoznanj.

Zelo dobro je, da smo prišli do teh spoznanj iz svojih lastnih izkušenj. Če bi se o tem le pogovarjali, oz. če bi o pomenu tega kje prebrali, ne bi bilo dovolj. Sedaj bomo vedeli za vse življenje.

Ko se začneš zavedati in razmišljati o učenju, o tem kaj se dogaja v tvojih mislih, učenje dobi večji smisel. Poleg neke teme se naučiš tudi nadzorovati svoje delo in iskati pot, na katero se želiš podati.

V zadnjem učnem sklopu sem spoznala, da je časovno načrtovanje zares pomembno.

Ko sam reguliraš svoje učenje, še posebej je bilo to dobro, ker sem glasbo ustvarjal sam, ti nihče ne govori, kaj moraš storiti to in kaj naslednjo uro. To mi je bilo zelo všeč. Poskrbel sem, da sem se najprej umiril, globoko vdihnil in razmislil, kaj je tisto, kar želim v določeni uri doseči.

Samoregulacijo pa učenci niso videli le kot področje, ki osmišlja učenje, temveč tudi kot področje, ki ga je mogoče prenesti v svoje življenje, o čemer govori naslednji citat:

Ko imaš sam opravka z načrtovanjem učnega procesa, organizacijo in časovno razporeditvijo svojega dela, učenje dobi večji smisel. Hkrati pa se zaveš, da ti bo vse to prav prišlo tudi kasneje v življenju.

V pogovoru so pozornost namenili tudi refleksiji, vrednotenju in razmišljanju o znanju, ki so ga potrebovali za realizacijo svojih načrtov, pri čemer so še posebej izpostavili prav pomen razmišljanja o znanju. O tem govorita naslednja citata:

Ko sam reguliraš svoje učenje, dejansko razmišljaš o tem, kakšno znanje potrebuješ in kaj se moraš naučiti. Če ti učitelj pove, kaj je to, se sicer to naučiš, vendar pa ne razmišljaš o tem, kako ti bo to služilo pri tvojem delu. Zato v tem ne zaznaš neke uporabne vrednosti.

Jaz pa vidim veliko vrednost predvsem v refleksiji. Ko pogledam nazaj, takrat hkrati gledam tudi v prihodnost in razmišljam, kaj bom naslednjič spremenila.

Zanimivo je, da se je v raziskavi pokazalo, da je učenec, ki je od prvega trenutka dalje vedel, kakšno glasbo bo ustvaril, načrtovanje svojega dela videl kot oviro, o čemer govori naslednji citat:

> *Na začetku drugega učnega sklopa, mi je učiteljica rekla, naj najprej načrtujem svoje delo in šele nato pričnem z ustvarjanjem glasbe v impresionističnem stilu. Povedal sem ji, da mi je to povsem jasno, vendar pa se mi načrtovanje ni zdelo smiselno, ker sem v sebi razločno slišal glasbo, ki sem jo želel ustvariti. In res, brez načrtovanja, sem ustvaril izvrstno skladbo.*

Ko so učenci naredili primerjavo med prvim in drugim akcijskim ciklom ter spregovorili o tem, kaj se je na področju samoregulacije po njihovem mnenju dogajalo, sem ostala brez besed.

6 ZAKLJUČEK

Na koncu šolskega leta sem učence povprašala o morebitnih spoznanjih iz raziskave, ki so se jim zdela še posebej dragocena. Učenci so izrazili navdušenje nad ustvarjanjem glasbenih vsebin, ki omogoča izražanje notranjih doživetij in občutkov, ter znanjem, ki so ga pridobili pri pouku. Učenci so posebno vrednost videli v dejstvu, da se pri pouku niso bali povedati resnice. Veliko jim je pomenilo, da so lahko iskreno vrednotili delo drug drugega. Te izkušnje so bile zanje zelo dragocene in jih bodo ponesli s seboj v življenje. Učenci so izrazili posebno zadovoljstvo tudi ob dejstvu, da je ustvarjanje glasbe lahko zelo zabavno.

Reference

[1] Boekaerts, M. (1997). Self-regulated learning: A new concept embraced by researcher, policy makers, educators, teachers and students. *Learning and Instruction, 7* (2), 161–186.

[2] Demetriou, A. & Efklides, A. (1989). The person's conception of the structures of developing intellect: Early adolescence to middle age. *Genetic, social and general psychology monographs*, *115*(3), 371 – 423.

[3] Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive developmental inquiry. *American Psychologist, 34*, 906–911.

[4] Hacker, D. J. (1998). Definition and empirical foudations. V D. J. Hacker, J. Dunlosky, A. C. Graesser (ur.), *Metacognition in Educational Theory and Practice* (str. 1–23). Mahwah, New Jersey: LEA Publishers.

[5] Kluwe, R. H. (1982). Cognitive knowledge and executive control: Metacognition. V D. R. Griffin (Ed.), *Animal mind -- human mind* (str. 201–224). New York: Springer-Verlag.

[6] McPherson, G. E. in Renwick, J. M. (2001). A Longitudinal Study of Self-regulation in Children's Musical Practice. *Music Education Research, 3*(2), 169–186.

[7] Paris, S. G. in Paris, A. H. (2001). Classroom applications of research on self-regulated learning. *Educational Psychologist, 36,* 89–101.

[8] Pokay, P. in Blumenfeld, P. C. (1990). Predicting achievement early and late in the semester: the role of motivation and use of learning strategies. *Journal of Educational Psychology, 82,* 41–50.

[9] Schunk, D. H. (2005). Self-regulated learning: The educational legacy of Paul R. Pintrich. *Educational Psychologist, 40,* 85–94.

[10] Sloboda J. in Davidson, J. (1996). The young performing musician. V I. Delige in J. Sloboda (ur.), *Musical Beginnings: Origins and development of musical competence.* New York: Oxford University Press.

[11] Smolej-Fritz, B. (2006). Motivacijski, kognitivni in metakognitivni vidiki samoregulativnega učenja pri nauku o glasbi. Doktorska disertacija, Univerza v Ljubljani, Filozofska fakulteta.

[12] Sperling R. A. Howard, B. C., Staley, R. in DeBois, N. (2004). Metacognition and self-regulated learning constructs. *Educational Research and Evaluation, 10*, 11–139.

[13] Tancig, S. (2004). Razvoj metakognicije - naučiti se učiti. Nekaj v pomoč učiteljem: vodnik za poučevanje skupine učencev z učnimi težavami, ki počasneje usvajajo znanja. Marija Kavkler, Nataša Končnik Goršič (ur.), (str. 45–50).

[14] Winne, P. H. (1995). Inherent details in self-regulated learning. *Educational Psychologist, 30,* 223–228.

[15] Zimmerman, B. J. (1994). Dimensions of academic self-regulation: A conceptual framework for education. V B. J. Zimmerman in D. H. Schunk (ur.). *Self-regulation of learning and performance: Issues and educational applications* (str. 3–20). New York: Lawrence Erlbaum Associates.

[16] Zimmerman, B. J. (2001). Theories of Self-Regulated Learning and Academic Achievement: An Overviev and Analysis. V B. J. Zimmerman in D. H. Schunk (ur.). *Self-regulated learning and academic achievement: theoreical perspective.* (str. 1–37). Mahwah: Lawrence Erlbaum Associates.

[17] Zimmerman, B. J. (1998). Developing Self-Fulfilling Cycles of Academic Regulation: An Analysis of Examplary Instructional Models. V D. H. Schunk in B. J. Zimmerman (ur.), Self-Regulated Learning: From Teaching to Self-Reflective Practice (str. 1–19). New York: The Guilfor Press.

KAKO RAZISKOVATI PLACEBO?
(metodološki razmislek)

Urban Kordeš
Univerza v Ljubljani, Pedagoška fakulteta
Kardeljeva pl. 16, Ljubljana
urban.kordes@guest.arnes.si

Povzetek:
Kot kaže lahko smiselno razpravljamo o placebo učinku samo, če sprejmemo dejstvo, da je trivialna (mehanicistična) slika človeka le približek in da za razmislek o placebu ne zadostuje. Treba se bo sprijazniti s tem, da sta pri placebu vzrok in posledica morda na različnih straneh *mind-body* prepada. Še več: pri raziskovanju placeba bi bilo verjetno treba vzeti v poštev možnost, da tu ne gre za proces in/ali učinkovino, ki bi delovala na fiziologijo, ampak da placebo vznikne kot posledica določenega doživljajskega gestalta, ki se zgradi in ohranja znotraj dialoške situacije. Ta gestalt je neponovljiv in individualen. Zaradi tega je bistveno raziskovati prvoosebno doživljajsko pokrajino.

1 UVOD

Znanost se redko sprašuje kakšne so epistemološke osnove njenega početja. Kibernetik Heinz von Foerster je nekoč izjavil: "Bolj temeljen je problem, ki ga ignoriramo, večje so možnosti za slavo in uspeh." Razmišljanje o "velikih" (navadno globokih epistemoloških) problemih je torej recept za blodnjavo kroženje po labirintih neodgovorljivih vprašanj; recept za nepraktičnost. Sodobna znanost dolguje svoj uspeh v veliki meri spretnemu zanemarjanju temeljnih ugank: fiziki so našli pot, kako zaobiti očiten vpliv zavesti na merjenje, psihologi so (z zelo podobno metodo: statistiko) zaobšli intrinzično individualnost posameznikovih doživljajskih pokrajin in tako dalje. Nenazadnje: kognitvni nevroznanstveniki se z ignoriranjem "velikega" oziroma "težkega" problema (odnos med telesnim in duševnim) lahko mirno posvetijo opazovanju delovanja možganov.

Tu in tam pa se primeri, da pojav, ki ga raziskujemo ne dopušča ignoriranja temeljnih vprašanj. Vse bolj jasno se kaže, da je placebo učinek en od takšnih primerov.

2 PLACEBO IN PLACEBO UČINEK

Placebo učinek se navadno opredeljuje kot objektivno (fiziološko merljivo) ali subjektivno izraženo izboljšanje zdravstvenega stanja, ki izhaja iz placebo posega oz. različnih vidikov kliničnega srečanja, pri čemer ni bila uporabljena nobena aktivna učinkovina oz. druga vrsta medicinske terapije.

Navadno ločujemo med placebom in placebo učinkom: placebo naj bi bil poseg ali učinkovina, ki je sprožila placebo učinek. Takšno poimenovanje je zavajajoče, saj predpostavlja, da vemo kaj pripelje do placebo učinka. Način razmišljanja, ki ga je tako uspešno posvojila medicinska znanost (in, ki ga jemlje kot povsem samoumevnega) predpostavlja, da je človekovo delovanje možno modelirati s prenosno funkcijo

$$y = f(x)$$

Izhodi (vedenje, stanje fiziologije itd., y) niso nič drugega kot odziv (f) na vhodne dražljaje (zunanji stimulus, učinkovina itd., x). V našem primeru bi enačba izgledala takole:

$$placebo\ učinek = f\ (placebo)$$

3 TRIVIALNI IN NETRIVIALNI SISTEMI/POJAVI

Tovrsten model omogoča razumevanje, opis, napovedovanje in spreminjanje mnogih delov sveta. Fiziki lahko z njim uspešno pojasnjujejo dogodke v fizičnem svetu (vsaj v velikem delu) in nenazadnje, tudi medicina lahko s tem modelom velik del človeškega stanja. Na tem delu je celo tako uspešna, da – kot je pravilno napovedal von Foerster – ignorira vse ostalo. Mehanicistični pogled na delovanje človeškega telesa omogoča razumevanje, modeliranje in celo napovedovanje širokega spektra procesov. Sodobna medicina, utemeljena na tem modelu, uspešno raziskuje in zdravi množico zdravstvenih problemov. Delovanje človekovega telesa je tako razstavljeno na množico spremenljivk, ki jih lahko manipuliramo s terapevtskimi posegi in/ali učinkovinami.

Vendar pa ta model ne omogoča celostnega razumevanja človekovega stanja, procesov, ki pripeljejo do bolezni in tistih, ki omogočajo spremembe in ozdravitev. Mehanicističen model težko dopušča sistematičen in metodološko veljavni razmislek o procesih zdravljenja, ki ne spadajo v shemo »učinkovina > fiziološka in/ali duševna posledica«. Procesi, ki se začno v duševnosti, so slepa pega takšnega modela. Natančneje: v tem modelu sploh ne obstajajo. Prav tako mehanicistični model zelo težko vključuje komunikacijske in socialne aspekte življenja pacienta (in zdravnika) – aspekte, ki morda bistveno vplivajo na razvoj bolezni oz. napredovanje zdravljenja.

Kaj je tisto, kar mehanicistični model zanemarja?
1. Nekateri pojavi so intrinzično nedeljivi. To pomeni, da edukcija na posamezne spremeljivke ne omogoči več razumevanja pojava. Dva primer: življenje, zavest.
2. Dialoškost oziroma krožnost. Nekaterih pojavov/procesov ne moremo razumeti, če jih opazujemo izolirane od okolja in brez upoštevanja preteklih interakcij z okoljem (osebne zgodovine). Takšni pojavi vzniknejo skozi dialog. Če se dialog prekine, pojav izgine.
3. Že omenjeno v prejšnjem odstavku: doživljanje. Človekovo izkustvo je tipičen primer zgornjih dveh lastnosti: je dialoško in obstaja kot nedeljiv gestalt. Če ga opazujemo, ga s tem spreminjamo.

Iz zgoraj zapisanega je jasno zakaj se večina medicinskih raziskav placeba poskuša osredotočiti na fiziološki del zgodbe. Medicinska metodologija namreč predvideva edino takšne raziskave. Znanih je mnogo raziskav fizioloških procesov, ki se sprožijo ob placebo učinku (dobro povzeto v Kaptchuk, 2008), kar je seveda zelo zanimiva tema, nima pa pretirane zveze z raziskovanja placebo učinka. Ob vse številnejših raziskavah nevrofizioloških korelatov placebo učinka se med raziskovalci placebo učinka krepi zavedanje, da standardni, medicinski metodološki pristopi niso najbolj ustrezni za raziskovanje in razumevanje tega pojava. Ne le to – tudi sama definicija pojava ni zadostna. Standard medicinskih raziskav – dvojno slepi poskusi – temeljijo na uporabi placeba (inertne snovi/posega) kot mehanizma preverjanja delovanja (aktivne) učinkovine/terapije. Ni pa povsem jasno, kako naj bi preiskovali placebo učinek sam zase. Primerjava s skupino, ki ni bila deležna nobenega tretmaja (»no-treatment«), seveda tudi ni zadovoljiva, saj (kot rečeno) sploh ne vemo kaj je tisto kar sproži placebo učinek. To pomeni, da je možno, da vsak angažma pacientov vpliva na potek zdravljenja.

Placebo učinek ni opis določenega fiziološkega sistema ampak je dogodek, kjer se sprožijo fiziološke posledice *brez* (očitnega) fiziološkega vzroka. To pomeni, da je vzročno-posledični princip pojasnjevanja, na katerem temelji znanostven diskurz, nemogoč. Natančneje: nemogoč je vzročno-posledičen način pojasnjevanja, če:
1. se osredotočimo samo na fiziologijo in ne upoštevamo človekovega doživljanja kot možnega vzroka za fiziološke posledice;
2. je naš sistem opazovanja samo individuum (in ne interakcija oz. dialog individuuma z okolico, z zdravnikom) in ne upoštevamo individuumove osebne zgodovine.

Če se lotimo raziskovanja placeba na enak način, kot raziskujemo delovanje klasičnih medicinskih učinkovin/terapij zanemarimo očitno dejstvo, da je izvor placbo učinka v doživljanju in da je bistvo človeškega doživljanja interakcija. V našem primeru interakcija med pacientom in zdravnikom, interakcija okolja in pacienta in vpliv osebne zgodovine (zgodovine interakcij) na doživljajsko pokrajino pacienta. Dialoga in doživljanja ne znamo raziskovati. Zakaj? Ker sta intrinzično netrivialna: opazovalec vpliva na raziskovan proces že s samim opazovanjem. Če bi bil človek trivialen sistem, potem bi veljalo, da na določen dražljaj (barva/oblika tablet, izjava zdravnika itd.) vedno odgovori na enak način (to pomeni, da bi ga lahko opisovali z zgoraj omenjeno prenosno funkcijo $y=f(x)$). Če bi bilo temu tako, bi lahko predpisali ustrezen postopek, ki bi pripeljal do placebo učinka. Ali pa bi kar določili kateri fiziološki sistemi se sprožijo pri placebu in bi s "pravimi" učinkovinami sprožili le-te.

Kot kaže lahko smiselno razpravljamo o placebo učinku samo, če sprejmemo dejstvo, da je trivialna (mehanicistična) slika človeka le približek in da za razmislek o placebu ne zadostuje. Treba pa se bo sprijazniti s tem, da sta pri placebu vzrok in posledica morda na različnih straneh *mind-body* prepada. Še več: pri raziskovanju placeba bi bilo verjetno treba vzeti v poštev možnost, da tu ne gre za proces in/ali učinkovino, ki bi delovala na fiziologijo, ampak da placebo vznikne kot posledica določenega doživljajskega gestalta, ki se zgradi in ohranja znotraj dialoške situacije[1].

4 UPOŠTEVATI NETRIVIALNO

Kako se torej lotiti raziskovanja placebo učinka?

Kot prvo, se je potrebno soočiti z dejstvom, da nas ta tem raziskovalnem polju čakajo "velika" vprašanja: Kako sta povezana doživljanje in fiziologija? Kako raziskovati pojave, na katere vsako opazovanje (raziskovanje) bistveno vpliva?

Pri raziskovanju placebo učinka se je potrebno osvoboditi predpostavk trivialnega vzročno-posledičnega modela. Kot rečeno: ne vemo kaj je "placebo" – ne vemo kaj povzroči placebo učinek. Še več: verjetno sploh ne moremo pričakovati, da bomo odkrili "povzročitelja". Že zdaj je jasno, da je pri nekaterih ljudeh placebo učinek povezan z avtoritativnostjo zdravnika/zdravilca, pri drugih pa s prijaznostjo; da nekateri rabijo vero v metodo, drugi spet ne; itd. Nekaterim se ozdravljenje "zgodi" brez prisotnosti drugih, drugim pa samo ob prisotnosti bizarnih ritualov in ikonografije...

Vse to kaže na nesmiselnost iskanja "vzroka" ozrioma "povzročitelja" placeba. Morda (pa še to pogojno) lahko govorimo o sprožilcu. Vendar z zavedanjem, da je ta od individuuma do individuuma različen.

Pozornost raziskovanja se torej ne bi smela toliko usmerjati v okoliščine ob pojavu placebo učinka. Veliko bolj zanimivo (in obetavno) bi bilo opazovati doživljanje ljudi, ki se jim je zgodilo ozdravljenje. Kateri doživljajski procesi so se

[1] Kot rečeno, vidim pojem "dialoška situacija" v najširšem možnem kontekstu. Kot so-ustvarjalni proces, pri katerem sodelujejo osebna zgodovina, okolica in (predvsem) osebe v okolici.

odvijali, kakšni dialogi in na katerih ravneh so se vrteli? Takšno raziskovanje najverjetneje ne bi privedlo do univerzalne metode sprožanja samozdravljenja. Bi pa morda opolnomočilo posameznike, da bi uvideli kako njihovo doživljanje vpliva na fiziologijo oziroma kako sta njihovo telo in duh povezana. Morda bi skozi raziskovanje doživljanja lahko odkrili kako lahko avtonomno navigirajo po svojih doživljajskih pokrajinah. Morda celo brez potrebe, da jih zdravnik pretenta z lažjo o sestavi tabletke, ki jo je predpisal.

5 VIRI

1) Kaptchuk, T. J., Kelley, J. M., & Conboy, L. A., et al. (2008). Components of placebo effect: randomised controlled trial in patients with irritable bowel syndrome. *British Medical Journal*, 336: 999–1003.
2) Kordeš, U. (2005). *Od resnice k zaupanju*. Studia Humanitatis, Ljubljana.
3) von Foerster, H. (1972). Responsibilities of Competence*, Journal of Cybernetics*, let. 2, št. 2, str.1-6

RAZISKOVANJE NEALOPATSKEGA ZDRAVLJENJA: ZAKAJ, KAKO IN S KAKŠNIM NAMENOM?

Metka Kuhar
Fakulteta za družbene vede
Kardeljeva pl. 5, 1000 Ljubljana, Slovenija
e-mail: metka.kuhar@guest.arnes.si

POVZETEK
V prispevku opredelimo nekaj osnovnih značilnosti modernega medicinskega modela zdravja in zdravljenja ter nanizamo ključne izzive temu modelu, ki izhajajo iz sodobnih družbenih trendov in znanstvenih dognanj. Na podlagi tega pregleda formuliramo pereča raziskovalna vprašanja ter raziskovalne nastavke v smeri nujnosti fokusa proučevanja konstitutivnih vidikov zdravljenja na doživljajska stanja in procese ter z njimi povezane kontekstualne dejavnike.

1 UVOD

Zmanjševanje človeškega trpljenja, doseganje zdravja, dobrega počutja in celostnosti (wholeness) so plemeniti in brezčasni cilji zdraviteljev, ki so v različnih zgodovinskih in kulturnih kontekstih poskušali razvozlati izmikajoče, izjemno kompleksno vprašanje, kako zdraviti. Poznani načini zdravljenja so izjemno raznoliki. Pri tem se že v izhodišču poraja vprašanje, kako sploh razumeti in konceptualizirati zdravje oz. bolezen. V zahodnem svetu je najbolj znano definicijo zdravja postavila Svetovna zdravstvena organizacija, po kateri je zdravje »stanje *popolnega* telesnega, duševnega in socialnega dobrega počutja in ne samo odsotnost bolezni in tegob« (opomba M.K.) (WHO, splet). Na tak celovitejši način so zdravje že v preteklosti definirali in ga tudi danes definirajo mnogi zdravilski sistemi in/oz. poznani posamezniki. Recimo, v prejšnjem stoletju je indijski modrec Aurobindo zdravje definiral kot celokupno bivanjsko stanje, pri čemer naj bi se zdravje merilo s kakovostjo bivanjskega izkustva človeka.

2 KONVENCIONALNI MEDICINSKI MODEL ZDRAVJA IN ZDRAVLJENJA

Kljub temu pa je v sodobni medicini prevladujoč mehanicistični pogled na delovanje človeškega telesa, pri katerem se na zdravje gleda kot na odsotnost znakov in simptomov fizične bolezni. Osnovna načela sodobne medicinske znanosti, pa tudi znanosti na sploh, temeljijo na Descartesovem (1637) razlikovanju med dušo oz. umom (res cogitans) in telesom (res extensa) – pri čemer je le slednje dostopno objektivnemu, znanstvenemu merjenju. Fizično telo se tako proučuje kot sistem, ki ga lahko opredelimo z odnosom med vhodnimi spremenljivkami (učinkovino oz. terapijo) in izhodnimi spremenljivkami (sprememba fizioloških in/ali biokemičnih lastnosti telesa), kar omogoča razumevanje, modeliranje in napovedovanje širokega spektra procesov. Sodobna medicina, utemeljena na tem modelu, uspešno raziskuje in zdravi množico zdravstvenih problemov.

Vendar pa tak medicinski model ne omogoča celovitejšega razumevanja človekovega stanja, procesov, ki pripeljejo do bolezni, in tistih, ki omogočajo spremembe in ozdravitev. Duševni procesi in doživljajska stanja so slepa pega takšnega modela. Zanimivo je, da je na to slepo pego opozoril že Platon pred več kot 2400 leti: »Zdravljenja posameznega dela telesa se ne lotevaj brez zdravljenja njegove celote. Ne zdravi telesa, če obenem ne zdraviš duše. … Velika napaka današnjega časa pri zdravljenju telesa je, da zdravniki ločijo telo od duše«. Mehanicistični model prav tako zelo težko vključuje komunikacijsko-socialni kontekst zdravljenja v smislu interakcij med zdravnikom in pacientom, pacientovega socialnega konteksta itd.

3 SODOBNI IZZIVI KONVENCIONALNEMU MEDICINSKEMU MODELU

Vendar pa lahko znotraj sodobne medicinske znanosti in prakse že zasledimo težnje k celovitejši obravnavi pacientov, s poudarkom na upoštevanju duševne in duhovne dobrobiti pacientov, njihovega življenjska stila, socialnih dejavnikov, terapevtskega odnosa in komunikacije ipd. Kot prvo, krepi se proučevanje pomena komunikacije, interakcije, odnosa, medicinskih »ritualov« v zdravstvenem kontekstu in drugih »mehkih«, kontekstualnih dejavnikov zdravljenja (npr. Ule, 2003), pa tudi proučevanje povezave med duševnimi procesi in zdravjem/boleznijo (recimo, obstajajo že samostojna znanstvenega področja, kot so psihosomatika, psihonevroimunologija itd.); pa tudi med zdravniki narašča zavedanje o pomenu obeh komponent ter zavestno upoštevanje obojega v praksi kljub nenaklonjenim

ekonomskim okvirom sodobnih zdravstvenih sistemov. Kot drugo, vse pogosteje se pojavljajo pozivi k t.i. integralni medicini in tudi medicinske raziskave in prakse (npr. Kligler & Lee, 2004; Schlitz & Amorok, 2005), ki ubirajo celovitejše pristope v smislu posvečanja obsežne pozornosti bolezni, pacientu in zdravniku; ne samo bolezni ali samo bolni osebi.

Poleg tega se krepijo različni trendi, ki demonstrirajo po eni strani nezadovoljstvo z obstoječim konvencionalnim modelom zdravljenja in padajočo družbeno verodostojnost obstoječega zdravstvenega sistema (podobno kot pri drugih modernih sistemih) (Malnar, 2002); po drugi strani pa odsevajo porast proaktivnega (in individualiziranega) zanimanja za zdravje in ukvarjanja z njim, saj zdravje postaja merilo uspešnosti, »pravilnega« življenjskega stila, učinkovitosti itd. Tako beležimo porast trenda t.i. alternativne medicine, tako na ravni različnih zdraviteljske ponudbe; povpraševanja in zaupanja s strani naraščajočega števila uporabnikov (glej recimo Passarelli, 2008); kakor tudi na ravni znanstvenega proučevanja tovrstnih pristopov (kar nakazujejo indikatorji, kot so ustanovitev Office of Alternative Medicine pri ameriškem National Institute of Health-a, ustanovitev in uglednost znanstvenih revij, kot so The Journal of Alternative and Complementary Medicine, Alternative Therapies, Advances itd.). Krepijo se tudi preventivni zdravstveni trendi, kot so wellness, sodobna nutricionistika, fitness in jogijsko gibanje itd., ter trend samozdravljenja na različne načine, recimo z zelišči itd. (glej recimo Klemenc-Ketiš, 2011).

4 DOPRINOS RAZISKAV PLACEBO UČINKA K RAZUMEVANJU ZDRAVLJENJA

Obstoječi konvencionalni medicinski model izzivajo tudi zlasti v zadnjem desetletju vse pogostejše raziskave *placebo učinka*, ki se zdi pomemben dokaz »samozdravilnih« sposobnosti organizma – katere imajo načeloma v konvencionalnem medicinskem modelu slabšalno konotacijo[1]. Kot drugo pa ta fenomen zaradi »zunanjih«, »neučinkovinskih« dejavnikov sprožitve izboljšanja bolezenskega stanja jasno izpostavlja pomen psiholoških, komunikacijskih oz. interakcijsko-odnosnih in socialno-kulturnih – skratka nealopatskih dejavnikov zdravljenja.

Placebo učinek lahko opredelimo kot objektivno (fiziološko merljivo) ali subjektivno izraženo izboljšanje zdravstvenega stanja, ki izhaja iz (fiziološko) »lažnega«, »praznega« posega. Standard medicinskih raziskav za dokazovanje placebo učinka so eksperimenti, kjer se primerja več skupin: skupine, ki je prejela aktivno učinkovino oz. je bila deležna posega; take, kjer sta bila prejeta snov oz. izpeljana procedura inertni (npr. sladkorna tableta ali injekcija fiziološke raztopine); ter skupine, ki ni bila deležna nobene terapije (»no-treatment« skupine za ugotavljanje t.i. spontane remisije bolezni). Večina raziskav placebo učinka je usmerjenih na nevrofiziološke korelate ozdravitve, kjer velja za prelomnico Benedettijevo (2009) odkritje sproščanja endogenih opiatov pri placebo analgeziji, ki je znanstveno kredibilno nakazalo obrise bioloških mehanizmov, na katerih naj bi temeljil preiskovani učinek.

Vendar pa se vse bolj krepi zavedanje, da za razumevanje oz. pojasnitev placebo učinka iskanje nevrofizioloških korelatov nikakor ni dovolj. Pomemben mejnik v raziskovanju placebo učinka predstavlja tako tudi študija Kaptchuka in njegove raziskovalne skupine (2008), izvedena na vzorcu pacientov s t.i. sindromom razdražljivega črevesa, ki je eksplicirala pomen konteksta medosebnega zdravljenja. Omenjena študija je pokazala, da obsežnejša, suportivna, empatična komunikacija med zdravnikom in pacientom, komunikacija, ki spodbuja pacientova pozitivna pričakovanja, in zdravnikovo izražanje zanimanja za pacienta v primerjavi s skopo, okrnjeno komunikacijo (oboje ob placebu tretmaju) prispeva k večjemu izboljšanju simptomov in boljši zaznani kakovosti življenja. Prav tako velja izpostaviti kot prelomno eno od novejših študij istega vodilnega avtorja, tudi na pacientih s sindromom razdražljivega črevesa (Kaptchuk idr., 2010), v kateri so eksperimentalni skupini neprikrito, odprto povedali, da bodo prejeli »placebo tablete, ki vsebujejo inertne substance – kot sladkorne tabletke – in za katere klinične študije kažejo, da signifikantno izboljšajo sindrom razdražljivega črevesa prek telesno-duševnih samozdraviteljskih procesov« (str. 1) ter pri tej skupini dosegli pomembno izboljšanje stanja na več dimenzijah v primerjavi s primerjalno skupino, ki ni bila deležna nobenega tretmaja. V tem kontekstu je zanimiva še karakterizacija placeba kot “the meaning responsa” (odziva na podlagi opomenjanja), ki jo na podlagi pregleda številnih raziskav placebo učinka zagovarja medicinski antropolog Moerman (2002)[2]. To poimenovanje izpostavlja vlogo kulturno specifičnega simboličnega osmišljanja bolezni, postopkov itd., ki obsega med drugim pomene, ki jih pacient

[1] Pri tem se pozablja, da je Hipokrat, ki ga moderna medicina pogosto navaja kot svojega očeta, izjavil: »Vsak ima zdravnika v sebi; moramo mu zgolj pomagati pri njegovem delu. Naravna zdravilna sila v vsakem izmed nas je največja sila pri izboljševanju zdravja.«

[2] Tudi živali lahko manifestirajo placebo učinke, česar ne moremo pojasniti v okviru opomenjanja (glej recimo McMillan, 1999).

pripiše medicinskemu ritualu (celotni proceduri, kontekstu) in odnosu z zdravnikom. Placebo učinek po Moermanu ni proces, ki ga je sprožila inertna substanca, ampak je proces samozdravljenja, ki poteka na podlagi (ne nujno zavestnega) projiciranja samo-zdravilnih sposobnosti v nekaj in/ali nekoga. Ta pogled razširja dosedanje najpogostejše psihološke načine pojasnjevanja placebo učinka kot fenomena zavestnega prepričanja ali nezavednega pogojevanja, da »nekaj« (tableta, postopek ipd.) deluje zdravilno.

5 POZNANI PSIHOLOŠKI, ŽIVLJENJSKO-STILNI IN DRUGI KONTEKSTUALNI DEJAVNIKI ZDRAVLJENJA

Na nefiziološke dejavnike ozdravitev so opozorile tudi številne druge študije, med bolj poznanimi so recimo študije Antonowskega s poudarkom na pomenu splošne življenjske usmeritve (»občutek koherentnosti«) za zdravje in zdravljenje; študije o vplivu spiritualnih občutkov in praks na fizično in mentalno zdravje (glej npr. Hill & Pargament, 2008) itd. Posebej pa velja izpostaviti študijo t.i. »spontanih« remisij rakastega stanja brez alopatskega zdravljenja ali z alopatskim zdravljenjem, ki po medicinskih pričakovanjih naj ne bi bilo dovolj za spremembo (O'Regan, B. & Hirshberg, 1993), v kateri se je izkazalo, da je tovrstna remisija tesno povezana s številnimi duševnimi in socialnimi dejavniki (ter prepletom le-teh); naj jih navedem samo nekaj: sprejetje krize, bolečine, žalosti, obupa, odkritje lastne moči in najdenje novega pomenljivega, izpolnjujočega načina življenja; sprememba preživelih, omejujočih, neustreznih stališč in prepričanj, prevzem nadzora nad lastnim življenjem; sprejemanje in izražanje svojih pozitivnih in negativnih čustev/občutij, želja, potreb, samosprejemanje, sposobnost postavljanja meja; vsaj en močan, ljubeč odnos oz. povezavo z osebo, aktivnostjo, organizacijo; sprememba obstoječih medosebnih odnosov na pozitivni način; motivacija po pomoči drugim itd.

6 ZAKLJUČNI PREMISLEK

Ob izpostavljenem vse večjem zavedanju pomena raznolikih duševnih in kontekstualnih dejavnikov, ki očitno pripomorejo k zdravljenju, da ne govorimo o vsaj navidezni raznolikosti metod, posebej na področju alternativne medicine, se zdi ključno vprašanje interakcije, sovplivanja temi dejavniki ter samozdraviteljskimi in tehnološkimi procesi zdravljenja. Oziroma, če gremo korak naprej, kakšna/katera doživljajska stanja in procesi (stanja zavesti) »sprožajo« zdravljenje?

V osnovi gre seveda za (še vedno) izjemno zagonetno, kompleksno vprašanje, kako pride do (fizične) ozdravitve oziroma kaj konstituira proces (fizičnega) zdravljenja. To vprašanje pa je v zgodovini moderne medicine dandanes aktualno na nov način, tako zaradi naraščajočih legitimizacijskih težav konvencionalne medicine; zaradi (pogosto nekritične in dezorientirane) fascinacije nad »alternativnimi« oblikami zdravljenja in porasta le-teh; zaradi aktivnejše vloge ljudi v ohranjanju zdravja in zdravljenju; in, ne nazadnje, zaradi trenda interdisciplinarnega raziskovalnega pristopa – v tem primeru povezovanja medicine in farmakologije s psihološko, nevro-kognitivno, komunikološko, sociološko-kulturološko, antropološko ipd. vednostjo.

Prispevek zato zaključujem s pozivom k raziskovanju, ki bi nas lahko v teh družbenih in znanstvenih pogojih približalo pojasnjevanju zadanih vprašanj. Glede na pregled literature se zdi smotrna raziskava, usmerjenja na doživljajski vidik procesa zdravljenja ter na različne kontekstualne dejavnike, ki potencialno prispevajo k ozdravitvi, v primerih, ko potencialen rezultat dokazljivo ni povezan z učinkovanjem kemijskih ali drugačnih fiziološko usmerjenih terapij; potencialno izboljšanje zdravstvenega stanja pa je izmerjeno fiziološko, in ne ostaja zgolj na ravni izpovedanega subjektivnega doživljanja.

Literatura

1. Antonovsky A. (1997). Salutogenese. Zur Entmystifizierung der Gesundheit. Tuebingen: DGVT Verlag.
2. Benedetti, F. (2009). Placebo effects: Understanding the mechanisms in health and disease. Oxford: Oxford University Press.
3. Hill, P. C. & Pargament, K. I. (2008). Advances in the conceptualization and measurement of religion and spirituality: Implications for physical and mental health research. Psychology of Religion and Spirituality, 1, 3-17.
4. Kaptchuk, T. J., Friedlander, E., Kelley, J.M., Sanchez, M. N., Kokkotou, E, idr. (2010). Placebos without deception: A randomized controlled trial in irritable syndrome. PLoS ONE, 5, 12: e15591. doi:10.1371/journal.pone.0015591.
5. Kaptchuk, T. J., Kelley, J. M., & Conboy, L. A., et al. (2008). Components of placebo effect: randomised controlled trial in patients with irritable bowel syndrome. British Medical Journal, 336: 999–1003.
6. Klemenc-Ketiš, Z. (2011). Kaj bolnike spodbuja k samozdravljenju in kje so meje? Farmacevtski vestnik, 2, 110-11.
7. Kligler, B. & Lee, R. (2004). Integrative medicine: Principles for practice. McGraw-Hill Professional.
8. Malnar, B. (2002). Sociološki vidiki zdravja. V: Toš, N., Malnar, B. (ur.). Družbeni vidiki zdravja

(str. 3-32). Dokumenti SJM. Ljubljana: FDV, IDV, CJMMK.
9. McMillan, F. D. (1999). The placebo effect in animals. Journal of American Veterinary Medical Association, 215, 7: 992-999.
10. Moerman, D. (2002). Meaning, medicine and the "placebo effect". Cambridge: Cambridge University Press.
11. O'Regan, B. & Hirshberg, C. (1993). Spontaneous remission: An annotated bibliography. Institute of Noetic Sciences. http://www.noetic.org/library/publication-books/spontaneous-remission-annotated-bibliography/
12. Passarelli, T. (2008). Complementary and alternative medicine in the United States. Report. Board on population health and public health practice. www.cwru.edu/med/epidbio/mphp439/complimentary_meds.pdf
13. Schlitz, M. M. & Amorok, T. (ur.) (2005). *Consciousness and Healing: Integral Approaches to Mind-Body Medicine.* St Louis (MO): Churchill Livingston/*Elsevier*
14. Ule, M. (2003). Spregledana razmerja: o družbenih vidikih medicine. Maribor: Aristej.
15. WHO. http://www.who.int/suggestions/faq/en/index.html

PREDTEORETSKI INTUICIJI IN NJUN VPLIV NA RAZISKOVANJE ZAVESTI

Olga Markič
Filozofska fakulteta, Univerza v Ljubljani
Aškerčeva 2, 1000 Ljubljana, Slovenija
e-mail: olga.markic@guest.arnes.si

POVZETEK

V članku sta predstavljeni dve temeljni predteoretski intuiciji, ki se porajata ob pojmu zavest, to sta vzročna in fenomenalna intuicija zavesti. Ti dve intuiciji sta zaznamovali začetek psihologije kot posebne znanosti o duševnih procesih, vplivata pa tudi na sodobne teorije zavesti tako v filozofiji kot v kognitivni znanosti. .

1 UVOD

Zavest je ena od največjih skrivnosti. Po eni strani nam je najbolj domača, saj imamo vsi zavestne izkušnje, po drugi strani pa ne vemo, kako bi zavestno doživljanje uskladili z znanstvenimi spoznanji o svetu. Ko skušamo odgovoriti na vprašanje "Kako lahko vznikne zavest iz nevronskih procesov v možganih?", se znajdemo v zadregi T.H. Huxley je leta 1866 to zadrego ponazoril z zdaj zelo znanim komentarjem: "*Kako lahko nekaj tako izrednega kot je stanje zavesti nastane kot posledica draženja živčnega tkiva, je prav tako nedojemljivo, kot je nedojemljiv pojav duha, potem ko Aladin podrgne po svoji svetilki.*"

Ko se sprehajatmo po gorski stezi in opazujete cvetoče travnike, poteka v vaših možganih vrsta procesov. Fotoni zadanejo retino in električni impulzi potujejo po optičnih živcih in med različnimi predeli možganov. Vsi ti procesi so zelo zapleteni in jih še ne poznamo do potankosti, vendar so relativno razumljivi in jih lahko raziskujemo z objektivnega stališča znanstvenika. Kaj pa subjektivni vidik? Ko pogledamo nežno cvetlico doživljamo barve in vonje, ki tako vstopajo v naše notranje psihično življenje. Morda to vzbudi čustva, izzove misli, spomine... Vsa ta doživetja tvorijo zavest, ki nam je neposredno dostopna.

V filozofskih in psiholoških razpravah lahko razpoznamo več različnih vidikov oziroma komponent zavesti. V središču razprave so tako enkrat ene in drugič druge značilnosti zavesti, kar razumevanje problema le še otežuje. V nadaljevanju se bom osredotočila predvsem na dve intuiciji zavesti, ki pomembno oblikujeta razmišljanja in raziskovanja zavestnih procesov. Pomagata nam razumeti, zakaj pogosto prihaja do nesporazumov in ekvivokacij.

2 POJEM ZAVESTI

Težave pri raziskovanju se začnejo že pri opredelitvi pojma zavesti. Tako je včasih poudarjen družbeni vidik, skupno vedenje, ki ga ima določena skupnost ljudi (npr. razredna zavest pri marksistih), drugič pa zavest uporabljamo predvsem v psihološkem pomenu, ki se nanaša na posameznike. V nadaljnji razpravi nas bo zanimal ta drugi vidik, zavest posameznika, brez posebnih etičnih ali političnih podtonov. Vendar je zavest, tudi če se omejimo na slednje, izraz z veliko različnimi pomeni. Kot ugotavlja ameriški psiholog Miller (1962), pomeni zavest, odvisno od izbrane figure govora, lahko stanje, proces, mesto, epifenomen, emergentno lastnost snovi, ali edino pravo resničnost.

Zavest pri človeku bi na primer lahko opredelili na naslednja načina:

(1) stanje biti zavesten kot pogoj za vse misli, občutke in voljo,

(2) stanje biti zavesten glede na običajne pogoje zdravega budnega življenja.

Prva je tranzitivna zavest, zavest o nečem. Gre za zavest kot intencionalno zavest. Druga pa je intranzitivna zavest, ki je bolj temeljna in je pogoj za prvo[1].

Gornja razdelitev je samo ena izmed mnogih, saj izraza "zavest" in "zavesten" zajemata širok spekter duševnih pojavov in se uporabljata v različnih pomenih in dometih. Robert van Gulick v preglednem prispevku o zavesti razlikuje med zavestjo bitja, kjer se nanašamo na celoten organizem, in zavestjo stanja, kjer se nanašamo na določeno duševno stanje ali process (van Gulick, 2010). Za vsako od teh dveh opredelitev pa obstajajo še nadaljnje bolj podrobne različice pomenov.

Menim, da je za razumevanje tako filozofskih razprav o zavesti kot za poskuse znanstvenega raziskovanja pomembno

[1] razen če ne zahtevamo, tako kot na primer fenomenologi, da je zavest vedno zavest o nečem.

prepoznati dve temljni predteoretski intuiciji, ki sta oblikovali metode in razprave o raziskovanju zavesti v zadnjih sto letih

3 PREDTEORETSKI INTUICIJI O ZAVESTI

Güzeldere je intuiciji povzel v naslednjih dveh trditvah (Güzeldere, 1995, str. 36):

1. Zavest je to, kar zavest počne.
   (Consciousness is as consciousness does)

2. Zavest je to, kar se zavesti dozdeva.
   (Consciousness is as consciousness seems).

Pri prvi gre za vzročno intuicijo, ki kot temeljno za naše duševno življenje jemlje vzročno vlogo zavesti. Pri drugi, fenomenalni intuiciji, pa je poudarek na tem, kako občutimo naše duševno življenje, t.j. za pojavne kvalitete, ki označujejo naše zaznave, bolečine in druga duševna stanja.

Ti dve intuiciji pogosto vlečeta vsaka v svojo smer. Zato se zgodi, da tisti, ki gradijo na osnovi ene intuicije, zanemarijo drugo in ne zajamejo njenih značilnosti. Güzeldere je prepričan, da diskusije pogoste zaidejo v slepo ulico zaradi še ene, po njegovem mnenju napačne intuicije. To je intuicija bistva (esencializma). Ta intuicija govori, da če označujemo zavest kot bistveno vzročno, potem mora biti bistveno ne-fenomenalna in obratno, če je bistveno fenomenalna, potem je bistveno ne-vzročna. Güzeldere sam se zavzema za nasprotno intuicijo, po kateri to, kar zavest počne, ne moremo določiti, če je odsotna fenomenalna zavest, oziroma še bolj pomembno, kako se zavesti dozdeva ne moremo pojmiti ob odsotnosti tega, kar zavest počne (Güzeldere, 1995, str. 36).

Prevlada ene ali druge intuicije je pomembno oblikovala pristop k raziskovanju duševnih procesov. Ko se je na prelomu 19..in 20. stoletja psihologija osamosvajala od filozofije, je takrat vodilna nemška psihološka šola poudarjala introspekcijo kot temeljno znanstveno metodo psihologije. Njihova temeljna hipoteza je bila, da je psihologija raziskovanje "fenomenologije" človekove duševnosti, zato so si prizadevali podati popoln opis zemljevida duševnosti, kot se prikazuje subjektu. Njihov cilj je bil podati "atome" osnovnih čutnih vtisov, ki jih je mogoče razlikovati. Na ta način naj bi po vzoru kemije zgradili nekakšno tabelo atomov človeške duševnosti. Najbolj znani šoli in laboratorija za introspekcijske raziskave sta bila Cornell šola v Ithaci (ZDA), ki jo je vodila Titchener, in Würzburška šola, ki jo je v Leipzigu (Nemčija) vodil Külpe. Problem tega pristopa je bil, da so bili zbrani podatki o "čutnih atomih" odvisni od občutljivosti in sposobnosti "introspekcionistov". Laboratorija sta tako navajala precej različne rezultate, kar je zbujalo dvom v objektivnost raziskav. Poleg tega pa je manjkala tudi splošno sprejeta metoda za falsificiranje rezutatov.

K zatonu introspekcionizma pa je poleg "notranjih" vzrokov pripomogla predvsem sprememba v raziskovalni klimi v Evropi in ZDA. Prevladovati je začela pozitivistična naravnanost tako v znanosti kot v humanistiki. Psihologi, ki so želeli pokazati svojo znanstveno naravnanost in se dokončno odtrgati od filozofije, so se začeli izogibati vsemu "duševnemu" in se skušali približati naravoslovnim znanostim. Zavest je bila izrinjena iz psiholoških raziskav, prav tako pa tudi metoda introspekcije, ki naj bi bila neznanstvena in preveč obtežena z vprašljivo metafiziko. O zavesti in njenem mestu v behavioristični psihologiji veliko povejo besede enega najvidnejših predstavnikov te smeri, J. Watsona: "Behaviorizem trdi, da zavest ni jasen in uporaben pojem. Behaviorist, ki se je izšolal kot eksperimentator meni, da izvira prepričanje v obstoj zavesti iz starodavnih časov praznoverja in magije. (Watson, Behaviorism, cit. po Güzeldere, 1995, str. 39) Ker izraz zavest ne nastopa v naravoslovnih znanostih, mu ni mesta niti v znanstveni psihologiji. Poleg metodološke predpostavke, da mora psihologija kot znanstvena disciplina preučevati le tisto, kar je mogoče raziskovati z opazovanjem, je šel behaviorizem še dlje. Kot temeljno predpostavko je sprejel prepričanje, da tudi sam pojav, ki ga psihologija preučuje, ni nič drugega kot opazovano vedenje.

Prav kritika behaviorizma je v drugi polovici 20. stoletja spodbudila nastanek nove raziskovalne paradigme, kognitivne psihologije. Psihologi so začeli raziskovati spoznavne (kognitivne) procese, ker so menili, da so ravno ti procesi odločilni za razumevanje človekovega vedenja. Temeljna hipoteza je postalo prepričanje, da je človekovo spoznavanje nekakšen informacijski sistem, ki dobiva informacije iz okolja, jih obdeluje, rezultati pa se kažejo v njegovem vedenju. Pri raziskovanju so pogosto (ne pa vedno) imeli pomembno vlogo računalniški modeli. Z njihovo pomočjo so skušali razložiti delovanje spomina, učenje, reševanje problemov, itd. Osrednje metodološko orodje je postala funkcionalna analiza in psihologi so si tako kot računalničarji pomagali z risanjem diagramov poteka.

Čeprav je bil tak pristop uspešen pri razlagi mnogih kognitivnih procesov, se je sčasoma izkazalo, da je za celovitejše razumevanje duševnosti potrebno vključiti tudi zavest. Pokazalo se je, da je prav zavest nerazjasnen pojav v kogntitivni paradigmi, zato so ponovno vzklila stara vprašanja v zvezi z zavestjo, ki so bila zakopana v dobi behaviorizma.

Opozoriti bi želela še na en vidik,, ki je prav tako odigral pomembno vlogo pri našem razumevanju zavesti. Gre namreč za raziskovanje nezavednega. Pri raziskovanju zavesti se kot pomembno kaže razlikovanje med zavestnim in nezavednim. Angleški psiholog Johnson – Laird meni, da je prav delitev med zavestnimi in nezavednimi procesi najboljši ključ za razumevanje strukture naše duševnosti. Za Descartesa je bila zavest bistvena sestavina duševnosti, biti "zavesten" je pomenilo, da se zavedaš svojih lastnih duševnih stanj. Ta pogled je prevladoval vse do Freuda, ki je prvi sistematično, znotraj svojega teoretskega okvirja, pokazal na vlogo nezavednega. Pri Freudu nezavedno obstaja zaradi preteklih dogodkov, ki jih lahko razložimo z mehanizmi potlačitve.

Nezavedno je sicer nedostopno za običajno introspekcijo, vendar pa ni *v principu* nedostopno. Dostop do zavesti je omogočen s posebno terapijo, psihoanalizo. Po drugi strani pa danes govorimo o "kognitivno nezavednem", ki obstaja zaradi same narave mehanizmov našega zaznavno-spoznavnega sistema in leži v principu zunaj našega zavednega dostopa. V sodobni kognitivni psihologiji je raziskovanje nezavednih duševnih procesov, ki so osnova za višje kognitivne procese, zelo pogosto in prinaša zanimive rezultate. Spomnimo se le na pravila Chomskijeve gramatike, na računske mehanizme, ki omogočajo vid (Marr) in na procese pri reševanju problemov in načrtovanju. Te raziskave kažejo, da je za razumevanje duševnosti prav tako kot raziskovanje zavesti pomembno raziskovanje nezavednih procesov.

4 RAZISKOVANJE ZAVESTI V SODOBNI FILOZOFIJI DUHA IN KOGNITIVNI ZNANOSTI

V sodobnih filozofskih in psiholoških razpravah lahko razpoznamo več različnih vidikov oziroma komponent zavesti. Zelo odmevno analizo zavesti je prispeval ameriški filozof Ned Block (1995). Najprej opredeli dva pojma zavesti:

Fenomenalna zavest (*Phenomenal consciousness*):
Duševno stanje *D* je fenomenalno zavestno zaradi fenomenalno zavestnih lastnosti, kot so "načini, kako občutimo doživljaje", "kvalitativni značaj", "qualia", "neposredne fenomenološke kvalitete", itd. Gre za subjektivno zavedanje, občutenje, doživljanje, kot so npr. doživetja ob gledanju, poslušanju, dotikanju, vonjanju, itd.

Zavest kot dostop (*Access consciousness*):
Duševno stanje *D* je zavestno dostopno, če so zadoščeni naslednji trije pogoji, ki so skupno zadostni, niso pa vsi nujni:
vsebina *D* je taka, da jo lahko uporabimo
(1) kot premiso v razmišljanju,
(2) za racionalni nadzor govora,
(3) za racionalni nadzor dejanja.
Pojem zavest kot dostop uporabimo za tista stanja, v katerih se ravnokar nahajamo, lahko pa tudi za spomine, na primer, prepoznavanje obraza, spominjanje preteklega dogodka.

Block je mnenja, da gre za dve različni vrsti zavesti, ki ju ne smemo zamenjevati in ki nista zvedljivi ene na drugo. Nato doda še dve kognitivni pojmovanji zavesti, ki sta različni od zgoraj omenjenih. To sta:

Samozavedanje (*Self-consciousness*):
Samozavedanje vsebuje pojem jaza in zmožnost uporabe tega pojma v razmišljanju o samem sebi. Kot primeren test samozavedanja mnogi štejejo prepoznavanje samega sebe v ogledalu, česar so sposobni nekateri višji primati in otroci, starejši od leta in pol.

Zasledovanje (*Monitoring*):
Gre za notranje sledenje, ki ima lahko več oblik. Lahko je to notranje zaznavanje fenomenalnih stanj ali jaza, nadzorovanje zaznavnih informacij ali pa gre za metakognitivni pojem o mislih višjega reda, kjer stanje spremlja misel, da je nekdo v tem stanju.

Pojma samozavedanja in zasledovanja sta lahko različna, na primer, ni nujno, da vsako zasledovanje vsebuje zavedanje o jazu. Vendar večina samozavedanja vsebuje zasledovanje tega, kar nekdo dela.

Seveda bi natančnejša analiza pokazala, da je tudi znotraj teh štirih kategorij potrebno razlikovati različne oblike zavesti. Tako se zdi povsem jasno, da so znotraj fenomenalne zavesti razlike med fenomenalno zavestjo pri gledanju, pri poslušanju, pri čustvovanju, itd. Prav tako je pri zavesti kot dostopu razlika, na primer, med tem, da se spomnimo zagovora diplome ali prepoznamo Ravelov *Bolero*.

Precej očitno je, da se Block v svoji analizi oprira na predteoretski intuiciji. Zavest kot dostop in v veliki meri tudi zasledovanje sta po naravi funkcionalni in ustrezata temu, kar smo opisali kot vzročna predteoretska intuicija. Fenomenalna zavest pa se, kot pove že ime, opira na fenomenalno intuicijo.

Predteoretski intuiciji zavesti sta oblikovali tudi znamenito razdelitev na “lahke” probleme zavesti in na “težak” problem zavesti, ki jo je prispeval ameriški filozof David Chalmers (1995). “Lahki” so ti problemi imenovani zato, ker so takšne narave, da se jih lahko loteva kognitivna znanost, predvsem kognitivna psihologija in nevroznanost. Raziskovalci iščejo odgovore na vprašanja kot so: Kako človek razlikuje čutne dražljaje in na njih reagira? Kako možgani usklajujejo informacije iz različnih izvorov in uporabijo te informacije za nadzor vedenja? Na kakšen način oseba ubesedi svoja notranja stanja? Pri tem seveda ne misli, da gre za dejansko lahke in zanemarljive probleme, vendar po njegovem obstaja realno upanje, da jih je mogoče razrešiti v nekam doglednem času.

Različne teorije zavesti, kot sta na primer Baarsova teorija globalnega delovnega prostora (the global workspace theory) (Baars, 1988) in Dennetova teorija mnogoterih osnutkov (multiple drafts model) (Dennett, 1991), so se omejile predvsem na raziskovanje funkcionalne vloge in vzročne povezanosti različnih duševnih / nevronskih stanj in procesov. Nevrobiološki pristop, ki ga zagovarjata Crick (1994) in Koch (2004), pa raziskuje nevrološke kolerate zavesti. Po Chalmersovi razdelitvi bi se vse naštete teorije ukvarjale z “lahkimi” problemi zavesti.

A tak pristop, ki se opera zgolj na vzročno intuicijo, ki je prevladovala in še prevladuje v psiholoških in nevroznanstvenih raziskovanjih, ne more v celoti zaobjeti pojma zavesti. Po Chalmersovem mnenju ne more dati zadovoljivega odgovora na najbolj zanimivo vprašanje iz uvodnega dela tega prispevka: “Kako lahko vznikne zavest iz nevronskih procesov v možganih?”. Zavest je v tem primeru

razumljena na osnovi fenomenalne intuicije. Tej intuiciji so zavezani predvsem filozofi, ki izhajajo iz fenomenološke tradicije (Brentano, Husserl, Merleau-Ponty) in naprej razvijajo fenomenološko raziskovanje.Vendar si vprašanja o kvalitativnih vidikih zavesti in o prvoosebnem raziskovanju doživljanja zastavljajo tudi bolj analitično usmerjeni filozofi, ki se ukvarjajo s filozofijo v kognitivni zanosti. .
(Gallagher, Zahavi, 2008).

Vprašanje, kako lahko fizični proces v možganih privede do subjektivnih doživetij, je ravno tisto, kar po Chalmersu predstavlja "težak" problem zavesti. Ta je po njegovem mnenju prava skrivnost duševnosti in je tisto, kar dela problem telesa in duha tako težak. Niti vzročne teorije niti fenomenološka analiza naj ne bi mogle preseči razlagalne vrzeli, ki se razpira med subjektivnim doživljanjem in tretjeosebnim raziskovanjem (Levine, 1983).

5 ZAKLJUČEK

Vzročna intuicija, ki je prevladovala in še prevladuje v psiholoških raziskovanjih, ne more v celoti zaobjeti pojma zavesti.. Posledično se tudi vsem teorijam zavesti, ki temeljijo na njej, izmika celovit vpogled v zavest. Vprašanje je, ali je kakšno upanje, da bi lahko oblikovali teorijo, ki bi integrirala tako vzročno kot fenomenalno intuicijo in odstranila potrebo bo posebnem "težkem" problemu.. Nickolas Humphrey (2011) v svoji najnovejši knjigi *Soul dust: The Magic of Consciousness* predlaga zanimivo rešitev.

Literatura

Block, N. (1995). "On a confusion about a function of consciousness", *Behavioral and Brain Sciences* 18 (2): 227-287.

Chalmers, D. (1995). "Facing up to the problem of consciousness", *Journal of Consciousness studies,* 3 (1), str. 200-219.

Crick, F.C. (1994). *The Astonishing Hypothesis. New York: Charle Scribners' Sons*
.

Dennett, D. (1991). *Consciousness explained,* Boston: Little & Brown.

Gallagher, S., Zahavi, D. (2008). *The Phenomenological Mind: An Introduction to Philosophy of Mind and Cognitive Science.* London, New York: Rautlage.

Güzeldere, G. (1995). " Consciousness: What it is, how to study it, what to learn from it's history", *Journal of Consciousness studies,* 2 (1), str. 30-51.

Humphrey, N. (2011). *Soul dust: The Magic of Consciousness*. Princeton, Oxford: Princeton University Press.

Koch, C. (2004). *The Quest for Consciousness,* Engelewood: Roberts and Company Publishers.

Levine, J. (1983). "Materialism and qualia: the explanatoy gap", *Pacific Philosophical Quarterly* 64, str. 354-361.

Miller, G. (1962). *Psychology: The science of mental life*, Harper & Row, New York.

Van Gulick, Robert, "Consciousness", *The Stanford Encyclopedia of Philosophy (Summer 2011 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/sum2011/entries/consciousness/>.

UVID V BOLEZEN IN NJEGOV POMEN ZA ZDRAVLJENJE PRI PACIENTIH Z BIPOLARNO MOTNJO RAZPOLOŽENJA

Tatjana Novak, Lilijana Šprah
Družbenomedicinski inštitut, Znanstvenoraziskovalni center SAZU
Novi trg 2, 1000 Ljubljana, Slovenija
Tel: +386 1 470 6445; fax: +386 1 426 1493
e-pošta: tatjana.novak@zrc-sazu.si

POVZETEK
Proučevanje uvida je pomembno z diagnostičnega in prognostičnega vidika. Okrnjen uvid je pogosto prisoten pri bipolarni motnji razpoloženja (BMR) in se povezuje z vrsto kliničnih dejavnikov poteka in izida. Raziskave kažejo, da je uvid pri BMR okrnjen predvsem v fazah manije in se izboljša z umikom maničnih simptomov. Prisotnost psihoze v katerikoli afektivni epizodi se povezuje s slabšim uvidom. Stopnja uvida se je pokazala kot neodvisna od trajanja in poteka BMR. Ohranjenost uvida je vsaj delno odvisna od ohranjenih izvršitvenih funkcij frontalnega režnja, čeprav povezava med uvidom in kognitivnimi rezultati ni konsistentna. Povečanje uvida pri pacientih BMR je mogoče doseči z vključevanjem svojcev v proces diagnostike, s poudarki na načinih spoprijemanja s stigmo in kognitivnih treningih v psihosocialnih intervencah in z dopuščanjem alternativnih razlag za nastanek in zdravljenje bolezni pri posameznikih, ki niso nujno v skladu z biopsihosocialnim modelom.

1. OPREDELITEV UVIDA

Samozavedanje odraža zmožnost posameznika, da se lahko osredotoči, kodira in prikliče informacije, povezane z njegovimi zmožnostmi, potezami in stališči, ki vodijo njegova vedenja, izbire in socialne interakcije (Johnson in sod., 2002). Z nevrološkega vidika funkcije samozavedanja podpirajo (dorzolateralna) frontalna kortikalna in parieto-temporalna področja (Amador in sod., 1994; Varga in sod., 2006; Orfei in sod., 2008).
Čedalje več raziskovalne pozornosti je namenjene okrnjenenemu samozavedanju, saj spremlja različna zdravstvena stanja, kot so možganska kap, travmatske poškodbe glave, shizofrenijo in motnje razpoloženja. Nezavedanje bolezni odraža nezmožnost prepoznavanja lastnega stanja kot bolezenskega in nezmožnost pripisovanja pravilnega pomena primanjkljajem, simptomom in posledicam, ki jih to stanje pušča na posameznikovem funkcioniranju (Orfei in sod., 2008).
Na podlagi kliničnih opisov pacientov iz 19. in začetkov 20. stoletja, ki so bili nezmožni prepoznati lastno delirantno stanje ali nevrološke simptome bolezni (npr. hemiparezo, afazijo), so izšli izrazi, ki nakazujejo na pomanjkljivo zavedanje, kot je npr.

anozognozija, ki v sodobni rabi označuje stanja posameznika, ki zaradi možganske poškodbe ne zmore (primerno) prepoznati resnosti svojih okvar na zaznavnem, gibalnem, čustvenem ali miselnem področju, ki jih sicer opazijo zdravniki in svojci (Orfei in sod., 2008). Po drugi strani pa se termin *uvid* nanaša na zavedanje duševnih motenj. Po Jaspersu (Jaspers, 1963; v Orfei in sod., 2008) obstaja razlika med *zavedanjem bolezni*, tj. občutek biti bolan in spremenjen, od *uvidom*, ki nakazuje na oceno narave in resnosti obolenja (Baier in sod., 1998). V literaturi izstopata vsaj dve uveljavljeni definiciji uvida v bolezen. David (1990) tako predvideva tri medsebojno prekrivajoče se dimenzije vpogleda: zmožnost opredeliti določene duševne dogodke kot patološke, prepoznati pri sebi duševno motnjo, prepoznati potrebo po zdravljenju in sprejeti zdravljenje. Druga uveljavljena multidimenzionalna definicija uvida Amadorja in sodelavcev (1994) poleg zavedanja simptomov vključuje še razloge nastanka simptomov (atribucijo). Te dimenzije uvida se medsebojno prekrivajo in so dinamične, saj se v teku bolezni lahko spreminjajo (Cuesta in sod., 2006).

2. POMEN EMPIRIČNEGA PROUČEVANJA UVIDA

Na relevantnost proučevanja vpogleda pri psihiatričnih motnjah opozarjajo številni empirični izsledki, ki okrnjen uvid povezujejo s slabšim sprejemanjem zdravljenja (Yen in sod., 2007; Staring in sod., 2009), slabšim psihosocialnim delovanjem (Yen in sod., 2008), redkejšim obiskom storitev na področju duševnega zdravja (Staring in sod., 2009), in s samomorilnim vedenjem (Yen in sod. 2007). Vendar ima dober uvid tudi temne plati, kajti v raziskavah ga povezujejo med drugim z višjo stopnjo disforičnega razpoloženja, obupa, slabšo samopodobo, manjšim blagostanjem in slabšo kvaliteto življenja (Staring in sod., 2009). Neugodne učinke ohranjenega uvida lahko razložimo s posrednim vplivom stigmatizirajočih prepričanj, ki deluje na povezavo dobrega uvida z depresivnim razpoloženjem, negativno samopodobo in slabšo kvaliteto življenja (Staring in sod., 2009), saj je ta povezava pomembna le, če so pri pacientih prisotna ponotranjena stigmatizirajoča prepričanja o duševni motnji (Staring in sod., 2009).
Proučevanje uvida ima diagnostično vrednost z vidika pravilne postavitve diagnoze in prognostično vrednost (Amador in sod., 1994) z vidika napovedi poteka motnje in uspešnosti sodelovanja v zdravljenju in rehabilitaciji. S konceptom uvida se močno povezuje tudi koncept mentalne nesposobnosti (Cairns in sod., 2005), ki se nanaša na nezmožnost za sprejemanje odločitev v zvezi z

zdravljenjem. Slednja pa korelira s potrebo po prisilni hospitalizaciji, s prisotnostjo psihoze, in s tipom diagnoze, posebno tistim s slabšo prognozo (npr. shizofrenija) (Owen in sod., 2011).

3. RAZISKAVE UVIDA PRI PACIENTIH Z BIPOLARNO MOTNJO RAZPOLOŽENJA (BMR)

Ob primerjavi stopnje uvida različnih kliničnih skupin se je pokazalo, da so imeli pacienti z veliko depresijo najboljši uvid, po stopnji uvida so jim sledili pacienti z bipolarno depresijo in mešano manijo (ki obe vključujeta depresivne simptome v diagnostične kriterije) (Dell'Osso in sod., 2002). Ko so primerjali uvid pri pacientih z BMR glede na paciente s shizofrenijo, se je pokazalo, da je bila stopnja uvida v akutni fazi primerljiva med skupinama (Pini in sod., 2001), v fazi remisije pa so imeli pacienti s shizofrenijo slabši uvid v bolezen in njene socialne posledice kot pacienti z BMR, čeravno bi lahko na omenjene razlike v uvidu vplivale starostne razlike med skupinama (Braw in sod., 2011). Ti izsledki potrjujejo tudi prejšnje raziskave, in sicer, da je okrnjenost uvida prehodno stanje pri BMR, ki je vezano predvsem na manično fazo in se izboljša ob umiku akutne simptomatike, medtem ko gre pri shizofreniji in unipolarni depresiji za bolj stabilen fenomen, saj je pri prvi uvid konstantno okrnjen, pri drugi pa je ohranjen in se celo poveča ob poslabšanju razpoloženja (Ghaemi in Rosenquist, 2004; Pini in sod., 2001). Nasplošno je narejenih manj raziskav o uvidu pri BMR glede na raziskave pri shizofreniji ali unipolarni depresiji, čeprav je okrnjen uvid pogost fenomen pri BMR (Cassidy, 2010) in predstavlja napovedni dejavnik sodelovanja v zdravljenju (Yen in sod., 2007).

Glede na tip afektivne epizode pri BMR se je pokazalo, da imajo pacienti v maniji slabši uvid kot pacienti v depresiji, mešani fazi ali v remisiji (Peralta in Cuesta, 1998; Yen in sod., 2007; Cassidy, 2010) in da je prisotnost psihoze ne glede na tip afektivne epizode povezana s slabšim uvidom (Peralta in Cuesta, 1998; Cassidy, 2010). Psihomotorični nemir in razdražljivost sta bila tista vidika manije, ki sta napovedovala okrnjenost uvida, medtem ko se hedonistični vidiki manije (evforija, grandioznost, hiperseksualnost in humor) niso izkazali kot pomembni prediktorji (Cassidy, 2010). To je skladno z ugotovitvami, da je raven disforičnega oz. depresivnega razpoloženja boljši napovednik prizadetega uvida kot pa evforično razpoloženje (Ghaemi in Rosenquist, 2004; Cassidy, 2010). Še več, depresivni simptomi so se pokazali kot najpomembnejši prediktor ohranjenega uvida ne glede na diagnozo (unipolarno veliko depresijo, shizofrenijo ali BMR) (Peralta in Cuesta, 1998; Cassidy, 2010).

Raziskave na pacientih z BMR tudi kažejo, da se uvid verjetno nikoli povsem ne povrne kljub kliničnemu okrevanju (Varga in sod., 2006; Yen in sod., 2007). V eni od raziskav je tako 47% pacientov z BMR v remisiji je ohranilo slabši uvid v primerjavi s 94% pacientov z ugotovljenim slabim uvidom v akutni fazi BMR (Varga in sod., 2006). Druga raziskava pa je opozorila na neugodni učinek ponavljajočih se maničnih epizod na ohranjenost uvida, saj pacienti s ponavljajočimi se maničnimi epizodami po dveh letih niso dosegli takšne ravni uvida kot ob začetku raziskave, medtem ko se je uvid povrnil na izhodiščno raven pri pacientih z eno manično epizodo kljub poslabšanju v manični epizodi (Yen in sod., 2007). Uvid se je torej slabšal s ponavljajočimi se maničnimi epizodami, ne pa tudi s ponavljajočimi se depresivnimi epizodami.

Prav tako je videti, da sta uvid in resnost bolezni dva samostojna konstrukta in da uvid ni zgolj sekundarna mera resnosti bolezni. Okrnjenost uvida namreč ni neposredna vezana na prisotnost afektivnih ali psihotičnih simptomov, poleg tega se je uvid pokazal kot neodvisen od trajanja in poteka bolezni, saj ni koreliral z longitudinalnimi kazalniki resnosti bolezni (trajanje motnje, število hospitalizacij) (Cassidy, 2010), niti z nekaterimi psihosocialnimi dejavniki, kot je starost pacientov (Dias in sod., 2008b). Izkazalo se je namreč, da starejši pacienti z BMR in tisti z več hospitalizacijami niso imeli boljšega uvida, čeravno bi lahko pričakovali, da bodo imeli zaradi daljše in ponavljajoče se izkušnje bolezni več znanja o bolezni in njenih posledicah. Višja izobrazba se je pokazala kot pomemben dejavnik ohranjenega uvida, verjetno zaradi večjega dostopa in izkoristka informacij (Dias in sod., 2008b).

Ohranjenost uvida je vsaj delno odvisna od ohranjenih izvršitvenih funkcij frontalnega režnja (Orfei in sod., 2008), kar je podkrepljeno s korelacijami med uvidom na eni strani in rezultati na različnih testih kognitivnih sposobnosti pri pacientih z BMR v evtimiji (Dias in sod., 2008b), bolj specifično z delovnim spominom pri pacientih z akutno afektivno epizodo (Varga in sod., 2007) in odločanjem pri pacientih v maniji (Adida in sod., 2008) na drugi strani. Vendar povezava med uvidom in kognitivnimi rezultati ni konsistentna, saj nekatere raziskave niso potrdile povezanosti uvida z rezultati na kognitivnih testih (Yen in sod., 2002; Varga in sod., 2006).

Okrnjenost uvida ne vpliva na zaznano kvaliteto življenja pri pacientih z BMR, saj se zdi, da ne glede na stopnjo uvida v lastno simptomatiko in okrnjenost zavedanja socialnih posledic bolezni pacienti dajejo podobne ocene kvalitete življenja in so navkljub slabšemu uvidu vseeno zmožni prepoznati in poročati o nižji kvaliteti življenja v primerjavi z zdravimi kontrolami (Dias in sod., 2008b).

4. NAČINI POVEČEVANJA UVIDA PRI BMR

Ohranjanje in povečevanje uvida pri pacientih je ključno tako z vidika postavitve diagnoze kot z vidika uspešnosti obvladovanja duševne motnje, kot je BMR. Zaradi različnih vplivov razpoloženjskih faz na posameznikovo zavedanje motnje, tj. okrnjenega uvida v maniji in povečanega uvida v depresiji, namreč prihaja do napačnih diagnoz z unipolarno depresijo in premajhne prepoznanosti BMR (Ghaemi, 2007). Zdravniki se tega fenomena pogosto pomanjkljivo zavedajo in v diagnostični proces premalo vključujejo mnenja svojcev in bližnjih, za katere ugotovitve kažejo, da so prepoznali dvakrat več vedenjskih simptomov manije kot pacienti (Ghaemi, 2007). Z vidika posameznika pa je dober uvid ključen za sprejemanje zdravljenja v akutnih fazah in v

obdobjih okrevanja, ter pri prepoznavanju zgodnjih opozorilnih znakov poslabšanj (Cassidy, 2010). Sistematično učenje prepoznavanja opozorilnih znakov motnje in izboljšanje uvida ne le pripomore k večji učinkovitosti terapevtskih tehnik (Orfei in sod., 2008), temveč je tudi boljša alternativa za posameznika kot učenje uvida preko ponavljajočih se izkušenj poslabšanj in izboljšanj epizod.

Pokazala se je tudi potreba po oblikovanju in vključevanju kognitivnih treningov v psihosocialne intervence, saj je sodeč po raziskavah pomanjkljiv uvid povezan tudi s kognitivnimi primankljaji na področju izvršitvenih funkcij (Adida in sod., 2008; Yen in sod., 2009). S treningom načrtovanja in izvajanja kompleksnih nalog, reševanja problemov in spremljanja uspešnosti izvedbe, bi lahko zagotovili bolj učinkovito učenje in rabo tehnik prepoznavanja zgodnjih znakov poslabšanj BMR in drugih strategij spoprijemanja.

Redke raziskave, ki so se ukvarjale specifično z učinki psihoedukacije na povečevanje uvida pri pacientih z BMR, kažejo, da je povečan uvid pomembno prispeval k boljšemu socialnemu funkcioniranju pacientov v remisiji (Lam in Wong, 1997). Učinke psihoedukacije so povezovali predvsem z drugimi merami izida pri BMR, kot je longitudinalno zmanjševanje frekvence poslabšanj in izboljšan klinični potek bolezni (Colom in sod., 2009), boljše sodelovanje v zdravljenju (Sajatovic in sod., 2009), ki posredno kažejo na izboljšanje uvida pri pacientih. Po drugi strani pa nekateri izsledki kažejo, da pacienti z BMR v remisiji po zaključeni edukacijski intervenci sicer brez težav povezujejo znake lastne motnje z opisi iste motnje pri drugih, čeprav so na začetku imeli različna, tudi medsebojno protislovna, prepričanja o bolezni, ki so bila večinoma nekompatibilna z medicinskim modelom bolezni (Matthew in sod., 2010). Tudi podobne raziskave na področju shizofrenije opozarjajo, da so pacienti in svojci prepoznali vzroke za bolezen v skladu z uveljavljenim biomedicinskim modelom, vendar kljub temu ni prišlo do pomembnih sprememb v njihovih razlagalnih modelih iskanja pomoči in oblik zdravljenja po zaključenem izobraževanju (Das in sod., 2006). Omenjene terapije, kot je psihoedukacija, pridobitev uvida vežejo na sprejemanje medicinskega modela razlage bolezni, ki je večkrat v nasprotju s posameznikovimi kulturno pogojenimi, vendar prav tako veljavnimi razlagami, ki pa nimajo mesta v sprejetih multidimenzionalnih definicijah uvida in jim zato nekateri sociološko orientirani raziskovalci očitajo evrocentričnost (Saravanan in sod., 2004). Posledično lahko lažje razumemo ugotovitve, da pacienti kljub dobremu uvidu slabo funkcionirajo (npr. Staring in sod., 2009), tisti s slabim uvidom pa niso nujno slabo funkcionalni (Mishra in sod., 2009), saj se izognejo soočenju z diagnozo duševne motnje in s tem povezanimi stigmatizirajočimi prepričanji (Saravanan in sod., 2004). Zato je še zlasti pomembno, da so v terapevtske intervence poleg medicinskega modela razlage bolezni vključeni tudi načini spoprijemanja s stigmo in normalizacije bolezni, ki spodbujajo občutek obvladovanja bolezni pri posamezniku (Staring in sod., 2009), tudi s tem, da ne omalovažujejo alternativnih virov iskanja pomoči (Matthew in sod., 2010). Kronični potek motnje, pogosta poslabšanja in nezadostno izboljšanje s predpisanim psihofarmakološkim zdravljenjem bi lahko bili razlogi, zakaj imajo posamezniki različna prepričanja o vzrokih in zdravljenju motnje (Matthew in sod., 2010), in zato kljub vključenosti v psihiatrično zdravljenje iščejo še alternativne poti pomoči. Dobro bi bilo upoštevati tudi širše vidike uvida, saj zavedanje težav, sprejemanje potrebe po zdravljenju in iskanje lokalno dostopnih zdravljenj že nakazuje na prisotnost uvida pri posamezniku, čeprav za sedaj ni vključeno v obstoječe modele in lestvice za oceno uvida.

5. ZAKLJUČEK

Raziskave kažejo, da je za okrnjen uvid značilno zanikanje prisotnosti patološkega procesa, napačno pripisovanje vzrokov za simptome, slabo sodelovanje v terapevtskem zdravljenju in povezanost z maničnimi in/ ali psihotičnimi simptomi. Stopnja uvida se tekom motnje lahko spreminja. Strategije prepoznavanja zgodnjih opozorilnih znakov motnje, spoprijemanja z notranjo stigmo in kognitivni treningi bi lahko pripomogli k povečevanju uvida pri pacientih z BMR.

6. LITERATURA

[1] Adida M, Clark L, Pomietto P, Kaladjian A, Besnier N, Azorin J-M, Jeanningros R, Goodwin GM (2008). Lack of insight may predict impaired decision making in manic patients. Bipolar Disorders;10; 829–837.

[2] Amador XA, Flaum M, Andreason SA, Clark SC, Gorman JM (1994). Awareness of illness in schizophrenia and schizoaffective disorders and mood disorders. Archives of General Psychiatry, 51;826–836.

[3] Baier M, Murray R, McSweeney M (1998). Conceptualization and measurement of insight. Archives of Psychiatric Nursing 12; 32-40.

[4] Braw Y, Sitman R, Sela T, Erez G, Bloch Y, Levkovitz Y (2011). Comparison of insight among schizophrenia and bipolar disorder patients in remission of affective and positive symptoms: Analysis and critique. European Psychiatry, doi:10.1016/j.eurpsy.2011.02.002.

[5] Cairns R, Maddock C, Buchanan A, David AS, Hayward P, Richardson G, Szumkler G in Hotopf M (2005). Prevalence and predictors of mental incapacity in psychiatric in-patients. British Journal of Psychiatry, 187; 379-385. 379-385.

[6] Cassidy F (2010). Insight in bipolar disorder: relationship to episode subtypes and symptom dimensions, Neuropsychiatric Disease and Treatment, 6; 627–631.

[7] Colom F, Vieta E, Sanchez- Moreno J, Palomino-Otiniano R, Reinares M, Goikolea JM in sod. (2009). Group psychoeducation for stabilised bipolar disorders: 5 -year outcome of a randomised clinical trial. The British Journal of Psychiatry, 194, 260-265.

[8] Cuesta MJ, Peralta V, Zarzuela A, Zandio M (2006). Insight dimensions and cognitive function in psychosis: a longitudinal study, BMC Psychiatry 6; 26.

[9] Das S, Saravanan B, Karunakaran KP, Manoranjitham S, Ezhilarasu P, Jacob KS (2006). Effect of a structured educational intervention on explanatory models of relatives of patients with schizophrenia (2006). The British Journal of Psychiatry, 188; 286-287.

[10] David AS (1990). Insight and psychosis. British Journal of Psychiatry, 156;798-808.

[11] Dell'Osso L, Pini S, Cassano GB, Concettina M, Seckinger RA, Saettoni M, Papasogli A, Yale SA, Amador XF (2002). Insight into illness in patients with mania, mixed mania, bipolar depression and major depression with psychotic features. Bipolar Disorders, 4; 315-322.

[12] Dias VV, Sofia Brissos S, Frey BN, Kapczinski F (2008a). Insight, quality of life and cognitive functioning in euthymic patients with bipolar disorder. Journal of Affective Disorders, 110;75–83.

[13] Dias VV, Brissos S, Carita AI (2008b). Clinical and neurocognitive correlates of insight in patients with bipolar I disorder in remission. Acta Psychiatrica Scandinavica, 117;28-34.

[14] Ghaemi S, Rosenquist K (2004). Is insight in mania state-dependent? A meta-analysis. The Journal of Nervous and Mental Disease, 192;771-775.

[15] Ghaemi SN (2007). Feeling and Time: The Phenomenology of Mood Disorders, Depressive Realism, and Existential Psychotherapy. Schizophrenia Bulletin, 30(1); 122–130.

[16] Johnson SC, Baxter LC,Wilder LS, Pipe JG, Heiserman JE, Prigatano GP (2002). Neural correlates of self-reflection. Brain, 125; 1808-14.

[17] Lam D, Wong G (1997). Prodromes, coping strategies, insight and social functioning in bipolar affective disorders. Psychological Medicine, 27; 1091-1100.

[18] Matthew AJ, Samuel B , Jacob KS (2010). Perceptions of illness in self and in others among patients with bipolar disorder. International Journal of Social Psychiatry, 56(5); 462-470.

[19] Mishra DK, Alreja S, Sengar KS, Singh AR (2009). Insight and its relationship with stigma in psychiatric patients. Industrial Psychiatry Journal, 18(1); 39-42.

[20] Orfei MD, Robinson RG, Bria P, Caltagirone C, Spalletta G (2008). Unawareness of illness in neuropsychiatric disorders: phenomenological certainty versus etiopathogenetic vagueness. Neuroscientist, 14(2);203-222.

[21] Owen GS, Ster C, David AS, Szumkler G, Hayward P, Richardson G, Hotopf M (2011). Regaining mental capacity for treatment decisions following psychiatric admission: a clinico-ethical study. Psychological Medicine, 41; 119-128.

[22] Peralta V, Cuesta MJ (1998). Lack of insight in mood disorders. Journal of Affective Disorders,49(1);55-58.

[23] Pini S, Cassano GB, Dell'Osso L, Amador XF (2001). Insight Into Illness in Schizophrenia, Schizoaffective Disorder, and Mood Disorders With Psychotic Features. American Journal of Psychiatry, 158;122-125.

[24] Sajatovic M. Ignacio R, West JA, Cassidy KA, Safavi R, Kilbourne AM (2009). Predictors of non-adherence among individuals with bipolar disorder receiving treatment in a community mental health clinic. Comprehensive Psychiatry, 50 (2);100-107.

[25] Saravanan B, Jacob KS, Prince M, Bhugra D, David AS (2004). Culture and insight revisited. British Journal of Psychiatry, 184; 107-109.

[26] Staring AB, Van der Gaag M, Van den Berge M, Duivenvoorden HJ, Mulder CL (2009). Stigma moderates the asssociations of insight with depressed mood, low self-esteem, and low quality of life in patients with schizophrenia spectrum disorders. Shizophrenia Research, 115; 363-369.

[27] Varga M, Magnusson A, Flekkoy K, David AS, Opjordsmoen S (2007). Clinical and neuropsychological correlates of insight in schizophrenia and bipolar I disorder: does diagnosis matter? Comprehensive Psychiatry, 48(6);583-591.

[28] Varga M, Magnusson A, Flekkoy K, Ronneberg U, Opjordsmoen S (2006). Insight, symptoms and neurocognition in bipolar I patients. Journal of Affective Disorders, 91(1);1-9.

[29] Yen CF, Chen CS, Ko CH, Yen JY, Huang CF (2007). Changes in insight among patients with bipolar I disorder: a 2-year prospective study. Bipolar Disorders,9; 238-242.

[30] Yen CF, Cheng CP, Huang CF, Ko CH, Yen JY, Chang YP, Chen CS (2009). Relationship between psychosocial adjustment and executive function in patients with bipolar disorder and schizophrenia in remission: the mediating and moderating effects of insight. Bipolar Disorders, 11; 190-197.

VEDENJSKE IN ČUSTVENE TEŽAVE UČENCEV V ŠOLI (BEHAVIORAL AND EMOTIONAL PROBLEMS OF PUPILS AT SCHOOL)

Damjana Oblak
Ekonomska gimnazija in srednja šola Radovljica
Gorenjska cesta 13 4240 Radovljica
Telefonska številka: 031 415 002
Elektronski naslov: damjana.oblak@gmail.com

Izvleček:
Mladostniki z vedenjskimi in čustvenimi težavami predstavljajo za učitelje trd oreh, pri delu z njimi največkrat nismo uspešni. Ne poznamo bistvenih značilnosti teh težav, nismo dovolj strokovno usposobljeni. A šola je idealen prostor za manifestacijo omenjene vedenjske problematike, tu se učenci stalno dokazujejo v odnosu do avtoritet kot tudi v rivalstvu z vrstniki. Šolski neuspeh pomeni nezadovoljstvo, ki sili v ponovne neustrezne akcije in razočaranja. Mladostnikovo problematično vedenje v največji meri 'vzdržujejo' njegova zgodovina, aktualna družinska situacija, individualne posebnosti, širše socialno okolje, vrstniki, šola in prijatelji. Ti faktorji so pogosto spregledani in zato smo tolikokrat neuspešni v šoli oziroma vsaj mislimo, da smo. A kljub naštetemu moramo otroku v šoli ustvariti optimalne pogoje, v katerih bo, glede na svojo težavo, lažje funkcioniral. Sodobni pristopi pomoči so usmerjeni k vzpostavljanju takega učnega okolja, ki bo sposobno spodbuditi in razvijati učenčevo dejavno sodelovanje, dejavno izražanje misli in idej, spodbuditi in razvijati učenčeve interese, nadarjenost ter močna področja, in si hkrati prizadevajo razvijati uspešno sodelovanje med družino in šolo. Temeljni namen projekta pomoči je raziskovanje in soustvarjanje razmer, ki omogočajo učenčevo optimalno udeleženost. Naš odnos do omenjene populacije mora temeljiti na sprejemanju, upoštevanju, spoštovanju ter prilagajanju se njihovi različnosti.

Summary:
Adolescents with emotional and behavioral problems represent the teachers a hard nut to crack, at work with them most of the teachers are often not successful. We do not know the essential characteristics of these problems, we are not sufficiently skilled. But school is an ideal place for the manifestation of these behavioral problems, students there constantly demonstrate in relation to the authorities as well as the rivalry with their peers. School failure is a disappointment which leads into a re-inappropriate actions and disappointments. Adolescent's problematic behavior mostly 'maintain' its history, current family situation, individual characteristics, the broader social environment, peers, school and friends. These factors are often overlooked and why we so often fail at school or at least we think we do. But despite all these elements teachers at school need to create optimal conditions for pupils and students, in which they will have a chance to function better. Modern methods of aid are directed to establishing such a learning environment that will be able to encourage and develop the pupils' and students' active participation, active expression of their thoughts and ideas, encourage and develop students' interests, talents and strengths, and simultaneously tries to develop a successful cooperation between family and school. The basic purpose of the project of such an aid is the research and creating such conditions which enable the pupil's and student's optimal participation. Our relationship with the population mentioned must be based on acceptance, respect and adaption to their diversity.

1. Uvod

Mladostniki z vedenjskimi težavami predstavljajo za učiteljice ter učitelje trd oreh, pri delu z njimi največkrat nismo uspešni. Pogosto slišimo mnenje, da je takih otrok, mladostnikov vedno več, a hkrati pozabljamo, da so otroci s tovrstnimi težavami vedno bili in tudi takrat jih ni bilo malo. Največja težava pa je v tem, da jih še danes v šoli pogosto niti ne opazimo kot take, ne poznamo njihovih težav in vzrokov zanje. Ne zavedamo se, da je njihovo neprimerno vedenje večkrat le klic na pomoč. V množici mladostnikov v razredu jih vidimo le kot moteče, smo večkrat jezni nanje, ker nam onemogočajo dobro izpeljavo pouka in jih na hitro odpravimo. Učitelji lahko s kognitivno-vedenjskim pristopom veliko pripomoremo k zmanjšanju opisanih težav. Pomagamo pri uravnavanju in samouravnavanju vedenja ter s tem posledično pri spreminjanju motečega vedenja. S tem pristopom pripomoremo k razvoju učenčeve morale, ki je ključna za njegovo ravnanje in samoregulacijo (Hribar, 2007).

2. Teoretska izhodišča

Učitelji ne poznamo bistvenih značilnosti težav, nismo strokovno usposobljeni za to. Zaradi neupoštevanja narave težav in napačno razumljene učiteljeve odgovornosti so naši cilji ponavadi previsoki. Ker svojih previsokih ciljev pogosto ne moremo doseči, smo nezadovoljni, nestrpni do mladostnikov, kar poveča problematično vedenje in to vodi v začaran krog. Oblikovanje realnih ciljev je prvi korak k uspešnemu delu z mladostniki s tovrstnimi težavami. Ponavadi si določimo cilj, da se težavno vedenje preneha in mladostnik izpolni vsaj minimalne zahteve glede učenja, a ta

cilj je pogosto prezahteven. Vedenjska problematičnost se stopnjuje že sama po sebi, čim starejši je otrok in čim bolj neuspešen je. Zato je pomemben in realen cilj že to, da ohranimo njegovo težavno vedenje na istem nivoju.

S pomočjo ankete sem učence spraševala po njihovih mnenjih na temo vedenjskih težav v razredu in ugotovila, da so tudi mladostniki do svojih motečih vrstnikov dokaj netolerantni in kritični do njihovega vedenja. Le nekateri razumejo njihovo neprimerno vedenje kot posledico stisk in težav.

Discipliniranost učencev se kaže v redu, ki vlada med poukom v razredu in je nujen pogoj za učinkovito učenje. Večina oblik neprimernega vedenja je blažje narave. Najpogostejše so klepetanje, skakanje v besedo, povzročanje hrupa, nepozornost, nagajanje sošolcem, sprehajanje po razredu brez vzroka in jih je mogoče omejiti z obrazložitvijo splošnih pravil obnašanja. Resnejše oblike (besedno napadanje, preklinjanje, predrznost, neubogljivost, nepriznavanje avtoritete, telesno obračunavanje) so redkejše in se ponavadi pojavijo, če dopustimo, da se blage oblike sprevržejo v navado, lahko pa so tudi odziv na osebne in učne težave, razlaga Kyriacou (1997).

Pedagogika otrok s težavami v vedenju in osebnosti se je v preteklem obdobju razvijala predvsem v smeri prevzgoje otrok in mladostnikov v vzgojnih zavodih, fazi nastajanja težav pa se ni posebej posvečala. Žal to prevečkrat počnemo še danes. Čakamo, da potencialno vedenjsko in osebnostno težaven otrok 'dozori' v prestopnika, da bi se z njim splačalo ukvarjati (Škoflek, 1983). Omenjeni avtor je v svoji raziskovalni nalogi *Motnje vedenja šolskih otrok* želel prestopiti mejo med vzgojnim zavodom in šolo ter pomagati razvijati šolam preventivno pedagoško delo z učenci, ki se nagibajo ali pa že izražajo razne vedenjske težave in ki živijo v takih okoliščinah, ki neugodno vplivajo na njihov uspeh, razvoj in dozorevanje.

V ta namen je raziskoval vedenje učencev pri pouku in ugotovil, da so tako učitelji kot učenci za najbolj 'moteče' oblike vedenja določili naslednja dejanja:

- klepetanje;
- dajanje pripomb med poukom, ki niso z njim v zvezi;
- ne slediti učiteljevi razlagi, ampak početi ta čas kaj drugega,

Za preprečevanje 'motečega' vedenja pa bi bilo po njihovem mnenju najučinkovitejše:

- nazorno, razumljivo, zanimivo posredovanje učne snovi;
- aktivno sodelovanje učencev pri pouku;
- učiteljeva strpnost ter človeški posluh za doraščajočega človeka.

Zelo podobne (identične) vzroke in tudi rešitve pa navajajo učenci še danes.

Jože Horvat (1997) opiše srednješolce z motnjami v vedenju in osebnosti (oz. mladostnike s posebnimi potrebami v socialni integraciji) kot osebe, ki kažejo čustveno neprilagojenost, povečan nevroticizem, nezadovoljstvo s samim seboj in svetom, ki jih obdaja, negativno samopodobo, nezaupanje v lastne sposobnosti, spontano in reaktivno agresivnost, izrazito družabnost, težnjo po dogodivščinah ter težnjo po izrazitem zadovoljevanju potreb. Vse te značilnosti so posledica čustvene zavrtosti, ki se oblikuje v najzgodnejšem otroštvu zaradi neustreznega odzivanja okolja, najizraziteje pa se pojavijo v obdobju adolescence, ko so potrebe po potrjevanju, uveljavljanju lastne identitete in dokazovanju na prvem mestu. Posledica je opazno nesprejemljivo vedenje za širši socialni prostor, ki se kaže kot impulzivnost, verbalna in fizična agresivnost, nastopaštvo pred vrstniki, uporništvo do avtoritet, zavračanje konvencionalnih vrednot, popolna pasivnost, brezvoljnost ter v skrajnih primerih tudi avtoagresivnost.

Prav šola je idealen prostor za manifestacijo omenjene vedenjske problematike, tu je potrebno stalno dokazovanje v odnosu do avtoritet kot tudi v rivalstvu z vrstniki. Šolski neuspeh pomeni nezadovoljstvo, ki sili v ponovne neustrezne akcije in razočaranja.

Myschker (1988) opisuje aktualne trende pedagoškega dela v Nemčiji. Mnogi strokovnjaki se strinjajo, da je število otrok in mladostnikov z vedenjskimi težavami vedno večje in da se socialno-emocionalne motnje kažejo čedalje bolj zgodaj. Vedno pogostejši simptomi v šolah so agresivni izbruhi, manjkajoče samozaupanje, izrazita nedisciplina, nestrpnost, na drugi strani pa umik, nezainteresiranost, čezmerna občutljivost, nezaupljivost, depresivnost in psihosomatske težave. V Nemčiji se je uveljavil berlinski model 'ambulantnih učiteljev'. Posebej uspešni in angažirani učitelji se nekaj ur na teden individualno ali v paru ukvarjajo z učenci, ki kažejo posebne težave (učne, vzgojne, razvojne, osebnostne), v svoje delo pa vključujejo tudi pogovore s starši in drugimi učitelji, ki učijo te otroke. Učitelj v uvajalni fazi sledi rednemu pouku, opazuje učenca in postopoma s pogovori in sproščenimi dejavnostmi navezuje stik z otrokom. Potem se pomoč diferencira glede na individualne potrebe vsakega otroka. Vsebuje lahko vedenjsko-modifikacijske programe, socialne, učno podporne tehnike, treninge za večjo učinkovitost socialnega vedenja, premagovanje konfliktov. V te programe se lahko vključijo tudi drugi učenci v razredu, že zaradi večjega razumevanja in sprejemanja različnosti pa tudi redukcije lastnih vedenjskih težav.

Bečaj (1987) se sprašuje, zakaj smo pri obravnavanju vedenjskih težav v šoli pogosto neuspešni. Po njegovem prepričanju je glavni vzrok za to neupoštevanje in nepoznavanje bistvenih značilnosti težav. Gre za simptomatiko, ki je za šolski pa tudi širši socialni prostor problematična in praviloma posledica dalj časa trajajočega nezadovoljevanja osnovnih psihosocialnih potreb, torej je to v bistvu čustvena težava, za katero je značilno:

- disocialna simptomatika (izostajanje od pouka, agresivnost, vandalizem, laganje, odklanjanje avtoritete),
- šolska neuspešnost,
- pomanjkanje interesov (pasivno preživljanje prostega časa),
- slabe delavne navade,

- pomanjkanje stikov z neproblematičnimi vrstniki (nima prijateljev med njimi, druži se s takimi, kot je sam),
- pomanjkanje stikov z odraslimi,
- izločenost iz socialnega okolja (klub, organizacija).

Odklonilne reakcije drugih, če so dovolj pogoste, začnejo otroka dodatno ogrožati v izpolnjevanju njegovih osnovnih psihosocialnih potreb (sprejetost, varnost, potrditev), to pa povzroči čustvene težave, ki se neposredno odrazijo tudi v vedenju. Tako se ustvari začaran krog, v katerem otrok zaradi svojih posebnosti povzroči okolju negativne reakcije, ki vzvratno ojačajo težavnost vedenja in ta spet odklanjanje okolja. Ustvari se več začaranih krogov (v šoli, doma, med vrstniki, v širšem družbenem okolju), ki stabilizirajo težavo. Tudi Bluesteinova (2008) podobno poudarja pomen zadovoljitve primarnih potreb učenca in povezanost nezadovoljenih potreb s krogom vedenjskih težav, neuspeha in kazni učenca.

Bečaj (1987) je prepričan, da otrokovo in mladostnikovo problematično vedenje v največji meri 'vzdržujejo' otrokova zgodovina, aktualna družinska situacija, individualne posebnosti, širše socialno okolje, vrstniki, šola in 'klapa'. Ti faktorji so pogosto spregledani in zato smo tolikokrat neuspešni v šoli oziroma vsaj mislimo, da smo. Ti faktorji bodo problematičnost še naprej podpirali, ne glede na to, kaj z otrokom v šoli počnemo. A kljub naštetemu moramo otroka oziroma mladostnika v šoli individualno opazovati, ne smemo biti že vnaprej prepričani, kako in kaj se bo zgodilo z njim. Ustvariti mu moramo optimalne pogoje, v katerih bo, glede na svojo težavo, lažje funkcioniral. Upoštevati moramo družino in okolje, iz katerega izhaja, njegove sposobnosti. Šola je institucija, ki vedenjsko problematiko praviloma povečuje zaradi tekmovalnosti, togosti in kriterija, po katerem se učence vrednoti (t.j. primernost vedenja ter učna uspešnost). Vse to pa je odraz splošnih norm in vrednot v celotnem kulturnem prostoru.

Otrok oziroma mladostnik pogosto prav dobro ve, kaj je prav in kaj ne. Potrebuje le oporo v situaciji sami in to mu lahko najbolje nudi učitelj, ki je z njim skupaj vsak dan, z njim preživlja težavne situacije. Pri tem naj mu pomagajo sodelavci, šolski svetovalni delavec, vodstvo šole. Za omenjeno populacijo so pomembna tudi jasna, pregledna in prožna pravila (kaznovanje, nagrajevanje), skupno planiranje, ureditev prostora in spregled kakšne malenkosti ter tudi vključevanje v razne organizacije (športne, gasilci, skavti), letovanja (na morje, v planine), izleti in ekskurzije, kjer so ti otroci in mladostniki sposobni konkurirati in se lažje uveljaviti kot na učnem področju.

V današnjem času se vse bolj poudarja, da bi morali imeti otroci, mladostniki več možnosti, da se sami zagovarjajo in ne da so pri tem odvisni od drugih (učiteljev, staršev). Imeti morajo moč nad tem, kako se bodo izobraževali in živeli, le tako so lahko tudi odgovorni za rezultate učnega procesa ter za uspehe in neuspehe v svojem življenju. Mojca Peček (1999) zagovarja vključevanje otrok, mladostnikov s posebnimi potrebami (tudi s čustvenimi in vedenjskimi težavami) v procese odločanja, saj bodo tako lažje postali neodvisni, odgovorni in si pridobili samozaupanje. Gre namreč za populacijo, ki jo največkrat dojemamo kot nemočno, neodgovorno, ki ni sposobna obvladovati svojega vsakodnevnega življenja, kaj šele odločati o sebi in svoji prihodnosti.

Pečkova (1999) poudarja tudi pomen raziskav, ki v ospredje postavljajo vprašanje, kaj lahko otroci povedo o sebi, svojem doživljanju sebe, okolice in vzgoje. Takšne raziskave bi zagotovile učiteljem, da bi bolje razumeli to populacijo. Pečkova meni, da tudi raziskovalci, profesionalci premalo vedo, kako priti do otrokovega, mladostnikovega mnenja. Nekatere raziskave so že pokazale, da jih vključeni otroci niso razumeli kot pomoč, ampak so bili osramočeni in stigmatizirani. Bistveno je sprejeti dejstvo, da ima vsak svojo enkratno in individualizirano predstavo sveta, ki je zasnovana na lastnih izkušnjah, čustvih in občutkih. Upoštevanje otrokove samopredstavitve lahko razširi njihovo odgovornost za lastno življenje, hkrati pa je vse bolj jasno, da je brez upoštevanja tega vzgojno načrtovanje neustrezno, saj ostaja le na površju in nikakor ne more biti individualizirano.

Kognitivno-vedenjski pristop temelji na predpostavki o spreminjanju motečega vedenja, ki jo skušamo doseči s spremembo prepričanj/kognicij učencev in z vedenjskimi postopki, kar rezultira v spremembi vedenja. Pri tem pristopu se v različni razsežnosti kombinirajo tako kognitivne kot tudi vedenjske tehnike. Kognitivno usmerjeni avtorji se bolj nagibajo v smer, da so prepričanja/kognicija tista, ki povzročijo spremembe v vedenju, vedenjsko usmerjeni avtorji pa menijo, da je pri tem odločilno vedenje, ki ga primerno podkrepimo oziroma kaznujemo. Pristop je usmerjen k specifično usmerjenim težavam učencev in poudarja, da so učenci sami sposobni kontrolirati svoje misli in dejanja. Veliko vlogo odigra motivacija, ki pripomore k aktivnosti samouravnavanja, v primeru, da je motivacija nizka, pa je učitelj tisti, ki učenca poskuša motivirati. Vendar pa je sprva zagotovo odločilna vloga učitelja, ki preverja iracionalna prepričanja učencev, jih uči novih spretnosti in jih spremlja pri osvajanju le-teh ter nudi povratne informacije (Hribar, 2007).

Sodobni pristopi pomoči (Magajna idr., 2008) upoštevajo učenčevo celovito osebnost, njegove vsakdanje življenjske razmere v šoli in doma ter življenjsko perspektivo; usmerjeni so k vzpostavljanju takega učnega okolja, ki bo sposobno spodbuditi in razvijati učenčevo dejavno sodelovanje, dejavno izražanje misli in idej, spodbuditi in razvijati učenčeve interese, nadarjenost ter močna področja, in si hkrati prizadevajo razvijati uspešno sodelovanje med družino in šolo. Temeljni namen projekta pomoči je raziskovanje in soustvarjanje razmer, ki omogočajo učenčevo optimalno udeleženost. Najpomembnejši sistemski dejavniki, ki omogočajo doseganje učinkovitega učnega okolja in učinkovite učne pomoči, so:

- dobra strokovna usposobljenost učiteljev in drugih, ki delajo z učenci z učnimi težavami (tudi poznavanje ustreznih zakonskih dokumentov);
- zagotavljanje in organizacija časa, prostora in ustreznih didaktičnih sredstev;
- vodenje šole za spodbujanje in omogočanje interdisciplinarnega timskega dela.

Pri oblikovanju načrta pomoči je potreben celovit, kompleksen pristop. Na področju ovir in primanjkljajev učenca se je dobro vprašati, ali smo z učencem vzpostavili ustvarjalni dialog, ali je tudi on povedal, kako razume svojo oviranost, kako jo opredeljuje, kaj mu pomaga in kako, kdaj ne zmore, kaj po njegovem potrebuje. Na temo učenčevih močnih področij moramo izvedeti, kako jih vidi in opredeli on sam in z vidika perspektive moči se sprašujmo, ali smo vzpostavili tak pogovor z njim, da nam in sebi tudi sam pomaga s svojim razumevanje in videnjem.

Šola je eden izmed pomembnih varovalnih dejavnikov psihosocialnega razvoja otroka oz. mladostnika. Varovalni dejavniki so pojavi, dogajanja, procesi, katerih vpliv na posameznika, izpostavljenega ogrožujočim dejavnikom, zmanjšuje ali izniči učinke neugodnih okoliščin (Kos Mikuš, 1992). Čeprav ponavadi, ko povezujemo šolo in odklonsko vedenje, mislimo na neustrezne zahteve, ki marginalizirajo veliko število učencev in na druge stresogene vplive, je šola pomemben socializacijski sistem. Šola je prvo delovno mesto otroka, prostor socialnih odnosov, pisanih in nepisanih pravil, psihosocialne klime, ki so ji izpostavljeni ljudje v občutljivem razvojnem okolju. Šola vpliva tudi na vedenjske težave in delikventnost. Varovalni vplivi šole imajo največji pomen pri otrocih, ki odraščajo v neugodnih družinskih in socialnih okoljih. Zanimanje raziskovalcev, poudarja Čačinovič Vogrinčičeva (2008), se je v devetdesetih letih končno premaknilo od preiskovanja dejavnikov tveganja za problematično vedenje k raziskovanju značilnosti, ki otroka varujejo, značilnosti odpornosti oziroma moči okrevanja otroka. Omenjena avtorica navaja raziskavo B. Benard iz leta 1997 (prav tam), kjer so se pokazali kot najpomembnejši varovalni dejavniki naslednji:

- socialna kompetentnost (odzivnost, prožnost, skrb, sposobnost empatije, komunikacijske veščine, razne oblike prosocialnega vedenja in smisel za humor),
- veščine za odpravljanje problemov oziroma veščina sposobnosti za konflikt (sposobnost, da učenec misli abstraktno, prožno, da reflektira, je sposoben razmišljati o alternativnih rešitvah za spoznavne in socialne potrebe),
- neodvisnost (občutek moči, kompetentnosti in nadzora, samospoštovanje in učinkovitost),
- občutek smisla in usmerjenost v prihodnost (prepoznati pomen in kompetentnost načrtovati ter spreminjati prihodnost),
- velika pričakovanja (pozitivna naravnanost šole na uspeh učencev, pristop iz perspektive moči, ki je korak k dobremu in uspešnemu delu, hkrati pa k zmanjšanju neželenega vedenja in čustvenih problemov),
- udeleženost oziroma vpetost, soustvarjanje (moč odločanja o smeri in poteku lastnega življenja).

Angleški raziskovalci (po Kos Mikuš, 1992) so proučevali razlike med pogostostjo delinkventnega vedenja na različnih šolah v londonskih mestnih četrtih s podobno populacijsko sestavo učencev. Raziskave so pokazale, da razlik v vedenju učencev ni mogoče pojasniti le z razlikami v socialnih, kulturnih in etničnih razlikah družin. Pokazale so se velike razlike med šolami. Razlike so ostale tudi, ko so primerjali skupine otrok s podobnimi individualnimi karakteristikami. To je bil očiten dokaz o vplivu šolskega okolja, kulture šole oz. etosa. Med šolskimi varovalnimi dejavniki so se pokazali kot najpomembnejši:

a) splošni šolski uspeh in uspeh pri enem predmetu ali področju
b) dobri odnosi z učitelji in vrstniki
c) priložnost za socialno učenje
d) nadzor nad dijaki
e) zapolnjen čas (interesne dejavnosti).

Vzrok za neustrezno vedenje so pogosto tudi čustvene težave. Mladostniki se z njim lahko podzavestno borijo proti mučni depresivnosti (Tomori, 2000). Močna čustvena nihanja so ena izmed značilnosti depresije, vendar pa je pri nekaterih mladostnikih tako zelo huda, da jim zagreni življenje na vseh področjih. Nikakor se ne morejo opomoči od kakega neuspeha ali pretresa. Pri njih potrtost traja še tedaj, ko ni več nobenega zunanjega razloga zanjo. V šoli začnejo pešati, pri pouku so brezvoljni, dejavnosti, pri katerih so dotlej uživali, postanejo nezanimive.

Liplinova (2003) poudarja, da družba od učiteljev (ki so tudi sami večkrat žrtve neprilagojenega vedenja) pričakuje, da se bodo aktivno soočali s čustvenimi težavami in asocialnim vedenjem otrok in mladostnikov. A po raziskavi slovenskih avtorjev (Magajna idr., 2005) se učitelji čutijo slabo strokovno usposobljeni za težave, povezane s hiperaktivnim, nemirnim vedenjem, in za čustvene težave. Premalo znanja na tem področju pa imajo učitelji tudi po prepričanju Pšundrove (2005), kar je pokazala njena raziskava. Pri delu s to populacijo bi morala v šoli učiteljem pomagati vrsta za to usposobljenih strokovnjakov (socialnih in specialnih pedagogov, psihologov, sociologov), potrebne so številne spremembe na različnih nivojih. Do takrat pa je naloga vsakega izmed nas učiteljev, da se po svoji moči trudi razumeti, poslušati vsakega otroka, mladostnika, še posebej tistega, ki ima težave. Imeti moramo voljo in spretnosti, da se naučimo razumeti čustva otroka, mladostnika, ki so podlaga njegovemu vedenju. Premalo se zavedamo, da je otrokovo neprimerno vedenje posledica vrste neugodnih okoliščin, za katere so v veliki meri odgovorni tudi drugi, ne le otrok sam. Naš odnos do njih mora temeljiti na sprejemanju, upoštevanju, spoštovanju ter prilagajanju se njihovi različnosti.

Izrabimo vsako priložnost, da izboljšamo njihovo samopodobo in imejmo pozitivna pričakovanja do njih.

3. Sklep

Prispevek zaključujem s predlogi, kako naj ravnamo z otroki, mladostniki s čustvenimi in vedenjskimi težavami, avtorja Krajnčana (2007), ki jih v veliki meri podpiram:

- imejmo pozitivna pričakovanja;
- izrabimo vsako priložnost, da pri mladih izboljšamo samopodobo (pohvale, nagrade);
- poslušajmo jih;
- iščimo rešitve (in ne vzroke) za neustrezno vedenje;
- spodbujajmo jih za prevzem odgovornosti za svoje vedenje;
- bodimo kritični do vedenja mladih;
- zastavimo si realne, uresničljive cilje;
- iščimo področja, kjer so uspešni (močna področja);
- dajajmo jim pomoč, ki jo potrebujejo, ko jo potrebujejo in le toliko, kot jo potrebujejo;
- izogibajmo se tistemu, kar bi jih stigmatiziralo.

V šolo umeščajmo kognitivno-vedenjski pristop kot teoretsko podlago celotnemu vzgojnemu modelu, kot teoretsko podlago svetovalnim pristopom ali kot vir inspiracije za uporabo posameznih tehnik uravnavanja disciplinskih problemov.

LITERATURA

Bečaj, J. (1987). Problem uspešnosti pri obravnavanju vedenjskih motenj na osnovni šoli. V: Bergant. M. (ur.). *Družina in vzgoja 6. Vedenjske motnje mladostnikov v sodobnem času.* 78-103. Ljubljana: Zveza prijateljev mladine Slovenije.

Bluestein, J. (2008). *The win-win classroom. A fresh and positive look at Classroom management*. Corwin Press: Thousand Oaks.

Čačinovič Vogrinčič, G. (2008). *Soustvarjanje v šoli. Učenje kot pogovor.* Pridobljeno 30. 6. 2009, iz http://www.mss.gov.si/fileadmin/mss.gov.si/pageuploads/ministrstvo/Publikacije.

Horvat, J. (1997). Integracija srednješolcev z motnjami v vedenju in osebnosti. V: Destovnik, K.(ur.).*Uresničevanje integracije v praksi. Vzgoja in izobraževanje otrok in mladostnikov s posebnimi potrebami.* 389-396. Ljubljana: Društvo defektologov Slovenije.

Hribar, H. (2007). *Spreminjanje motečega vedenja v razredu s kognitivno-vedenjskim pristopom.* Diplomsko delo. Ljubljana: Pedagoška fakulteta.

Krajnčan, M. (2007). *Osnove doživljajske pedagogike*. Ljubljana:Pedagoška fakulteta.

Kyriacou, C. (1997). *Vse učiteljeve spretnosti*. Radovljica: Regionalni izobraževalni center Radovljica.

Liplin, N. (2003). Reševanje konfliktov pri mladostnikih. V: Štritof, J., Liplin, N. (ur.). *Zbornik svetovalnih delavcev ljubljanskih osnovnih šol,* 25-29. Ljubljana: Osnovna šola Dravlje.

Magajna, L., Pečjak, S., Peklaj, C., Bregar, G. K., Čačinovič Vogrinčič, G., Kavkler, M., Tancig, S. (2005). *Učenci z učnimi težavami v osnovni šoli– razvoj celovitega sistema učinkovite pomoči* (raziskovalno poročilo, projekt št. VP-0830). Ljubljana: Razvojno-raziskovalni inštitut Svetovalnega centra.

Magajna, L., Kavkler, M., Čačinovič Vogrinčič, G., Pečjak, S., Bregar Golobič, K. (2008). *Učne težave v šoli. Koncept dela.* Ljubljana: Zavod Republike Slovenije za šolstvo.

Myscher, N. (1988). Aktualni trendi pedagoškega dela z vedenjsko motenimi otroki in mladostniki. *Ptički brez gnezda 13* (26), 51-65.

Peček Čuk, M. (1999). Od odvisnosti k odgovornosti. Vključevanje učencev s posebnimi potrebami v procese odločanja. *Socialna pedagogika, 3* (4), 357-376.

Pšunder, M. (2005). How effective is school discipline in preparing students to become responsible citizens? Slovenian teachers' and students' views. *Teaching and Teacher Education, 21* (4), 273-286.

Škoflek, I. (1983). Motnje vedenja šolskih otrok. Ljubljana: Pedagoški inštitut pri Univerzi Edvarda Kardelja v Ljubljani.

Tomori, M. (2000). Psihosocialni dejavniki pri mladoletniškem prestopništvu. V: Šelih, A. (ur.). *Prestopništvo in odklonsko vedenje mladih – vzrok, pojavi, odzivanja*. 89-112. Ljubljana: Bonex.

MEHANICISTIČNI POGLED NA VOLICIJO: NEVROETIČNA PERSPEKTIVA

Toma Strle
Univerza v Ljubljani, oddelek za Kognitivno znanost
Kardeljeva ploščad 16, 1000 Ljubljana, Slovenija
Tel: +386 31 338 268
email: toma.strle@guest.arnes.si

POVZETEK

V članku bom prvič govoril o tem, kaj mehanicistična (nevroznanstvena) razlagana duševnih procesov (predvsem volicijskih procesov) pomeni za razlago naše zavestne volicije – predstavil bom eksperiment J. D. Haynesa (2011), ki govori o predvidevanju naših namer in odločitev in Wegnerjev (2002) model mentalne vzročnosti. Drugič, glede na vedno bolj prisotno razlago duševnih procesov z nezavednimi možganskimi mehanizmi, se postavi nevroetično vprašanje ali lahko percepcija sveta in duševnosti kot deterministične (Nichols in Knobe 2007) ali mehanicistične (Nahmias et al. 2007) vpliva na moralne sodbe in pripisovanje moralne odgovornosti. In tretjič, na koncu bom poskusil pokazati, da se čutimo ogrožene in mislimo, da je ogrožena vloge naše zavestne volje s strani mehanicistične razlage duševnosti, zaradi naših intuicij o tem, kaj zavestna volja in mi sami smo v povezav s svetom in možgani.

1 ZAVESTNA VOLJA IN NEZAVEDNI MOŽGANSKI MEHANIZMI

Ali eksperimenti Libeta (1985) in eksperimenti Haynesa in kolegov (2007, 2008, 2011) res kažejo na to, da je naš občutek, da smo iniciatorji dejanj in se zavestno odločamo za dejanja, iluzoren. Ali res pomenijo, da ljudje nimamo zavestnega vpliva (zavestne volje) na lastna dejanja? Najprej moramo razločiti med dvema okvirjema, v katerih poteka diskusija o določenosti (zavestne) volicije (svobodne volje) z nečim zunanjim. Tradicionalna filozofska diskusija se ukvarja z vprašanjem združljivosti zunanjega determinizma dogodkov in naravnih zakonov, psihologija in nevroznanost pa večinoma z vprašanjem ali lahko vzročno moč zavestne volicije ohranimo navkljub temu, da se zdi, da naša mentalna stanja in posledično dejanja določajo in povzročajo nezavedni možganski in psihološki mehanizmi. Klasični filozofski problem združljivosti svobodne volje in determinizma je pravzaprav širši problem, vendar gre v obeh za nekakšno določanje zavestne volje z nečim zunanjim[1] (kasneje bomo videli, da je pogled na zavestno volicijo kot nekaj zunanjega možganskim mehanizmom morda napačen). V članku me bo predvsem zanimalo, kaj mehanicistična razlaga duševnosti pomeni za našo volicijo in kako deterministični in mehanicistični (nevrološki) pogled na duševne procese vpliva na naše moralne sodbe in pripisovanje moralne odgovornosti.

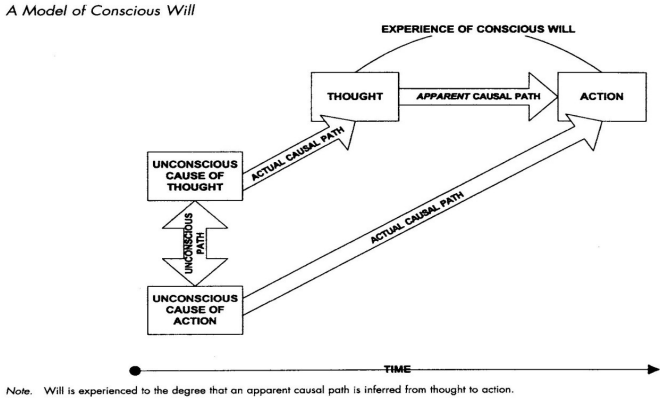
1.1 Nezavedne odločitve in namere

V nadaljevanju dela Libeta (Libet 1985), je npr. Haynes (2011) s pomočjo funkcionalne magnetne resonance (fMR) pokazal, da lahko iz prostorskega vzorca fMR signala v frontopolarnemu (BA10) in parietalnem korteksu predvidimo enostavne motorične namere/odločitve ljudi približno 7 sekund pred samo zavestno namero/odločitvijo za motorično dejanje v približno 65-70% primerih. V eksperimentu so udeleženci ležali v fMR skenerju, kjer so gledali v sredino zaslona, na katerem so se naključno prikazovale črke, med tem pa so se lahko svobodno odločili ali bodo pritisnili levi ali desni gumb. Morali so poročati, katera črka je bila prikazana na zaslonu, ko so naredili zavestno odločitev. Haynes pravi, da »kar kažejo ti izsledki je, da se kaskada nezavednih možganskih procesov, ki se začnejo v prefrontalnem in parietalnem korteksu, odvija preko nekaj sekund in pripravlja subjektivno svobodne in samo-tempirane odločitve.« (Haynes 2011) Seveda priznava, da taki eksperimenti ne zajemajo dolgoročnih odločitev, ki bi vključevale kompleksno mišljenje. Poleg tega pravi, da je možna interpretacija eksperimenta tudi, da nikoli ne bomo sposobni natančno predvideti odločitev in namer, saj področja možganov, ki jih je preučeval, predstavljajo le delni vpliv (*bias*) na oblikovanje odločitve in da na končno odločitev v poznejših časovnih fazah vplivajo tudi drugi procesi. Zanesljivo »branje« namer in misli sicer trenutno ni možno in postavlja se veliko epistemoloških in metodoloških dvomov, da bomo to sposobni kadarkoli storiti, vendar taki eksperimenti kažejo na to, da igrajo nezavedni možganski mehanizmi večjo vlogo pri naših odločitvah, kot smo mislili v preteklosti.

[1] Moramo pa se zavedati, da tudi, če bi ugotovili, da ima zavestna volja vpliv na dejanja in odločitve, to še ne bi razrešilo tradicionalnega problema združljivosti svobodne volje in zunanjega determinizma, saj je mogoče, da bi bila prav ta zavestna volja determinirana s preteklimi dogodki in naravnimi zakoni in v tem smislu ne bi bila svobodna. Poleg tega nevroznanost ne more odgovoriti na vprašanje ali živimo v determinističnem svetu ali ne. Lahko govori le o tem ali so možgani deterministični sistem (Roskies, 2006) in preučuje mehanizme duševnosti.

1.2 Wegnerjeva teorija mentalne vzročnosti

Nezavedni mehanizmi (genetski, nezavedni, kognitivni, nevrološki, emocionalni, socialni), ki jih raziskuje psihologija in ostale znanosti, po Wegnerju izključujejo človeško zavestno poseganje in s tem tudi tradicionalno predstavo o ljudeh kot zavestnih dejavnikih. Wegnerjeva teorija navidezne mentalne vzročnosti tako pravi, da »Izkušnja hotenega dejanja izvira iz interpretacije lastnih misli kot vzroka za dejanje.« (Wegner 1999, str. 480). Zavestna volja tako po Wegnerju nima nikakršne vzročne moči; občutek nam pravi nekaj, kar ni resnično. Zavest v tem primeru le spremlja rezultat nezavednih mehanizmov (Wegner zagovarja pozicijo modularnega epifenomenalizma), katerim v resnici pripada vzročna moč. »Ponavadi se zdi, da zavestno odločamo o svojih voljnih dejanjih, a to je iluzija.« (Wegner 2002, str. 1).



Slika 1: Model mentalne vzročnosti in vpliva zavestne volje na dejanja (Wegner 1999, str. 483).

Nezavedni mehanizmi v možganih, ki so vzročno odgovorni za dejanje, so različni od (še vedno nezavednega) mehanizma, ki je odgovoren za nastanek zavestne misli o tem dejanju. Zavestna misel o dejanju se pojavi 'ob pravem času', vendar to ne pomeni, da je kakorkoli vzročno prispevala k dejanju. Tako lahko izkusimo zavestno voljo tudi tedaj, ko naša misel ni mogla ničesar prispevati k dejanju in napačno sklepamo, da smo »intencionalno povzročili« dejanje, v katerega smo bili v resnici »prisiljeni«, nezavedni mehanizmi pa ustvarijo »… tako zavestno misel o dejanju kot proizvedejo tudi občutek volje, ki ga izkusimo …« (Wegner 2002, str. 98)

1.3 Pet pogledov na volicijo

Volicija oz. volja je nejasen, nedefiniran in mnogoter koncept, ki ga različni filozofi in znanstveniki pojmujejo različno (Roskies, 2010). Roskies v svojem članku navaja, da nekateri volicijo npr. razumejo kot neko osnovo za notranje (endogeno) dejanje, avtonomijo ali izbiro. Nekateri razlikujejo voljna dejanja od refleksnih ali takšnih, ki so določena z okoljem. Drugi trdijo, da je volicija notranje mentalno dejanje odločanja ali formiranja namere (npr. Searle 1983 ali Zhu 2004a; iz Roskies 2010) ali pa da je volicija odločitev ali namera sama (Adams & Mele 1992; iz Roskies 2010). Ker Roskies zanima predvsem nevroznanstveni pogled na volicijo, le to razdeli v pet kategorij, glede na eksperimentalne pristope nevroznanosti: volicija kot začetek dejanja, volicija kot namera, volicija kot odločanje oz. sposobnost izbire dejanja, volicija kot izvršna kontrola, volicija kot inhibicija in kontrola ter volicija kot občutek. Različni eksperimentalni pogledi na volicijo kažejo, da volicija pravzaprav ni enotna sposobnost oz. funkcija možganov, ampak je skupek mnogoterih sistemov, od katerih vsak po svoje prispeva k temu, čemur pravimo npr. (zavestna) volja ali volicija. To pa ima za posledico, da preučevanje enostavnih odločitev in namer (npr. Haynes 2011) ne podaja celotne slike o zavestni volji (voliciji). Za boljši odgovor bi npr. potrebovali metodološki pristop, ki bi omogočil bolj celostno preučevanje volicije.

1.4 Vprašanji jaza in zavesti pri preučevanju volicije

Problemov in vprašanj je pri preučevanju volicije, posebej zavestne, zelo veliko, vendar se mi zdi, da moram omeniti dva, ki sta iz diskusije ponavadi izpuščena oz. postavljena na obrobje. Prvi problem je vprašanje, kdo je ta »jaz«, ki naj bi se odločal, izbiral in tvoril namere. Ko govorimo npr. o odločanju za neko dejanje in npr. podamo odgovor, da so nezavedni mehanizmi vodili našo odločitev, mi pa smo imeli le iluzoren občutek, da smo se odločili mi, predpostavljamo, da vemo, kaj ta »jaz« je. Trditev, »Jaz nimam zavestnega vpliva na svoje odločitve« prazna, če ne vemo, kaj ta »jaz« je. Mislim, da se raziskovalci volicije temu vprašanju poskušajo velikokrat izogniti ali pa »jaz« enostavno postavijo ven iz deterministčnega univerzuma in/ali možganskih mehanizmov in tako uspejo ohraniti nek poseben status tega »jaza«. Ker pa imamo na voljo vedno več mehanicističnih razlag duševnosti, se zdi, da temu jazu ostaja vedno manj prostora in posledično vedno manj vpliva na npr. odločitve. Zdi pa se , da za ohranitev vpliva volje potrebujemo tudi nek vzročni vpliv zavesti. Nevroznanost je sicer izzvala naše intuicije in tradicionalne poglede na razmerje med zavestjo in dejanjem (Roskies 2010), vendar se moramo zavedati, da na vprašanje, kaj je zavest, nimamo odgovora in zato vprašanje vpliva zavesti na naše odločitve, mentalna stanja in dejanja ostaja odprto.

2 MORALNE SODBE V DETERMINISTIČNEM IN MEHANICISTIČNEM SVETU

Vprašanje, ki so si ga zastavili eksperimentalni filozofi, je ali ima deterministični (Nichols in Knobe 2007) in/ali mehanicistični (Nahmias et al. 2007) pogled na duševnost vpliv na naše dojemanje moralne odgovornosti in moralne sodbe? Deterministična in mehanicistična razlaga duševnih pojavov (ožje volicije) in je nevroetično vprašanje. Prvič v okviru nevroznanosti etike (Roskies 2002), zato, ker nevroznanost poskuša razlagati etično in moralno (preučuje volicijo, moralne sodbe ipd.). In drugič, v okviru etičnih implikacij nevroznanosti (etika nevroznanosti) (Roskies 2002). Mehanicistična razlaga, kot bomo videli, vpliva na naše intuicije o vplivu zavestne volje ter posledično vodi v manjše pripisovanje moralne odgovornosti drugim osebam (in morda tudi sebi) in tako morda vpliva tudi na družbene prakse, ki se navezujejo na vprašanje moralne odgovornosti

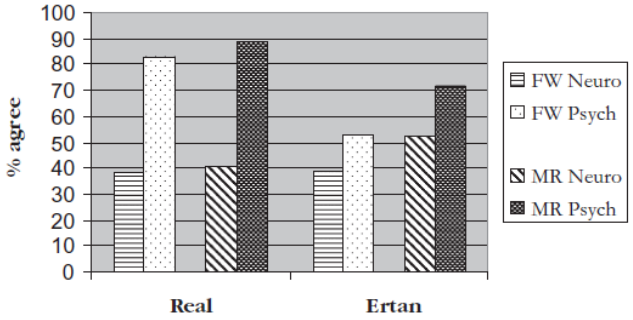
(zadnje vprašanje presega obseg članka). V okviru etike nevroznanosti, so eksperimenti Nicholsa in Knobeja (2007) pokazali, da predstava deterministočnega sveta na moralne sodbe v primeru nizko-čustvenega scenarija ne vpliva, vpliva pa na moralne sodbe v primeru visoko-čustvenega scenarija. Po drugi strani pa so Nahmias et al. (2007) pokazali, da predstava deterministočnega sveta z jezikom nevroznanosti (mehanicistični opis) zmanjša pripisovanje moralne odgovornosti.

2.1 Moralne sodbe v deterministočnem svetu

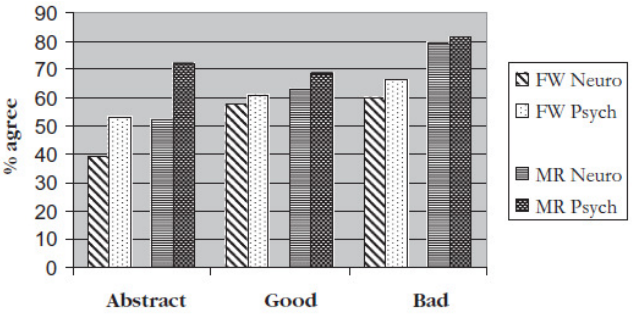
V eksperimentu Nicholsa in Knobeja (2007) so subjekti dobili opis deterministočnega in indeterministočnega vesolja in vprašanje katero vesolje se jim zdi bolj podobno našemu. Večino subjektov (preko 90%) je reklo, da je indeterministično vesolje bolj podobno našemu. Potem sta dala subjektom visoko-čustven ali nizko-čustven scenarij v deterministočnem ali indeterministočnem vesolju (4 scenariji). V visoko-čustvenem scenariju sta udeležence vprašala naslednje: »Kot je Bill storil že velikokrat v preteklosti, zasleduje in posili tujko. Je možno, da je Bill popolnoma moralno odgovoren za posilstvo tujke?« V nizko-čustvenem scenariju pa: »Kot je Mark storil že velikokrat v svoji preteklosti, goljufa pri plačevanju davkov. Je možno, da je Mark popolnoma moralno odgovoren za goljufanje pri plačevanju davkov?« V deterministočnem scenariju je manjšina (23%) udeležencev odgovorila, da je Mark za goljufanje davkov moralno odgovoren, a po drugi strani je večina (64%) odgovorila, da je Bill popolnoma odgovoren za posilstvo. V indeterministočnem scenariju je večina osebkov moralno odgovornost pripisala tako goljufu (89%) kot posiljevalcu (95%). Nichols in Knobe rezultate razlagata s tem, da imajo čustva pomemben vpliv na naše sodbe o svobodi in moralni odgovornosti in zato naše moralne sodbe v čustvenih scenarijih niso odvisne od deterministočnega ali indeterministočnega pogleda na svet. Roskies (2006) tako sklepa, da tak pogled, če je pravilen, »kaže na to, da dejanski psihološki procesi, ki so vključeni v vsakodnevne moralne sodbe o odgovornosti, najbrž večinoma delujejo neodvisno od teoretičnih pogledov na determinizem in mehanizem.« (Roskies 2006, str. 422). Nahmias et al. (2007) so kritizirali opisani eksperiment, češ, da deterministočnega sveta ni predstavil mehanistično – v svojih eksperimentih so poskusili pokazati, da mehanicistični (nevrološki) opis sveta (in osebe v njem), močneje vpliva na moralne sodbe kot psihološki opis.

2.2 Moralne sodbe v mehanicističnem svetu

Nahmias et al. (2007) so predpostavili, da če subjektom eksperimenta predstavijo deterministični svet tako, da bodo ljudje privzeli mehanistično razlago sveta in oseb v zgodbi, potem bodo osebam pripisovali manj svobodne volje in moralne odgovornosti. V enem scenariju so svet in odločanje osebe v zgodbi opisali v smislu nevroznanstvenih, mehanicističnih procesov (nevro-scenarij) v drugem pa v smislu psiholoških, intencionalnih procesov (psiho-scenarij). V prvem setu scenarijev so opisali alternativni svet Erto, za katero so udeležencem rekli, da je zelo podobna našemu svetu, in da naj si predstavljajo (čeprav mislijo, da je v našem svetu drugače), da je to kar pravi scenarij resnično na Erti. Ta dva osnovna scenarija (nevro in psiho) so predstavili še v treh različicah: Erto in osebo na Erti so opisali ali abstraktno ali osebo povezali z dobrimi ali s slabimi dejanji. Zadnji abstraktni paragraf zgodbe, ki je opisoval dejanje osebe, so nadomestili z dobrim (donirati denar) ali salbim (ubiti ženo) dejanjem: Tako so skupaj dobili 6 različnih scenarijev (nevro-abstraktni in psiho-abstraktni, nevro-dober in psiho-dober ter nevro-slab in psiho-slab scenarij, slika 3). V drugem setu scenarijev pa so Erto primerjali z resničnim svetom, spet z nevro in psiho opisom determinizma, vendar samo z različico abstraktnega opisa sveta in osebe v zgodbi (skupaj štirje dodatni scenariji, slika 2). 1124 dodiplomskih študentov so preko spletnega vprašalnika prosili, naj povejo, koliko se strinjajo ali ne-strinjajo (na lestvici od 1-6) s trditvami, npr. o tem, da so osebe iz zgodb zmožne odločanja iz svoje lastne svobodne volje, o tem, da imamo lahko osebe iz zgodbe za moralno odgovorne za svoje odločitve … (skupaj 10 vprašanj).



Slika 2: Veliko manj udeležencev je osebam v abstraktni različici nevro-scenariju v nasprotju s psiho-scenarijem pripisalo svobodno voljo (FW) in moralno odgovornost (MR).



Slika 3: Procenti udeležencev, ki so se strinjali s tem, da se osebe odločajo iz svoje svobodne volje (FW) in s tem, da so moralno odgovorne za svoje odločitve (MR). V nevro-scenariju so udeleženci osebam večinoma pripisovali manj FW in MR v primerjavi s psiho scenarijem.

Ti eksperimenti kažejo na to, da nevrološki, mehanicistični opis sveta v primerjavi s psihološkim, intencionalnim, na splošno zmanjšuje naše pripisovanje svobodne volje in moralne odgovornosti drugim osebam, vendar mislim, da je

to posledica tega, kako pojmujemo sebe, zavestno volicijo in nasploh mehanicistično razlago duševnih procesov.

3 ZAKLJUČEK: STRAH PRED MEHANICISTIČNO RAZLAGO DUŠEVNIH PROCESOV

Zdi se, da tako, kot nas je bilo strah tega, »da zemlja ni središče vesolja, temveč droben delec nekega svetovnega sistema« (Kopernik) in da je »biološko raziskovanje spodilo domnevno človeško prednost v stvarstvu in mu odkazalo izvor v živalskem svetu« (Darwin), nas je strah tega, da niti Jaz »ni gospodar v lastni hiši, pač pa se mora zadovoljiti z bornimi sporočili o tem, kar se nezavedno dogaja v njegovem duševnem življenju.« (Freud) (S. Freud; iz: Bahovec 1990, str. 3). Zdi se, da razlaga duševnih procesov in najbolj »intimnih« delov nas samih z nezavednimi mehanizmi pomeni konec tiste čarobnosti, ki jo včasih občutimo, ko se zazremo sami vase ali ko zremo v svet. Občutek nam pravi, da čarobnost lahko ostane le, če sebe in zavest ohranimo kot nekaj posebnega, nekaj ločenega od »krutega« sveta zunanjega determinizma in »ohromljujoče« moči nezavednih mehanizmov. Mislim, da ta predstava izhaja iz naših intuicij o tem, kaj zavestna volja in mi sami smo v povezav s svetom in možgani. Mislim, da nevroznanstvena spoznanja in mehanicistične razlage ne kažejo na izgubo vloge zavestne volje in ne bi spreminjale našega pogleda na moralno odgovornost in moralne sodbe, če bi sami sebe in zavestno voljo pojmovali drugače:

a) Naše volje ne moremo več videti kot neke nepovzročene moči, preko katere imamo popolno vzročno kontrolo nad sabo, ampak kot proces, ki je le del vzročne verige, ki vodi do nekega dejanja. Kot proces, ki ni edini vzrok za neko dejanje ali stanje, a vendar nadvse pomemben pri kreiranju tega, kar smo.

b) Velikokrat raziskovalci volicije predpostavljajo, da smo mi, naša zavest in naš jaz nekaj zunanjega ali determinističnemu svetu ali možganskim mehanizmom in posledično ne moremo biti svobodni oz. nimamo zavestnega vpliva na svoja lastna dejanja. Dennett npr. pravi: »… vso dejavnost, ki jo opravlja namišljen homunkul v kartezijanskem gledališču, moramo razbiti in porazdeliti po prostoru in v času v možganih. … Če se napravimo zelo majhne, lahko pozunanjimo skoraj vse.« (Dennett 2003, str. 237- 38). Mislim, da problem lahko razrešimo tako, da sebe ne izvzamemo ne iz deterministicnega vesolja, ne iz mehanicistično-materialnega sveta možganov, ampak se postavimo nazaj kot intrinzičen del le teh. Skratka, sebe (tega »jaza«, ki se npr. odloča) ne smemo razumeti kot ločenega od mehanicistično-deterministicnega sveta.

c) Za ohranitev vpliva zavestne volje bi morali razrešiti tudi vprašanje, kako nastanejo, kaj so in kakšno vlogo igrajo zavest in subjektivni občutki, česar še nismo uspeli storiti. Morda pa je stvar bolj preprosta in bomo čez čas zavest jemali kot del materialnega sveta tako samoumevno kot danes jemljemo to, da zemlja kroži okrog sonca in da se bomo na takšen način govora o zavesti preprosto navadili. Naj zaključim z mislijo J. Levineta: »Zdi se, da mentalne lastnosti ne moremo videti enake njenemu fizičnemu korelatu. A čeprav nas naša nezmožnost, da bi to videli, resnično bega, to ne dokazuje in ne more dokazati, da v resnici nista ista stvar. Kajti naša predstava o stvari ne more jamčiti za to, kakšna je stvar v resnici.« (Levine 1999) Vprašanje o naravi zavesti ostaja odprto.

LITERATURA

Bahovec E. D. 1990. *Kopernik Darvin Freud, Študije iz filozofije in zgodovine znanosti*. Društvo za teoretsko psihoanalizo (zbirka Analecta), Ljubljana

[1] Dennett C.D. 2003. *Freedom Evolves*. London, Penguin Books

[2] Haynes J.D. 2011. Decoding and Predicting Intentions. *Annals of the New York Academy of Sciences 1224*, 9-21

[3] Haynes J.D. et al. 2007. Reading Hidden Intentions in the Human Brain. *Current Biology 17*, 323-328.

[4] Levine J. 1999. Conceivability, Identity, and the Explanatory Gap, *Towards a Science of Consciousness 3*: cognet.mit.edu/posters/TUCSON3/Levine.html

[5] Libet B. 1985. Unconscious cerebral initiative and the role of conscious will in voluntary action. *The Behaviour and Brain Sciences 8*, 529-566.

[6] Nahmias E., Coates J.D., Kvaran T. 2007. Free Will, Moral Responsibility, and Mechanism: Experiments on Folk Intuitions. *Midwest Studies in Philosophy*, XXXI, 214-242

[7] Nichols S., Knobe J. 2007. Moral responsibility and determinism: The cognitive science of folk intuitions. *Nous 41*:4, 663–685

[8] Roskies A. 2010. How does neuroscience affect our conception of volition? *The Annual Review of Neuroscience 33*, 109-130.

[9] Roskies A. 2006 Neuroscientific challenges to free will and responsibility. *Trends in Cognitive Science 10*(9), 419-423.

[10] Roskies A. 2002. Neuroethics for the New Millenia. *Neuron 35*, 21-23.

[11] Soon et al. 2008. Unconscious determinants of free decisions in the human brain. *Nature Neuroscience 11*(5), 543-545.

[12] Wegner D.M. 2002. *The Illusion of Conscious Will*. London, Massachusetts, MIT Press (Bradford Books).

[13] Wegner D.M., Wheatley T. 1999. Apparent Mental Causation, Sources of the Experience of Will. *American Psychologist*, July, 480-492

RAZGRADNJA IZKUŠNJE SPROŠČENE TELESNE DRŽE IZ TAIJI QUANA

Jernej Sever

Filozofska fakulteta, oddelek za filozofijo
Raziskavo delno sofinancira Evropska unija, Evropski socialni sklad

IZVLEČEK

V tem članku bomo predstavili zastavitev triletnega raziskovalnega dela, ki raziskuje povezavo med sproščeno človeško držo, čustveno stabilnostjo in samopodobo. Hipoteza, da taka povezanost obstaja, izhaja iz mnogih izkušenj različnih meditativnih in telesnih tehnik. Poskušali bomo nakazati, kako je mogoče, s pomočjo fenomenološke metode kompleksno izkušnjo sproščene in optimalno uravnotežene drže pri taiji quanu, prevesti na merljive in empirično preverljive parametre. Dobljeni parametri bodo določali empirični okvir znotraj katerega se imenovana izkušnja lahko zgodi in nam nudili pomoč pri razvoju novih telesnih tehnik, metod in naprav, ki bi omogočale hitrejše, učinkovitejše in nadzorovano približevanje omenjeni izkušnji.

1. UVOD

Mnogo tradicij, ki se ukvarjajo s telesnimi ali meditativnimi tehnikami ter holistični načini zdravljenja, poudarjajo tesno povezanost telesnih in kognitivnih procesov. (W. Myers - rolffing metoda, Alexandrova tehnika, Zheng Manqing - taiji quan). Te izkušnje vedno bolj potrjujejo tudi moderne nevrološke raziskave (A. Damasio, Le Deux) in jih prevezama tudi kognitivna filozofija (R. W. Gibbs, A. Clark). V našem raziskovalnem delu bi radi na podlagi empirično preverljivih metod pokazali, da lahko že vzpostavljanje sproščene optimalno uravnotežene drže in ohranjanje te drže v gibanju vpliva na našo izkušnjo, samopodobo in zavest. Ustrezni empirični parametri bi nam lahko pomagali pri oblikovanju učinkovitih vadbenih modelov in naprav, ki bi na preverljiv in kontroliran način omogočili omenjeno izkušnjo. Ta izkušnja je pomembna zato, ker nam omogoča izbiro optimalnejših vedenjskih strategij na nivoju telesa in višjih kognitivnih procesov, s tem pa lahko s pomočjo lastne volje vplivamo na naše zdravje, počutje in tudi našo medosebno interakcijo. V primeru taiji quana, gre pri izkušnji sproščene drže in čustvene stabilnosti, za zamenjavo določenih vedenjskih strategij, ki se dogajajo na nivoju motoričnih in čustvenih procesov. Zamenjava je možna samo v primeru, da zagotovimo ustrezno okolje, ki jo omogoča. Okolje v tem primeru v prvi vrsti predstavlja učitelj in skupina, ki tako znanje goji in prenaša naprej. Uspešnost učenja je tako zelo odvisna od izkušenj in znanj posameznega učitelja, ta pa se lahko zelo razlikujejo. Menimo, da lahko z ustreznim poznavanjem mehanizmov, ki določajo in omogočajo sproščeno in pravilno poravnano držo, oblikujemo optimalnejše okolje, ki omogoča bolj nadzorovano in učinkovitejše učenje. V ta namen smo razvili inovativno obliko sproščanja, ki ima glede na prve izkušnje zelo dobre učinke in je primerna za uporabo pri eksperimentih.

2. OSNOVNA HIPOTEZA IN UTELEŠENJE

Koncept utelešene kognicije se nanaša na interakcijo med možgani, telesom in biološkim/fizičnim in kulturnim okoljem (Gibbs, 2006). Naše zavedanje, samopodoba in mehanizmi, ki obvladujejo naša vedenja, so vpeti v to interakcijo. Gibbs poskuša na različnih primerih pokazati, kako se lahko na osnovi ustreznih telesnih izkušenj oblikujejo tudi abstraktni simboli in pomeni. Celo popolnoma abstraktni koncepti kot so ravnovesje in svoboda imajo svoje reference v konkretnih situacijah, relacijah, stvareh in občutkih. Vsak posameznik si na osnovi lastnih izkušenj in vpetosti v njemu dane socialne, kulturne in druge mreže na svoj način uteleša dane pomene. Zato posamezni simboli ne drsijo le po mreži simbolnih pomenov, kot bi to ponazorila psihoanaliza (Lacan, 1996), ampak drsijo po mreži kulturnih, socialnih, bioloških (naše zaznave in občutki) svetov, ki se križajo v vsakem posamezniku. Ravnotežje je eden izmed takih konceptov, ki je v prvi vrsti zgrajen na osnovi naše telesne izkušnje, ki vsebujejo mrežo vizualnih, avditornih, kinestičnih in taktilnih informacij. Na osnovi te izkušenj se lahko oblikuje metaforični pomen, ki v kombinaciji z različnimi konteksti tvori najrazličnejše pomene ter relacije »ravnovesje sil«, »politično ali socialno ravnovesje«, »je neuravnovešena osebnost«. Na osnovi tega pogleda lahko predvidimo, da nam nova izkušnja optimalnega telesnega ravnotežja in poravnave, ohranjanja tega ravnotežja v gibanju in borbi, ponudi novo vsebino, ki

lahko na novo razporedi simbole in pomene, ki se nanje pripenjajo.

To pa pomeni, da tudi naša zavest nastaja na križanju vseh teh nivojev. Človek, ki je imel spiritualno ali transcendentalno izkušnjo ima drugačen pogled na svet in samega sebe, podobno velja za avtista ali druga patološka stanja zavesti. V našem primeru se preko izkušnje iz tai ji quana, ki ji bom poskušal določiti ustrezen empirično preverljiv okvir, lahko naučimo vzpostavljati optimalno telesno držo, ta je povezana s pravilno poravnavo telesa in sprostitvijo kronično napetih mišičnih skupin. Na ta način se, vsaj glede na izkušnje in izpovedi, oblikuje nova izkušnja ravnotežja, ki vpliva tudi na naše zaznavanje samega sebe in okolice. Spremenjen način odzivanja je še bolj očiten v stresnih situacijah, kjer za obrambo pred napadalcem ne uporabimo več avtomatičnih prirojenih odzivov, ampak ta odziv zamenjamo z drugačno, bolj optimalno gibalno strategijo.

Na osnovi tega lahko postavimo osnovno hipotezo, ki jo bomo poskušali potrditi v našem raziskovalnem delu.

Vzpostavljanje ozaveščene optimalne sproščene telesne drže lahko vpliva na čustveno stabilnost in na našo zavest.

3. METODOLOGIJA

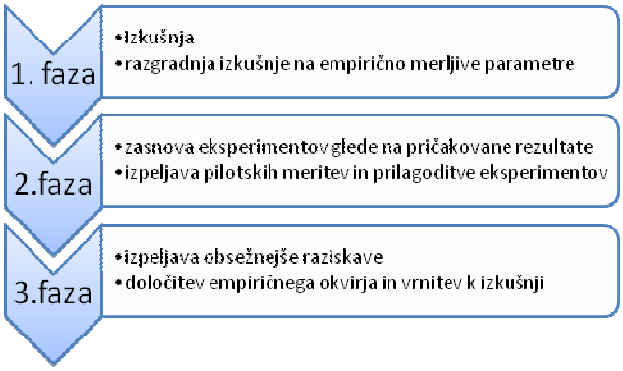
Pri poizkusu konceptualizacije zavesti lahko izpostavimo dva osnovna in nasprotujoča pristopa. Eden je fenomenološki, ki zavest raziskuje predvsem iz vidika prvo osebne izkušnje. Drugi je empiričen, pretežno nevrološki, kjer se raziskovanje zavesti osredotoča predvsem na opazovanje empirično merljivih procesov v telesu in možganih (Northoff, 2004). Pri raziskovalnem delu bomo poskušal upoštevati oba omenjena pristopa, izhajali bomo iz osebnih izkušenj, ki so v taki ali drugačni obliki tudi zapisane in poskušali najti empirično merljive okvirje, znotraj katerih se take prvo osebne perspektive lahko oblikujejo.

3.1. Fenomenološka metoda

Husserlovo maksimo »nazaj k stvarem« lahko razumemo predvsem kot postavitev izkušnje na prvo mesto. Iz vidika fenomenologije nas zanima, kako stvari izkusimo in jih doživljamo (Zahavi, 2008). Preprosto povedano, s pomočjo fenomenološke metode neko izkušnjo vzamemo kot relevantno samo po sebi in jo poskušamo razgraditi na elementarne parametre, oziroma na tisto, kar omogoča samo to izkušnjo (Zahavi, 2008). Zahavi omenja »front loading« fenomenologijo, kjer na njeni osnovi vzpostavimo predpostavke, na osnovi katerih lahko zastavimo in izpeljemo ustrezne eksperimente. Ena izmed podkategoriji take uporabe fenomenologije je tudi nevrofenomenologija. Na podoben način bomo tudi sami zastavili naše raziskave. Ideja za samo raziskovanje in tudi zastavitev eksperimentov se bo oblikovala na osnovi lastnih izkušenj in izkušnje nekaterih učiteljev in mojstrov, ki se ukvarjajo s tem področjem. Nekatere izkušnje so tudi zapisane, vendar se pri njihovem opisu navadno uporablja metaforične jezik, ki izvira iz drugačnega kulturnega konteksta in ne uporablja znanstvene terminologije, zato jih je brez lastne izkušnje težko prevesti na konkretne empirične parametre.

3.2. Empirična metoda

Empirične raziskave nam bodo omogočila preverjanje predpostavk, ki smo jih postavili na osnovi fenomenološke razgradnje izkušnje. Empirični del raziskave je razdeljen na različna znanstvena področja. Pri tem prevladujeta predvsem kineziologija in delno nevrologija, za merjenje optimalne telesne poravnave, ravnotežja in gibanja ter psihologija za merjenje čustvene stabilnosti, spremembe v samopodobi in učinka na stresno izpostavljenost. S pomočjo nevrologije bom poskušali preverjati direktne učinke, ki jih ima lahko sproščanje določenih mišičnih skupin na delovanje možganov. Na nivoju kineziologije bomo merili učinkovitost poravnave na dane sile. Z obstoječimi in prilagojenimi psihološkimi vprašalniki bomo merili dolgotrajnejše učinke izbranih vadbenih metod na čustveno stabilnost, anksionost in samopodobo.



skica 1: shema raziskovalnega procesa

4. PRIMER RAZGRADNJE IZKUŠNJE SPROŠČENE TELESNE DRŽE PRI TAIJI QUANU

Poznavanje izkušnje je eden izmed poglavitnih delov zastavljene raziskave. Prav poznavanje omogoča njeno razgradnjo, ustrezno izpeljavo raziskav, prilagoditve eksperimentov in ob zaključku postavitev empiričnega

okvirja, ki to izkušnjo do neke mere pogojuje. Na nivoju interpretacije kompleksne subjektivne izkušnje, čustvene stabilnosti in notranjega miru ima borilna veščine, kot je taiji quan, določeno prednost pred drugimi meditativni in telesnimi tehnikami, ki se ukvarjajo s podobno tematiko. Pri njej je najdenje notranjega miru, vedno povezano s praktično učinkovitostjo, ki se mora izkazati ali pa se jo vedno lahko preveri v borbeni situaciji. Ustrezno odzivanje v borbeni situaciji je povezano s čisto določenim čustveno-motoričnim stanjem, ki ima svoje psihološke in fiziološke razsežnosti, glede učinkovitost odzivanja, pa je mogoče postaviti empirični okvir v katerem se ta izkušnja zgodi.

a1) **Metaforični opis** (Zheng Manqing, 1985): Celo telo se mora sprostiti. Če nisi sproščen je tvoje telo kot boksarska vreča. Če se lahko sprostiš popolnoma, potem je ostalo preprosto. Ne smeš dovoliti niti najmanjše napetosti. Če se nekdo sprosti popolnoma, se sprostijo tudi vezi in žile in telo se potopi navzdol. Potopitev (Chen) pomeni ne plavajoč. Če nam uspe potopiti še »či«, potem se zbere Shen.

a2) Komentar: Telo lahko sprostimo takrat, ko ga pravilno poravnamo. Če telo ni pravilno poravnano, rabimo odvečno moč, da se upiramo gravitaciji. Posamezne mišice se aktivirajo zato, da kompenzirajo napake v drži. Naša drža naravno ni optimalna, ker pa se v njej počutimo domače, je težko ozavestiti napetosti in jih sprostiti, zato sprostitev ni preprost proces in zanjo potrebujemo ustrezno okolje, ki omogoča in dopušča učenje. Ustrezno okolje lahko predstavlja učitelj z ustrezno izkušnjo, lahko pa bi bila tudi naprava, ki nudi pravo povratno informacijo. Koncept čija je zelo širok koncept, ki zaznamuje fiziološke in psihološke učinke. V tem primeru lahko to metaforo dokaj ustrezno zamenjamo s pozornostjo. Šele, ko smo pozorni na optimalno poravnavo in je vsak delček našega telesa sproščen, naša pozornost pa poteka od pet pa do temena glave, potem steče či. V primeru, da pozornost izgubimo, lahko pride do odvečnih mišičnih napetosti in či se zaustavi. Ko se pozornost združi z sproščenostjo in potopitvijo dobimo Shen (duh, bog ali bolje figurativno zavest, ali mogoče celo bolje gonilna moč), ko ta občutek dobimo, ga ohranjamo v gibanju in v borbi. Borbena situacija ne prinese nobene spremembe

a3) **Kineziološka interpretacija:** Ustrezen kineziološki koncept, ki bi ustrezal tej izkušnjo bi lahko bil optimalna poravnava. » Idealna poravnava pri stoji omogoča telesu, da ohranja ravnovesje z uporabo minimalne notranje energije« (Massion, 2004). Telesna poravnava lahko zmanjša efekt gravitacijske sile, ki nas vleče iz centra. Informacije o velikosti teh sil centralni živčni sistem dobiva po vsej verjetnosti iz golgijevega tetivnega organa, ki meri napetost posamezne motorične enote. Te receptorje Massion imenuje gravireceptorji in jih obravnava kot četrto kategorijo senzorjev, ki skrbijo za ohranjanje ravnotežja. Meni, da gravireceptorji merijo silo, ki jo posameznik uporabi, da se upira zunanjim silam. V stoječem položaju ti receptorji merijo vektor sile na vsakem sklepu, ki omogoča posamezniku, da se upira sili gravitacije, ta informacija pa prispeva k notranji reprezentaciji vertikalne osi.

5. VZPOSTAVITEV OPTIMALNEGA UČNEGA OKOLJA

Predvidevamo lahko, da se notranja reprezentacija vertikalne osi razlikuje od človeka do človeka in da imajo ljudje navadno občutek, da je njihova poravnava pravilna, čeprav lahko zelo odstopa od nekih optimalnih vrednosti. Taiji quan poskuša oblikovati okolje, kjer lahko našo držo izboljšamo, to izboljšano držo pa uporabimo tudi v gibanju in borbi. Ker pa je metoda odvisna predvsem od znanja in izkušnje učitelja, učenje ni preveč učinkovito. Tudi dober učitelj lahko učenca učinkovito uči le v primeru, da nenehno dela z njim in mu tako daje konstantno povratno informacijo o učinkovitosti njegovega gibanja. Učenje taiji quana traja mnogo let. Iz vidika metode in dolgotrajnosti taijiquan ni preveč primeren za uporabo pri eksperimentih. Zato smo poskušali razviti tehniko, s katero bi omenjene učinke dosegli hitreje. Osnovni cilj taiji quana je doseči sproščeno in tekoče gibanje brez odvečnih in nenadnih kontrakcij posameznih mišičnih skupin. To je možno le, če so mišice, ki obdajajo sklepe sproščene. Mojstri taiji quana to stanje včasih imenujejo »odprti sklepi«. Pri metodi sproščanja smo se zato osredotočili predvsem na sproščanje najbolj problematičnih sklepov pri vadbi taiji quana, ramen in kolkov.

5.1. Opis postopka sproščanja

Postopek je preprost. Masiranec se usede na stol, med dve stojali ali med dve mizi, ki sta v taki višini, da roki sproščeno počivata na stojalu. Roki postavimo na deske , ki so opremljene s skodeličnimi ležaji. Ti omogočajo tekoče gibanje brez trenja v vse strani. Masiranec poskuša ves čas ohranjati pokončno poravnavo. Maser začne krožiti z njegovo roko po podstavku. Kroži z velikimi počasnimi krogi. Na začetku se roka masiranca upira kroženju, na nekaterih mestih se mišice napnejo in zaustavijo gibanje. Na mestu kjer se mišice napnejo se maser zaustavi in z rahlim tresenjem ali pritiskanjem aktivirane mišice začne mišico sproščati in jo podaljševati (masiranec pri tem zavestno sodeluje), ko se mišica sprosti, sledi naslednja in tako naprej, dokler v roki ne nastane konstanten rahel upor brez

nenadnih kontrakcij. Podoben postopek ponovimo tudi s kolčnimi sklepi.
Glede na učinke menimo, da na nivoju golgijevaga tetivnega organa in mišičnega vretena vnesemo nove vrednosti o napetosti in dolžini posameznih motoričnih enot. Te vrednosti se vzpostavljajo glede na optimalno in sproščeno poravnavo telesa. Mišice in tkiva ki obkrožajo sklepe so po sproščanju bolj sproščenie ob sklepih ni odvečnih napetosti, kar telesu omogoča, da se poravna bolj optimalno. Na subjektivnem nivoju je učinek mogoče testirati na tak način, da v stoječem položaju masiranca pritisnemo z dodatno silo v smeri gravitacije. Po masaži v telesu masiranca ni opaznih skoraj nobenih sprememb, medtem ko pred masažo telo ob dodanem pritisku ni tako stabilno in lahko rahlo zaniha.

6. ZASTAVITEV OSNOVNEGA EKSPERIMENTA IN UPORABA PSIHOLOŠKIH VPRAŠALNIKOV

Izpostavil bom predvsem področje kineziologije in psihologije. Raziskovalno delo pa bo vključevalo tudi druge eksperimente. Na področje kineziologij bomo s pomočjo dveh pritiskovnih plošč, emg-ja, 3d scanerjev ter merilcev sile, merili odzivanje telesa v stoječem položaju v primeru dodatne sile v smeri gravitacije, pred in po masaži. Izvedli bomo pilotske meritve, po potrebi eksperimente ustrezno prilagodili in izbrali najbolj primerne metode za objektivno, ponovljivo in občutljivo vrednotenje optimalne telesne poravnave. Glede na izkušnje in subjektivne oblike testiranj, bi lahko pričakovali, da ima optimalno poravnano telo manjša nihanja COG na pritiskovni plošči, predvsem pa, da so horizontalne sile na tla ob dodatni obremenitvi manjše. Prav tako lahko pričakujemo, da so pri optimalni poravnavi navori na sklepih manjši, oziroma minimalni, kar pomeni potrebo po manjši mišični aktivaciji za uspešno upiranje gravitaciji in dodani sili. Nepravilno poravnano telo mora nepravilnosti kompenzirati z dodatnimi stabilizacijskimi strategijami, ki pa povzročijo nihanje sklepov in COG.

Na nivoju psihologije bomo na večji populaciji s pomočjo ustreznih vprašalnikov merili psiholoških učinke dolgotrajnejše vadbe na čustveno stabilnost, samopodobo in izpostavljenost stresu. Vprašalnike bomo uporabljali sočasno z izvajanjem izbranih merjenj kvalitete poravnave. Meritve se bodo izvajale sočasno s prenosom izbranih eksperimentov na večje skupine testirancev, prostovoljcev ali izbranih skupin znotraj podjetji. Na osnovi statističnih modelov bomo obdelali dobljene rezultate in ugotavljali povezanost med sproščeno telesno držo in čustveno stabilnostjo.

7. PRIČAKOVANI REZULTATI

Glede ne izkušnje in poznavanje telesnih tehnik in izbora testiranja pričakujemo, da bi se morala osnovna hipoteza potrditi. Največ težav pričakujemo s pilotskimi meritvami na nivoju optimalne poravnave in vzpostavitvijo najbolj primerne metode za njeno merjenje. Na praktičnem nivoju bo mogoče rezultate raziskave uporabiti za oblikovanje empirično preverljiv vadbenih modelov in naprav, na področju boljšega soočanja s stresom, boljšega zdravja in počutja.

LITERATURA

1. Clark, A. (1998).*Being There: Putting Brain, Body, and World Together Again*. MIT Press.

2. Damasio, A. (1999). *The Feeling of what Happens: Body and Emotion in the Making of Consciousness*. New York: Harcourt Brace and Co,.

3. Gibbs, R. W. (2006). *Embodiment and Cognitive Science*. Cambridge: Cambridge University Press.

4. Gallagher, S., Zahavi, D. (2008). *The Phenomenological Mind: An Introduction to Philosophy of Mind and Cognitive Science.* London, New York: Rautlage.

5. Lacan, J. (1996): *Štirje temeljni koncepti psihoanalize*. Ljubljana: Analecta.

6. Bronstein, A. M.(2004). *Clinical disorders of balance, posture and gait*. London:. Arnolod cop.

7. Zheng Manqing (1985). *Zheng Tzu's Thirteen treatises on taiji quan*. Berkely:. Blue snake books.

NEKATERE ZNAČILNOSTI POVEZAVE MED UMOM IN TELESNIMI SISTEMI

Maja Smrdu[1]

[1]Psihiatrična klinika Ljubljana, Studenec 48, Ljubljana
e mail:maja.smrdu@psih-klinika.si

Povzetek
V klinični praksi se vse bolj soočamo z različnimi somatskimi boleznimi, pri katerih je vedno bolj razviden vpliv psiholoških dejavnikov, zlasti način razmišljanja in čustvovanja. Vedno več je potrditev, da tovrstni procesi vplivajo na endokrini, imunski in nevrološki sistem, kar se izraža v pojavnost bolezni, poteku bolezni (dolgotrajnosti in intenzivnosti) ter tudi v posledični smrtnosti. V prispevku so predstavljene različne meta študije, ki dokazujejo tovrstno sovplivanje ter okvirna predstavitev mehanizmov, ki pripomorejo k temu.
Glede na obseg prispevka so predstavitve in razlage bazične ter poenostavljene.

Preden se lahko spustimo v integracijo osnovnega problema prispevka, je pomembno, da definiramo posamezne koncepte, ki so ključni za razpravo o odnosu med umom in telesom.

Definicija uma
Za Damasia » um pomeni, da organizem formira nevrološke reprezentacije, ki lahko postanejo podobe, se jih lahko manipulira v procesu imenovanem misel in se z njimi postopno vpliva na vedenje s predvidevanjem prihodnih akcij, njihovega načrtovanja in izbire naslednje aktivnosti.« [1, str.90] Skladno z medicino, ki se osredotoča na povezavo uma in telesa pa se lahko definira um kot zavedne in nezavedne miselne vzorce, ki vključujejo podobe, zaznave in namere, ter jih generira funkcionalna mreža nevroloških centrov v možganih. Vsebujejo tudi homeostatične reprezentacije, ki omogočajo kontekst za posameznikovo samozavedanje in čustveno doživljanje [2,3].

Mehanizmi od zgoraj navzdol (top-down) in od spodaj navzgor (bottom-up)
Mehanizmi od zgoraj – navzdol so tisti, ki se začnejo z mentalnim procesiranjem na nivoju cerebralnega korteksa. V primeru predstavljanja ali meditacije, se primarno nanašajo na zavestno in namensko mentalno aktivnost, čeprav so nezavedni nevrološki procesi ravno tako vključeni. V nasprotju s tem pa se mehanizmi od navzgor- navzdol začnejo s stimulacijo različnih somato-, visceralno- in kemo-senzornih receptorjev [3], nato vplivajo na centralno živčno procesiranje in mentalne aktivnosti, ki vodi iz periferije do hrbtenjače in cerebralnega korteksa.

Kljub temu, da se razlikuje dvoje vrst mehanizmov, pa je ključno poudariti, da večinoma potekajo oboji mehanizmi istočasno. To upoštevajo tudi terapije, ki so usmerjene na povezavo uma in telesa. Npr. progresivna mišična relaksacija vključuje poti od navzdol-navzgor, ki so aktivirane s perifernimi senzornimi aferentnimi odgovori in vplivajo na različne visceralne aktivnosti (npr. znižana mišična napetost in krvni pritisk) ter poti od zgoraj- navzdol, ki so aktivirane s parasimpatičnim živčevjem [4].
Optimalno zdravje je odvisno od učinkovite izmenjave med perifernim in centralnim živčnim sistemom ter je močno povezano z eferentnim vagalnim tonusom. Komunikacija med vagalnimi potmi se lahko začne, olajšajo ali inhibira s hormoni, imunsko delujočimi proinflamatornimi citokini in/ali različnimi nevropeptidi ter nevrotransmiterji. Spremembe v duševnih procesih (npr. osredotočena pozornost, dojemanje stresa) se hitro izrazijo v telesu, preko upada avtonomnega, nevroimunskega in nevroendokrinološkega sistema. [5]
Ta okvir dvojne izmenjave poudarja več nivojev nevroloških osi na katerih se odvija interakcija, predvsem vlogo specifičnih frontotemporalnih kortikalnih regij pri reprezentaciji in kontroli neugodnih simptomov, ki pa so recipročno v interakciji s subkortikalnimi strukturami. Te so vključene v telesno homeostazo in odgovore na stres [5]. Te ideje so skladne z Damasijevo hipotezo somatskih markerjev [1], ki poudarja hierarhično kontrolo in reprezentacijo znotraj možganov in recirpročno interakcijo med nevronskimi mrežami.
Negativna čustva kot sta depresija in anksioznost imata direkten učinek na celice imunskega sistema in vplivata na sekrecijo proinflamatornih citokinov. Poleg tega negativna čustva tudi pripomorejo k podaljšanju ali kroničnosti infekcij ter podaljšajo čas celjenja ran.
Endokrini sistem služi kot centralni sistem, preko katerega vplivajo psihološki dejavniki na zdravje; stres in depresija namreč lahko izzoveta sprostitev pituitarnih in adrenalinskih hormonov, kar ima večplasten učinek na imunske funkcije [6]. Dodatno pa imajo ljudje pod stresom tudi za zdravje neugodna vedenja, ki jih pripeljejo k večjemu tveganju, npr. neurejen in pomanjkljiv spanec, večjo nagnjenost k uživanju alkohola in zlorabi drog, slabšo prehrano, manj vadbe in s tem imunske in endokrinološke posledice [7].

Vpliv na različne bolezni nevrološko-fiziološke spremembe

Kaj tovrstne spremembe različnih sistemov in vpliv uma, pomeni za posamezne bolezni? Začetek astme in njeno poslabšanje se običajno pojavita ob povečanih psiholoških stresorjih, katerih pogost spremljevalec je depresija. Remisija depresije pa je povezana z izboljšanjem astmatičnih simptomov in zmanjšano uporabo zdravil. Veliko je že bilo napisanega o odnosu med astmo in čustvi, vedno bolj pa se to povezavo razširja še z nevrološkim področjem. Sinteza literature s tega področja poudarja, poleg že znanih področij za regulacijo čustev, pomembnost vpliva insularnega in anteriorno cingularnega korteksa. Oba sta lahko povezana s telesnimi spremembami povezanimi z astmo in vplivata na pojavnost in vztrajanje simptomov [8] .

Za kardiovaskularne bolezni velja podobno kot za astmo. Meta analize kažejo, da imajo depresivni bolniki s koronarno srčno boleznijo dvakrat večje tveganje za smrtnost kot bolniki brez depresivnih simptomov. [9].

Na področju rakovih obolenje je obširna meta analiza 25 neodvisnih študij pokazala, da se stopnja mortalitete bolnikov z rakom, ki imajo depresivne simptome, poveča za 25% in se ob težji depresiji poveča na 39% [9]. Zadnja odmevna raziskava raka na dojkah podobno dokazuje, da je zmanjšanje simptomov depresije v prvem letu po diagnozi, statistično pomembno povezano s podaljšano življenjsko dobo. V skupini, kjer so se simptomi depresije zmanjšali v prvem letu je v povprečju ta znašala 53,6 mesecev, pri kontrolni skupini brez zmanjšanja depresije pa 25,1 mesec. Ob tem avtorji poudarjajo, da nobeden od kontroliranih in široko izbranih medicinskih in demografskih parametrov, ne razloži tovrstne spremembe [10].

Tako kot lahko opisana interakcija izzove in poslabša bolezni, pa lahko ta isti sistem uporabimo tudi za izboljšavo zdravstvenega stanja. Meditacija kot redna mentalna vadba vpliva na funkcijo in strukturo možganov. Meta analiza je pokazala, da vse študije poročajo o spremembi volumna in gostoti sive substance v možganskih predelih, ki regulirajo vzburjenje in čustvovanje: insuli, hipokampusu, prefrontalnem korteksu in hrbtenjači. Poleg tega ob takšni praksi ni najti običajnega starostnega upada sive možganovine [11]. Zlasti Damasio in raziskave, ki jih povzema Craig [3] sugerirajo da reprezentacije višjega reda visceralnih stanj ali vzorci vkodirani v insularnem in prefrontalnim korteksom [1] in povezani fiziološki odgovori ter vzorci misli reprezentirani znotraj anteironega cingularnega korteksa, so lahko tista točka, preko katere je mogoča sprememba funkcije tudi z uporabo terapij usmerjenih na interakcijo uma in telesa.

Tovrstne terapije se že v uporabi in dokazujejo svojo učinkovitost ravno preko uporabe opisanih mehanizmov. Primer tega je obravnava sindroma kronične bolečine, ki preko teorije kontrole vrat [12], modificira dejavnike, ki vplivajo na količin in vrsto informacij posredovanih v možgane. Pri tem poudarja zlasti vpliv misli; pri čemer negativne misli odprejo vrata in s tem spustijo več bolečinskih informacij, medtem ko pozitivne misli zaprejo brata in omejijo bolečinska sporočila.

Tabela 1: Dejavniki, ki odprejo vrata [12].

Telesni	Degenerativne spremembe, mišična napetost, zloraba drog
Kognitivni	Pozornost na bolečino, misli o ne kontroliranosti bolečine, prepričanja o skrivnostnosti bolečine, katastrofično mišljenje
Čustveni	Depresivnost, strah/ anksioznost, jeza
Aktivnost	Preveč ali premalo aktivnosti, slaba prehrana, neuravnoteženost med delom, socialno in rekreacijsko aktivnostjo
socialni	Majhna podpora družine in prijateljev, osredotočanje drugih na bolnikovo bolečino, pretirana želja drugih po zaščiti bolnika

Naše misli, predstave, čustvovanje, vedenje in fiziološke ter telesne reakcije so stalno povezane in soodvisne. Ta proces se dogaja v kontekstu okolja, ki dodatno in specifično vpliva na posameznika. Ob negativnih prepričanjih o bolečini in samemu sebi, pride do sprememb v čustvenem stanju, kar privede do izogibanja aktivnostim (zaradi strahu pred dodatnimi poškodbami, bolečinami ali občutki neadekvatnosti) ter posledičnega upada splošne telesne kondicije. Kar je pomembno ob temu je, da se s tako sproženo krožnostjo odprejo vrata za povečan dotok bolečinskih informacij [12]. Torej prispevek k povečanju bolečine je dvojen.

Pomembnost kognitivnih in čustvenih procesov se kaže tudi z uporabo placeba, ob katerem se ravno tako aktivirajo notranji mehanizmi, ki modificirajo fiziologijo. Poudarja se zlasti vpliv opiodnega in dopaminskega nevrotransmiterskega sistema na vključenost anteriorno cingulatornega področja, dorsolateralnega preforntalnega in orbitofrontalnega korteksa, insule, amigdale in medialnega talamusa [13].

Zaključek

V literaturi, ki obravnava odnos med umom in telesom, se vse bolj izpostavlja dvojna povezava delovanja mehanizmov od zgoraj navzdol (top-down) in od spodaj navzgor (bottom-up), ki potekajo med možgani in perifernimi tkivi, ter vključujejo kardiovaskularni in imunski sistem. Predstavljene so nekatere osnovne komponente integrativnega psihofiziološkega okvirja, ki poskuša pojasniti bazične substrate interakcije med umom in

telesom. Dvojna povezava avtonomnih in nevroendokrinoloških poti usmeri informacijo med CŽS in periferijo ter izzove afektiven, hormonalen in imunski odgovor. Natančen odnos med vsemi temi sistemi še ni znan, tako obstajajo zgolj posamezni modeli, ki se večinoma osredotočajo zgolj na povezavo uma z enim od omenjenih sistemov (npr. kot pri psihoimunologiji). Kljub temu pa se pojavlja vse več tendenc k združevanju te kompleksne celote. Hkrati, ko se spoznava interakcije med sistemi pa se razvijajo tudi razne terapije, ki poskušajo novo znanje uporabiti za izboljšanje zdravstvenega stanja in življenjske kvalitete bolnikov. Nedvomno pa je to področje, ki dobiva vse več veljave in raziskovalnega zanimanja.

Literatura

1. Damasio, A.R. *Descartes' Error: Emotion, Reason and the Human Brain.* New York, NY: Putnam. 1994.
2. LeDoux, JE. *Synaptic Self: How Our Brains bcome Who we are.* New York, NY: Viking. 2002.
3. Craig, A.D. How do you feel?Interoception: the sense of the physiological condition of the body. *Nat Rev Neurosci*. 3, pp.655-666. 2002.
4. Taylor, A.G., Goehler,L.E., Galper,D.I., Innes,K.E., Bourguignon, C. Top-down and bottom-up mechanisms in mind-body medicine: developmental of integrative framework for psychophysiological research. *Explore.* 6, pp. 29-41. 2010.
5. Gold, PW, Chrousos, GP. Organisation of the stress system and its dysregulation in melanholic and atypical depression: high vs low CRH/NE states. *Mol Psychiatry*. 7, pp. 254-275. 2002.
6. Rabin, B.S. *Stress, immune function, and health: The connection*. New York: Wiley-Liss & Sons. 2009.
7. Kiecolt-Glaser,J.K., McGuire, L. *Journal of Consul and Clinical Psy*. 20. pp 537-547. 2002.
8. Rosenkranz, M.A., Davidson, R.J. Affective neural circuitry and mind–body influences in asthma. *NeuroImage.* 47, 972-980. 2009
9. Satin, J.R. Linden, W. Phillips,M.J. Depression as a Predictor of Disease Progression and Mortality in Cancer Patients. *Cancer.* 15, pp. 5349-5361. 2009.
10. Giese-Davis, J**.,** Giese-Davis, J., Collie, K., Rancourt, K., Neri, E., Kraemer, H. C., & Spiegel, D. Decrease in depression symptoms is associated with longer survival in metastatic breast cancer patients: A secondary analysis. *Journal of Clinical Oncology* 29(4):413-420. 2011.
11. Walach,H., Schmidt, S., Jonas, W.B. *Neuroscience, consciousness and spirituality*. Springer. New York: 2011.
12. Otis, J.D. *Managing chronic pain.* Oxford university press. Oxford. 2007.
13. Zubieta,J, Stohler, C.S. Meurobiological mechanisms ofplacebo responses. *Annual N Y Acad Sci*. 1156, pp. 198-210. 2009.

VPLIV NAVIDEZNEGA OKOLJA NA ČUSTVENO STANJE IN OBČUTEK PRISOTNOSTI V TEM OKOLJU

Katja Žibrek, Anja Podlesek, Matjaž Mihelj, Domen Novak
Filozofska fakulteta
Oddelek za psihologijo
Aškerčeva 2, 1000 Ljubljana, Slovenija
katjazibrek@gmail.com

POVZETEK

Navidezna resničnost je medij, ki zaradi možnosti potopitve posameznika v umetno ustvarjeno okolje vzbuja čustva in občutek prisotnosti v tem okolju. Nekatere raziskave so povezale večji občutek prisotnosti z doživljanjem čustev v navideznem okolju in to predpostavko tudi potrdile. Ugotavljanje povezanosti med omenjenima pojavoma je bil glavni cilj tudi naše raziskave. Dvainštirideset udeležencev, pretežno študentov, je najprej izpolnilo vprašalnike PANAS in STAI X-1 in s tem poročalo o svojem začetnem čustvenem stanju, nato pa v naključnem vrstnem redu izkusilo tri različna navidezna okolja (nevtralno, sproščeno in anksiozno), pri čemer so po vsaki izkušnji poleg istih vprašalnikov rešili še vprašalnik prisotnosti SUS. Rezultati so potrdili, da navidezna okolja lahko vzbudijo različna čustva v predvideni smeri, nevtralno okolje pa je vzbudilo občutja dolgočasja, zaradi česar nismo mogli ugotoviti, ali se večji občutek prisotnosti povezuje s čustvenimi okolji ali samo z večjo stopnjo vzburjenja.

1 UVOD

Navidezna resničnost je medij, ki je širšemu raziskovalnemu svetu že poznan in postaja vedno bolj zanimiv tudi na področju psihološkega raziskovanja. Zaradi možnosti konstrukcije visoko kontrolirane in realno neogrožujoče situacije se navidezna resničnost uporablja predvsem za zdravljenje raznovrstnih anksioznih motenj, v zadnjem času pa je aktualno predvsem raziskovanje občutka prisotnosti v navidezni resničnosti ter dejavnikov, ki na ta občutek vplivajo.

Občutek prisotnosti (angl. *presence*) je posameznikov občutek, da se nahaja v navideznem okolju oz. da ima zaznavno iluzijo o neposredovanju [1], torej dojema navidezno okolje kot resnično in se na dražljaje iz njega tudi odziva. Obstaja načinov merjenja tega občutka, poznamo pa predvsem dva pristopa, in sicer subjektivne ter objektivne mere prisotnosti [2].

Pod subjektivne mere spadajo predvsem vprašalniki, pri čemer sta najbolj razširjena dva: vprašalnik Witmerja in Singerja (Presence Questionnaire - PQ, [3]) ter vprašalnik SUS [4]. Slednji se osredotoča zgolj na občutek prisotnosti pri posamezniku, vprašalnik PQ pa naj bi prepletal značilnosti posameznika z značilnostmi navidezne resničnosti oz. dejavniki potopitve v navidezno okolje [5].

Pri objektivnih merah prisotnosti najdemo predvsem vedenjski realizem, držo telesa, preizkus dvojne naloge; najbolj razširjene pa so fiziološke meritve, kjer raziskovalci spremljajo bitje srca, dihanje, prevodnost kože in temperaturo kože. Vseeno pa fiziološki pokazatelji niso direktna mera prisotnosti, zato služijo bolj kot dopolnilo drugim meram prisotnosti [2].

Na občutek prisotnosti vpliva več dejavnikov, predvsem pomembno vlogo pa igrajo značilnosti medija oz. navidezne resničnosti, ki s svojimi tehničnimi značilnostmi vpliva na to, da se posameznik potopi v navidezno okolje (angl. *immersion*) [6]. Pomembne so predvsem značilnosti zaslona (ali gre za zaslon, pritrjen na glavo ali prikaz na platnu), ozvočenje (*surround* sistem), haptični dražljaji, reprezentacija telesa v prostoru in telesna vključenost v okolju. Zanimivo je, da realizem vidnih dražljajev ni pomemben dejavnik – slika na zaslonu je lahko fotografsko natančna ali pa narejena iz preprostih črt – v obeh primerih bo posameznik imel zaznavno iluzijo o neposredovanju.

Eden najbolj pomembnih dejavnikov, ki naj bi vplival na to, ali se bo počutil posameznik prisotnega ali ne, je mera, do katere navidezna resničnost uspe vzbuditi čustva. Znano je, da različne vrste medijev lahko vzbudijo čustva pri posameznikih (npr. [7]) in enako velja za navidezno resničnost. Postavilo se je torej vprašanje, kako sta povezana občutek prisotnosti in čustva. Da bi podrobneje raziskali to povezavo, so strokovnjaki izdelali različna navidezna okolja, ki naj bi vzbujala čustvene odzive pri udeležencih in ki bi hkrati omogočala dobro mero potopitve v navidezna okolja, da bi lahko izzvala občutek prisotnosti v teh okoljih. V okviru tega projekta je nastalo kar nekaj pomembnih raziskav [8,9], ki potrjujejo povezanost med občutkom prisotnosti ter čustvi. V svoji raziskavi so uporabili tri različna navidezna okolja, prikazana na zaslonu, pritrjenem

na glavo – eno okolje, ki naj ne bi spremenilo čustvenega stanja udeleženca, in dve okolji, ki naj bi vplivali na čustveno stanje udeleženca (sproščeno in anksiozno okolje). Te razlike v okoljih so variirali predvsem z vremenom v navideznem parku, zvokom in barvami. Ugotovili so, da navidezno okolje lahko vzbudi čustva v predvideni smeri (torej anksiozno okolje vzbuja bolj anksiozna čustva, sproščeno pa bolj pozitivna čustva; v primeru nevtralnega okolja pa čustva ostanejo razmeroma enaka poročanim čustvom v začetnem stanju) in kar je še bolj presenetljivo, v primeru »čustvenih« okolij se poveča tudi posameznikov občutek prisotnosti. Niso pa odkrili nobenih razlik v prisotnosti glede na tip čustev (ni bilo pomembno, ali gre za sproščujoče ali anksiozno okolje). Avtorji so želeli še nadalje ugotavljati, kakšna je smer vplivanja, vendar so analize pokazale krožno povezanost – zdi se, da čustva vplivajo tako na občutek prisotnosti kot obratno.

V naši raziskavi smo se soočili s podobno nalogo, kot so si jo zadali Riva in sodelavci [9], vendar smo zaradi tehničnih omejitev namesto navideznih okoljih, osnovanih s pomočjo projekta EMMA, uporabili preprostejša okolja, medtem ko smo ohranili enake inštrumente merjenja čustev in prisotnosti. Iz tega vidika nas je zanimalo tudi, ali lahko »manj realistično« okolje, ki smo ga uporabili v naši raziskavi, vzbudi občutke prisotnosti in pripadajoča čustva tako, kot to ugotavljajo pri bolj realističnih preizkušnjah navideznega parka. Ker realizem vidnih dražljajev za občutek prisotnosti ni pomemben in ker naj bi bili zadoščeni tudi drugi vidiki potopljenosti v navidezno okolje, smo sklepali, da bodo navidezna okolja vzbudila čustva in občutek prisotnosti. Cilj naše raziskave je bil torej, da se čim bolj približamo raziskavam v okviru projekta EMMA z uporabo drugačnih navideznih okolij in bolj omejenimi tehničnimi zmogljivostmi našega laboratorija.

V okviru raziskave smo obikovali tri hipoteze:

- navidezna okolja lahko vzbudijo čustva v predvideni smeri;
- različna navidezna okolja se na področju vzbujanja čustev med seboj razlikujejo;
- občutek prisotnosti bo v čustvenih navideznih okoljih večji.

Glede na prvo hipotezo naj bi udeleženci v primerjavi z začetnim čustvenim stanjem po vključenosti v anksiozno okolje poročali o višji stopnji negativnih in anksioznih čustev; po sproščenem okolju pa pričakujem nižjo stopnjo negativnih in anksioznih čustev; po izkušnji nevtralnega okolja ne pričakujemo sprememb v poročanih čustvih glede na začetno stanje.

V zvezi z drugo hipotezo naj bi se pojvile razlike med poročanimi čustvi po izkušnji različnih navideznih okolij.

Pri zadnji hipotezi pa bi bil v nevtralnem okolju občutek prisotnosti manjši kot v primeru sproščenega in anksioznega okolja.

2 METODA

V eksperimentu je sodelovalo 42 udeležencev različne izobrazbe, prevladovali so študenti. Med udeleženci je bilo 23 žensk in 19 moških, starih od 19 do 35 let. Udeleženci, ki so zaradi čustvenih težav obiskali strokovno pomoč, so bili iz vzorca izločeni.

Za ugotavljanje čustvenega stanja udeležencev pred in po izkušnjah navideznega okoljasmo upoabili Vprašalnik negativnega in pozitivnega čustvovanja PANAS [10], vprašalnik STAI (State-Trait Anxiety Inventory, [11]), in sicer lestvico STAI X-1 s 20 postavkami.

Da bi preverili učinkovitost navideznega okolja pri vzbujanju čustev, so bili vsi udeleženci izpostavljeni trem različnim navideznim okoljem. Izvajali smo torej ponovljene meritve za tri različne scenarije: nevtralnega, anksioznega ter sproščenega. Zaporedje scenarijev se je pri udeležencih spreminjalo, in sicer je za tri scenarije (1, 2, 3) možnih 6 različnih zaporedij: 1-2-3, 1-3-2, 2-1-3, 2-3-1, 3-1-2 ter 3-2-1. Naključna razporeditev scenarijev je bila potrebna, da bi se izognili učinku zaporedja eksperimentalnih pogojev.

Navidezno okolje je bilo projicirano na platno dimenzije 3 × 2 metra, ki je bilo nameščeno približno en meter pred tekočim trakom, po katerem je hodil udeleženec med projekcijo navideznega okolja. Za zvočne efekte je bil nameščen *surround* sistem ozvočenja, z zvočniki, ki so se nahajali pred udeležencem, na vsaki strani platna in za udeležencem. Navidezni scenariji so bili izdelani pri Hacomi (Volketswil, Švica). Udeleženec je med sprehodom skozi navidezno okolje hodil po tekočem traku, katerega hitrost je bila prilagojena hitrosti gibanja skozi navidezno okolje.

Eksperiment je bil razdeljen na dva dela. V prvem delu so udeleženci z vprašalnikoma PANAS in STAI X-1 ocenjevali svoje začetno stanje pozitivnega in negativnega čustvovanja ter anksioznosti.

V drugem delu smo udeležence pospremili v prostor s projekcijskim platnom in jim naročili, naj stopijo na tekoči trak tako, da bodo s pogledom usmerjeni proti platnu. Vključili smo tekoči trak in pustili, da se navadijo na hojo po njem. Nato smo predstavili tri navidezne scenarije v naključnem vrstnem redu, ki so trajali vsak po 4 minute. Po vsakem končanem scenariju so udeleženci sedli za mizo in ponovno rešili enaka vprašalnika kot na začetku, hkrati pa dobil še tretji vprašalnik, ki je bil namenjen preverjanju občutka prisotnosti v navideznem okolju. Po treh preizkušnjah smo jih prosili še za njihov komentar.

Scenariji so bili trije: nevtralni, sproščeni in anksiozni scenarij. Vsi trije so prikazovali sprehod skozi mesto in čez

kanjon po mostu. Scenarije smo variirali s pomočjo elementov, prikazanih v tabeli 1.

Nevtralno mesto Nevtralno vreme in atmosfera Širok pločnik Brez avtomobilov Brez zvoka	Nevtralni kanjon Nevtralno vreme in atmosfera Širok most z ograjo Brez reke pod mostom Brez zvoka
Sproščeno mesto Sončni zahod Normalna širina pločnika Malo avtomobilov Prijetna glasba	Sproščeni kanjon Sončni zahod Normalna širina mostu z ograjo Lepa, čista reka Prijetna glasba
Anksiozno mesto Temačna atmosfera Dež, grom in bliskanje (zvok) Ozek pločnik Veliko avtomobilov	Anksiozni kanjon Temačna atmosfera Dež, grom in bliskanje (zvok) Ozka brv brez ograje Umazana reka

Tabela 1: *Elementi v scenarijih treh navideznih okolij.*

3 REZULTATI

3.1 Preverjanje prve hipoteze: začetno čustveno stanje v primerjavi s stanjem po vključenosti v navidezna okolja

Najprej smo preučevali razlike med začetnim čustvenim stanjem udeležencev in poročanih čustvih po izkušnji treh navideznih okolij. V parih smo primerjali rezultate udeležencev na treh različnih področjih (pozitivnega čustvovanja, negativnega čustvovanja ter anksioznosti), in sicer začetno stanje (i) s stanjem po nevtralnem, (ii) s stanjem po sproščenem in (iii) s stanjem po anksioznem okolju. Če so bile porazdelitve razlik med podatki, zbranimi v različnih pogojih normalne, smo uporabili *t*-test za odvisne vzorce s 5-odstotno ravnjo zaupanja, če porazdelitve niso bile normalne, pa neparametrični test (Wilcoxonov test ekvivalentnih parov).

Razlike v čustvovanju v začetnem stanju in po izkušnji nevtralnega okolja so bile statistično pomembne pri pozitivnih čustvih, $t(41) = 3{,}819$, $p = {,}000$, pri čemer je velikost učinka $r = {,}51$, moč testa pa $1 - ß = {,}96$, in sicer so udeleženci poročali o manjši stopnji pozitivnega čustvovanja kot na začetku. Razlike so bile pomembne tudi pri negativnih čustvih, $Z = -3{,}17$, $p = {,}002$, $r = {,}35$, $1 - ß = {,}88$, in sicer so udeleženci poročali o manjši stopnji negativnega čustvovanja kot na začetku. Anksioznost v začetnem stanju pa se ni razlikovala od tiste po izkušnji nevtralnega okolja, $Z = -0{,}68$, $p = {,}494$, $r = {,}07$, $1 - ß = {,}54$.

Razlike med začetnim stanjem in po izkušnji sproščenega okolja so bile statistično pomembne pri negativnih čustvih, $Z = -3{,}37$, $p = {,}001$, $r = {,}37$, $1 - ß = {,}90$, in sicer so udeleženci poročali o manjši stopnji negativnih čustev po izkušnji sproščenega okolja. Udeleženci so poročali tudi o manjši stopnji anksioznosti kot na začetku meritev, $t(41) = 2{,}08$, $p = {,}044$, $r = {,}31$, $1 - ß = {,}80$. Pozitivna čustva po izkušnji sproščenega okolja niso bila statistično pomembno drugačna od tistih na začetku meritev, $t\ (41) = 0{,}93$, $p = {,}536$, $r = {,}14$, $1 - ß = {,}59$.

Razlika med izraženostjo pozitivnih čustev v začetnem stanju in po predvajanju anksioznega okolja je bila statistično pomembna, $t(41) = 2{,}176$, $p = {,}035$, $r = {,}32$, $1 - ß = {,}75$, saj so udeleženci poročali o manjši stopnji pozitivnih čustev po izkušnji anksioznega okolja. Statistično pomembno večja pa je bila tudi anksioznost po anksioznem okolju, $Z = -2{,}45$, $p = {,}014$, $r = {,}66$, $1 - ß = {,}88$. Doživljanje negativnih čustev se po izkušnji anksioznega okolja ni statistično pomembno razlikovalo od tistega pred meritvami, $Z = -0{,}40$, $p = {,}688$, $r = {,}04$, $1 - ß = {,}51$.

3.2 Preverjanje druge hipoteze: razlike v čustvih po izkušnji navideznih okolij

Nadalje so nas zanimale razlike v čustvovanju pri udeležencih po izkušnji posameznih navideznih okolij. Razlike med čustvi po treh različnih izkušnjah smo preverjali z analizo variance, če so bile v različnih eksperimentalnih pogojih porazdelitve dosežkov na vprašalnikih normalne in variance dovolj podobne, sicer smo uporabili Friedmanov test.

Prav na vseh čustvenih področjih so bile statistično pomembne razlike med stanji po treh eksperimentalnih pogojih: za PANAS pozitivna čustva $F(1{,}694,\ 69{,}446) = 6{,}17$, $p = {,}005$, $MSE = 19{,}54$, $\eta^2 = {,}13$, $1 - ß = {,}84$, za PANAS negativna čustva $\chi^2(2) = 9{,}04$, $p = {,}011$, $W = {,}11$, parcialni $\eta^2 = {,}11$; za STAI X-1 $F(2,\ 82) = 14{,}69$, $p = {,}000$, $MSE = 22{,}56$, $\eta^2 = {,}26$, $1 - ß = 0{,}99$.

Zanimalo nas je še bolj specifično, ali se čustveni okolji (sproščeno in anksiozno) razlikujeta od nevtralnega okolja po vzbujanju čustev pri udeležencu; torej, ali se podatki, dobljeni s ponovljenimi meritvami po različnih navideznih okoljih, med seboj pomembno razlikujejo. Za primerjave sem uporabila *t*-test za odvisne vzorce oz. Wilcoxonov test ekvivalentnih parov. Ker so bile te primerjave izvedene po analizi variance in je šlo torej za *post hoc* test, sem hipoteze preverjala z Bonferronijevim popravkov, tj. pri ravni *p*, enaki ,017.

Po vključenosti v sproščeno okolje so bila pozitivna čustva statistično pomembno bolj prisotna kot po nevtralnem okolju, $t(41) = -2{,}89$, $p = {,}006$, $r = {,}38$, $1 - ß = {,}94$, opaziti je bilo tudi statistično pomembno manjšo stopnjo anksioznosti po izkušnji sproščenega okolja, $Z = -2{,}94$, $p = {,}003$, $r = {,}32$, $1 - ß = {,}92$. Negativna čustva se niso spremenila, $Z = -0{,}31$, $p = {,}756$, $r = {,}03$, $1 - ß = {,}53$.

V primerjavi med izkušnjo nevtralnega in anksioznega okolja je prišlo do statistično pomembnih razlik na vseh

področjih čustvovanja; po predvajanju anksioznega okolju so udeleženci poročali o bolj pozitivnih, $t(41) = -3{,}06$, $p = {,}004$, $r = {,}43$, $1 - ß = {,}82$, pa tudi bolj negativnih čustvih, $Z = -2{,}75$, $p = {,}006$, $r = {,}30$, $1 - ß = {,}88$; večja pa je bila tudi anksioznost v primerjavi z izkušnjo nevtralnega okolja, $t(41) = -3{,}11$, $p = {,}003$, $r = {,}44$, $1 - ß = {,}91$.

Statistično pomembno različna pa so bila tudi čustva po anksioznem in po sproščenem okolju. Po anksioznem okolju so udeleženci poročali o večji stopnji negativnih čustev, $t(41) = -3{,}11$, $p = {,}003$, $r = {,}44$, $1 - ß = {,}92$, ter večji anksioznosti, $t(41) = -4{,}77$, $p = {,}000$, $r = {,}60$, $1 - ß = {,}98$, medtem ko se pozitivna čustva po izkušnjah obeh okolij niso statistično pomembno razlikovala, $t(41) = 1{,}08$, $p = {,}288$, $r = {,}16$, $1 - ß = {,}61$.

3.3 Preverjanje tretje hipoteze: prisotnost v različnih navideznih okoljih

Zanimalo nas je še, kako so udeleženci ocenjevali svoj občutek prisotnosti v različnih navideznih okoljih in če je prihajalo do razlik med ocenami glede na tri eksperimentalne pogoje.

Primerjave s *t*-testom so pokazale, da se je občutek prisotnosti v čustvenem in nevtralnem okolju razlikoval. V primerjavi z nevtralnim okoljem je bil občutek prisotnosti v sproščenem okolju višji, $t(41) = -2{,}52$, $p = {,}016$, $r = {,}37$, $1 - ß = {,}85$, in prav tako je bil višji v anksioznem okolju, $t(41) = -3{,}05$, $p = {,}004$, $r = {,}43$, $1 - ß = {,}90$. Med obema čustvenima okoljema pa ni bilo statistično pomembnih razlik, $t(41) = -0{,}21$, $p = {,}832$, $r = {,}03$, $1 - ß = {,}50$.

Ali obstajajo tudi korelacije med poročanimi občutki prisotnosti, merjenimi z vprašalnikom SUS, in poročanimi čustvi, ki smo jih preverjali z vprašalnikom PANAS in STAI X-1? Statistične analize so pokazale, da je bila statistično pomembna le ena korelacija, in sicer so pozitivna čustva po sproščenem scenariju pomembno korelirala z občutkom prisotnosti po istem scenariju.

4 INTERPRETACIJA

S pomembnimi razlikami med čustvi v začetnem stanju in po različnih navideznih okoljih smo potrdili, da lahko navidezna okolja vzbujajo čustva pri udeležencih in glede na tip čustvovanja lahko zasledimo predvidena čustva – po anksioznem okolju bolj anksiozna, po sproščenem pa znižanje negativnih in anksioznih čustev.

Nadaljna analiza je pokazala, da so se udeležencem skozi eksperimentalne pogoje čustva spreminjala, razlike med primerjanjem čustev po izkušnji navideznih okolij pa dokazujejo, da so različna okolja vzbudila različna čustva v predvideni smeri – sproščeno okolje je znižalo negativna in anksiozna čustva v primerjavi s čustvi, ki jih je vzbudilo anksiozno okolje.

Nazadnje pa smo preverjali prisotnost v povezavi s čustvenimi okolji in tudi tukaj dobili potrditev, da so se v čustvenih okoljih udeleženci počutili bolj prisotne, v nevtralnem okolju pa so poročali o manjšem občutku prisotnosti.

Če pa rezultate pogledamo nekoliko pobližje, se nam razkrijejo določene nepričakovane značilnosti: tudi nevtralno okolje je vzbudilo spremembo čustev, čeprav tega nismo pričakovali. S primerjavo drugih čustvenih odzivov smo ugotovili, da je nevtralno okolje udeležence dolgočasilo, kar pomeni, da to okolje ni bilo ustrezno za našo raziskavo. S to ugotovitvijo pa se pojavi še ena težava. In sicer v zadnji hipotezi, kjer smo preverjali občutek prisotnosti glede na čustveno oz. nečustveno okolje, ne moremo zaključiti, da je prišlo do višje stopnje občutka prisotnosti zaradi okolja, ki je vzbujalo čustva, temveč zaradi okolja, ki je vzbujal višjo stopnjo aktivnosti, kar lahko zožimo na količino dražljajev, ki jih je udeleženec v določenem okolju prejel. Da bi lahko torej preverjali, ali so v ozadju večjega občutka prisotnosti res čustva, bi morali izenačiti število dražljajev v nevtralnem in čustvenih okoljih.

Navidezna resničnost vsekakor medij, ki ponuja nove in razburljive poglede na človekovo duševnost, zato je smiselno nadaljevati z uporabo tega medija v raziskavah.

Literatura

[1] Lombard, M., & Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of Computer Mediated Communication, 3*(2), 34-45.

[2] IJsselsteijn, W.A., Freeman, J., Avons, S.E., Ridder, H. (2003). Presence: Concept, determinants and measurement. *Presence: Teleoperators and Virtual Environments, 12* ,340-352.

[3] Witmer, B.G., & Singer, M.J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments, 7*, 225 - 240.

[4] Slater, M. in Usoh, M. (1994). Representation systems, perceptual position and presence in virtual environments. *Presence: Teleoperators and Virtual Environments, 2*, 221–234.

[5] Slater, M. (1999a). Measuring presence: A response to the Witmer and Singer presence questionnaire. *Presence: Teleoperators and Virtual Environments, 8*, 560-565.

[6] Slater, M., Usoh, M., & Chrysanthou, Y. (1995). The influence of dynamic shadows on presence in immersive virtual environments. V: Goebel, M. (ur.), *Virtual Environments '95*. (str. 8-21), New York: Springer Computer Science.

[7] Lang, P.J. (1996). Negative video as structure: emotion, attention, capacity and memory. *Journal of Broadcasting and Electronic Media*, *40*, 460–477.

[8] Baños, R., Botella, C., Liaño, V., Guerrero, B., Rey, B., Alcañiz, M. (2004). Sense of Presence in Emotional

Virtual Environments. *Presence: Teleoperators and Virtual Environments, 10*(4), 156-159.

[9] Riva, G., Mantovani, F., Capideville, C.S., Preziosa, A., Morganti, F., Villani, D., Gaggioli, A., Bottela, C., Alzaniz, M. (2007). Affective interections using virtual reality: the link between presence and emotion. *Cyberpsychology and Behaviour, 10*(1), 45-56.

[10] Watson, D., Weber, K., Assenheimer, J.S., Clark, L.A., Strauss, M.E., McCormick, R.A., (1995). Testing a Tripartite Model: II. Exploring the symptom structure of anxiety and depression in student, adult and patient samples. *Journal of Abnormal Psychology, 104*(1), 3-14.

[11] Spielberger, C.D., Gorsuch, R.L, Lushene, R.E. (1970). *STAI manual for the State-Trait Inventory*. Palo Alto, CA: Consultnig Psychologists Press.

Zbornik 14. mednarodne multikonference

INFORMACIJSKA DRUŽBA – IS 2011

Proceedings of the 14th International Multiconference

INFORMATION SOCIETY – IS 2011

Robotika

Robotics

Uredili / Edited by

Jadran Lenarčič, Leon Žlajpah, Andrej Gams

http://is.ijs.si

14. oktober 2011 / October 14th, 2011
Ljubljana, Slovenia

PREDGOVOR

Pod okriljem multikonference »Informacijska družba« tretjič zapored organiziramo tudi konferenco Robotika, s katero nadaljujemo tradicijo raziskovalne robotike v Sloveniji. Kljub finančno in raziskovalno zahtevnem področju se je v Sloveniji uveljavilo več raziskovalnih skupin, ki ne samo, da sledijo trendom sodobne robotike, temveč jih soustvarjajo.

Robotika se je v zadnjih desetletjih utrdila v zavesti ljudi kot nepogrešljiv del moderne tehnologije. Brez robotskih manipulatorjev si ne znamo več predstavljati sodobnih industrijskih procesov, npr. varjenja avtomobilskih ohišij, vstavljanja obdelovancev vstiskalnice, razpršilnega barvanja ter izdelovanja drugih visokotehnoloških produktov. Ne presenečajo nas kirurški roboti ali servisni mobilni roboti, ki bodo kmalu čistili naša stanovanja. Znanstvena fantastika in pa zabavna industrija pa silita ideje še naprej. Na nekaterih področjih je raziskovalna robotika pravzaprav že ujela ideje znanstvene fantastike. Tudi vse večje število kvalitetnih mednarodnih robotskih strokovnih revij dokazuje, da postaja robotika priznana tudi kot samostojna znanost.

V zborniku so zbrani prispevki raziskovalcev iz Slovenije in tujine. Veseli, da je med avtorji velik del mladih raziskovalcev, ki so še dokaj na začetku svoje ustvarjalne poti, kljub temu pa so njihovi prispevki svetovno primerljiva konferenčna dela. Tako kot je raznoliko raziskovanje v robotiki, so raznoliki prispevki na letošnji konferenci. Razdelimo jih lahko v več sklopov.

Prvi sklop obravnava prispevke s področja učenja v robotiki. Pri tem ne gre samo za učenje gibanj robotskih manipulatorjev temveč tudi učenje predmetov ter navigacije.

Temu sledi drugi sklop prispevkov, kjer je tematika klasično robotsko generiranje trajektorij, kar se lahko uporablja za zagotavljanje stabilnosti, ob interakciji s silo ter za telemanipulacijo.

Posebne pozornosti pa so v zadnjem časudeležni rehabilitacijski roboti, ki se vse več uporabljajo kot »pametni« funkcionalni rehabilitacijski pripomoček. Rehabilitacijski robot lahko razbremeni fizioterapevte težkegafizičnega dela,hkrati pa poskrbi za učinkovito terapijo poškodovanca. To je tudi tematika prispeveka, ki opisuje uporabo robotike za rehabilitacijo pacientov po možganski kapi.

Zadnji sklop sestasvljajo prispevki s temo konstrukcije in načrtovanja mehanizmov oz senzorjev.

Eden izmed namenov te robotske konference je, da vzpodbudi interdisciplinarno debato med znanstveniki, ki seukvarjajo z različnimi področiji robotike. Letošnji prispevki predstavljajo različne inovativne raziskave sodelavcev treh skupin na katerih temelji prihodnost raziskovalnega dela na področju robotike v Sloveniji in svetu.

Jadran Lenarčič, Leon Žlajpah, Andrej Gams

ROBOTSKO UČENJE IGRE KENDAMA

Rok Vuga, Bojan Nemec, Aleš Ude
Odsek za avtomatiko, biokibernetiko in robotiko
Institut Jožef Stefan
Jamova 39, 1000 Ljubljana, Slovenia
e-mail: rok.vuga@ijs.si, bojan.nemec@ijs.si, ales.ude@ijs.si

ABSTRACT

An autonomous robot, which makes decisions based on previous experience and feedback from its sensors is one of the challenges in robotics. Among the most promising framework to bring traditional robotics towards true autonomy is reinforcement learning. However, reinforcement learning in high dimensional spaces usually required to encode robot tasks is extremely difficult. The key idea to speed up the reinforcement learning is to limit the potentially huge search space of the policy by using previous experience. In the paper we evaluate this idea on learning of ball-in-a-cup playing robot. We propose learning which relies on generalizing to new policies from similar cases. It is shown that this approach dramatically reduces number cycles needed to learn the appropriate policy.

1 UVOD

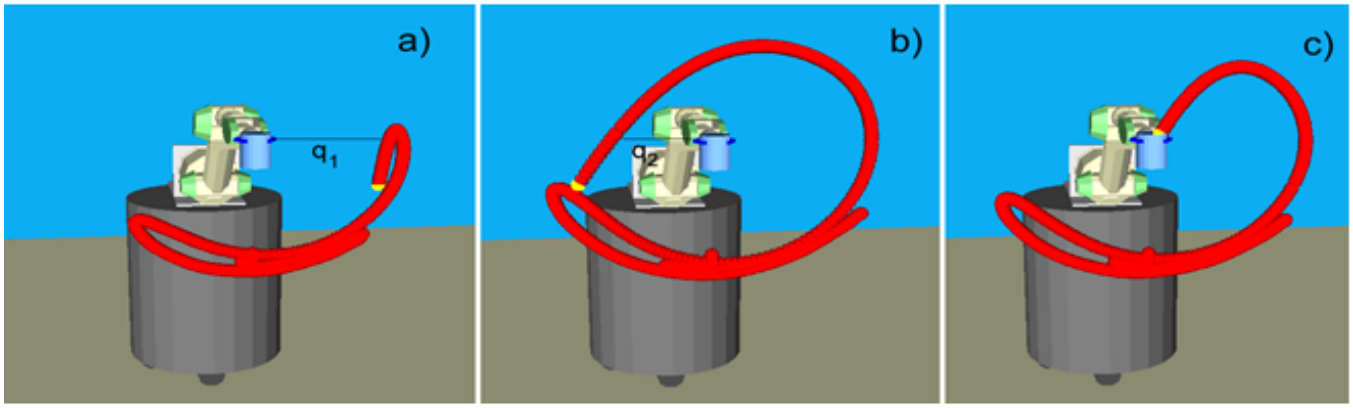
Robotsko učenje nalog je zahtevno, ker je prostor, v katerem iščemo ustrezne parametere za opis zadane naloge, zelo velik [11]. Dimenzija tega prostora je odvisna ne samo od števila prostostnih stopenj robota, ki je lahko zelo veliko pri modernih humanoidnih robotih, temveč tudi od načina predstavitve naloge ter od okolja, v katerem naloga poteka. En največjih izzivov v sodobni robotiki je zato iskanje načina, kako zmanjšati dimenzijo prostora za iskanje potencialnih rešitev, ne da bi s tem bistveno zmanjšali uspešnost učenja. Najbolj obetavni pristopi za reševanje tega problema so učenje s posnemanjem (imitation learning), pri katerem zagotovimo približek giba s pomočjo predhodno demonstriranih gibov [11,2], ter spodbujevano učenje (reinforcement learning) [7,8], kjer robot iterativno izboljšuje izvedbo naloge, dokler ne doseže želenega cilja.

Cilj tega prispevka je pospešiti proces učenja z združitvijo idej spodbujevanega učenja in statističnega posploševanja [13]. Statistično posploševanje generira trajektorijo za izvedbo želene naloge z uporabo trajektorij za izvedbo podobnih nalog. Torej, ko robotu pokažemo, kako naj se odzove na nekaj podobnih, a ne enakih, situacij, lahko znanje posploši na vse primere, ki so znotraj območja pokazanih primerov. Ker pa statistična generalizacija ne temelji na fizikalnem modelu, je dobljen rezulat samo približek optimalne rešitve dane situacije. Rezultat posplošitve je enolična določitev trajektorij s pomočjo t.i. iskalnih parametrov, ki opisujejo želeno nalogo. Število teh parametrov je običajno dosti manjše od splošnega zapisa trajektorije. V primeru igre kendama je cilj naloge določen s položajem lončka, v katerega mora žogica pasti. Če predpostavimo, da se ta lonček nahaja nekje v ravnini X-Z, za zapis njegovega položaja potrebujemo dve koorinati. Po drugi strani pa je zapis trajektorije, ki je tak met povzročila, sestavljen iz veliko večjega števila parametrov. Z ozirom na to predlagamo učenje v dveh korakih. Najprej z uporabo učenja s posnemanjem posplošimo začetno znanje za naš iskalni parameter, določen s ciljem naloge. Nato s pomočjo spodbujevanega učenja izboljšamo dobljeni rezultat tako, da s spreminjanjem iskalnega parametra kompenziramo morebitno napako, storjeno v prvem koraku.

2 UČENJE KENDAME S POMOČJO SPODBUJEVANEGA UČENJA

Pri igri kendama imamo lonček, na katerega je z vrvico pritrjena žogica, kot je prikazano na sliki 1. Lonček moramo zanihati tako, da žogica zleti vanj. V literaturi se učenje kendame pogosto uporablja za preizkušanje algoritmov strojnega učenja [5,4,6,12,9,10]. Problem je zelo zahteven, saj je rezultat odvisen od celotnega poteka trajektorije. Le majhna sprememba v izvedbi kateregakoli dela trajektorije povzroči, da žogica zaniha povsem drugače. Vsi avtorji začnejo učenje z približkom, ki je že podoben končni izvedbi naloge. Na tej podlagi je glede na uporabljeni algoritem potrebnih od 60 do 100 poizkusov spodbujevanega učenja, da se robot nauči žogico tako zanihati, da pade v lonček.

Slika 1: Kendama: žogico moramo zanihati tako, da pristane v lončku.



Slika 2: Tri izvedbe kendame v simulaciji. Sliki a) in b) prikazujeta demonstracijski izvedbi. Naučeno izvedbo prikazuje slika c).

3 POSPLOŠEVANJE PREDHODNO PODANEGA ZNANJA

Imamo zbirko demonstracijskih trajektorij s pripadajočimi parametri, ki opisujejo dano nalogo.

$$Z = y_d^i(t_{i,j}), \dot{y}_d^i, \ddot{y}_d^i(t_{k,ji}) : q_i \mid i = 1,...M, \quad j = 1,...T_i \qquad (1)$$

kjer so $y_d^i(t_j), \dot{y}_d^i(t_j), \ddot{y}_d^i(t_j)$ izmerjene pozicije, hitrosti in pospeški trajektorije i, M je število demonstracijskih primerov, T_i pa število vzorcev vsake trajektorije. Indeksiranje prostostnih stopenj je iz enačbe (1) izpuščeno zaradi enostavnosti. $q_i \in R^i$ so parametri, ki opisujejo dano nalogo. Povezujejo demonstracijsko trajektorijo s pripadajočo rešitvijo. Na primer, pri metu žoge je to cilj, v katerega vrže robot žogo pri izvedbi določene trajektorije. Pri posploševanju uporabimo te parametre za poizvedbo v bazo demonstracijskih trajektorij. Potrebujemo algoritem, ki bo za vsako novo poizvedbo **q**, ki v splošnem ni ena od demonstracijskih nalog q_i, generiral trajektorijo na podlagi znanja iz baze ***Z***.

Da lahko posplošimo demonstracijske primere za novo situacijo **q**, potrebujemo funkcijo

$$G(Z,q) \rightarrow [w,\tau,g] = [\theta] \qquad (2)$$

kjer so **θ** parametri trajektorije. V našem primeru so bili podani v zapisu dinamičnih primitivov gibanja (DMP) [6]. V DMP notaciji τ in g predstavljata trajanje in cilj trajektorije, uteži w pa njeno obliko. V splošnem funkcijska povezava (2) ni znana. G lahko postane funkcija samo, če so posplošitve trajektorij podobne demonstracijskim. Na primer, žogo lahko vržemo v isto tarčo na mnogo različnih načinov. Funckijsko odvisnost med poizvedbo **q** ter demonstracijskimi trajektorijami Z zagotovimo le tako, da omejimo posplošene mete, ki so rezultat poizvedbe q_i, na mete podobne demonstracijskim. Ker pa je v splošnem težko najti model, ki bi globalno opisoval odvisnost G, se identifikaciji izogibamo in za posploševanje gibov raje uporabimo regresijske metode.

4 UČENJE V ZMANJŠANI DIMENZIJI

V splosnem je cilj učenja z uporabo gradientnih metod najti take parametre $\theta \in R^i$, ki maksimizirajo pričakovani donos nagradne funkcije

$$J(\theta) = E\{\sum a_k r_k\} \qquad (3)$$

kjer je k časovni korak, a_k so uteži odvisne od časovnega koraka, H je trajanje (horizont), ki je lahko neskončno, r_k pa je nagrada, dobljena v posameznem časovnem koraku. Robotski met je primer epizodne naloge - vsak met je ena epizoda v učenju. Splošno pravilo za posodobitev parametrov pri gradientnih metodah je

$$\theta_{m+1} = \theta_m + \alpha_m \nabla J(\theta) \qquad (4)$$

kjer je α_m hitrost učenja. V vsaki iteraciji parametri sledijo največji spremembi nagrade.

Glavna težava gradientnih metod učenja je v estimaciji vrednosti gradienta $\nabla_\theta J(\theta)$. Če bi poznali deterministični model okolja, bi ga lahko računali po enačbi

$$\nabla J = \frac{\partial \sum a_k r_k}{\partial \theta} \qquad (5)$$

Običajno takega modela nimamo, zato se za estimacijo običajno poslužujemo metod, kot so končne diference [7], razmerje verjetnosti [14], naravni gradient strategije [3] in druge. Te metode estimirajo vrednost gradienta s pomočjo več ponovitev akcije v danem stanju. Težava nastane, ker se z večanjem dimenzionalnosti parametrov, na katerih vršimo učenje, veča tudi število potrebnih ponovitev akcije, da dobimo dovolj dober približek gradienta. Če pa

imamo dovolj predhodnega znanja, da nad njim naredimo statistično posplošitev, pa lahko parametre **θ** predstavimo s parametri **q**, ki imajo mnogo manj dimenzij. Enačba (4) se spremeni v

$$q_{m+1} = q_m + \alpha_m \nabla_q J(q),$$

$$\nabla_q J \approx \frac{\Delta \sum a_k r_k}{\Delta q} \qquad (6, 7)$$

Pri našem eksperimentu lovljenja žogice v lonček je funkcija G estimirana s pomočjo demonstracijskih parametrov q_i, ki so bili dobljeni s pomočjo demonstracijskih trajektorij, ki so povzročile let žogice v okolici iskalnega parametra **q**, ki je v resnici želen cilj žogice. Iskalni prostor učenja je tako omejen s q_i.

Ker je število dimenzij iskalnega parametra **q** majhno in smo z demonstracijskimi primeri q_k pokrili celoten prostor, v katerem želimo vršiti učenje, smo za določitev vrednosti gradienta ∇J uporabili metodo finite gradient. V stanju, v katerem se nahajamo, L-krat spremenimo parametre, ki se jih učimo, za majhne inkremente $\Delta q_l, l = 1...L$. Za vsako variacijo $q_m + \Delta q_l$ izračunamo približek spremembe nagradne funkcije $\Delta J_l = J(q_m - \Delta q_l) - J(q_m + \Delta q_l)$, izvedemo eksperiment ter dobimo vrednost nagrade. Gradient parametrov se nato izračuna po enačbi

$$\nabla J = \Delta Q^{+} \Delta J \qquad (8)$$

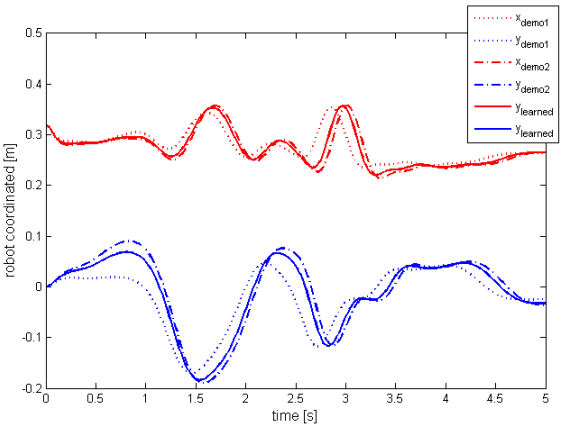
kjer je $\Delta Q = [\Delta q_{1,..}, \Delta q_L], \Delta J = [\Delta J_{1,..}, \Delta J_I]$ in + označuje Moore-Penroseov psevdoinverz.

Z združitvijo pristopov učenja s posnemanjem in spodbujevanega učenja torej predlagamo naslednji postopek:

1. Posnamemo nekaj demonstracijskih gibov, ki uspešno rešijo podane naloge. Če je možno, to znanje razporedimo čez celotni prostor možnih nalog.
2. Dobljeno znanje posplošimo za novo situacijo, v kateri se robot znajde (glej 3. poglavje).
3. Rezultat izboljšamo s pomočjo spodbujevanega učenja; učimo iskalne parametre, ki jih uporabljamo za posploševalni algoritem.
4. Naučeno rešitev dodamo v bazo znanja. Ob soočenju z morebitno novo nalogo proces učenja ponovimo začenši v točki 2).

5 REZULTATI SIMULACIJE IN EKSPERIMENTOV

Predlagani pristop smo preverili na primeru učenja igre kendama, tako v simulacijskem okolju, kot tudi na realnem robotu. Bazo znanja smo sestavili iz dveh trajektorij. Ena od njiju je žogico zanihala premočno, druga pa premalo, kot prikazuje slika 2. Dobili smo ju s pomočjo snemanja človekove izvedbe naloge. Za ta namen smo uporabili Optotrak, sistem za optično sledenje markerjem. Parameter **q**, ki določa cilj naloge, smo definirali kot razdaljo med vrhom lončka in žogico v trenutku, ko je ta na višini lončka in leti navzdol. Za demonstracijski trajektoriji sta ti vrednosti znašali $q_1 = -0{,}27\ m$ in $q_2 = 0{,}19\ m$. Premer lončka je bil 0.08 m. Ker želimo žogico spraviti v lonček, je začetni iskalni parameter enak **q**=0. Cilj učenja je najti tak iskalni parameter **q**, da bo naloga pravilno izvršena.



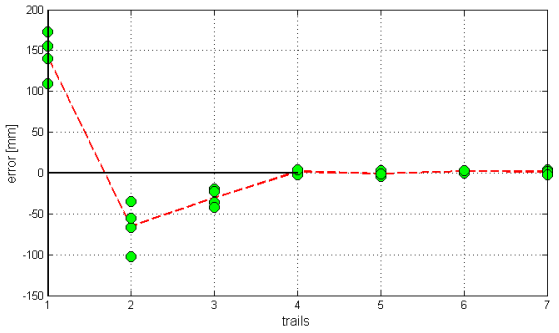
Slika 3: Demonstracijski in naučena trajektorija za izvedbo kendame.

Algoritem z iskanjem prave vrednosti enodimenzijskega parametra zelo hitro najde rešitev. S primerno izbranima začetnima trajektorijama se robot nauči zadeti lonček že po povprečno treh poizkusih. Trajektorija, izvedena v ravnini X-Z, je parametrizirana z 32 vrednostmi za vsako koordinato. To pomeni, da bi z uporabo klasičnega učenja v polni dimenziji morali iskati pravilne vrednosti za 64 parametrov. Zaradi take razlike v številu parametrov prostora učenja je predlagani algoritem mnogo učinkovitejši od ostalih pristopov iz literature [5,1,4,6]. Poudariti je potrebno, da pristop deluje le, če sta začetni - demonstracijski trajektoriji izbrani smiselno. Žogico je namreč možno zanihati na neskončno mnogo načinov. Dve poljubni začetni trajektoriji ne vodita k uspešnemu učenju. Izbrati ju moramo tako, da sta si med seboj "podobni". Na primer, ne smemo izbrati ene take, ki žogico zaniha s pomočjo dveh nihajev, druge pa take, ki to naredi z enim samim nihajem. Potek trajektorij, ki smo jih uporabili za naš eksperiment, prikazuje slika 3.

Rezultate smo potrdili tudi s pomočjo preizkusa na realnem robotu. Žogica je v lonček zletela po štirih poizkusih. Konvergenca napake za eksperiment je prikazana sliki 4. Slika 5 prikazuje postavitev eksperimenta. Algoritem je bil uspešen tudi v primeru, če smo spremenili dolžino vrvice. Če smo jo skrajšali iz 0.37 m na 0.35 m, je robot za uspešno izvedbo naloge potreboval 7 poizkusov.

6 ZAKLJUČKI

V tem prispevku smo predstavili robotsko učenje igre kendama. Naš pristop temelji na združevanju učenja s posnemanjem in spodbujevanega učenja, ki poteka na dveh nivojih. Najprej s pomočjo posploševanja začetnega znanja izračunamo začetni približek izvedbe za rešitev našega primera. Nato rezultat izboljšamo s pomočjo učenja parametrov učenja s posnemanjem. S tem bistveno zmanjšamo iskalni prostor spodbujevanega učenja. Predlagani pristop smo preverili na primeru učenja igre kendama, tako v simulaciji kot v realnem okolju. Rezultati kažejo, da je pristop hitrejši od klasičnega spodbujevanega učenja.



Slika 4: Konvergenca napake pri eksperimentu igre kendama na realnem robotu. Prekinjena črta označuje povprečje štirih eksperimentov.

LITERATURA

[1] H. Arisumi, K. Yokoi, and K. Komoriya, "Proceedings of the ieee/rsj int. conf. intelligent robots and systems," in Kendama game by casting manipulator, Edmonton, Canada, 2005, pp. 3187–3194.

[2] R. Dillmann, "Teaching and learning of robot tasks via observation of human performance," Robotics and Autonomous Systems, vol. 47, no. 2-3, pp. 109– 116, 2004.

[3] S. A. Kakade, "Natural policy gradient," Advances in neural information processing systems, vol. 14, pp. 1531–1538, 2002.

[4] J. Kober, J. Peters, "Policy search for motor primitives in robotics," Neural Information Processing Systems (NIPS), 2008.

[5] H. Miyamoto, S. Schaal, F. Gandolfo, H. Gomi, Y. Koike, R. Osu, E. Nakano, Y. Wada, and M. Kawato, "A kendama learning robot based on bidirectional theory," Neural Networks, vol. 9, no. 8, pp. 1281–1302, 1996.

[6] B. Nemec, M. Zorko, and L. Zlajpah, "Learning of a ball-in-a-cup playing robot," in Proc. of 19th International Workshop on Robotics in Alpe-Adria-Danube Region RAAD, Budapest, Hungary, June 2010.

[7] J. Peters and S. Schaal, "Reinforcement learning for parameterized motor primitives," in International Joint Conference on Neural Networks, 2006, pp. 73–80. [Online]. Available: http://www-clmc.usc.edu/publications/P/peters-IJCNN2006.pdf

[8] ——, "Reinforcement learning of motor skills with policy gradients," Neural Networks, vol. 21, pp. 682–697, 2008.

[9] T. Sakaguchi and F. Miyazaki, "Dynamic manipulation of ball-in-cup game," in Robotics and Automation, 1994. Proceedings., 1994 IEEE International Conference on, may 1994, pp. 2941 –2948 vol.4.

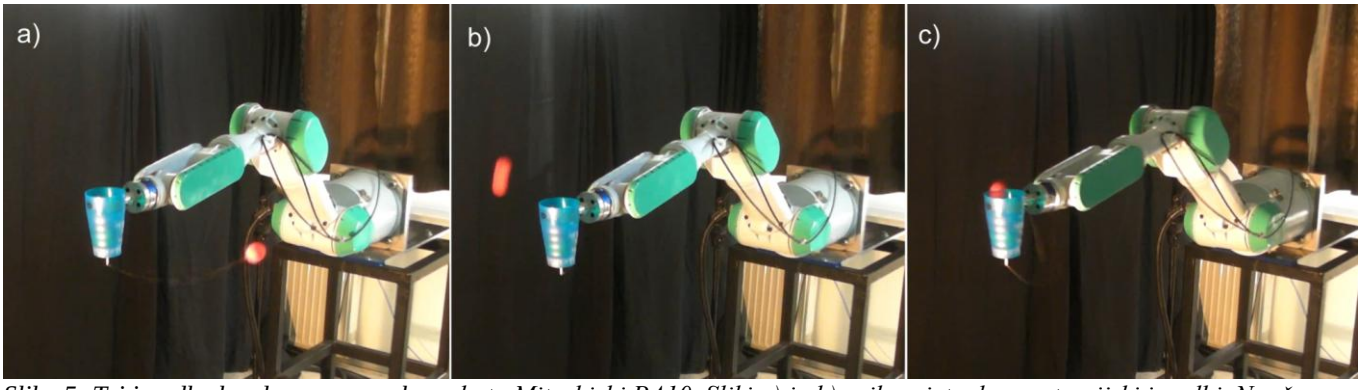
[10] S. Sato, T. Sakaguchi, Y. Masutani, and F. Miyazaki, "Mastering of a task with interaction between a robot and its environment : "kendama" task." Transactions of the Japan Society of Mechanical Engineers., vol. 59, no. 558, pp. 487—-493, 1993.

[11] S. Schaal, "Is imitation learning the route to humanoid robots?" Trends in Cognitive Sciences, vol. 3, no. 6, pp. 233–242, 1999.

[12] K. Takenaka, "Dynamical control of manipulator with vision : "cup and ball" game demonstrated by robot," Transactions of the Japan Society of Mecha- nical Engineers., vol. 50, no. 458, pp. 2046—-2053, 1984.

[13] A. Ude, A. Gams, T. Asfour, and J. Morimoto, "Task-specific generalization of discrete and periodic dynamic movement primitives," IEEE Trans. Robotics, vol. 26, no. 5, pp. 800–815, 2010.

[14] R. J. Williams, "Simple statistical gradient-following algorithms for connectionist reinforcement learning," Machine Learning, vol. 8, no. 23, 1992.



Slika 5: Tri izvedbe kendame z uporabo robota Mitsubishi PA10. Sliki a) in b) prikazujeta demonstracijski izvedbi. Naučeno izvedbo prikazuje slika c).

INTEGRATION OF DIFFERENT ROBOT MOVEMENTS AND GENERALIZATION IN REAL-TIME

Denis Forte, Aleš Ude, Andrej Gams
Department of Automation, Biocybernetics and Robotics
Jožef Stefan Institute
Jamova 39, 1000 Ljubljana, Slovenija
Tel: +386 1 4773419; fax: +386 1 4251038
E-mail: denis.forte@ijs.si

ABSTRACT

In this paper we present a new methodology to learn and integrate different movement primitives in real-time. Our approach starts from a library of example trajectories for each primitive movement, which serves as a basis for the generation of a complete representation for the trained movement primitives by statistical generalization. To enable fast switching between different movement primitives, it is essential that on-line calculations needed to initialize and switch to a new movement primitive are done in real-time. We show that by converting the initial trajectory data into dynamic systems, we can switch to a new movement primitive within a real-time sensory feedback loop. Experimentally we also show that the accuracy of the generalized movements is sufficient to realize tasks such as feed-forward grasping.

1 INTRODUCTION

In this paper we investigate the problem of real-time, goal-directed trajectory generation using a database of example movements [1], [2]. It has often been studied as part of programming by demonstration systems [3]. Our primary interest is in real-time synthesis of new trajectories using local methods.

It has been shown by Ude et al. [2] that it is possible to generalize the movements collected in an example database to new situations by utilizing the goal of an action as a query point into the database. In [2] movement generalization was implemented by employing a combination of locally weighted regression [4] and Gaussian process regression [5], where raw trajectory data was used as input for generalization. The proposed approach was applied to generalize various behaviors including reaching, throwing, and drumming. While this approach can take into account external perceptual feedback to generalize example movements to different situations, its computational cost is prohibitive for use in a real-time feedback loop. The goal of this paper is to provide an approach that is efficient enough to be applied in such a loop.

The approach described in [2] uses Dynamic Movement Primitives (DMPs) [6] as the basic representation for the encoding of robot movements. DMPs have many useful properties such as a built-in ability to react to perturbations without introducing discontinuities in the resulting robot motion. As an autonomous representation, they are not directly dependent on time, which makes it easy to stop the execution of movement without extensive bookkeeping of time evolution. DMPs can also be extended to include capabilities such as obstacle avoidance and avoidance of joint limits. All these adaptations can be done in real-time, which enable the robot to react to external sensory feedback. In this paper we expand on such built-in abilities by providing a methodology for real-time generation of DMPs based on a trajectory database.

One DMP can encode one specific robot trajectory. In case of point-to-point (discrete) movements, the trajectory of each robot degree of freedom y (given either in joint or in task space) is described by the following system of nonlinear differential equations

$$\tau \dot{z} = \alpha_z \left(\beta_z (g - y) - z \right) + f(x), \tag{1}$$

$$\tau \dot{y} = z, \tag{2}$$

$$\tau \dot{x} = -\alpha_x x, \tag{3}$$

where x is the phase variable and z is an auxiliary variable. α_x, α_z, β_z and τ need to be specified in such a way that the system converges to the unique equilibrium point $(z, y, x) = (0, g, 0)$. The nonlinear term f contains free parameters that enable the robot to follow any smooth point-to-point trajectory from the initial position y_0 to the final configuration g

$$f(x) = \frac{\sum_{k=1}^{N} w_k \Psi_k(x)}{\sum_{k=1}^{N} \Psi_k(x)} x, \quad \Psi_k(x) = \exp\left(-h_k (x - c_k)^2\right). \tag{4}$$

Here c_k are the centers of radial basis functions distributed along the trajectory and $h_k > 0$. Weights w_k are estimated so that the DMP encodes the desired trajectory. For robots with many degrees of freedom, each degree of freedom is represented by its own equation system (1) - (2), but with a common phase (3). We used the algorithm described in [2] to determine the placement, width and number of radial basis functions $\boldsymbol{\Psi_k}$.

2 APPROXIMATION OF MOVEMENTS WITH GAUSSIAN PROCESS REGRESSION

Lets assume that we have a set of robot movements $\mathbf{M}_i$, $i = 1,\ldots,NumEx$, which all result in a successful execution of a given task in different situations. As example we consider a set of reaching movements towards different targets in 3-D space. We denote the parameters characterizing the task by $\mathbf{q}_i \in \mathbb{R}^m$, $i = 1,\ldots,NumEx$, m being the dimensionality of these parameters, which we also call query points. Every movement $\mathbf{M}_i$ is encoded by a sequence of trajectory points $\{y_{ij},\ \dot{y}_{ij},\ \ddot{y}_{ij} \in \mathbb{R}^{dof}\}$, measured at times t_{ij}, $j = 1,\ldots,n_i$, $t_{i1} = 0$. Here n_i denotes the number of samples on trajectory $\mathbf{M}_i$, while *dof* denotes the number of degrees of freedom encoded by the example trajectories. We have experimented both with end-effector trajectories (in this case y_{ij} are points in the Cartesian space) and with robot joint trajectories (in this case y_{ij} are the joint angles stemming from the active degrees of freedom). The problem is to compute a trajectory for any given query point $\mathbf{q}$. For example, in case of reaching, a query point is given by the desired target position and we need to compute the associated reaching trajectory $\mathbf{M}$. Example movements $\mathbf{M}_i$ can be acquired either by kinesthetic guiding or by imitation.

To become able to accomplish a task in any situation, the robot needs to learn a function that maps the parameters describing the task $\mathbf{q}$ into the parameters describing the desired trajectory $\mathbf{M}$, i. e.

$$G : q \mapsto M. \qquad (5)$$

In general, $\mathbf{G}$ is not a function. For example, in the case of reaching movements, there are many different ways to reach towards a desired destination. However, we can impose an additional constraint that synthetic reaching trajectories should be similar to the example reaching trajectories. The closer the desired query point $\mathbf{q}$ is to the query point $\mathbf{q}_j$, the more similar the generated trajectory $\mathbf{M}$ should be to the trajectory $\mathbf{M}_j$ associated with query point $\mathbf{q}_j$. With this additional constraint, $\mathbf{G}(\mathbf{q};\{\mathbf{M}_1,\ldots,\mathbf{M}_{NumEx}\})$ becomes a function that can be learned.

To reduce the amount of data that we need to process for action generalization, we first encode each of the example movements $\mathbf{M}_i$ as a dynamic movement primitive (DMP). Any of the standard methods proposed in the literature can be used for this purpose.

The initial raw trajectory data $\mathbf{M}_i$ can be converted into DMPs, i. e. $\mathbf{M}_i \mapsto (\mathbf{w}_i, \mathbf{g}_i, \tau_i)$, where $\mathbf{w}_i \in \mathbb{R}^{N \times dof}$ are the weights of DMPs for all degrees of freedom, $\mathbf{g}_i \in \mathbb{R}^{dof}$ are the final configurations on the example trajectories, i. e. $\mathbf{g}_i = y_{ini}$, and $\tau_i \in \mathbb{R}$ are the time durations of example trajectories, i. e. $\tau_i = t_{ini}$.

2.1 Trajectory generalization using Gaussian process regression

The conversion of raw example trajectories into DMPs results in a significant data reduction. For example, a four second trajectory sampled at 500 Hz contains 2000 data points, which can typically be reduced to a DMP defined by a few tens of radial basis functions. In this section we propose to synthesize new movements directly from the estimated DMP parameters. In this case function (5) becomes

$$G\left(\{w_i, g_i, \tau_i; q_i\}_{i=1}^{NumEx}\right) : q \mapsto (w, g, \tau). \qquad (6)$$

Gaussian Process Regression (GPR) can be applied to estimate function (6). Gaussian processes are based on Bayesian probability modeling [5]. The resulting models have an interesting and useful feature that, besides output values, they also predict confidence in these values. GPR exhibits good generalization performance and the predictive distribution can be used to measure the uncertainty of the estimated function. It has been demonstrated that this technique outperforms other regression methods on problems such as estimating inverse dynamics of a seven degrees of freedom robot arm [7].

Technically, a Gaussian process is defined as

$$g(q) \sim GP\ (m(q), k(q, q')), \qquad (7)$$

where $m(\mathbf{q}) = \mathbb{E}[g(\mathbf{q})]$ is the mean function and $k(\mathbf{q},\mathbf{q'}) = \mathbb{E}[(g(\mathbf{q}) - m(q))(g(\mathbf{q'}) - m(\mathbf{q'}))]$ the covariance function of the process. Lets assume that we have – as when estimating function (6) – a set of noisy observations $\{(\mathbf{q_i};\ y_i) | i = 1,\ldots,NumEx\}$, $y_i = g(\mathbf{q_i}) + \epsilon$, $\epsilon \sim \mathcal{N}(0, \sigma_n^2)$. Subtracting the mean from the training data, we can further assume that $m(\mathbf{q}) = 0$. Given a set of query points $g(\mathbf{q^*})$, the joint distribution of all outputs is estimated by

$$\begin{bmatrix} y \\ y^* \end{bmatrix} \sim N\left(0, \begin{bmatrix} K(Q,Q) + \sigma_n^2 I & K(Q,Q^*) \\ K(Q^*,Q) & K(Q^*,Q^*) \end{bmatrix}\right), \qquad (8)$$

where $\boldsymbol{Q}$, $\boldsymbol{Q^*}$, y, y^* respectively combine all inputs and outputs and $\mathbf{K}(\cdot,\cdot)$ are the associated joint covariance matrices calculated according to Eq. (6). It can be shown [5] that the expected value $\bar{y}^*$ associated with the new query points $\mathbf{q^*}$ is given by

$$\bar{y}^* = K(Q^*,Q)\left[K(Q,Q) + \sigma_n^2 I\right]^{-1} y, \qquad (9)$$

One commonly used covariance function is

$$k(q, q') = \sigma_f^2 \sum_{i=1}^{m} \exp\left(-\frac{1}{2}\frac{(q_i - q_i')^2}{l_d^2}\right), \qquad (10)$$

which results in a Bayesian regression model with an infinite number of basis functions. m denotes the dimension of the query point space. See [5] for more details.

With GPR new estimates are calculated using equation (9). The most computationally expensive part is the calculation of $[\mathbf{K}(\boldsymbol{Q},\boldsymbol{Q}) + \sigma_n^2\mathbf{I}]^{-1}$, but since this matrix depends only on the training data, the necessary calculations can be done off-line using for example the Cholesky decomposition. The dimension of this matrix is equal to the number of data points. In our case, this is equal to the

number of example movements *NumEx*, which is typically not too large (at most a few hundred).

Note that by writing

$$z = \left[K(Q,Q) + \sigma_n^2 I \right]^{-1} y, \qquad (11)$$

equation (9) and the estimated parameter $\bar{y}*$ associated with the query $\boldsymbol{Q}^* = \mathbf{q}^*$ can be written as

$$\bar{y}* = \sum_{i=1}^{NumEx} k(\mathrm{q}*, \mathrm{q}_i) z_i, \qquad (12)$$

where in our experiments $\bar{y}*$ stands for $\bar{\tau}*$, $\bar{g}_l*$, and $\bar{w}_{kl}*$. Thus the training data are weighted based on the distance between the training query points and the current query point. This means that nearby training points influence the result more.

To generate a new movement, the robot is given a desired query point $\mathbf{q}^*$, e. g. the desired reaching location. For each of the parameters defining a generalized DMP (τ, g_l, and w_{kl}), which encodes a suitable motion trajectory for this task situation, we need to calculate (12) on-line, whereas (11) can be stored in memory.

2.2 Comparison of Gaussian process regression and locally weighted regression

In our following set of experiments we compare the performance of the proposed approach and the performance of a method that uses complete trajectories without an intermediate trajectory conversion step (as proposed in [2]) for the purpose of task-specific generalization of DMPs. In this section we examine the advantages and disadvantages of both approaches and their suitability for on-line generalization.

In contrast to our new approach, which converts the training data into DMPs, the approach proposed in [2] keeps complete trajectories in memory and generalizes to new DMPs without the intermediate trajectory conversion step. In this case locally weighted regression (LWR) instead of GPR has to be used in some calculations that are needed for trajectory generalization. LWR is a memory-oriented, non-parametric method for statistical approximation. The basic idea is to compute local models using data from a small neighborhood of the desired query point. Since raw trajectories are used for estimation, the resulting systems of linear equations are much too large to be resolved on-line. Unlike this approach, which defers most of the calculations to the moment when a query needs to be answered, our new, GPR-based method performs most of the calculations off-line once all of the training data have been acquired. The most expensive off-line calculations are needed for the calculation of (11) and for the estimation of hyper-parameters l_i, σ_f and σ_n as defined in (9) and (10). After the end of learning the training data can be discarded and only simple calculations shown in (12) are needed to answer a new query or in other words generalize to new situations.

Here we note that in case of LWR we need to specify one additional parameter, i. e. influence radius, which determines how many nearby trajectories will be taken into account for generalization by LWR. Some approaches for the selection of the optimal radius can be found in [2]. Our new, GPR-based approach does not require such a parameter. In the following we call the approach proposed in this paper MPG (Movement Primitives Generalization) and the approach from [2] RTG (Raw Trajectories Generalization).

Our experiment was performed with the 7 DOF KUKA Light-Weight Robot arm. The task was to reach towards an object and grasp it. The object can be grasped either from its right or from its left side. Thus the robot needs to learn how to grasp the object from both sides. We demonstrated 144 reaching movements (72 from the right side and 72 from the left side of the object), which were all acquired by kinesthetic guiding of the arm (see figure 1).

Figure 1: *The acquisition of reaching movements by kinesthetic guiding of KUKA Light-Weight Robot arm.*

To test the MPG and RTG in real experiments, where the correct trajectories are not known, we applied the leave-one-out cross validation (L1OCV) to determine the number of necessary basis functions. In L1OCV, each of the demonstrated trajectories is taken out from the training data and re-estimated by generalization from remaining trajectories. The generalized and the skipped trajectory are then compared to determine an average error over the entire trajectory. The L1OCV score is given by an average error over all trajectories in the training data. We also test how the approximation by different number of weighted radial basis functions effects the average error in joint space:

$$error_k = \frac{1}{T_{end,k}} \sum_{i=1}^{T_{end,k}} \left\| y_{i,k,gen} - y_{i,k,ideal} \right\| \;\middle|\; k = 1,\ldots,NumEx, \qquad (13)$$

$$error_{W_n} = \frac{1}{NumEx} \sum_{k=1}^{NumEx} error_{k,n} \;\middle|\; n = 3,\ldots,MaxN. \qquad (14)$$

where $y_{i,k}$ are the points on the trajectory k, given in joint space.

Figure 2 represents the comparison of MPG and RTG by approximation with different number of radial basis functions and two radiuses for RTG. Results in figure 2 show that RTG is more stable then MPG, but there is a big difference in computation times for MPG and RTG based generalization at larger weights numbers. MPG based approach generalizes much faster and is therefore more suitable for on-line calculation. These results also show that 8 radial basis functions are enough to approximate the

reaching trajectories in our experiment. With more than 8 weights the average error does not change significantly regardless of the selected method.

The number of radial basis functions needed to approximate the trajectories depends on the type of movement. With the MPG-based approach we need to use the same number of basis function to estimate all training trajectories, otherwise it is not possible to apply GPR. The longer and more complex the trajectories are, the more radial basis functions are needed to approximate the movements. The automatic selection of the number of basis functions is discussed in [8].

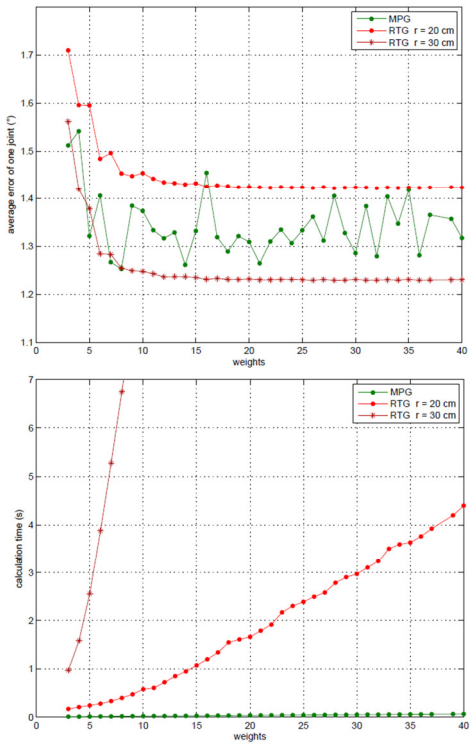


Figure 2: *The results of RTG and MPG were compared with the taken trajectories in joint space over the entire trajectories by L1OCV method. In the top graph the red line with dots shows an error of generalization by RTG at 20 cm radius, red line with stars at 30 cm radius and green line represents an error of generalization by MPG in joint space. The bottom graph shows calculation times that are needed for generalization process of RTG and MPG in dependency of weights number.*

3 SWITCHING BETWEEN TWO DIFFERENT MOVEMENT PRIMITIVES

The main goal was to demonstrate on-line generalization of trajectories, which can be accomplished only with our new, MPG-based approach while the RTG approach is too slow. We show the performance of real-time generalization in a task in which the robot switches between two different types of
reaching movements (for left- and right-side grasps) in case of perturbations.

The KUKA arm was controlled in stiffness mode. While reaching the stiffness is high enough to properly perform the generalized reaching move. The external joint torques are monitored during execution. If joint torques exceed a threshold (determined empirically), the algorithm switches to a lower stiffness mode, knowing that a physical disturbance occurred. During low stiffness mode the robot is compliant enough to move in the direction of push. Meanwhile, new generalized reaching movements are constantly calculated (every 0.02 seconds) based on the current position of the robot's endeffector. When the perturbation stops, the newest generalized reaching movement starts being executed. If for example the robot starts reaching from the right side of the object and the perturbation causes it to move to the left side, the algorithm switches from right- to left-side reaching movement. The object is grasped once the robot reaches the end-position on the reaching trajectory. We used a BarrettHand BH-8 Series attached at the top of the arm for grasping. The described reaching process is shown in Figure 3.

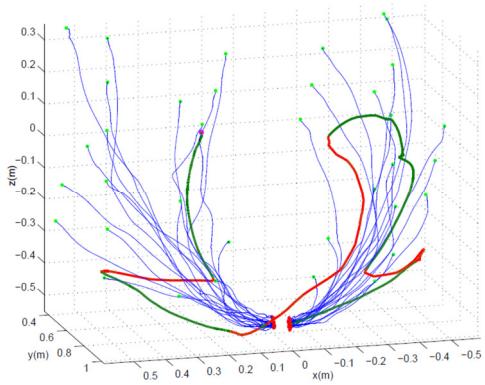


Figure 3: *3-D graph illustrates movement of the Light-Weight Robot arm. Green curves are generalized trajectories while blue curves are some of the demonstrated trajectories. Red curves represent robot movement under physical disturbance. Beginning of the third green curve shows the reshaping of the robot when passing to different type of reaching moves.*

4 CONCLUSION

We developed a new approach for on-line generalization of discrete movements based on Gaussian process regression. Unlike previous generalization approaches, which either required significant on-line calculations [2] or global optimization [1] prone to local minima, the proposed approach can avoid both. Our experiments have shown that despite significant data reduction, which provides the basis for real-time implementation of the proposed approach, the generated movements remain close to the ideal movements. The real-time implementation allowed us to realize tasks such as on-line switching between movement primitives based on perceptual feedback, which would not be possible with previous memory-based approaches.

References

[1] E. Gribovskaya, S. M. Khansari-Zadeh, and A. Billard, "Learning nonlinear multivariate dynamics of motion in robotic manipulators," The International Journal of Robotics Research, vol. 30, no. 1, pp. 80–117, 2011.

[2] A. Ude, A. Gams, T. Asfour, and J. Morimoto, "Task-specific generalization of discrete and periodic dynamic movement primitives," IEEE Trans. Robot., vol. 26, no. 5, pp. 800–815, 2010.

[3] R. Dillmann, “Teaching and learning of robot tasks via observation of human performance,” Robotics and Autonomous Systems, vol. 47, no. 2-3, pp. 109–116, 2004.
[4] C. G. Atkeson, A. W. Moore, and S. Schaal, “Locally weighted learning,” AI Review, vol. 11, pp. 11–73, 1997.
[5] C. E. Rasmussen and C. Williams, Gaussian Processes for Machine Learning. Cambridge, MA: MIT Press, 2006.
[6] A. J. Ijspeert, J. Nakanishi, and S. Schaal, “Movement imitation with nonlinear dynamical systems in humanoid robots,” in Proc. IEEE Int. Conf. Robotics and Automation, Washington, DC, 2002, pp. 1398–1403.
[7] D. Nguyen-Tuong, M. Seeger, and J. Peters, “Model learning with local Gaussian process regression,” Advanced Robotics, vol. 23, pp. 2015– 2034, 2009.
[8] S. Schaal and C. G. Atkeson, “Constructive incremental learning from only local information,” Neural Computation, vol. 10, no. 8, pp. 2047– 2084, 1998.

ROBOTIC ROPE TURNING IN COOPERATION WITH HUMAN

Tadej Petrič, Andrej Gams, Leon Žlajpah
Department of Automation, Biocybernetics and Robotics,
Jozef Stefan Institute
Jamova 39, 1000 Ljubljana, Slovenia
e-mail:{tadej.petric, andrej.gams, leon.zlajpah}@ijs.si
URL: abr.ijs.si

ABSTRACT

The paper describes a control algorithm for rotating the rope, i.e. a soft and pliable object that has no support from below. To perform such tasks the control algorithms are usually complex, because the gravity can easily affect the shape of the object. As an example we have selected the task of rotating the rope with the robot in cooperation with a human. In this case the robot must not only determine the frequency of the task, but also the direction of the rotation of the rope. The method we proposed is based on an adaptive frequency oscillator combined with a Fourier series representation in the feed-back loop. In this way it can extract the frequency and the phase from an unknown periodic signal, in real-time, and without any additional signal processing or pre-processing.

1 INTRODUCTION

Many robotic tasks depend on the ability to extract frequencies on-line, and to synchronize movement or actuation with the task object. For tasks where the dynamic behavior and response are crucial, approaches that adjust the rhythm and behavior of the robot, in order to achieve synchronization [9], have to be applied. Such tasks include swinging of different pendulums [2], playing with different toys [6, 7, 14], locomotion [4] or even tasks involving manipulation of object in air, such as juggling [11], lassoing [13] and rope turning [1, 5].

Performing these tasks with robots requires appropriate trajectory generation and foremost precise frequency tuning. However, the on-line frequency extraction for robotic tasks is often difficult because of the relatively low frequencies involved, i.e. frequencies vary from practically 0 Hz up to ~5 Hz. Higher frequencies are encountered only seldom or for special tasks, like the acceleration of a gyroscopic device [6]. The low frequencies of the tasks make on-line frequency extraction problematic, since at least a few periods have to pass to estimate the frequency. Such signal processing methods, like FFT with a sliding window have a delay in estimation, which influences the assessment of the phase of the observed signal.

When performing the rhythmic robotic tasks one has to assure not only the correct frequency, but also the correct output waveform. Not many approaches exist that allow both frequency extraction and waveform learning [5, 10]. Righetti et al. [10] proposed an approach based on a set of adaptive frequency oscillators in a feedback loop. Their approach allows extraction of a various number of frequency components. This was used for the first layer of a two-layered movement imitation system [5]. In their approach the waveform was encoded with weighted kernel functions as a part of a dynamic movement primitive (DMP). While the first layer extracts the fundamental frequency the second layer ensures the smooth referential trajectory. A modification of this system with a novel approach for the first layer was given in [7, 8]. They proposed an approach, which uses only one oscillator and an adaptive Fourier series in a feedback loop. The combination of an adaptive oscillator and the adaptive Fourier series allows extracting the fundamental frequency of the input signal and thus control of rhythmic robotic tasks.

The contribution of this paper is the extension of a two-layered control system introduced in [7, 8]. We augmented the two-layer system with another layer which has to provide the control strategy. This allows us to control different robotic tasks, where the robot must not only synchronize its movement but also decide what kind of movement is required to accomplish the desired task. For example, such task is turning the rope in cooperation with a human, where high-level control must determine the direction of rotation. This task has already been a subject of several studies, where in [5] the robot assisted a human who was turning the rope, but was unable to rotate the rope alone. In [1] they introduced a method for controlling the rotational center of an unsupported object while maintaining the rotation of an object.

The proposed method simplifies the synchronization between the rope and the movement of the robot, by determining the frequency from a measured signal. Augmented with the high-level control it allows successful cooperation with the human, even if the human starts to rotate the rope in an arbitrary direction, since the robot will always appropriately modify and adjust the movement to the current situation.

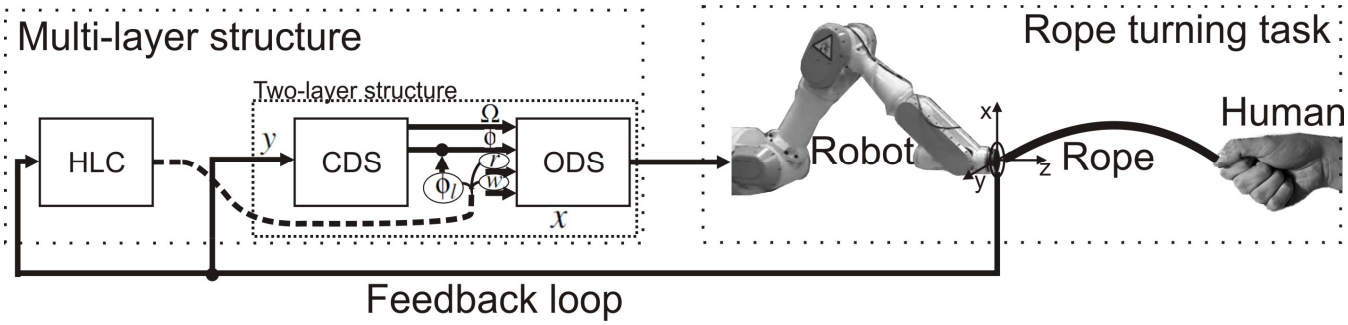


Figure 1: *Multi-layered structure of the proposed system. HLC is the high level control system, which selects the proper movement primitive, adjusts the phase lag parameter and the amplitude parameter for successful control of rhythmic robotic*

The paper is organized as follows. In section 2 we give a brief description of the control system. In section 3 we evaluate the proposed approach on a real-world experiment of a cooperative rope turning of a robot and a human. Conclusion and summary are given in section 4.

2 MULTI-LAYERED CONTROL SYSTEM

In a previous study by Gams et al. [3], a two-layered imitation system was presented. In their work the authors explained that the system can be used for extracting the frequency spectrum of the input signal, learning the waveform of one period, and imitating the desired waveform at an arbitrary frequency.

The first layer, i.e. the canonical dynamical system (CDS), is used for frequency extraction. It is based on a set of adaptive frequency oscillators in a feedback loop. The second layer is called the output dynamic system (ODS) and is used for learning and repeating the desired waveform. It is based on dynamic movement primitives (DMPs) [12].

Petrič et al. [7, 8] proposed an improvement of the first layer. They used a single adaptive frequency phase oscillator combined with a feedback structure based on an adaptive Fourier series. A short recap follows. The feedback structure of an adaptive frequency phase oscillator is governed by

$$\dot{\phi} = \Omega - K \cdot e \cdot \sin\phi, \tag{1}$$

$$\dot{\Omega} = -K \cdot e \cdot \sin\phi, \tag{2}$$

$$e = y - \hat{y}, \tag{3}$$

Where K is the coupling strength, ϕ is the phase of the oscillator, e is the input into the oscillator and y is the input signal. Since there is only one oscillator, the frequency Ω and phase ϕ are clearly defined, i.e. no logic algorithm is needed to select if from a pool of oscillators as it was requires in [3]. The feedback loop is now represented by the Fourier series

$$\hat{y} = \sum_{i=0}^{M} \left(\alpha_i \cos(i\phi) + \beta_i \sin(i\phi)\right), \tag{4}$$

where M is the size of the Fourier series. The amplitudes associated with the terms of Fourier series are determined by

$$\dot{\alpha}_i = \eta \cdot e \cdot \cos(i\phi), \tag{5}$$

$$\dot{\beta}_i = \eta \cdot e \cdot \sin(i\phi), \tag{6}$$

where η is the learning constant and $i=0...M$. As shown in Fig. 2, the oscillator receives the difference between the input signal and the adaptive Fourier series combined with the adaptive oscillator. Since a negative feedback loop is used, the difference approaches zero when the Fourier series representation approaches the input signal. Such a feedback structure performs an adaptive Fourier analysis.

This approach has the ability to adapt to the basic frequency Ω of the input signal y. The number of frequency components M it can accurately extract depends on how many terms of the Fourier series are used (the size of M). On the other hand, increasing the size of M would increase the computational costs.

Augmenting the system with the output dynamic system, makes it possible to synchronize the movement of the robot to an arbitrary measurable periodic quantity. Namely, the input signal and the frequency are encoded in the Fourier series and the desired robot trajectory is encoded in the output dynamic system. Since the adaptation of the frequency and the learning of the desired trajectory can be done simultaneously, all of the system time delays, e.g. delays in communication, sensor measurements delays, etc., are automatically included. Additionally, when a predefined motion pattern for the trajectory is used, the phase between the input signal and the output signal can be adjusted with a phase lag parameter ϕ_l (see also Fig. 1). This enables us to either predefine the desired motion or to teach the robot how to perform the desired rhythmic task online.

Even though the canonical dynamical system by itself can reproduce the demonstration signal, using the output dynamical system allows for easier modulation in both amplitude and frequency; learning of complex patterns without extracting all frequency components; it is more robust to perturbation; and acts as a sort of filter. Especially the greater robustness is crucial when performing fast, dynamic tasks. Moreover, when multiple output signals are needed, only one canonical system can be used with the individual output system which assures that the waveforms of the different degrees-of-freedom are realized appropriately.

Presented system can be used for control of rhythmic robotic tasks where the strategy of performing the task remains the same. However, when we are dealing with a robot and

human collaboration for performing rhythmic tasks, the control system must not only synchronize the motion of the robot, but also select a proper trajectory for the selected task. For example, when rotating a rope by a human and a robot, the robot must not only determine the frequency of the task, but it must also determine the direction of the rotation of the rope. Otherwise, the system could be in counter-phase, which prevents the rotation of the rope. To solve this, we proposed an additional higher layer, which acts as an intelligent control layer. It has to select a proper movement primitive (weight vector **w**, for output trajectory), set the correct phase between multiple DOF, and adjust the amplitude of the output trajectory. The block scheme of a multi-layered control system is presented Fig. 1. In order to correctly perform the desired task the high level control has to select the correct movement primitive, adjust the phase lag and the amplitude parameter. While the lower layers (i.e. two-layered system) synchronize the motion to an arbitrary input signal and are able to independently respond to the perturbations, the control strategy remains the same (i.e. high level control layer).

3 ROBOTIC ROPE TURNING

In this section we evaluate the proposed system. The results of controlling a periodic robotic motion in cooperation with a human are shown.

We performed the experiment on a Mitsubishi PA-10 robot. A JR-3 force/torque sensor was attached to the top of the robot to measure the torques and forces on a string. Additionally, an optical system (Optotrak Certus) was used for validation, i.e. for measuring the motion of a human hand. The multi-layered control system was implemented in Matlab/Simulink. The control scheme is presented in Fig. 1. The imitation system (the two-layered system), which is based on an adaptive frequency oscillator combined with a dynamic Fourier series in the first layer (CDS) and DMP in the second layer (ODS), provides the desired trajectory for the robot. The motion of a robot is constrained to up-down, left-right motion using inverse kinematics.

The movement primitive for the task of rotating the rope was predefined (circular motion) by selecting appropriate weights for vector **w** and amplitude r for the output dynamic system. The task of high level control (HLC) is to select an appropriate primitive for rotating in one direction and another primitive for rotating in the opposite direction and allowing the transition from one to another direction during a single experiment. The transition must be smooth and consistent.

The populating of the frequency spectrum is done without any signal processing. Namely, as the whole process of frequency extraction and adaptation of the waveform is completely embedded in the dynamics of the adaptive frequency oscillator combined with the adaptive Fourier series. Unless stated otherwise we use the following parameters for the two-layered system: $\eta = 2$, $K = 20$, $M = 10$, $r = 1$. The parameters were determined empirically.

Fig. 2a shows the results of frequency extraction (top plot) from the measured torque signal (second plot). As we can see the frequency of the imitated motion quickly adapted to the measured periodic signal. When the rotation of the rope is stable, the human stops swinging the rope and maintains the hand in a fixed position. The movement of the human hand is shown in the third plot. In the last plot we show the movement of the robot. By comparing last two plots in Fig. 2a, we can see that after 3 s the energy transition to the rope is done only by the motion of the robot.

Fig. 2b shows the behavior of our proposed system when the rotation direction of the rope is changing. The top plot shows the frequency extraction. The input signal (i.e. the feedback signal), is shown in the second plot. The third plot shows the movement of the human hand, where we can see that the rotational direction changes at approximately 5 s and 16 s. When the human changes the rotational direction of the rope the robot must not only synchronize to the opposite direction, but it must also change the movement primitive in order to be synchronize with the rope. By observing the bottom plot, we see that the transition from one movement primitive to another (i.e. rotating in one direction to rotating in opposite direction) is smooth and consistent. This shows that the robot is able to synchronize its motion to both rotational directions.

The frequency of the task depends on the parameters of the rope, i.e. weight, length, flexibility etc., and the energy which is transmitted in to the rope. The rotating frequency of the rope can be influenced by the amplitude of the motion, i.e. how much energy is transmitted to the rope. The amplitude can be easily modified with the amplitude parameter r. Using the high level control, the desired rotational speed of the rope can easily be achieved.

Fig. 2c shows results of the frequency tracking of a rope turning task. As we can see, the proposed system can easily track the desired frequency. This shows that the system is adaptable and robust.

6 CONCLUSION

We presented a new architecture which is an upgrade to the two-layered imitations system. The proposed system has an additional high level control layer, which selects the proper movement primitive for the desired task. The transition between movement primitives is also smooth and consistent.

The synchronization and movement generation is embedded in the lower layers, i.e. two-layered imitation system. The system can react to perturbations in the lower layers, while the strategy from the high level control remains the same.

The lowest layer of the system essentially implements an adaptive Fourier series of the input signal. It can extract the frequency and phase of an unknown periodic or pseudo periodic signal. This is done in real-time without any additional processing of the input signal. Combining this system with the dynamic motion primitive enables a simple and computationally inexpensive control of rhythmic tasks with at least one measurable periodic quantity.

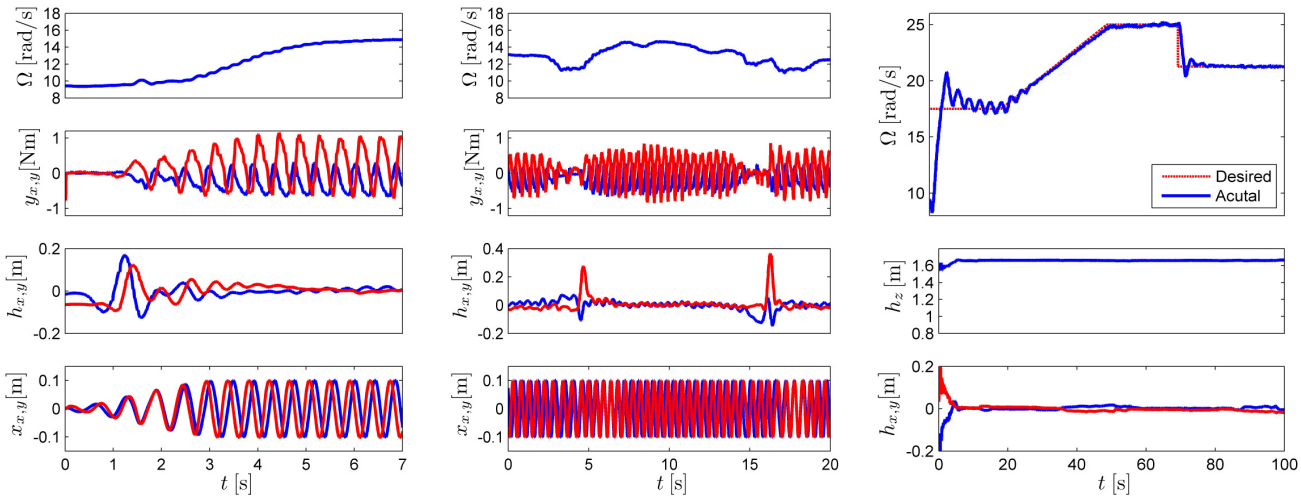


Figure 2: *a) The figure shows the initial frequency adaptation process (top plot) of the cooperative human-robot rope turning. The second plot shows the measured torque signal. The third plot shows the movement of a human hand, and the bottom plot shows the movement of a robot. b) The figure shows the behaviour of our proposed system when human changes the rotational direction of the rope. Frequency adaptation is shown in the top plot, and the second plot shows the measured torque signal. The third plot shows the movement of a human hand, and the bottom plot shows the movement of a robot. c) The figure shows the behaviour of our proposed system to track the desired rotational frequency of rope turning. Frequency adaptation is shown in the top plot, and the second plot shows the distance between human hand and the top end of the robot. The bottom plot shows the movement of a human hand.*

We also showed that the proposed approach is able to recognize the human motion, i.e. the rotational direction of the rope, and select proper movement primitive for successful human-robot cooperation in turning the rope.

References

[1] Chyon Hae Kim, Yonekura, K., Tsujino, H., Sugano, S., 2009. Physical control of the rotation center of an unsupported object rope turning by a humanoid robot, Humanoids 2009. 9th IEEE-RAS International Conference, pp.148-153.

[2] Furuta, K. 2003. Control of pendulum: from super mechano-system to human adaptive mechatronics. In Decision and Control, 2003. Proceedings. 42nd IEEE Conference on, volume 2, pages 1498 – 1507 Vol.2.

[3] Gams, A., Ijspeert, A. J., Schaal, S., and Lenarcic, J. 2009. On-line learning and modulation of periodic movements with nonlinear dynamical systems. Auton. Robots, 27(1):3–23.

[4] Ijspeert, A. J. 2008. Central pattern generators for locomotion control in animals and robots: A review. Neural Networks, 21(4):642 – 653.

[5] Maeda, Y., Takahashi, A., Hara, T., Arai, T. 2001. Human-robot cooperation with mechanical interaction based on rhythm entrainment-realization of cooperative rope turning, Robotics and Automation, 2001. Proceedings 2001 ICRA. IEEE International Conference on , vol.4, no., pp. 3477- 3482.

[6] Petrič, T., Curk, B., Cafuta, P., and Žlajpah, L. 2010a. Modeling of the robotic powerball: a nonholonomic, underactuated, and variable structuretype system. Mathematical and Computer Modelling of Dynamical Systems. 10.1080/13873954.2010.484237.

[7] Petrič T., Gams A., Žlajpah L. 2010, Frequency extraction based on adaptive fourier series: application to robotic yo-yo, ICINCO 2010.

[8] Petrič T., Gams A., Tomšič M., Žlajpah L. 2011, Control of rhythmic robotic movements through synchronization with human muscle activity, 2011 IEEE International Conference on Robotics and Automation, ICRA 2011, May 9-13, 2011, Shanghai, China, Proceedings, 2011.

[9] Pikovsky, A., Author, Rosenblum, M., Author, Kurths, J., Author, Hilborn, R. C., and Reviewer. 2002. Synchronization: A universal concept in nonlinear science. American Journal of Physics, 70(6):655–655.

[10] Righetti, L., Buchli, J., and Ijspeert, A. J. 2006. Dynamic hebbian learning in adaptive frequency oscillators. Physica D, 216(2):269–281.

[11] Ronsse, R., Lefevre, P., and Sepulchre, R. 2007. Rhythmic feedback control of a blind planar juggler. Robotics, IEEE Transactions on, 23(4):790 –802.

[12] Schaal, S., Mohajerian, P., and Ijspeert, A. 2007. Dynamics systems vs. optimal control – a unifying view, vol. 165 of Progress in Brain Research, pp 425 – 445.

[13] Shibata, T., Kaneko, T., Komoriya, K., Tanie, K., Kawasaki, H. 1994. Heuristic learning for lasso robot, From Perception to Action Conference, 1994., Proceedings, pp. 78- 87.

[14] Žlajpah, L. 2006. Robotic yo-yo: modelling and control strategies. Robotica, 24(2):211–220.

INTERACTIVE OBJECT LEARNING FOR HUMANOID ROBOTS

David Schiebener, Ales Ude

Department for Automation, Biocybernetics and Robotics
Jozef Stefan Institute
Jamova Cesta 39, 1000 Ljubljana, Slovenia
e-mail: david.schiebener@ijs.si

ABSTRACT

This paper describes a novel approach for learning of unknown objects by a humanoid robot. First, hypotheses about possible objects are generated and then validated by pushing them. This allows the segmentation and hence the learning of new objects in difficult cluttered environments. Visual features belonging to the object are accumulated over several pushing-and-verification iterations. We show that the learned model, in combination with the proposed segmentation process, allows robust object recognition.

1. INTRODUCTION

Segmenting an object using only visual information is in general a very difficult task. Passive vision systems have so far failed to get even close to the abilities of humans on this task [1][2]. The main reason for this is that no clear and comprehensive definition for the concept "object" has been found so far. For each principle that could be used to define the concept of object, e.g. closure, connectedness, etc., counterexamples can be found. Thus in general a sufficient criterion to decide if some part of an observed scene constitutes an object is not known.

Even though simple heuristics are in general not sufficient to find objects, they can give hints to generate hypotheses about their existence, which can then be tested using stronger criteria. As a robot is not constrained to passively observing a scene, he can use its manipulation abilities to physically interact with the hypothetical objects, and observe the outcome of its own actions. Like humans, the robot can use its (partial) control over the objects and the resulting visual input to observe - and learn about - the effects of its actions [3]. For example, moving an object can help to extract its boundaries [4].

Although grasping an object would be preferable in order to be able to move it in a controlled way, this is not easy if the object is unknown. A simpler alternative is to just push it. In our previous work [5] we showed that pushing can be useful for object segmentation. Here we extend this initial work by providing a methodology to discover more candidate surfaces that give hints about the existence of the object, and we developed a new approach that allows for reliable feature accumulation across a number of different snapshots. Based on these results we developed an object recognition system, which supports both autonomous object learning and object recognition.

2. OVERVIEW

The learning of new objects is done in four steps:

- **Generation of object hypotheses**: Visual features that seem to lie on a smooth surface patch are detected and grouped together.
- **Verification by pushing**: The hypothetical object is pushed. The resulting feature motion allows to verify which features belong to the object. Additional features are added if they move concurrently.
- **Feature accumulation**: The above step can be repeated arbitrarily many times to accumulate object features from multiple viewpoints.
- **Learning of visual appearance**: The appearance of the object is represented using a bag-of-features approach.

3. HYPOTHESIS GENERATION

The starting point for segmenting and learning unknown objects is to form hypotheses about possible objects. Here, only the visual information from the cameras of the robot is used. This visual information may be misleading, and therefore these hypotheses must then be examined further by pushing the hypothetical object and observing the induced feature motion.

Most objects in human-related environments consist of planar or curved surfaces. Hence it is reasonable to look for planar or cylindrical surface patches, which are mathematically simple to describe, to generate hypothesis about the existence of the objects.

We apply the Harris corner detector [6] to choose interest points in the camera images, which will be used both for hypothesis generation and object learning and recognition. The points determined by this detector are usually distinctive enough to allow for reliable matching in the two images from the stereo cameras. We can calculate the position of the corresponding 3-D point using the calibration of the camera pair [7]. There may be some incorrect points due to mismatches, but they are too few to affect the hypothesis generation.

Given a set of 3-D points, our goal is to find planes and cylinders that contain as many of these points as possible. For each surface patch, we have to expect that only a rather small part of all features belongs to it. To enable the detection of surface patches among many outliers, we apply the RANSAC algorithm [8], which enables us to find the parameters defining the surface patch that contains maximal subsets of feature points belonging to the parametric surfaces. RANSAC achieves this by randomly selecting a minimal number of points, which is sufficient to calculate the parameters of the sought for surface, and then counting how many points of the whole set lie within a tolerance of the defined surface.
The plane or cylinder containing the largest number of points is added to the list of hypotheses and its points are removed from the set. RANSAC can then be run again on the remaining points. This is repeated until no surface with more than a minimal number of points can be found.

3.1 Plane detection

A 3-D plane is defined by the equation $ax + by + cz + d = 0$ and contains all points (x,y,z) that fulfill this equation. The vector (a,b,c) is the surface normal. If it has unit length, then the above equation gives the distance of the point (x,y,z) to the plane (a,b,c,d). A plane is uniquely defined by three points that are not collinear. With this in mind, the implementation of RANSAC for planes is straightforward:

- repeat N_p times:
  - select 3 different points at random
  - calculate the plane parameters
  - check for each point if it lies within tolerance t_p of the plane, count the inliers
- return the parameters of the plane with maximal number of inliers

It can occur that a hypothesis extends to two or more objects which by chance contain points lying in the same plane. To avoid misled attempts of pushing in this case, we group the features of each plane using clustering. Single points that are far away (more than 2x standard deviation) from the cluster centers are discarded, because they are with high probability outliers. Sometimes a hypothesis containing a large object is accidentally divided by the above clustering process. However, this is not a serious problem for our system because the initial hypothesis will be expanded after the push (as other feature points on the object will move in unison with the initial hypothesis).

3.2 Cylinder detection

Finding cylinders in a point cloud is more complicated because the parameters of a cylinder cannot be determined so easily from a few points on its surface. We applied the algorithm proposed in [9], which uses a 2-stage RANSAC approach, first estimating the cylinder axis and then the appropriate radius and offset from the origin for that axis.

4. HYPOTHESIS VALIDATION BY PUSHING

By making the object move, visual features can be analyzed for coherent motion. This is a very strong evidence for deciding if they belong to the same object or not. While such information could not be obtained by passive observation only, inducing motion on the object resolves most of the ambiguities about object segmentation. We use simple pushing movements to verify the initial object hypotheses and to extend them to features that move coherently with the initial features. The initial hypotheses serve as a cue for promising points and directions of pushing.
A necessary prerequisite for the estimation of feature point motion is to be able to match the features before and after the push. For its descriptiveness and robustness to small rotations, we use SIFT descriptors [10] to find matches of the features in the images before the push and after it. For all initial features for which a corresponding feature is found, the new 3-D positions are calculated using stereo images.
To estimate the object motion, we use RANSAC again, as we have to expect feature mismatches, especially on the object that has moved because of possible occlusions and slightly different appearance. We make the assumption that the object moves as a rigid body. The parameters of such a motion can be obtained from three different pairs of corresponding points before and after the push [11]. If x_o is the initial position of a point, then its new position x_n is given by the transformation $x_n = R\, x_0 + t$, where R is a $3x3$ rotation matrix and t a translation vector.

Image 1: *The left image shows all Harris interest points, the middle and right one the generated hypotheses.*

After the object has been pushed, the initial hypothesis is evaluated to confirm whether the hypothetic feature points have moved as a rigid body or not. RANSAC is applied to estimate the transformation with which most of the points of the hypothesis concur. If the motion of a hypothesis is larger than a threshold, the features that moved coherently are considered validated, and those who did not are discarded. The hypothesis is ignored if the estimated parameters suggest that the hypothetical features did not move. If none of the generated hypotheses moved, another attempt to push one of them is made. If at least one of the hypotheses has moved, we assume to have found an object whose appearance needs to be learned.

5. OBJECT LEARNING AND RECOGNITION

To learn the appearance of the segmented object from multiple viewpoints, the object must be moved several times. At every step, new points are added to the hypothesis if they seem to belong to the object, and can be verified after the next push. As we use SIFT descriptors for feature matching between stereo image frames, it is an obvious choice to use these features for describing the object. However, it is possible to use any other desired local descriptor at locations of the confirmed points.

We use the "bag-of-features" approach [12] for representing the visual appearance of the learned objects. Essentially, a bag-of-features is a histogram of the feature descriptors present in the object. When learning a new object, we create such a histogram at every step.

When trying to recognize an object hypothesis, a BOF-histogram is calculated from the features belonging to it. This histogram can then be compared with the histograms of all known objects. As we have several histograms for each object, a k-nearest-neighbors decision can be used for recognition.

6. EXPERIMENTAL EVALUATION

We conducted experiments to evaluate the generation of object hypotheses in complex scenes, the segmentation and learning of unknown objects by pushing them repeatedly, and the recognition of objects using both our initial hypotheses and segmentation results that were improved by pushing the object several times.

To test the quality of the initial hypotheses, we carried out a number of experiments in differently complex scenes, each containing 5-8 objects that stand close together and partly occlude each other. 50% of the hypotheses approximately coincided with an object, 39% contained a part of an object, and 11% contained parts of two or more objects. In simple scenes the hypotheses are usually correct or contain a part of a large object.

We applied our system to the learning of 15 different objects. The number of features contained in each initial object hypothesis varied strongly between the different objects. For the evaluation of the object recognition system, in addition to the 15 test objects, another 25 objects were learned from presegmented images. Thus the complete database contained 40 objects. We tried to recognize the learned objects in complex scenes containing 5-8 objects.

On the average, the initial hypotheses lead to a recognition rate of 68%, which also gives an idea about their usefulness for segmentation. After the first push and the subsequent verification of the hypothetical feature points, the average recognition rate is 65%, which is -- somewhat surprisingly -- slightly lower than for the initial hypothesis. As now only the confirmed points are used for

Image 2: *Learning of an object over several iterations of pushing and verification.*

recognition, the effect of this first push was mainly to remove the features from the object hypothesis that did not move in unison with the majority of feature points, or were not found in the next image. By that, the number of features is reduced to around 32% of the size of the initial hypothesis. Apparently, this affects the recognition so strongly that the positive effect of eliminating the false features is voided. But after the second push, new confirmed features are added at each iteration, and now the positive effect is significant. The recognition rate immediately rises to 79% after the second push, 86% after the third and 92% after the fifth. It finally converges to a value between 92% and 95%.

7. CONCLUSION

In this paper, we presented a method for the segmentation and learning of unknown objects in unstructured environments. We generate initial object hypotheses from 3-D points, which were obtained through stereo vision, by detecting planar and cylindrical surfaces amongst them. The hypotheses are then verified, corrected and extended by pushing them repeatedly. Objects are learned using bag-of-feature histograms based on the SIFT descriptors of the points belonging to the object. We have shown experimentally that the objects learned this way can later be recognized, and that the segmentation by pushing can serve as a powerful methodology for recognition in complex scenes.

References

[1] **G. Kootstra, J. Ypma, and B. de Boer**, Active exploration and keypoint clustering for object recognition, *in: Proc. IEEE Int. Conf. Robotics and Automation, Pasadena, CA, 2008.*

[2] **G. Metta and P. Fitzpatrick**, Early integration of vision and manipulation, *Adaptive Behavior, vol. 11, no. 2., 2003.*

[3] **G. Metta and P. Fitzpatrick**, Grounding vision through experimental manipulation, *Philosophical Transactions of the Royal Society: Mathematical, Physical and Engineering Sciences, vol. 361, no. 1811, 2003.*

[4] **P. Fitzpatrick**, First contact: An active vision approach to segmentation, *in: Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, Las Vegas, Nevada, 2003.*

[5] **E. Stergarsek and A. Ude**, Object segmentation and learning through feature grouping and manipulation, *10th IEEE-RAS Int. Conf. Humanoid Robots, 2010.*

[6] **C. Harris and M. Stephens**, A combined corner and edge detector, *in: Alvey Vision Conference, pp. 147–151, 1988.*

[7] **R. Hartley and A. Zisserman**, Multiple View Geometry in Computer Vision, *Cambridge University Press, 2004.*

[8] **M. A. Fischler and R. C. Bolles**, Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography, *Communications of the ACM, vol. 24, no. 6, 1981.*

[9] **T. Chaperon and F. Goulette**, Extracting Cylinders in Full 3D Data Using a Random Sampling Method and the Gaussian Image, *in: Proc. Vision Modeling and Visualization Conference, 2001.*

[10] **D. G. Lowe**, Object recognition from local scale-invariant features, *in: Proc. Int. Conf. Computer Vision, Corfu, Greece, 1999.*

[11] **B. K. P. Horn**, Closed-form solution of absolute orientation using unit quaternions, *in: Journal Optical Society America A, vol. 4, 1987.*

[12] **G. Csurka, C. Dance, L. X. Fan, J. Willamowski, and C. Bray**, Visual categorization with bags of keypoints, *in: Proc. ECCV Int. Workshop on Statistical Learning in Computer Vision, 2004.*

Mobile Manipulation of Kitchen Containers

Christian Bersch, Dejan Pangercic, Benjamin Pitzer, Thomas Rühr, Bharath Sankaran,
Jürgen Sturm, Cyrill Stachniss, Michael Beetz, Wolfram Burgard
{pangercic, ruehr, sturmju, beetz}@cs.tum.edu
{Benjamin.Pitzer, Christian.Bersch}@us.bosch.com
bharath@seas.upenn.edu, {stachnis, burgard}@informatik.uni-freiburg.de

***Abstract*— In this work we report about our efforts to equip service robots with the ability to robustly operate articulated containers such as refrigerators and drawers in kitchen environments. We identified three important aspects for such systems: (1) the ability to detect fixtures on an articulated object, (2) to robustly open and close them and (3) to store and retrieve information about these objects in the map. In particular, we detect grasping fixtures such as handles and knobs in 3D point clouds using a RANSAC-based plane detection and subsequent clustering approach. Further, we developed two types of controllers to operate articulated objects: the first controller is model-free and incrementally opens articulated models. The second controller is model-based and estimates both the kinematic structure and the kinematic parameters. Finally, we store the inferred articulation models into our knowledge processing system *KnowRob* readily available for later interactions. All software components have been tested on two PR2 (Personal Robot 2) robots located at TUM and Bosch and are freely available on `ros.org`.**

I. Introduction

Autonomous mobile manipulators have demonstrated impressive new capabilities recently [1], [2]. Researchers have enabled robots in their labs to make pancakes, play pool, fold clothes, and more. The reasons behind this rapid progress are manifold: the growing interest in mobile manipulation, standardized platforms, open source code, and system integration, along with improvements in perception, control, learning, and planning. In this paper we report about the integrated effort of three PR2 Beta Sites on mobile manipulation of kitchen containers which includes opening and closing the doors and fetching the objects out of containers (see video `http://youtu.be/G3ygwaLflYg`). The effort gains maximum leverage from the open source algorithms in ROS and the standardized robotic platform PR2.

II. Detection of Furniture Fixtures

In the application pipeline we first proceed by detecting the container faces using RANSAC-based segmentation in 3D point clouds as put forth in [3]. In order to detect fixtures we first find point clusters that are within the polygonal prism of the container faces using euclidean distance measure and then fit RANSAC lines or circles to those clusters and thereby differentiate between handles and knobs. The geometric centroid is then used in order to grasp the fixture. To generate the 3D point clouds use head and torso mounted Kinect RGBD sensors.

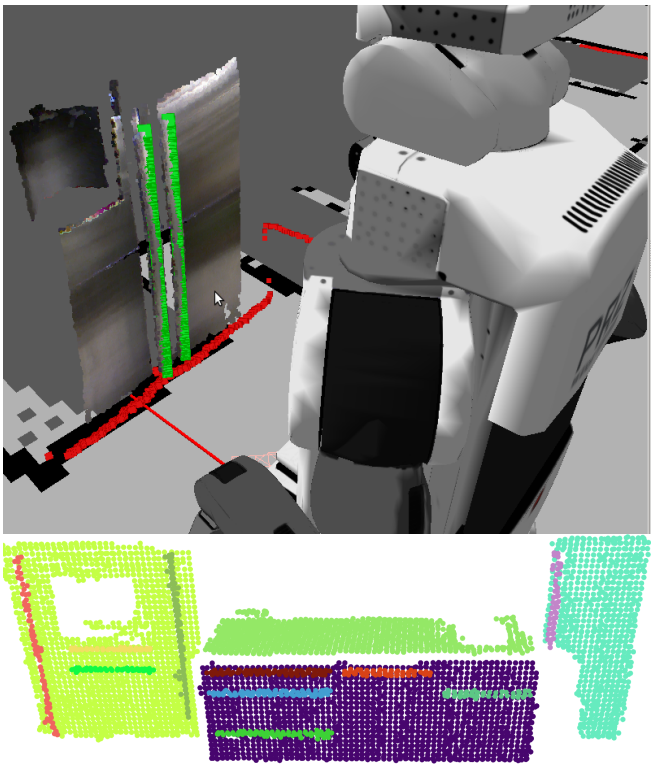


Fig. 1: Detection of handles at Bosch (top) and TUM (bottom).

In order to open the door we developed two types of controllers that assume no a priori knowledge of the articulation model of the door.

III. Generalized Container Opening Controller

Firstly we present a general controller (see Algorithm 1 and code [1]) that makes use of the compliance of the PR2 robot's arms and the force sensitive finger tip sensors. Since the arms lack force sensors, the algorithm uses the Cartesian error of the end effector (commanded vs. actual position) to determine when the maximum opening is reached. The algorithm relies on the grippers maintaining a strong grasp while the arms are compliant. This way the mechanism that is to be opened steers the arm in its trajectory even when there is a considerable difference between the pulling and the opening direction. The robot also adjusts its base position if

[1] `http://www.ros.org/wiki/ias_drawer_executive`

Algorithm 1: Controller for opening containers with unknown articulation model. Note: poses are stored as transformation matrices (translation vector and rotation).

Initialize p_0 = point on the handle candidate;
$p_1 = p_0 + n_{furnitureplane}; t = 0$
while *gripper_not_slipped_off AND Cartesian_error* $<$ *threshold* **do**
  if *d(p_{t+1},projection of robot footprint)* $< .1$ *m* **then**
    move_base(artif. workspace constr. for p_{t+1})
  move_tool(p_{t+1})
  stabilize_grasp() (see Figure 2)
  $Rel = p_0^{-1} * p_{curr}$ with current tool pose p_{curr}
  Extrapolate: Rel_s = scale $(Rel, (|Rel| + .05)/|Rel|)$
  $p_{t+2} = p_0 * Rel_s$
  t = t + 1
Return: Set of poses $P\{p_0 ... p_n\}$ representing the opening trajectory.

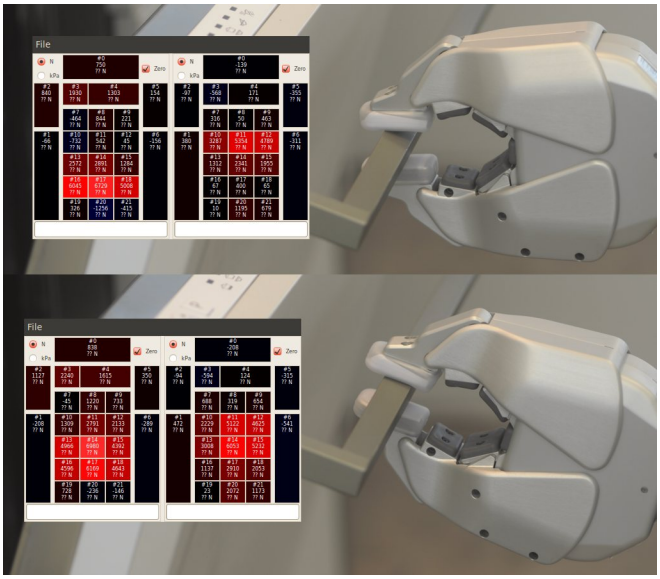


Fig. 2: The fingertip sensors are used to adjust the tool frame rotation to the rotated handle. Left part of the figure displays arrays of sensor cells on the PR2 robot's fingers. Asymmetry in the top-left part gives the measure of misalignment between the gripper and the handle.

the door mechanism requires this. The controller records a set of poses with the stable (aligned) grasps and stores them in KnowRob [2] system for later re-use. The controller works reliably as long as the force required to open the container is lower than the limit the friction of the gripper tips imposes.

A particular problem when opening unknown containers is the possibility of collisions of a container with the robot. This could occur e.g. when a low drawer is being opened and pulled into the robot's base. Since the articulation model is not known, an a priori motion planning step is not possible. We thus propose the following heuristic: we exclude tool

[2] http://www.ros.org/wiki/knowrob

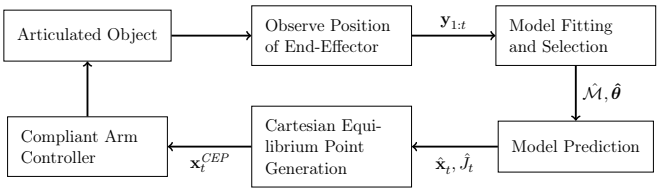


Fig. 3: Overall control structure. The robot iteratively estimates the kinematic model of the articulated object from the perceived trajectory of its end effector and evaluates it to generate the next Cartesian equilibrium point.

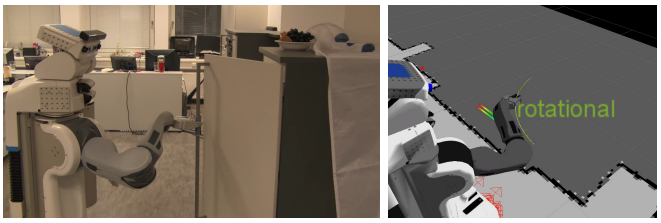


Fig. 4: The PR2 robot operates the fridge in the TUM kitchen and learns the kinematic model using Freiburg's articulation stack.

poses whose projections of the gripper to the floor fall close to or within the projection of the robot's footprint from the allowed workspace limit L of the gripper. This way, the robot tries to move backwards and prevents the collision.

IV. Kinematics-based Container Opening Controller

Second approach by Sturm et al. [4], [5] is depicted in Figure 3 which shows a block diagram of our controller for learning the kinematic model. We assume that the robot has already successfully grasped the handle of an articulated object and that a suitable initial pulling direction is known. The robot pulls in this direction using equilibrium point control (EPC) [6] and observes the resulting motion of its end effector. From this partial trajectory, it continuously (re-)estimates the kinematic model of the articulated object, that the robot uses in turn to predict the continuation of the trajectory. We deal with the workspace limits of the manipulator by a secondary controller that moves the omni-directional base of the robot so that the reachable volume of the manipulator is maximized. After the motion of the end effector has come to a rest, we estimate the range of valid configurations of the articulated object. In sum, this gives us the full kinematic model of the articulated object. Finally, we store the learned model in the *KnowRob* for later re-use as well. An example of this model learning step is visualized in Figure 4.

V. Pose Invariant Object Detection in Containers

After the containers are opened, pose information relating to the objects of interest needs to be discerned to enable a successful grasp. Given the constrained manipulation space within these containers, the detection and pose information

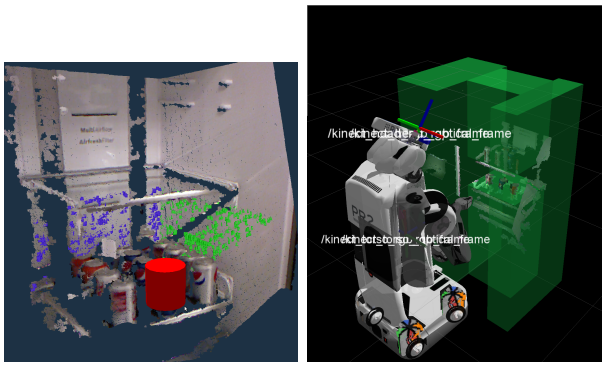


Fig. 5: **Left:** The red marker is used to represent the cluster extracted corresponding to the object of interest. The plane fits are illustrated by the colored overlay on the RGBD point cloud. **Right:** The collision model of the container is aligned with landmarks of the physical container.

need to be accurate within a small margin of error.
To this effect, we use a priori knowledge of the internal structure of the container as a prior in our object detection pipeline. In the detection pipeline, once the internal structure of the container is visible, planes are fit in the corresponding 3D point cloud. The intersection of perpendicular planes and their relative pose to the robot's base frame, gives us the relative pose of the container. This information, along with the knowledge of the planes is used as a prior for our object detection and clustering algorithms. We initially subtract the detected planes from the data to get the point cloud corresponding to the structure enclosed within the container walls. A SURF based nearest neighbor (FLANN [7]) template matching is performed to detect objects of interest in the scene around the location prior. To extract the pose and centroid of the cluster corresponding to the object of interest, the SURF matching results are used to augment the euclidean cluster segmentation of the point cloud (see Figure 5).

VI. Grasping Objects from inside Containers

In order to grasp detected objects from inside containers the robot has to avoid collisions with the container itself as well as with other objects while maneuvering its arm. For this, we use an RRT-based arm motion planner together with a collision map specifying obstacles for the robot to avoid. However, a perception-based collision map alone proved insufficient for computing collision-free arm motion trajectories as the robot cannot reliably detect a completely closed surface of the container, particularly when the container has reflective or transparent parts. Thus, we use a pre-specified collision model of the container, which is precisely aligned with the physical container object as visualized in Figure 5. This is achieved by locating the key landmarks such as the dominant planes of the container surfaces using 3D perception. Additionally, artificial collision obstacles are added to the collision map around the object to be grasped. This prevents the robot from colliding with objects that are hidden behind other objects and thus may have been missed during the object detection.

VII. Conclusions

The cooperation between different PR2 sites allowed us to join our efforts on building larger robotic applications. We combined our algorithms on door opening, perception of objects and planning, and extensively tested and validated our approaches at two different PR2 sites.

However, we also found that more generalization efforts are required so that these (existing) solutions become truly applicable in other environments. Therefore, we plan to direct our future research to generalize our approaches, for example, to deal with different shapes of handles, transparent doors and textureless objects. Authors would like to thank Willow Garage, Menlo Park, CA for providing and maintaining the PR2 robot and the ROS software.

References

[1] M. Cusumano-Towner, A. Singh, S. Miller, J. F. O'Brien, and P. Abbeel, "Bringing clothing into desired configurations with limited perception," in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA) 2011*, May 2011, pp. 1–8. [Online]. Available: http://graphics.berkeley.edu/papers/CusumanoTowner-BCD-2011-05/

[2] J. Bohren, R. B. Rusu, E. G. Jones, E. Marder-Eppstein, C. Pantofaru, M. Wise, L. Mosenlechner, W. Meeussen, and S. Holzer, "Towards autonomous robotic butlers: Lessons learned with the pr2," in *ICRA*, Shanghai, China, 05/2011 2011.

[3] R. B. Rusu, N. Blodow, Z. C. Marton, and M. Beetz, "Close-range Scene Segmentation and Reconstruction of 3D Point Cloud Maps for Mobile Manipulation in Human Environments," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, St. Louis, MO, USA, October 11-15 2009.

[4] J. Sturm, A. Jain, C. Stachniss, C. Kemp, and W. Burgard, "Operating articulated objects based on experience," in *IROS*, Taipei, Taiwan, 2010.

[5] J. Sturm, C. Stachniss, and W. Burgard, "Learning kinematic models for articulated objects," *Journal on Artificial Intelligence Research (JAIR)*, vol. 41, 2011.

[6] A. Jain and C. Kemp, "Pulling open doors and drawers: Coordinating an omni-directional base and a compliant arm with equilibrium point control," in *ICRA*, Anchorage, AK, USA, 2010.

[7] M. Muja and D. G. Lowe, "Fast approximate nearest neighbors with automatic algorithm configuration," in *International Conference on Computer Vision Theory and Application VISSAPP'09)*. INSTICC Press, 2009, pp. 331–340.

SKIING ROBOT NAVIGATION LEARNING

Luka Peternel, Tadej Petrič, Bojan Nemec
Department of Automation, Biocybernetics and Robotics
Jozef Stefan Institute
Jamova 39, 1000 Ljubljana, Slovenia
Tel: +386 1 4773419; fax: +386 1 4251038
e-mail: lukapeternel@gmail.com

ABSTRACT

The main focus of this research is to teach skiing robot how to navigate on a ski track based on examples provided by a human. We accomplished, tested and compared two methods of acquiring knowledge through generalization: artificial neural networks (NN) and Gaussian regression process (GPR). We tested several types of neural networks and compared how successful is each of the types in learning and function fitting. We also tested two approaches of Gaussian regression process. The first approach uses GPR to calculate output in each sampling moment based on the input sensory data and data from demonstrated examples. In the second approach we collect input sensory data only at the beginning of each turn and generated the trajectory for that entire turn. Resulting trajectories were presented with dynamic movement primitive (DMP) and generalized with Gaussian regression process in order to generate output trajectory for any kind of input sensory data. For comparison we also tested the navigation method using a mathematical model which is an example of non-learning method.

1 INTRODUCTION

In recent years there has been a considerable focus on research in humanoid robotics. In humanoid robotics we try to imitate the human body and its motion. The most common motion of a humanoid robot is walking. Inspiration for more complex motions can be found in certain sports. In case of skiing robot we found an inspiration in alpine skiing.

Humanoid robots are in most cases mobile robots. Since skiing robot is also a mobile robot we have to deal with the problem of navigation of mobile robots. In order for the skiing robot to successfully navigate in a certain environment, it must be able to sense the environment and determine its position. If obstacles are present the robot must detect them in order to prevent collision by avoidance. In case of skiing robot, ski gates are the obstacles. For detection of ski gates we use visual sensors and computer vision. Computer vision must be able to detect both position and color of the ski gates since skiing is not just avoiding obstacles but also avoiding them in a certain pattern.

Most of the robots nowadays conduct predefined operations in a controlled environment. If they face a slightly different situation they are usually unable to handle it. In this case the control algorithm needs to be modified or changed manually. Research in robotics tends toward machine learning which is much more robust and better in dealing with diverse situations. We should be able to show robot a few examples how to perform certain task in order to make it be able to adjust itself when encountering diverse situations. Artificial neural networks [3] and Gaussian regression process [2] are among the most common and suitable methods of machine learning.

2 NAVIGATION WITH MATHEMATICAL MODEL

For purpose of comparison we first realized navigation with mathematical model which was already used by Lahajnar [1] and introduced by Fajen and Warren. This model is described by equation 2.1.

$$\ddot{\phi} = -b\dot{\phi} - k_g(\phi - \psi_g)(e^{-c_1 d_g} + c_2) + \\ + k_o(\phi - \psi_o)(e^{-c_3|\phi - \psi_o|})e^{-c_4 d_o} \tag{2.1}$$

where Φ is angle of robot's current heading, Ψ_g is heading of goal, Ψ_o is heading of obstacle, d_g is distance to the goal and d_o is distance to the obstacle.

There are three components in equation 2.1: damping component, goal component and obstacle component. Goal component attracts robot's heading toward the goal. Attractive acceleration falls proportionally when robot is reducing the angle between its heading and the heading of the obstacle and increases exponentially when the distance to the goal is decreasing. Obstacle components reflects robot heading away from the obstacle. Reflective acceleration increases exponentially when robots is heading toward the obstacle to a certain point and then falls when Φ-Ψ_o is close to zero in order to allow the robot to pass the obstacle. Reflection increases exponentially when robot is closing the distance between itself and the obstacle since the reaction should be faster as we get near the obstacle.

Side of avoiding the obstacle in this model is random. We have to modify this model in a way that it will follow a certain pattern of obstacle avoidance. We defined that our skiing pattern will require robot to move around the green gate on the left side and around the red gate on the right

side. We achieved this by monitoring the color of the gates with an appropriate algorithm that changed the side of obstacle avoidance in accordance to our skiing pattern.
Since we control robot's turning rate with tilting it and consequentially tilting the carving skis, we have to realize a connection between equation 1.1 and angle of tilt. This connection can be found in angular velocity (equation 2.2).

$$r(t) = \frac{v(t)}{\dot{\phi}(t)} \tag{2.2}$$

where v is velocity of the robot and r is a turning radius. From turning radius we can then get the tilt through the model of carving skis. This also compensates impact of the velocity on the turn. We also included a more direct compensation by scaling the angular acceleration of the heading with robot velocity factor and included a low-pass filtration of velocity in order to eliminate the oscillation.

3 NEURAL NETWORKS

Artificial neural networks are simplified models of human brain which is made of biological neural networks. Most common use of artificial neural networks is in finding and learning the association between the given inputs and outputs in order to make an appropriate function to fit that association. Neural networks consist of neurons which are divided into layers and are connected with each other by weighted connections. The first layer is called the input layer, the last layer is called the output layer and the layers in-between are called hidden layers. A neuron is composed of the activation function and output transfer function. The activation function collects and sums the weighted connections that are connected to it. The result of the activation function is then feed to transfer function whose output is also the output of the neuron. The most commonly used transfer functions are step function, logistic function and Gaussian function.
Learning process of neural network is done by changing and adjusting weights of connection between the neurons. A common learning rule is called backpropagation of error or Delta rule. According to this rule weights has random values at the beginning of the learning process. We feed an input sample to the input layer of neural network and it then calculates the output. This output is compared to the output from the sample. Network first adjusts values of the weights between the connection of the last and prior to the last layer according to the error. Then it calculates the outputs of the prior to the last layer and compares them to the desired and adjusts weights accordingly. This process continues recursively to the input layer. Neural network repeats this process until output error is small.

3.1 Feed-forward neural networks

In feed-forward networks, information moves in only one direction (from input to output). We tested two feed-forward neural networks: a casual feed-forward network and a cascade-forward network. A casual feed-forward network has connections only between neighboring layers of neurons. Input of neuron in cascade-forward network is not connected only to the outputs of neurons in neighboring layer but also to the neurons in all previous layers.

4 GAUSSIAN REGRESSION PROCESS (GPR)

Gaussian regression process is based on generalization of data. It generates a probability function that fits the training data. Output value for a given input value can be described through Gaussian noise model:

$$y = f(x) + N(0, \sigma_n^2) \tag{4.1}$$

Regression is a search for $f(x)$ which is a mean value of normal distribution.
Equation 4.2 represents a covariance function that calculates the correlation between the two inputs (x_i and x_j).

$$k(x_i, x_j) = \rho_f^2 \exp\left[\frac{-(x_i - x_j)^2}{2l^2}\right] + \rho_n^2 \delta(x_i, x_j) \tag{4.2}$$

GPR calculates the covariance function 4.2 among all possible combination of inputs. These combinations then form the elements of matrix K.

$$K = \begin{bmatrix} k(x_1, x_1) & \cdots & k(x_1, x_n) \\ \vdots & \ddots & \vdots \\ k(x_n, x_1) & \cdots & k(x_n, x_n) \end{bmatrix} \tag{4.3}$$

Prediction of output y_* for a new input vector X_* can be modeled with the following expression:

$$\begin{bmatrix} Y \\ y_* \end{bmatrix} \sim N\left(0, \begin{bmatrix} K(X,X) + \rho_n^2 I & K(X, X_*) \\ K(X_*, X) & K(X_*, X_*) \end{bmatrix}\right) \tag{4.4}$$

The best estimate for y_* is mean value of this distribution.

$$\overline{y}_* = \mu(X_*) = K(X_*, X)\left[K(X,X) + \rho_n^2 I\right]^{-1} Y \tag{4.5}$$

4.1 Dynamic Movement Primitive (DMP)

There are several ways to describe trajectories of motion. The easiest method for describing a trajectory is to store points and the corresponding time values. This method requires a considerable amount of space and is impractical for processing. DMP offers us a much better method for presentation of trajectories where the entire trajectory can be described only with few parameters.
DMP is based on nonlinear dynamic system response and is described with the following system of equations:

$$\tau \dot{z} = \alpha_z(\beta_z(g - y) - z) + f(x) \tag{4.6}$$

$$\tau \dot{y} = z \tag{4.7}$$

where y is the current position, g is goal, τ is duration of trajectory and $f(x)$ is a nonlinear function which allows arbitrary shape of trajectory.

$$f(x) = \frac{\sum_{i=1}^{n} w_i \psi_i(x)}{\sum_{i=1}^{n} \psi_i(x)} \quad (4.8)$$

$$\psi_i(x) = \exp\left[-h_i (x - c_i)^2\right] \quad (4.9)$$

where ψ_i are radial base functions distributed among the trajectory. Parameters c_i are centers of these radial based functions and h_i their widths. In order to avoid time dependency of function f, we replace time variable with phase variable x whose dynamic is described by equation 4.10.

$$\tau \dot{x} = -\alpha_x x \quad (4.10)$$

Weight of function $f(x)$ is decreasing when time increases therefore system of equations 1.8-1.9 converges to the goal $[y\ z] = [g\ 0]^T$.

5 EXPERIMENTS AND RESULTS

5.1 Neural networks

We used information about position of ski gates obtained by computer vision and velocity of skiing robot as an input data for neural network. The output was the tilt of carving skis which translates into a turning radius. We determined that information about the closest and the second closest ski gate is sufficient information for neural network to successfully navigate through the track. Based on experimentation in simulation environment we determined that 25 hidden layers is an optimal number for the complexity of our problem. We used 30 examples where human manually demonstrates how to successfully navigate different tracks. Data during the demonstration was obtained with sampling frequency of 100 *Hz*. Tracks used for demonstration have to be diverse in order for trained neural network to be able to successfully navigate a variety of different tracks.

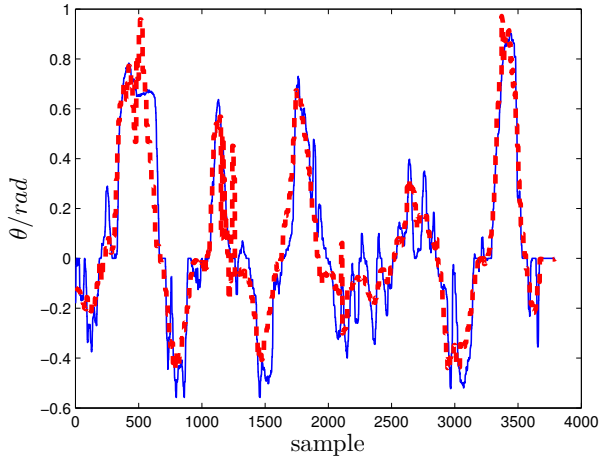


Figure 1: *Validation for neural network*

Once network finished the learning process we tested it by feeding it with validation data from a demonstration track that was not used in its training set. We compared the output generated by the neural network with the desired output from the validation data. This comparison is shown in figure 1. The blue curve is the desired output and the red curve is an output generated by the neural network. We can see that the response is satisfactory.

5.2 Gaussian regression process

We first tested the method where we use GRP in the same manner as neural network. The inputs and outputs are the same. We used the same training set of data as we did in case of neural network in order to be able to make a better comparison between the results of both learning methods.
GPR calculates ski tilt (output) by generalization of training data in each sampling moment. This is done by calculating correlation between sensory information obtained in that movement and vast amount of training data. During the simulations, it proved that amount of training data was too high for GPR to successfully calculate the correlation between all possible combinations. In order to give GPR enough time for calculation of the output, we were forced to reduce the sampling frequency of our system from 100 *Hz* to 10 *Hz*. We also had to reduce the amount of training data by reducing the examples of tracks from 30 to 15 and take only every 150th sample from those examples.

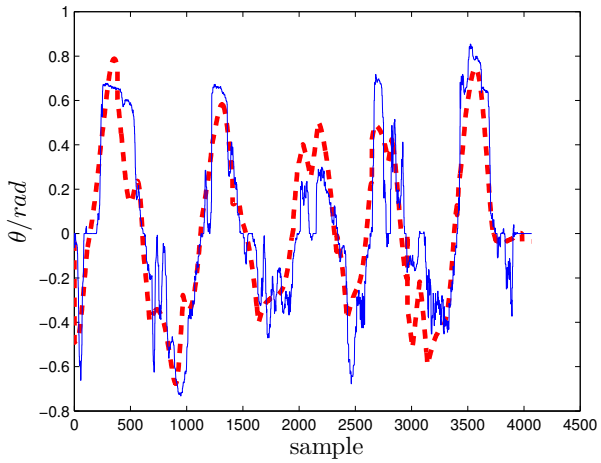


Figure 2: *Validation for Gaussian regression process*

Figure 2 shows a validation check for GPR method. The blue curve is the desired output from the validation example that was not used in the training set and the red curve is an output generated by GPR according to the training examples. The matching is sufficient for a successful navigation.

5.3 GPR used for trajectories presented with DMP

This method still uses GPR for generalization and function fitting but uses a different data and navigation approach. In contrast to the previous GPR method, this method does not calculate the tilt of robot in each sampling time. It calculates the trajectory of tilt for an entire turn based on the sensory information obtained only at the beginning of the turn. The

training examples consist of sensory information at the beginning of each turn and corresponding output trajectories which led to that particular turn.
The experiments proved that this method of navigation is less suitable for the skiing robot. The problem of skiing robot navigation is too complex and the calculated tilt trajectory for the entire turn based on sensory information obtained only at the beginning of the turn is in many cases not successful in correctly navigating around the gates. Robot velocity is constantly changing as a result of rough surface or different inclination of slope. To slightly compensate this problem we scaled the duration of trajectory τ in equation 1.8 with a velocity factor. On top of that, the connection between the ski tilt trajectory and the actual trajectory of robot position is complex.

5.4 Comparison between the methods of navigation

We compared the two machine learning methods of skiing robot navigation by giving them a task of navigating through the same track. We also tested navigation with mathematical model on the same track with parameter setting of equation 2.1 which allowed a similar skiing style as in case of neural networks and GPR. This was achieved by high parameter c_3 and low parameter k_o which contribute to lower obstacle reflection and therefor allows robot to ski closer to the gates.

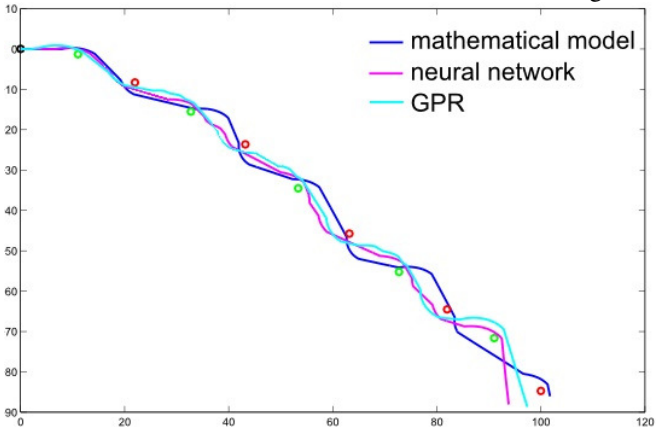


Figure 3: *Test of all three methods in simulation environment*

Figure 3 shows the results of navigation using all three methods. We can see that both method with neural network and method with GPR are successful. Method of navigation with mathematical model was successful for the most part of the track but robot gained too much speed at the end as a consequence of skiing close to the gates and was therefore unable to successfully navigate through the last 3 gates. In case of neural network robot completed the track in 24.0 *s* and achieved average velocity 6.09 *m/s*. GPR completed the track in 25.3 *s* and had an average velocity 5.86 *m/s*.
We determined that overall the best method for navigation of skiing robot is navigation with neural networks. We tested this method in real environment where we used skates instead of carving skis. The experiments are shown in figure 4.

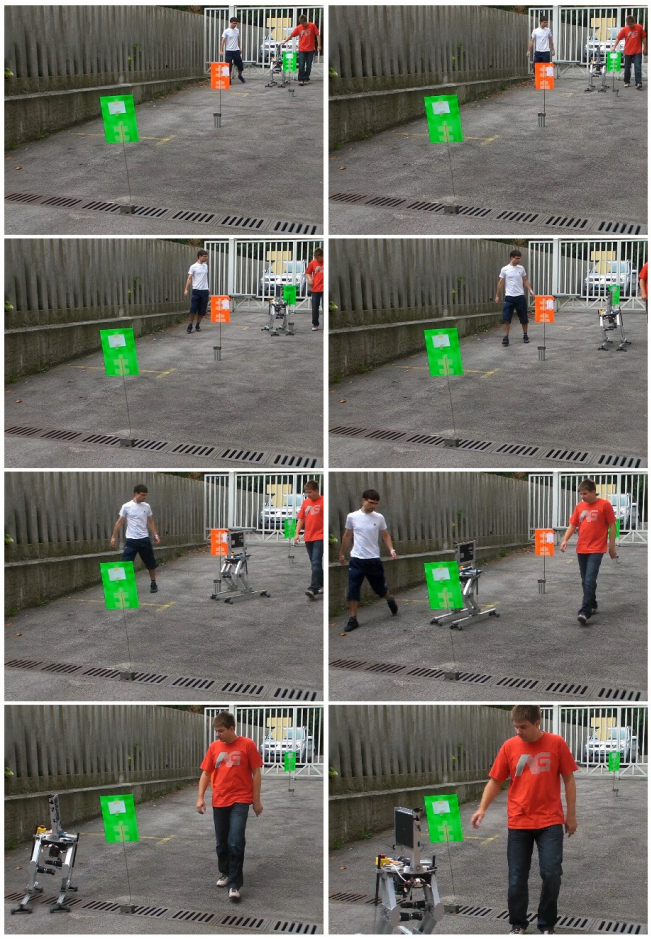


Figure 4: *Experimentation in real environment*

6 CONCLUSION

Our experiments proved that methods of machine learning (NN and GPR) are in most cases superior to the methods of navigation with a mathematical model. The greatest advantage of machine learning is that the robot can acquire knowledge through training examples which allows it to be successful in solving wide range of problems where mathematical model is only successful in solving limited range of problems. One disadvantage of machine learning is the lengthy and difficult process of acquiring training examples.
In future we will test the developed methods on a ski track where robot will be using carving skis instead of skates.

References

[1] L. Lahajnar. Avtonomno smučanje humanoidnega robota. *PhD thesis*. Faculty of electrical engineering and computer science. University of Ljubljana. 2009.

[2] D. Forte, A Ude, A. Kos. Uporaba Gaussove regresije pri strojne učenju prijemala. *Proc*. Odsek za avtomatiko, biokibernetiko in robotiko. Inštitut Jožef Štefan. In Slovenian.

[3] I. Kononenko, M. R. Šikonja.Inteligentni sistemi. *Book*. Založba FE in FRI, Ljubljana. 2010.

Robotsko posnemanje gibanja s prilagajanjem spremenljivi podlagi

Andrej Gams, Tadej Petrič, Jan Babič, Aleš Ude, Leon Žlajpah
Department for Automation, Biocybernetics and Robotics
Jozef Stefan Institute
Jamova cesta 39, 1000 Ljubljana, Slovenia
e-mail: andrej.gams@ijs.si

ABSTRACT

Directly imitating human movement does not guarantee the same behavior of the robot because of their different dynamical properties and perhaps different conditions under which the task is being performed. In the paper we show a method that constrains the extracted movement primitives when mapping them to a robot, taking into account a critical criterion of the task. Our approach is based on prioritized task control and allows direct movement transfer as long as the selected criterion is met. It only constrains the movement when the criterion approaches a critical condition. The critical condition thus triggers a reflexive, subconscious behavior, which has higher priority than the desired, conscious movement. We demonstrate the properties of the algorithm on a real, human-inspired leg robot developed in our laboratory, imitating squatting movement with changing ground inclination.

1 UVOD

Ko nekoga prosimo, naj posnema gibanje, ki ga izvajamo bo ta oseba najverjetneje zajela bistvo gibanja. Primer tega je npr. izvajanje počepa, kjer se oseba ne bo prevrnila čeprav ga izvaja na drugačni podlagi oz. čeprav ni povsem enaka demonstratorju. Direkten prenos gibanja sklepov z demonstratorja na robota ponavadi ne opravi želene naloge [1], torej se robot ob počepu prevrne. Razlog temu so različne dinamične lastnosti posameznih segmentov, možnosti aktuatorjev, ter zunanji pogoji. Gibanje, ki ga prenesemo na robota, mora biti spremenjeno, tako da upošteva te razlike.

Raziskovalci so preizkusili že veliko različnih načinov vodenja robotov s posnemanjem gibanja. Eden od pristopov je posnemanje gibanja v kartezičnih koordinatah, npr. gibanje vrha roke. To lahko popolnoma ignorira posneto gibanje sklepov demonstratorja. Gibanje sklepov robota nadzoruje algoritem inverzne kinematike. Gibanje v ničelnem prostoru naloge in prioritizirano vodenje nalog lahko dodatno uravnava gibanje posameznih sklepov kot sekundarna naloga [2, 3]. Tak način vodenja se pogosto uporablja za izmikanje oviram [4].

V prispevku predlagamo spremenjeno shemo prioritiziranega vodenja, kjer je posneto gibanje sklepov sekundarna naloga, vendar se primarna naloga, ki jo podaja izbrani kriterij, izvaja samo ob doseženem pragu taistega kriterija. Dokler se praga ne doseže, se izvaja samo sekundarna naloga. Ko enkrat dosežemo prag izbranega kriterija, primarna naloga prevzame vodenje in sekundarna naloga se izvaja samo še v njenem ničelnem prostoru. Prehod je gladek in zvezen. Prehod nazaj v izvajanje samo sekundarne naloge je prav tako gladek in zvezen. Na ta način posnemamo refleksne odzive nevroloških sistemov v ljudeh, ki imajo višjo prioriteto kot zavestno gibanje [5].

Uporabnost metode je v prispevku prikazana na nalogi stabilnosti robotske noge na podlagi s spremnljivim naklonom ob ohranjanju lege v napol skrčenem stanju ter ob izvajanju počepa. Robot ima v sagitalni ravnini podobne kinematične lastnosti kot človek [6]. Kriterij stabilnosti določa lega normalizirane točke ničelnega navora (ang. Zero Moment Point). Robot izvaja posneto gibanje po sklepih. Ko doseže prag stabilnosti, predlagana shema vodenja zvezno preklopi na primarno nalogo - zagotavljanje stabilnosti.

V naslednjem poglavju najprej opisujemo model robota ter določitev ničelne točke navora, ki določa kriterij stabilnosti. Tretje poglavje opisuje predlagano prioritizirano shemo vodenja. Rezultati eksperimentov so podani v četrtem poglavju. Na koncu sledijo še zaključki.

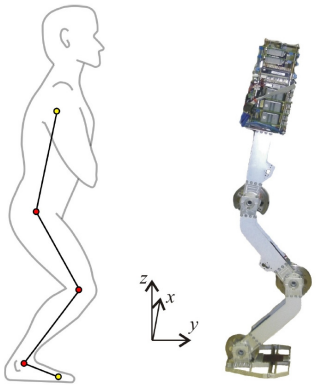
2 DOLOČITEV NIČELNE TOČKE NAVORA

Ob izvajanju počepa z nogama skupaj je podporni poligon omejen na dolžino stopal, kar precej omeji območje stabilnosti. Pogoj stabilnosti zahteva, da je točka ničelnega navora (ZMP) znotraj podpornega poligona. Stabilnost je eden glavnih problemov pri vodenju humanoidnih robotov. V našem primeru humanoidne robotske noge smo zaradi lastnosti robota omejeni na zagotavljanje stabilnosti v sagitalni ravnini (naprej-nazaj). Stabilnost v lateralni ravnini (levo-desno) zagotavlja široko stopalo. Poleg tega je robot planaren in gibanj v lateralni ravnini niti ne more izvajati.

Točko ničelnega navora je kot kriterij stabilnosti v robotiki definiral Vukobratovič [7]. Na kratko jo lahko definiramo kot točka na tleh okoli katere je vsota vseh navorov aktivnih sil enaka nič [8].

Humanoidno robotsko nogo, ki jo uporabljamo za demonstracijo uporabnosti predlaganega mehanizma, lahko

predstavimo kot trojno invertirano nihalo, ki ima večino mase v zgornjem segmentu. Ta modelira človeško telo. Taka struktura je v osnovi zelo nestabilna. Robot ter shema človeka sta predstavljena na sliki 1. Podobnost je očitna.



Slika 1: *Silhueta človeka in pa mehanizem robotske noge.*

Robot je najbolj stabilen, ko je ZMP na sredini podpornega poligona - stopala. To točko označimo z MSP (*ang.* Most Stable Point). Najbolj nestabilen je, ko je ZMP na robu podpornega poligona, spredaj ali zadaj. Ti točki označimo z LSP (*ang.* Least Stable Point). Bližje kot je ZMP LSP-ju, bolj nestabilen je robot. Ko ga enkrat doseže, se robot prevrne. Lokacijo ZMP točke računamo z dinamičnim modelom robota. *i*-temu segmentnu pripišemo maso m_i, v težišču $\mathbf{r}_i = [x_i, y_i, zi]^T$, vztrajnostni tenzor $\mathbf{I}_i$ in pa kotno hitrost ω_i. Zunanje sile in navore označimo z $\mathbf{F}_{i,k}$ in $\mathbf{M}_{i,j}$, kjer indeks *k* številči sile, *j* pa navore, ki delujejo na *i*-ti segment. Skupna enačba rotacijskega in translacijskega gibanja sistema v poljubni točki $\mathbf{p} = [x_p\ y_p\ 0]^T$ na ravnini $z = 0$ je podana z

$$\sum_i \{m_i(\mathbf{r_i}-\mathbf{p})\times(\ddot{\mathbf{r}}_i+\mathbf{g})+\mathbf{I_i}\dot{\omega}_i+\omega_i\times\mathbf{I_i}\omega_i\} +\mathbf{C}=\mathbf{M_p}, \quad (1)$$

kjer

$$\mathbf{C}=-\sum_j \mathbf{M}_j-\sum_k (s_k-p)\times\mathbf{F}_k \quad (2)$$

Tu je s_k vektor, ki kaže proti poziciji zunanje sile $\mathbf{F}_k$, ki deluje na robota. **g** je vektor težnostnega pospeška. Spreminja se s spremembo naklona podlage. $\mathbf{M}_p$ je rezultanta navorov v izbrani točki **p**. Po definiciji ZMP [7] samo navor $\mathbf{M}_p = [0\ 0\ \mathbf{M}_z]^T$ deluje v točki $\mathbf{p}_{zmp} = [x_{zmp}\ y_{zmp}\ 0]^T$ in če zanemarimo vztrajnosti $\mathbf{I}_i$ [9] ter brez zunanjega vpliva na robota, je gibanje ZMP v posameznih smereh podano z

$$x_{zmp}=\frac{\sum_j m_i\{x_i(\ddot{z}_i+g_z)-z_i(\ddot{x}+g_x)\}}{\sum_i m_i(\ddot{z}_i+g_z)}, \quad (3)$$

in

$$y_{zmp}=\frac{\sum_j m_i\{y_i(\ddot{z}_i+g_z)-z_i(\ddot{y}+g_y)\}}{\sum_i m_i(\ddot{z}_i+g_z)}. \quad (4)$$

Razmerje med pozicijo ZMP ter hitrostjo gibanja sklepov robota **q** podaja

$$u_z=\mathbf{J}_z\dot{\mathbf{q}}, \quad (5)$$

kjer

$$\mathbf{J}_z=\left[\frac{\partial y_{zmp}}{\partial q_i}\right] \quad (6)$$

Tu je u_z hitrost ZMP v odvisnosti od izbrane naloge. Za primer humanoidne robotske noge to postane $u_z = \dot{y}_{zmp}$ – hitrost ZMP v *y* smeri. Jacobijeva matrika $\mathbf{J}_z$ je funkcija trenutnih vrednosti posameznih sklepov. Povezuje vektor hitrosti v sklepih $\dot{q}$ in hitrost ZMP.

3 PRIORITIZIRANO VODENJE

Naloga humanoidne robotske noge je izvajati počep, kjer pa so gibi posameznih segmentov določeni vnaprej iz človeških demonstracij. Tako tudi ne upoštevajo razlik v dinamiki in kinematiki med demonstratorjem in mehanizmom. Gib naj bi mehanizem ponovil tudi ob drugačnem ter spreminjajočem naklonu podlage. Tako gibanje je smiselno le v primeru, da se robot ob tem ne prevrne. To pomeni, da je ohranjanje stabilnosti primarna naloga. V tem razdelku bomo pokazali, kako lahko ponovimo gibanje demonstratorja ter posnete gibe spremenimo le v primeru, ko bi se robot prevrnil. Razlika s klasičnim vodenjem s prioritetami nalog je v vodenju ZMP, kjer lega ZMP ni definirana, temveč se lahko prosto premika znotraj podpornega poligona, dokler ostane pod izbranim pragom. Dokler je pod pragom je gibanje sklepov robota povsem določeno s posnetim gibom demonstracije. Ko se ZMP približa izbranemu pragu blizu LSP, zagotavljanje stabilnosti zvezno in gladko prevzame nadzor ter omogoča samo še gibanje, ki robota ne bi prevrnilo. Želeno gibanje sklepov postane sekundarna naloga zagotavljanju stabilnosti. Na ta način deluje kot neke vrste refleksno obnašanje. Za zagotavljanje stabilnosti definiramo indeks stabilnosti kot funkcijo normalizirane ZMP na liho potenco, torej

$$z_n=\left(\frac{y_{zmp}-y_{MSP}}{|\,y_{LSP}-y_{MSP}\,|}\right)^{2n-1}, n=1,2,3,\ldots, \quad (7)$$

Vodenje stabilnosti podaja

$$\dot{y}_{zmp}=u_z=K_p e_z=K_p(z_d-z_n)=K_p(-z_n), \quad (8)$$

kjer je K_p proporcionalno ojačenje in z_d referenčno gibanje ZMP $z_d = 0$. Hitrost sklepov robota je s tem definirana kot

$$\dot{\mathbf{q}}=\mathbf{J}_z^{\dagger}u_z+\mathbf{N}_z^{'}\dot{\mathbf{q}}_n \quad (9)$$

kjer je $\dot{q}$ hitrost v sklepih, kot je bila določena v demonstraciji, $\mathbf{J}_z^{\dagger}$ je psevdo-inverz (6) in matrika $\mathbf{N}_z^{'}$ je podana z

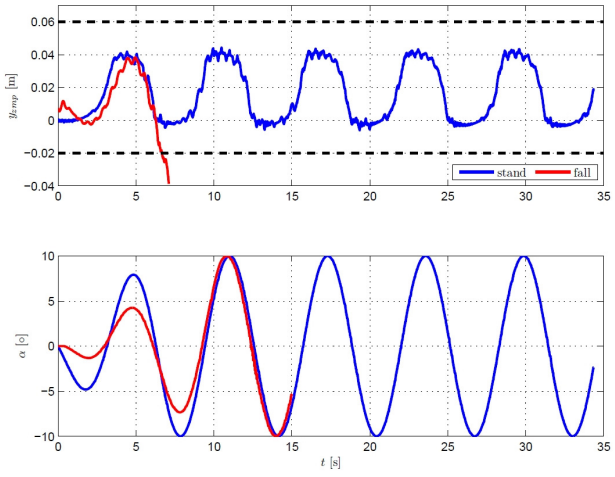
$$\mathbf{N}_z{}' = \mathbf{I} - |\, z_n \,|\left(\mathbf{J}_z^{\dagger}\mathbf{J}_z\right).. \qquad (10)$$

Zapis (10) ničelnega prostora $\mathbf{N}_z^{'}$ omogoča poljubno gibanje sklepov, če je le $|z_n|$ blizu 0 ($|z_n|\approx 0$). S tem se ZMP lahko prosto premika znotraj podpornega poligona. To deluje, ker je v enačbi (7) z_n na potenco $2n$-1 ter s tem praktično nič na večjem delu intervala -1 < z_n < 1. Ta interval je odvisen od velikosti n, višji kot je n, bliže LSP se lahko približa. Dokler je $|z_n|\approx 0$ vodenje stabilnosti nima vpliva na gibanje sklepov. Prvi element v enačbi (9) postane 0, ker je $u_z = 0$. Drugi element v enačbi (9) postane $\mathbf{I}\dot{q}_n$ To pomeni, da se posnete hitrosti iz demonstracije prenesejo na robota. Ko je $|z_n|$ blizu 1 ($|z_n| = 1$), ničelni prostor, ki ga podaja enačba (10), prevzame klasično obliko in omogoča samo gibanje v ničelnem prostoru primarne naloge - stabilnosti.

4 EKSPERIMENT NA REALNEM ROBOTU

V tem razdelku so predstavljeni rezultati eksperimentov na humanoidni robotski nogi, ki smo jo postavili na 6-stopenjsko Stewart-ovo ploščad, s katero smo spreminjali naklon podlage.

Želeno gibanje sklepov robota pri izvajanju počepa smo dobili iz posnetega gibanja demonstratorja. Gibanje smo zajeli z merilno opremo Optotrak. Direkten prenos posnetega gibanja na robota se zaključi s padcem robota, ki ne ohrani stabilnosti. To je lepo vidno na sliki 2, ki prikazuje gibanje normirane točke ZMP ob ohranjanju napol pokrčene stoječe lege.



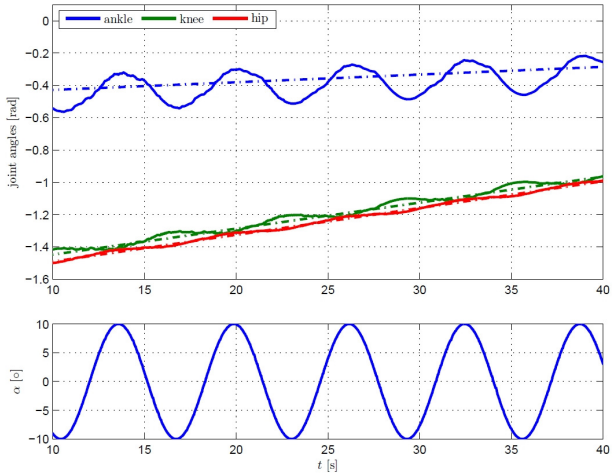
Slika 2: *Zgoraj: Gibanje normirane ZMP ob ohranjanju stabilnosti (modra) ter brez ohranjanja. Naklon podlage v spodnjem grafu.*

Nespremenjeno gibanje (rdeča črta) vodi ZMP izven območja stabilnosti in robot pade. Nasprotno, ob zagotavljanju stabilnosti, ko se ZMP približa meji, kar prikazuje črta oblike črta-pika, ohrani robota v stoječi legi.

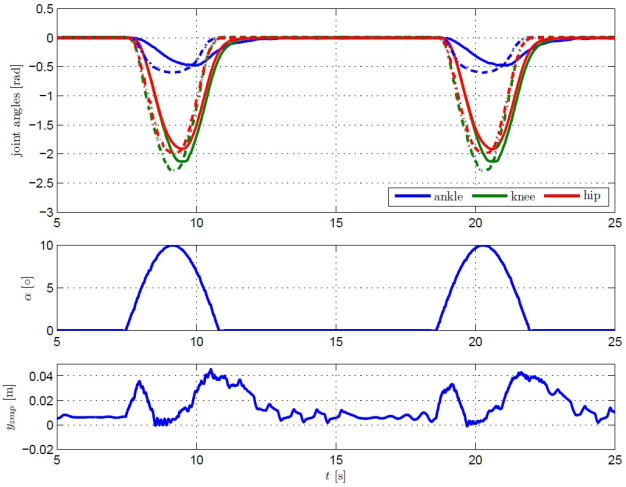
Slika 3 prikazuje gibanje posameznih sklepov ob zagovaljanju stabilnosti (polna črta) ter posneto gibanje demonstratorja (črtkana). Trajektorije so si pričakovano precej podobne.

Slika 4 prikazuje rezultate izvajanja počepa ob spreminjajočem naklonu podlage. Kot lahko vidimo, so so trajektorije podobne, vendar spremenjene toliko, da robot ohrani stabilnost.

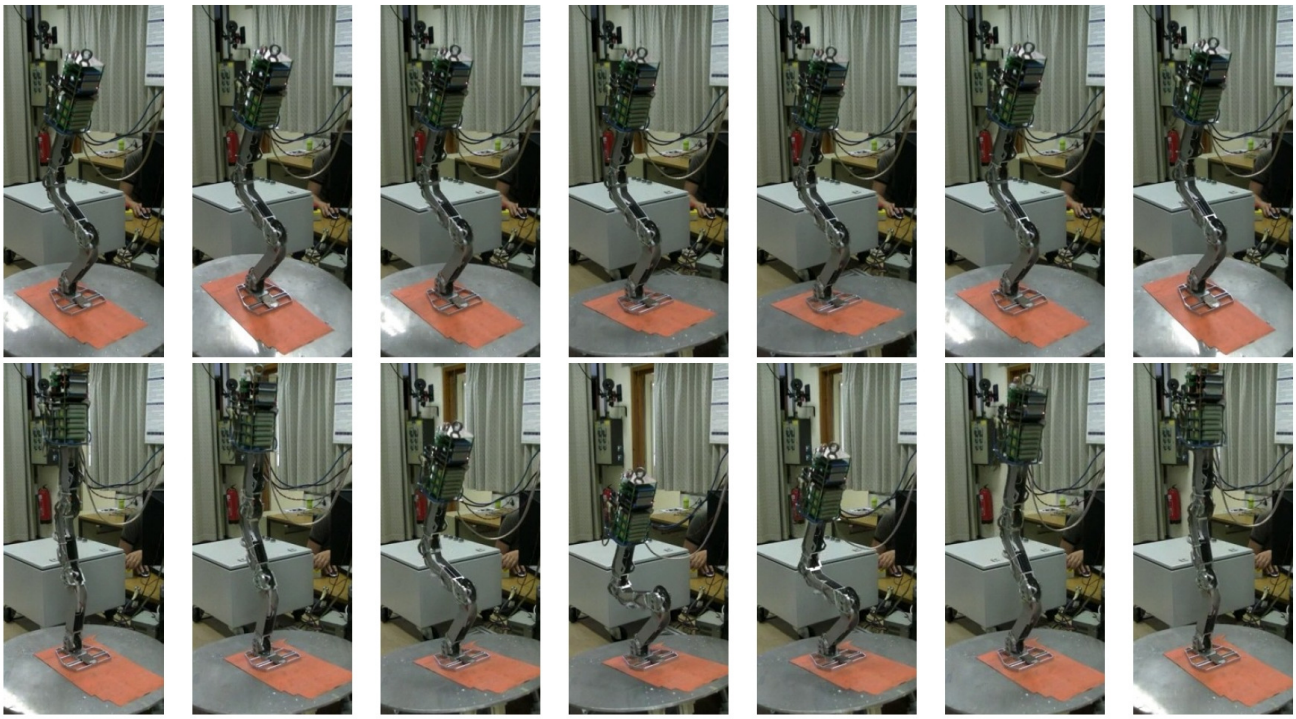
Slika 5 v zgornji vrstici prikazuje sekvenco slik, ko robot ohranja stabilno napol skrčeno lego ob spreminjajočem naklonu podlage. Spodnja vrstica prikazuje uspešno izvajanje počepa.



Slika 3: *Referenčno gibanje sklepov (črtkana) in dejansko, spremenjeno zaradi gibanja podlage, prikazanega v spodnjem grafu.*



Slika 4: *Referenčno (črtkana) in dejansko gibanje sklepov (zgoraj) ob izvajanju počepa ob spreminjanju podlage (na sredi). Gibanje normirane ZMP spodaj.*



Slika 4: *Sekvenca slik ob ohranjanju lege pri spreminjajočem naklonu podlage (zgoraj). Izvajanje počepa spodaj.*

5 ZAKLJUČEK

V prispevku smo prikazali uporabo spremenjenega prioritiziranega vodenja za izvedbo naloge ohranjanja stabilnosti humanoidne robotske noge. Predlagana metoda omogoča neposredno kopiranje merjenega gibanja demonstratorja na robota, ne da bi se to gibanje kar koli spremenilo, dokler ne doseže izbranega praga primarne naloge. Ko enkrat doseže izbrani prag, primarna naloga zvezno prevzame nadzor in dovoljuje vplivanje na gibanje posameznih segmentov samo še kot sekundarna naloga primarne naloge. Z demonstracijo na mehanizmu humanoidne robotske noge smo pokazali možnost uporabe algoritma za posnemanje gibanja in zagotavljanje stabilnosti. Rezultati so pokazali, da metoda deluje. S svojim prioritiziranim nadzorom nad gibanjem sklepov predlagana metoda deluje kot neke vrste podzavestno oz. refleksno gibanje, ki deluje samo takrat, ko je potrebno. V prihodnosti bomo preizkusili metodo tudi za druge naloge, npr. za nalogo izogibanja oviram.

References

[1] A. Ude, C. Atkeson, and R. M., "Programming fullbody movements for humanoid robots by observation," Robotics and Autonomous Systems, vol. 47, no. 2-3, pp. 93–108, 2004.

[2] O. Khatib, "A unified approach for motion and force control of robot manipulators: The operational space formulation," Robotics and Automation, IEEE Journal of, vol. 3, no. 1, pp. 43 –53, february 1987.

[3] L. Sentis, J. Park, and O. Khatib, "Modeling and control of multi-contact centers of pressure and internal forces in humanoid robots," in Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on, oct. 2009, pp. 453 –460.

[4] L. Žlajpah and B. Nemec, "Kinematic control algorithms for on-line obstacle avoidance for redundant manipulators," in Intelligent Robots and Systems, 2002. IEEE/RSJ International Conference on, vol. 2, 2002, pp. 1898 – 1903 vol.2.

[5] J. Babic, J. G. Hale, and E. Oztop., "Human ensorimotor learning for humanoid robot skill synthesis," Adaptive Behavior, 2011.

[6] J. Babic, L. Bokman, D. Omrcen, J. Lenarcic, and F. Park, "A biarticulated robotic leg for jumping movements: theory and experiments," Journal of mechanisms and robotics, vol. 1, no. 1, 2009.

[7] M. Vukobratovic and B. Borovac, "Zero-moment point - thirty five years of its life," I. J. Humanoid Robotics, vol. 1, no. 1, pp. 157–173, 2004.

[8] Q. Huang, K. Yokoi, S. Kajita, K. Kaneko, H. Arai, N. Koyachi, and K. Tanie, "Planning walking patterns for a biped robot," Robotics and Automation, IEEE Transactions on, vol. 17, no. 3, pp. 280 –289, jun 2001.

[9] T. Furuta, T. Tawara, Y. Okumura, M. Shimizu, and K. Tomiyama, "Design and construction of a series of compact humanoid robots and development of biped walk control strategies," Robotics and Autonomous Systems, vol. 37, pp. 81–100(20), 30 November 2001.

TWO-ARM ROBOT ASSEMBLY WITH HUMAN INTERACTION

Nejc Likar, Leon Žlajpah
Department of Automation, Biocybernetics and Robotics
Jozef Stefan Institute
Jamova 39, 1000 Ljubljana, Slovenia
e-mail: {nejc.likar, leon.zlajpah}@ijs.si

ABSTRACT

The paper presents the control of two-arm robot simple assembly task (peg-hole insertion). As the task is performed in a human environment, the robots must be compliant to make the operation human-safe. The human (operator) has higher priorities of controlling the task, so the robots take into account all the external perturbations (object displacements) caused by the operator and change the procedure of the assembly task. Using the impedance control scheme, the reference target model is realized which specify the interaction between robot and environment.

1 INTRODUCTION

In robotics, artificially built mechanical arms are used for realizing human skills and completing tasks better, more accurate and more repeatable than humans. So far, research on the control of robotic arms has been mainly aimed to single-arm robots, while many human tasks are done using two arms [1]. To have two arms on a robot is as important as on a man. The use of two arms on a robot makes it possible to perform various kinds of sophisticated tasks. If the load is heavier than the carrying capacity of a single robot, multiple robots can distribute the load among them.

Also using two robots instead of one, we can increase the workspace. In assembly tasks multiple robots can handle several objects at once increasing the speed of assembly. Assembly of objects in space is made easier where there is no fixed workbench on which to mount the single robot. The research area in a two-arm robot includes kinematics and dynamics of a two-arm mechanical system, force sensory feedback control for the coordination of the two arms, development of a programming system for the description of tasks done by the two-arm robot.

During recent years, many different controllers for cooperating multiple robots have been proposed. The most common is position/force, where the extra degrees of freedom of the multiple arm system are used to control internal force, which is exerted by the manipulators on the object. Impedance controller is the second choice. It is implemented by adding a force loop around a position controller. Sensed force is used to make corrections to the commanded trajectory via admittance/compliance relationship. The modified trajectory is the input to the position controller [1].

This paper presents the use of impedance controlled two robot arms, which are cooperating in a task of assembly. The task is designed in a way that human interaction is possible which is measured with torque sensors in robot joints. In simulation the absolute coordinate systems is used for describing the virtual object between the arms. And the relative coordinate system is used to describe motion between the arms (assembly task).

2 KUKA LIGHTWEIGHT ROBOT

Lightweight robot is especially designed for working in interaction with unknown environments and with humans. Its load/weight ratio and high motion velocity is similar to human arms. Further, they have integrated mechanical and electronic design with sensing and control capabilities enabling compliant interaction and human safe operation (figure 1).

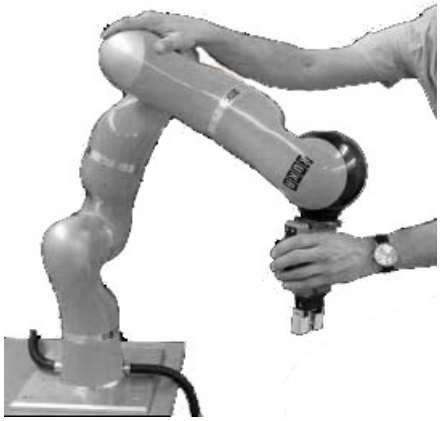


Figure. 1. *Kuka Lightweight robot*

Impedance controlled robot is programmed to act as a virtual mass-spring-damper system with changeable parameters. Instead of prescribing a position or a force, the dynamic relation between them is prescribed, while the actual force and position during interaction depends also on the environment properties.

Compliant behavior in Cartesian space is required in applications where robot's end-effector interacts with the environment. The aim of an impedance controller is to establish a mass-damper-spring relationship (1) between the Cartesian position $\Delta \boldsymbol{x}$ and the Cartesian force $\boldsymbol{f}$, given by:.

$$\mathbf{f} = \mathbf{M}\Delta\ddot{\mathbf{x}} + \mathbf{D}_\mathbf{k}\Delta\dot{\mathbf{x}} + \mathbf{K}_\mathbf{k}\Delta\mathbf{x} \tag{1}$$

Where $\boldsymbol{M}$, $\boldsymbol{D}_k$ and $\boldsymbol{K}_k$ are positive definite matrices representing the virtual inertia, damping and stiffness of the system.

Control parameters can be switched over within one control cycle (1 ms). In this way, it is possible to switch extremely quickly from a stiff, position-controlled mode to a compliant behavior [2].

3 IMPEDANCE CONTROL SCHEME

Controlling two-arm robots involve a problem related to the physical contact and interaction between arms (constrained motion and collision avoidance), between arms and environment in bimanual contact operations. Impedance control provides a common control approach to cope with uncertainties in robotic arms and environment, and maintain interaction forces within some desired level [3].

Compared with compliance control, which is designed to achieve a desired static behavior of the interaction, impedance control is designed to achieve a desired dynamic behavior of the end-effector, characterized by a given mass, damping and stiffness [4].

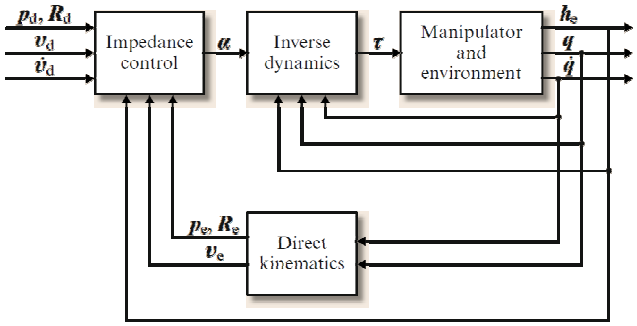


Figure 2. *Diagram of the impedance control.*

The impedance control (figure 2) computes the acceleration input (2) on the basis of the position $\mathbf{a_p}$ (3) and orientation $\mathbf{a_o}$ (4) feedback as well as the force and moment measurements $\mathbf{h_c}=[\boldsymbol{f}\ \boldsymbol{\mu}]$.

$$\boldsymbol{\alpha} = \mathbf{J}^{-1}(\mathbf{q})\left(\mathbf{a} - \dot{\mathbf{J}}(\mathbf{q},\dot{\mathbf{q}})\dot{\mathbf{q}}\right) \tag{2}$$

Where $\mathbf{a} = \begin{bmatrix}\mathbf{a}_p & \mathbf{a}_o\end{bmatrix}$

$$\mathbf{a}_p = \ddot{\mathbf{p}}_d + \mathbf{K}_{Mp}^{-1}\left(\mathbf{K}_{Dp}\Delta\dot{\mathbf{p}}_{de} + \mathbf{K}_{Pp}\Delta\mathbf{p}_{de} - \mathbf{f}\right) \tag{3}$$

$$\mathbf{a}_o = \mathbf{T}(\varphi_e)\left(\ddot{\varphi}_d + \mathbf{K}_{Mo}^{-1}\left(\begin{matrix}\mathbf{K}_{Do}\Delta\dot{\varphi}_{de} + \mathbf{K}_{Po}\Delta\varphi_{de} \\ \cdots - \mathbf{T}^T(\varphi_e)\boldsymbol{\mu}\end{matrix}\right)\right) + \dot{\mathbf{T}}(\varphi_e,\dot{\varphi}_e)\dot{\varphi}_e \tag{4}$$

Where

$$\mathbf{K}_M = \begin{bmatrix}\mathbf{K}_{Mp} & 0 \\ 0 & \mathbf{K}_{Mo}\end{bmatrix} = \begin{bmatrix}m\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}\end{bmatrix} \tag{5}$$

With mass m and $\boldsymbol{M}$ is a 3x3 inertia tensor (6).

$$\mathbf{K}_D = \begin{bmatrix}\mathbf{K}_{Dp} & 0 \\ 0 & \mathbf{K}_{Do}\end{bmatrix} \tag{6}$$

$\boldsymbol{K_D}$ is a damping matrix (7),

$$\mathbf{K}_P = \begin{bmatrix}\mathbf{K}_{Pp} & 0 \\ 0 & \mathbf{K}_{Po}\end{bmatrix} \tag{7}$$

and $\boldsymbol{K_P}$ is a stiffness matrix (6).

The matrix $\mathbf{T}$ is a transformation between angular velocity ω and Euler angle φ.

It is necesary to measure the force $\boldsymbol{f}$ *and* torque $\boldsymbol{\mu}$ of the environment interaction at the end-effector.

The inverse dynamics computes the torques for the joint actuators (8).

$$\boldsymbol{\tau} = \mathbf{J}^T\mathbf{h_c} = \mathbf{J}^T(\boldsymbol{\Lambda}(\mathbf{q})\boldsymbol{\alpha} + \boldsymbol{\Gamma}(\mathbf{q},\dot{\mathbf{q}})\dot{\mathbf{q}} + \mathbf{h_e}) \tag{8}$$

Where

$$\boldsymbol{\Lambda}(\mathbf{q}) = \left(\mathbf{J}\mathbf{H}(\mathbf{q})^{-1}\mathbf{J}^T\right)^{-1}, \qquad \mathbf{H} = \begin{bmatrix}\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{T}\end{bmatrix} \tag{9}$$

and

$$\boldsymbol{\Gamma}(\mathbf{q},\dot{\mathbf{q}}) = \mathbf{J}^{-T}C(\mathbf{q},\dot{\mathbf{q}})\mathbf{J}^{-1} - \boldsymbol{\Lambda}(\mathbf{q})\dot{\mathbf{J}}\mathbf{J}^{-1} \tag{10}$$

including centrifugal and Coriolis effects.

4 TWO-ARM ROBOT COOPERATION

When using multiple-arms, there are several control options: master/slave:, non-master/slave, coordinated motion and object motion [5].

In master/slave control scheme, the force controlled slave arm follows the position controlled master arm. The position of the slave arm is determined by the position of the master arm. Therefore, the impedance of the slave arm has to be very small to follow the motion of the master arm. This is one of the serious problems which this scheme has.

A non-master/slave scheme was proposed by [6]. Their method propose a hybrid position/force control scheme, where workspace position vector is defined as the combination of the absolute position/orientation of the object and the relative position/orientation between two robot arms.

In this paper, the non-master/slave scheme is used. In addition, both robot arms are impedance controlled, what provides a simple, powerful interface for direction motion and environmental interaction. Assembly task in our case is done by two arms together, with an equal contribution of task execution.

The arms work on the same object, therefore the workspace coordinates (figure 3) are defined uniquely on the object to be handled. With this unique coordinate system it is possible to describe all levels of tasks.

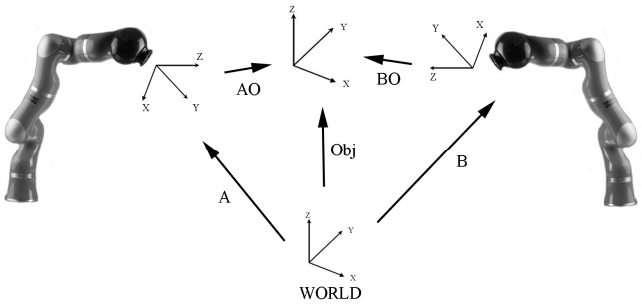


Figure 3: *Workspace coordinates for two-arm robot assembly task*

It is assumed that the center of object **Obj** coordinate frame is the assembled object. The transformation between robot end-effector and object is **AO** and **BO** for the right robot. The assembly task can be done with changing the position or orientation of **AO** or **BO** coordinate frame.

Because the Kuka LWR arms have torque sensors in joints, it is possible to detect the external perturbations in environment and human impacts when changing the pose of assembly task.

At the beginning of the assembly phase, the object (11) is located in the middle of two end effectors of the robots.

$$\mathbf{Obj} = \begin{bmatrix} 1 & 0 & 0 & x_A + \dfrac{x_A - x_B}{2} \\ 0 & 1 & 0 & y_A + \dfrac{y_A - y_B}{2} \\ 0 & 0 & 1 & z_A + \dfrac{z_A - z_B}{2} \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{11}$$

When human move the arms, or the arms interact with the environment, the stiffness of the arms changes and the arms become compliant. Moving the arms also recalculates the position of the object. The position of the other arm is recalculated according to the position of the object. In case of contact with arm **A**, the transformations are (12):

$$\begin{aligned} \mathbf{Obj} &= \mathbf{A} \cdot \mathbf{AO} \\ \mathbf{B} &= \mathbf{A} \cdot \mathbf{AO} \cdot \mathbf{BO}^{-1} \end{aligned} \tag{12}$$

Respectively when in contact with the right arm **B**, the transformations are (13):

$$\begin{aligned} \mathbf{Obj} &= \mathbf{B} \cdot \mathbf{BO} \\ \mathbf{A} &= \mathbf{B} \cdot \mathbf{BO} \cdot \mathbf{AO}^{-1} \end{aligned} \tag{13}$$

The stiffness of the arms changed from high to low, when the torques in joints were above the threshold value, which was in our case:

```
Torque_threshold =[5 5 4 3 2 2 1]
```

And gradually become high again when the torques were below the threshold value. This gradual change of stiffness prevents the fast movements of the robot. High stiffness and low stiffness values of the joints were:

```
Low  = [20 10 10 8 2 1 1]
High  = [1000 1000 1000 1000 1000 1000 1000]
```

Figure 4 shows the change of the joint stiffness, when external torques are above threshold. The external torques are caused by environmental interaction.

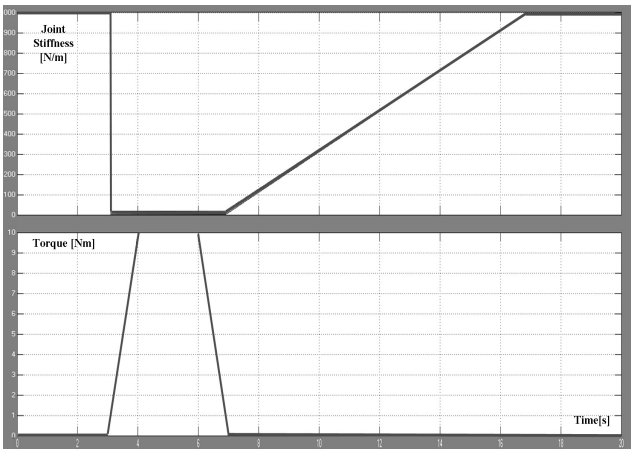


Figure 4: *Change of the joint stiffness*

The assembly task was simple peg-hole insertion. The direction of insertion was the *z*-axis of the robot end-effector (figure 5). To simplify the approach phase, before the start of the assembly, the robots were in initial position, with end-effectors aligned in z-axis.

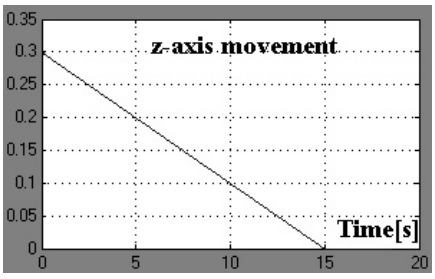


Figure 5: *The movement in z- axis in the assembly task*

In our experiment, the left (**A**) robot was moved by hand during the assembly phase. The assembly task was not influenced by the movement, because the right (**B**) robot adapted its pose with recalculating (12) its position regarding to the position of the virtual object **Obj**, which was moved together with the left arm.

The figure 6 shows the *x*, *y* and *z* coordinates of the left and right robot end-effector, and of the object during the assembly task.
After 3 seconds, the torques of the left arm were above the threshold, because the arm was moved by hand. The arm become soft and changed its position.

The position of the object and of the right arm was recalculated. During the assembly task, the distance between left (right) end-effector and the virtual object in the middle, is reduced. After 15 seconds, the assembly task was finished and the end-effectors of the robots and the virtual object merged together. Again when moving the robot, both robots moved as one.

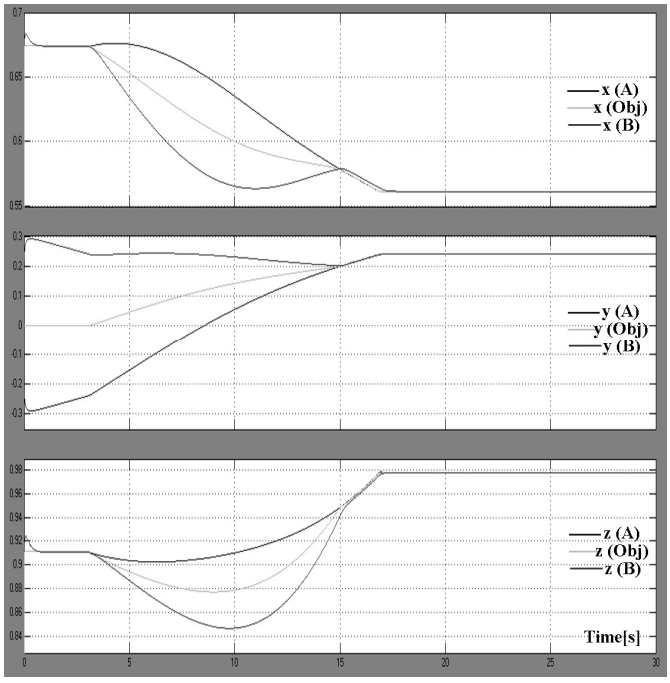


Figure 6: *The xyz coordinates of the **A**, **B** and **Obj** during the assembly task execution*

5 CONCLUSION

In this paper, a non-master/slave control scheme of two robot arms is proposed and tested with simple assembly task. The virtual object is added between two robot arms. And all the transformations are transferred from one arm to another through this object. The benefit of this approach is simple execution of the task done by two arms together, with an equal contribution between the two arms. It is also possible to control two arms independently in parallel.

In addition, the arms were impedance controlled, and equipped with torque sensors, which allowed us to detect environmental interaction, and detect the displacements of the arms or object caused by the human. This property is useful in environments, where robot and human cooperate in same tasks. This makes the work with robots safer.

References

[1] R. G. Bonitz , T.C. Hsia: Internal force-Based Impedance Control for cooperating Manipulators. IEEE Trans. Robotics and Automation, Vol. ***12(1),*** pp.78-89. 1993

[2] R. Bischoff, et al. : The KUKA-DLR Lightweight Robot arm – a new reference platform for robotics research and manufacturing. KUKA Roboter GmbH, 2010.

[3] Surdilovic et al.: Compliance Control with Dual-Arm Humanoid Robots: Design, Planning ans Programming. IEEE International Conference on Humanoid Robots Nashville, TN, USA, Dec. 6-8, 2010.

[4] L. Villani, J. De Schutter: Handbook of Robotics. Force Control. pp. 161-185, Springer 2008

[5] S. A. Schneider, R. H. Cannon : Object Impedance Control for Cooperative Manipulation: Theory and Experimental Results. IEEE Trans. Robotics and Automation, Vol. **8**(3), pp.383-394. 1992.

[6] M. Uchiyama, P. Dauchez: A Symmetric Hybrid Position/Force Control Scheme for The Coordination of Two Robots. IEEE International Conference on Robotics and Automation. Vol. 1, pp.350-356. 1988.

TELEMANIPULACIJA ROBOTSKE PRIJEMALKE BARRETT HAND

Matjaž Ogrinc, Andrej Gams, Aleš Ude
Department of Intelligent Systems
Jozef Stefan Institute
Jamova 39, 1000 Ljubljana, Slovenia
Tel: +386 1 4773419; fax: +386 1 4251038
e-mail: matjaz.ogrinc42@gmail.com

ABSTRACT

We present a telemanipulation method for a three finger robotic gripper called the BarrettHand. The operator's finger movement is captured using the OptoTrak system from NDI, which captures active markers mounted to the hand. The captured markers' positions are converted to four joint angle values, which are then scaled and set as desired joint positions of the robotic hand. Since the BarrettHand's kinematic configuration is not anthropomorphic, point-to-point mapping is the obvious choice. Instead, we present an algorithm where joint-to-joint mapping is applied, allowing an intuitive control of the hand. The operator can successfully perform various grasp types with the robotic mechanism.

1 UVOD

Pri teleoperaciji operater daljinsko upravlja z mehanskim manipulatorjem. Najpogosteje se uporablja enega od sledečih tipov preslikave gibanja: preslikava notranjih koordinat oz. kotov v sklepih (*joint to joint mapping*), preslikava lege (*pose mapping*), ali preslikava pozicije (*point-to-point mapping*) [9].

Preslikava sklepnih koordinat oziroma kotov v sklepih je pogosta pri telepanipulaciji antropomorfne robotske prijemalke [1]. Zaradi podobnosti kinematike človeške in robotske roke lahko kote preslikamo neposredno, brez dodatnih transformacij. V nekaterih primerih lahko podobnost dosežemo tako, da kinematiko človeške roke približamo kinematiki robotske tako, da jo omejimo z uporabo *eksoskeletona* [2].

Mapiranje lege poskuša najti konfiguracijo sklepnih koordinat robotske prijemalke tako, da rezultirajoče lege konic prstov ustrezajo legi prstov demonstratorja.

V zadnjem času se uporablja identifikacija prijemov, ki bazira na uporabi nevronskih mrež [3]. S pomočjo le-teh zaznamo tip prijema, ki se transformira v predefinirane sklepne koordinate robotske prijemalke. [4] uporablja skrite markove modele za identifikacijo tipa prijema, nevronska mreža pa določi ekvivalentne sklepne koordinate prijemalke.

Pri vodenju z demonstracijo se prav tako uporablja neke vrste mapiranje lege. [5] uporablja prepoznavanje prijema, ki temelji na merjenju sklepnih kotov roke demonstratorja. Preslikava notranjih koordinat se izvaja z uporabo knjižnice prijemov.

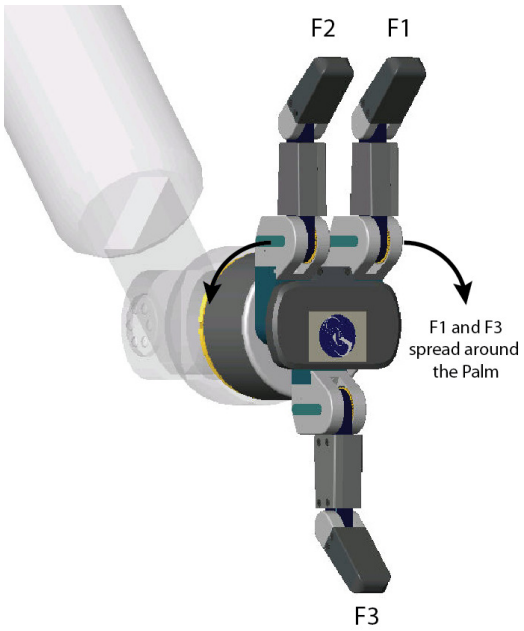
Pri uporabi pristopov, ki temeljijo na prepoznavi prijemov, je pomembna predvsem zanesljivost in stabilnost algoritmov. Znana in pogosta težava je nepredvideno preklapljanje med tipi prijemov, ki je posledica nekonsistentnega zaznavanja vzorcev. Zato običajno lahko razpoznavamo le omejeno število prijemov, kar je resna omejitev te metode.

Najbolj pogost pristop za preslikavo uporabnikovih prijemov na prijemalko je *point-to-point mapping* oz. preslikava pozicij konic prstov [6]-[8]. Za ta pristop je potrebno merjenje pozicij prstov demonstratorja, oziroma kotov prstov. Pri slednjem je potrebno uporabiti direktno kinematiko (DK) človeške roke. Za vodenje prijemalke pa je potrebno poznavanje inverzne kinematike (IK). Tak pristop je predvsem primeren za precezijsko prijemanje.

V tem eksperimentu predstavljamo vodenje triprstne robotske prijemalke BarrettHand. Kljub temu, da prijemalka ni antropomorfna oziroma se njena kinematika zelo razlikuje od človeške, želimo pokazati, da jo je možno voditi z uporabo preslikave notranjih koordinat. Pri tem vpeljemo omejitev kota enega od sklepov, in s tem njeno kinematiko približamo kinematiki človeške roke. Operater prijemalko upravlja s premiki treh prstov.

2 TELEMANIPULACIJA PRIJEMALKE

V eksperimentu je uporabljena triprstna prijemalka BarrettHand proizvajalca Barrett Technology Inc (slika 1). BarrettHand ima tri prste ter štiri aktivne stopnje prostosti. Vsak prst (F1, F2, F3) je sestavljen iz dveh segmentov ter dveh sklopljenih sklepov, katera poganja en motor. Dva od prstov sta sinhrono in simetrično gibljiva okoli dlani (sklep S). Nizko-nivojska hitrostna regulacija motorjev je izvedena v samem tovarniškem krmilniku, pozicijsko vodenje pa se izvaja v okolju MATLAB/ Simulink.



Slika 1: *Robotska prijemalka BarrettHand [11].*

Slika 2 prikazuje shemo vodenja. Operaterjevo gibanje zajemamo z uporabo sistema NDI OptoTrak. Posamezna enota (slika 3, levo) s pomočjo treh infrardečih kamer zajema pozicijo aktivnih markerjev (desno) v prostoru.

Slika 2: *Shema vodenja.*

2.1 Preslikava koordinat

Na roki (slika 4) je nameščenih 6 markerjev. Točke, ki predstavljajo 3D pozicije markerjev, so zapisane s krajevnimi vektorji $\boldsymbol{m}_i$, kjer je i zaporedno število markerja. Sklepne kote določimo iz pozicij markerjev. Inverzna kinematika je določena na sledeč način. Določimo orientacijo roke iz poziciji markerjev 1,3,5. Orientacija je definirana z rotacijsko matriko **R**:

$$\begin{aligned} \boldsymbol{i} &= \boldsymbol{p}_{13} \times \boldsymbol{p}_{15} \\ \boldsymbol{j} &= \boldsymbol{i} \times \boldsymbol{p}_{53} \\ \boldsymbol{k} &= \boldsymbol{i} \times \boldsymbol{j} \\ \mathbf{R} &= \begin{bmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \end{bmatrix} \end{aligned} \tag{1}$$

Matrika **R** predstavlja preslikavo med globalnim koordinatnim sistemom ter sistemom pripetim na dlan.

Slika 3: *Sistem zajemanja gibanja z aktivnimi markerji OptoTrak.*

Vektorji označeni z $\boldsymbol{p}_{ij}$ za i, j = 1,2,3...6, $i \neq j$, so določeni kot razlike krajevnih vektorjev markerjev:

$$\boldsymbol{p}_{ij} = \boldsymbol{m}_j - \boldsymbol{m}_i \tag{2}$$

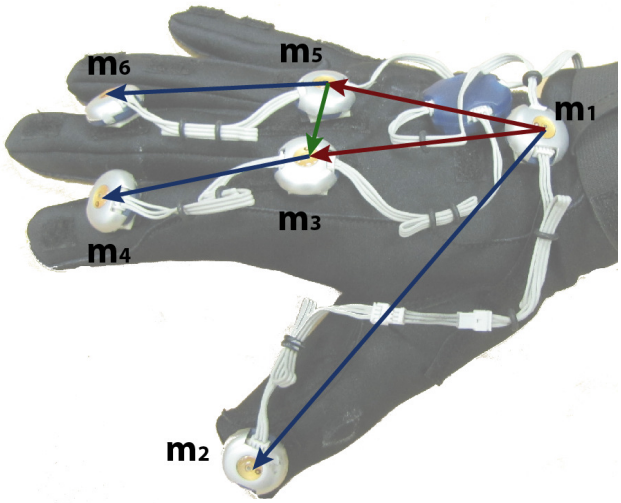
Vektorja kazalca in sredinca $\boldsymbol{p}_{34}$ in $\boldsymbol{p}_{56}$ transformiramo v koordinatni sistem roke:

$$\boldsymbol{p}_{ij}^{\boldsymbol{R}} = \mathbf{R} \cdot \boldsymbol{p}_{ij} \tag{3}$$

Kote v sklepih prstov robota (F1, F2, F3) označimo z q_{F1}, q_{F2}, q_{F3}. Kot sklepa S označimo s q_S. Definirani so kot funkcije vektorjev $\boldsymbol{p}_{ij}$ in posameznih komponent vektorjev $\boldsymbol{p}_{ij}^{\boldsymbol{R}}$, normiranih na dolžino 1, kar nakazuje n v indeksu vektorja. S k_S so označene skalirne konstante.

$$\begin{aligned} q_{\mathbf{F1}} &= k_1 \, arcsin \, p_{34\,n,x}^{R} \\ q_{\mathbf{F2}} &= k_2 \, arcsin \, p_{56\,n,x}^{R} \\ q_{\mathbf{F3}} &= k_3 \, arccos\left(\boldsymbol{p}_{13\mathbf{n}} \cdot \boldsymbol{p}_{12\mathbf{n}}\right) \\ q_{\mathbf{S}} &= k_{\mathbf{S}} \left(arcsin \, p_{34\,n,z}^{R} - arcsin \, p_{56\,n,z}^{R}\right) \end{aligned} \tag{4}$$

Območje kotov q_{Fi} je od 0 do 140 stopinj. Ker se hodi kotov človeških prstov razlikujejo od tistih pri robotski prijemalki, uporabimo skalirne faktorje k, da izkoristimo celotno območje prstov prijemalke. Motor S vrti prsta F1 in F2 okoli dlani v območju 0 do 180 stopinj. Z namenom zmanjšanja razlike med kinematikama človeške in robotske roke, vpeljemo omejitev kota q_S na maksimalno vrednost 60 stopinj. Vsi prsti so tedaj razmaknjeni na 120 stopinj, kar je primerno za prijemanje objektov sferičnih oziroma cilindričnih oblik.



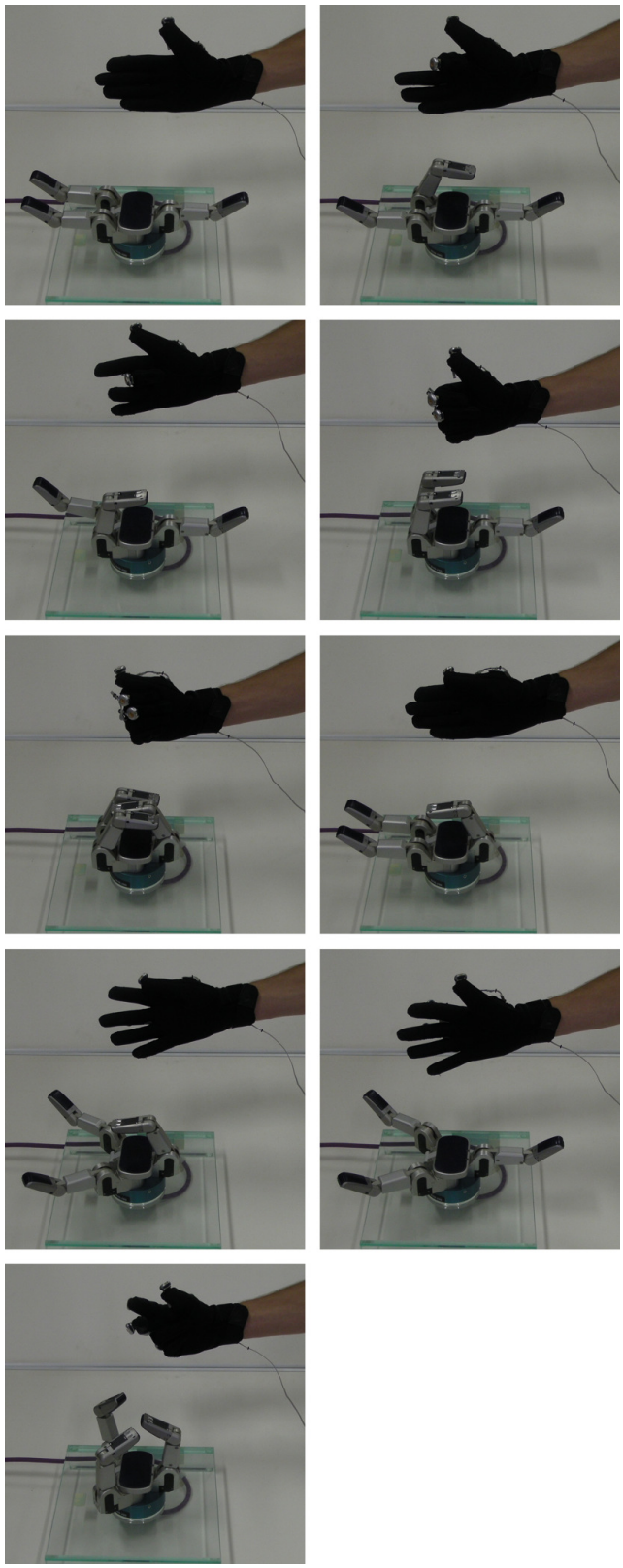
Slika 4: *Pritrditev aktivnih markerjev sistema OptoTrak, z označenimi markerji in vektorji*

3 EKSPERIMENTLNA EVALVACIJA

Algoritem vodenja smo preizkusili na realni napravi, robotski prijemalki BarrettHand. Sekvenca posnetkov vodenja je prikazana na sliki 5. Prikazuje demonstrirane lege človeške roke, ki so replicirane na robotski prijemalki. Intuitivnost vodenja je zadovoljiva, prav tako natančnost, saj je možno izvajati precizne prijeme. Kot pomanjkljivost takega sistema se izkaže odsotnost haptične povratne informacije.

4 ZAKLJUČEK IN NADALNJE DELO

Predstavljen eksperiment prikazuje telemanipulacijo z uporabo preslikave notranjih koordinat. Tako vodenje velja za neprimerno pri uporabi prijemalk ne-antropomorfne oblike. Kot smo že omenili, smo zato v kinematiko vpeljali omejitve. V [9] so avtorji za telemanipulacijo BarrettHand prijemalke vpeljali enako omejitev hoda sklepa S. Avtorji so uporabili vodenje s preslikavo pozicije, lego roke pa se merili z uporabo *dataglove*-a. Ta način zahteva poznavanje direktne kinematike človeške ter inverzne kinematike robotske roke. Ker imajo prsti BarrettHanda manj stopenj prostosti (DOF) od človeške roke (1 oz. 2 DOF), je inverzna kinematika preslikava treh dimenzij v dve oz. eno sklepno koordinato. Ta problem ni rešljiv, zato avtorji predstavljajo rešitev, ki več različnih leg demonstratorja preslikajo v iste notranje koordinate. Iz predstavljene izpeljave in primerjave sekvence slik je možno opaziti podobnost rezultatov, dobljenih z uporabo predlagane metode ter tiste iz [9]. Potrebo po uporabi DK človeške roke ter IK robotske to opažanje postavlja pod vprašaj.



Slika 5: *Lege roke in pripadajoče lege prijemalke*

Seveda pa bi bilo potrebno podrobneje analizirati rezultate metode opisane v omenjenem članku za konkretnejšo primerjavo.

Običajno telemanipulacija ponuja tudi haptično povratno informacijo operaterju. V primeru telemanipulacije robotske prijemalke se to izvede z uporabo eksoskeletona. Kadar se zanašamo le na vizualno povratno informacijo, želimo merjenje zunanjih ali notranjih koordinat roke izvesti čimbolj enostavno, brez dodatnih priprav in nameščanja pripomočkov. *Dataglove* ter metode zajemanja gibanja z markerji tega ne omogočajo.

V [10] avtorji predstavljajo zajemanje gibanja z uporabo naprave Microsoft Kinect. Naprava zajema barvno in globinsko sliko. Avtorji predstavljajo rešitev, ki globinsko sliko primerja z modelom roke. Roka je modelirana kot struktura z 27 stopnjami prostosti. Ker je tak način iskanja lege roke procesno zahteven, se pojavi problem hitrosti vzorčenja. Z uporabo Kinecta in s tem razvoja učinkovitega in zanesljivega algoritma za sledenje rok, bi dodatno izboljšali sistem v smislu enostavnosti uporabe, kar je pomembno za nadaljnje raziskave.

Literatura

[1] K. Kyriakopoulos, J. Van Riper, A. Zink, and H. Stephanou, "Kinematic Analysis and Position/Force Control of the Anthrobot Dextrous Hand," *IEEE Trans. on Systems, Man and Cybernetics, vol. 27, no. 1, pp. 95–104, February 1997.*

[2] A. Wright and M. Stanisic, "Kinematic Mapping between the EXOS Handmaster Exoskeleton and the Utah-MIT Dextrous Hand," *in Proc. of the IEEE International Conference on Systems Engineering, 1990, pp. 809–811.*

[3] T. Wojtara and K. Nonami, "Hand Posture Detection by Neural Network and Grasp Mapping for a Master Slave Hand System," *in Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems, vol. 1, Sept., Oct. 2004, pp. 866–871.*

[4] S. Ekvall and D. Kragic, "Interactive Grasp Learning Based on Human Demonstration," *in Proc. of the IEEE International Conference on Robotics and Automation, vol. 4, April, May 2004, pp. 3519–3524.*

[5] J. Aleotti and S. Caselli, "Grasp Recognition in Virtual Reality for Robot Pregrasp Planning by Demonstration," *in Proc. of IEEE International Conference on Robotics and Automation, May 15-19, 2006, pp. 2801–2806.*

[6] R. Rohling and J. Hollerbach, "Optimized Fingertip Mapping for Teleoperation of Dextrous Robot Hands," *in Proc. of the IEEE International Conference on Robotics and Automation, vol. 3, May 1993, pp. 769–775.*

[7] M. Fischer, P. van der Smagt, and G. Hirzinger, "Learning Techniques in a Dataglove Based Telemanipulation System for the DLR Hand," *in Proc. of the IEEE International Conference on Robotics and Automation, vol. 2, May 1998, pp. 1603–1608.*

[8] J. Hong and X. Tan, "Calibrating a VPL DataGlove for Teleoperating the Utah/MIT Hand," *in Proc. of the IEEE International Conference on Robotics and Automation, vol. 3, May 1989, pp. 1752–1757.*

[9] A. Peer, S. Einenkel and M. Buss, "Multi-fingered Telemanipulation - Mapping of a Human Hand to a Three Finger Gripper," *Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication, Technische Universität München, Munich, Germany, August 1-3, 2008*

[10] I. Oikonomidis, N. Kyriazis, A. A. Argyros, " Efficient Model-based 3D Tracking of Hand Articulations using Kinect," *Proceedings of the 22nd British Machine Vision Conference, BMVC'2011, University of Dundee, UK, Aug. 29-Sep. 1, 2011*

[11] Barrett Technology Inc, "BarrettHand BH8-Series User manual Firmware Version 4.4.x," *2010*

VPLIV DVOROČNEGA MOTORIČNEGA UČENJA NA ENOROČNO IZVAJANJE NALOG

Matic Trlep, Matjaž Mihelj, Marko Munih
Univerza v Ljubljani
Fakulteta za elektrotehniko
Tržaška 25, 1000 Ljubljana
e-mail: matic.trlep@robo.fe.uni-lj.si

ABSTRACT

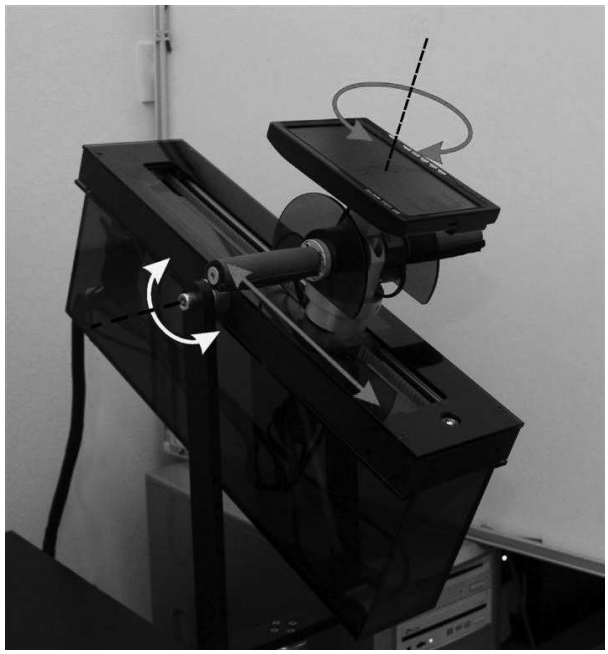
Bimanual training is a well established method of post-stroke rehabilitation. The transfer of learned patterns among limbs and between tasks mastered with both arms towards movements performed only with the impaired arm could make rehabilitation faster and more effective. With our bimanual system 35 healthy subjects made a series of movements using both arms or only their dominant arm. The movements were done under different conditions - bimanual or unimanual, with or without the influence of the force field on the movement, and using both arms equally or using one arm more than the other. By comparing the rotation errors the subjects made during different conditions, we can see that the subjects adapted to the changed dynamics of the movements. Different arm participation ratio in the combined bimanual movement has no significant influence on the learning. Bimanual training had proved to be beneficial also for single limb performance.

1 UVOD

Za opravilo številnih vsakodnevnih aktivnosti je pogosto potrebna koordinirana uporaba obeh rok. Roki lahko delujeta enakovredno ali pa se dopolnjujeta. Prav tako lahko veliko stvari, ki jih opravimo z eno roko, izvedemo bolje in hitreje z uporabo obeh rok. Ravno to hipotezo hitrejšega in učinkovitejšega učenja je potrebno raziskati tudi v rehabilitaciji po poškodbah mišično živčnega sistema.

V preteklosti je bilo obsežno raziskano učenje motoričnih vzorcev posameznega uda pod vplivom zunanjih motilnih sil [1, 2]. Podoben pristop so nato povzeli v številnih raziskavah, ki so se ukvarjale s prenosom naučenih gibov iz enega uda na drugega [3]. Manjše število raziskav se je lotilo tudi prenosa spretnosti naučenih dvoročno na en sam ud [4]. Pokazali so delni prenos iz dvoročnih gibov na enoročne gibe dominantnega ud. Predvsem pa so bile izpostavljene podobnosti prenosa med posameznima udoma in prenosa iz dvoročnega gibanja na enoročno gibanje le z dominantnim udom.

Raziskave potrjujejo nevronsko povezanost med udoma, kar lahko koristimo tudi pri rehabilitaciji. Tako se je kot ena izmed metod rehabilitacije uveljavila tudi dvoročna vadba [5], ki izboljšuje spretnost, moč prijema in funkcionalnost prizadetega uda [6]. Možganska polobla brez okvare predstavlja vzorec za ustrezne odzive obnovljenega živčnega omrežja [7]. V zadnjih letih se v proces rehabilitacije po možganski kapi uvajajo robotske naprave kot pomoč pri terapiji in za ocenjevanje vadbe. Več študij je raziskalo vplive robotske vadbe na povrnitev funkcionalnosti paretične roke [8, 9, 10].



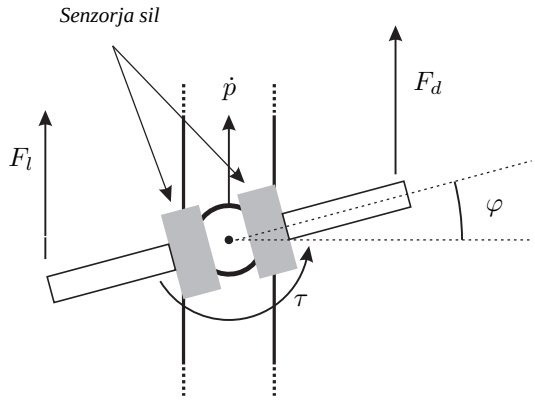
Slika 1: *Mehanizem za dvoročno vadbo. Sivi puščici označujeta aktivni prostostni stopnji, bela pa pasivno za nastavitev naklona mehanizma.*

Cilj raziskave je določiti prisotnost prenosa naučenih motoričnih vzorcev med dvoročno vadbo in enoročnim gibanjem dominantnega uda. Zanima nas tudi, kako je ta prenos odvisen od različnega prispevka posameznega uda k skupnemu gibu.

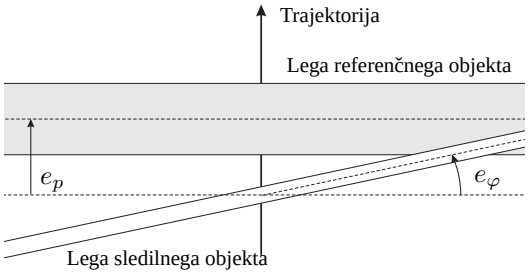
2 METODE

2.1 Strojna oprema

Predhodno delo na dvoročno vodenem haptičnem robotu je pokazalo nekatere slabosti komercialnih robotskih sistemov [11]. Potrebni bi namreč bili bolj preprosti, prilagodljivi in nenazadnje tudi cenejši interaktivni mehanski



Slika 2: *Zgornji pogled prijemala. Slika prikazuje sili leve F_l in desne roke F_d, zasuk priejamla φ in translacijsko hitrost prijemala $\dot{p}$.*



Slika 3: *Navidezna naloga sledenja. Prikazani sta napaka rotacije e_φ in napaka sledenja e_p.*

sistemi za dvoročno vadbo.

V članku je predstavljen preprost robotski sistem (slika 1) z dvema aktivnima prostostnima stopnjama (sivi puščici). Prvo os predstavlja translacija vrha, drugo pa rotacija vrha okrog osi pravokotne na translacijsko os. Dodatno nam pasivna rotacija omogoča nastavitev naklona celotne naprave (bela puščica).

Dvoročno prijemalo (slika 2) pritrjeno na vrhu mehanizma omogoča neodvisne meritve sil in navorov obeh rok, ki delujeta na prijemalo. Za to uporabljamo dva 6-osna senzorja sil (*JR3 50M31*). Na prijemalu je montiran manjši barvni zaslon za vizualizacijo nalog namenjenih urjenju.

2.2 Naloga

Oseba sedi za mehanizmom s pogledom na zaslon, na katerem se mu prikazuje navidezna naloga (slika 3) in z eno ali obema rokama drži prijemalo. Referenčni objekt (širši trak) se giblje po vnaprej določeni trajektoriji (po trapezni hitrostni krivulji enaki za vse gibe), njegova orientacija pa je vedno konstantna (vodoravna) in neodvisna od pozicije in orientacije prijemala. Testiranec mora s premikanjem vrha robota, ki ga predstavlja sledilni objekt (bela črta), slediti gibanju referenčnega objekta.

Za dosego želene aktivacije določenih mišic med gibanjem nudi mehanizem nastavljiv upor, ki nasprotuje uporabnikovim gibom. Poleg dvoročnega načina izvedbe nalog sistem omogoča tudi enoročno vadbo. Da je potrebna aktivacija mišic posamezne roke v obeh načinih primerljiva, je upor med enoročnim načinom manjši za polovico.

Ločili smo tudi dva primera glede na razmerje sil, ki jih k skupnemu gibanju prispevata obe roki. Tako v prvem primeru delujeta obe roki z enakim razmerjem ($50\% : 50\%$). V drugem primeru mora oseba z nedominantno roko izvajati večje sile kot z dominantno (razmerje $75\% : 25\%$).

2.3 Vodenje

Vodenje je bilo v celoti zasnovano kot model v okolju Matlab Simulink, ki ga poganjamo na računalniku z operacijskim sistemom Matlab xPC Target, saj nam le-ta zagotavlja delovanje sistema v realnem času.

Za način vodenja smo izbrali admitančni pristop – mehanizem vodimo tako, da na sistem delujemo z zunanjo silo. Silo vodenja določimo iz sil obeh rok in želenega razmerja sil, kot kaže spodnja enačba.

$$F_c = K_l F_l + K_d F_d \tag{1}$$

F_l in F_d sta skalarni vrednosti sile leve oziroma desne roke v smeri translacijske osi (slika 2). Faktorja K_l in K_d sta vrednosti s katerima skaliramo sile, da zagotovimo želeno razmerje sil med rokama. V kolikor je želeno razmerje $50\% : 50\%$, potem je $K_l = K_d = 1$, če pa naj leva roka prispeva večji del k skupnemu gibu, je $K_l = 0,5$ in $K_d = 1,5$. S skaliranjem zagotovimo, da mora roka za isti učinek izvesti večje/manjše sile, kot običajno.

V nekaterih fazah delovanja je na gibanje vplival tudi navidezni motilni navor. Vpliv motilnega navora se pozna na končni rotaciji, kot navor, ki poizkuša prijemalo zavrteti v smeri dominantne roke. Da je motilni navor τ_m odvisen od uporabnikovega delovanja in hkrati tudi ponovljiv, je njegova vrednost proporcionalna translacijski hitrosti gibanja prijemala $\dot{p}$

$$\tau_m = \begin{cases} 10\dot{p} & \text{za } \dot{p} > 0, \\ 0 & \text{za } \dot{p} \leq 0. \end{cases} \tag{2}$$

Med gibi, kjer motilni navor ni aktiven, je njegova vrednost $\tau_m = 0$.

Podobno kot silo vodenja izračunamo tudi navor vodenja iz navorov obeh rok (τ_l za levo in τ_d za desno roko) okrog navpične osi in navideznega motilnega navora

$$\tau_c = K_l \tau_l - K_d \tau_d + \tau_m. \tag{3}$$

Kot je bilo že omenjeno, je za vodenje obeh prostostnih stopenj mehanizma bil uporabljen admitančni pristop. Uporabili smo preprost admitančni model drugega reda

$$F_c = m\ddot{p}_r + b_t\dot{p}_r, \tag{4}$$

$$\tau_c = I\ddot{\varphi}_r + b_r\dot{\varphi}_r. \tag{5}$$

kjer predstavlja m navidezno maso, I vztrajnost ter b_p in b_φ viskozno dušenje sistema (med enoročnim načinom imata dušenji polovično vrednosti kot med dvoročnim). Referenčni poziciji p_r, φ_r in hitrosti $\dot{p}_r$, $\dot{\varphi}_r$ dobljeni iz enačb (4) in (5) se dalje uporabijo v proporcialno-diferencirnemu (PD) regulatorju končnih stopenj.

2.4 Merilni protokol

Cilj raziskave je določiti, v kolikšni meri se človeški motorični sistem med dvoročno vadbo lahko prilagodi na delovanje zunanjega polja motilnih sil. Posebno nas zanima prenos naučenih vzorcev gibanja med dvoročno vadbo na izvedbo gibov le z dominantno roko. Preverili bi radi tudi, v kolikšni meri je ta prenos odvisen od deleža, ki ga mora določena roka izvesti za uspešno izvedbo skupnega giba.

V raziskavi je sodelovalo 35 zdravih oseb starih 20 do 32 let brez motoričnih težav. Razdeljeni so bili v 5 skupin glede vrsto naloge, ki so jo morali opravljati – razlike v razmerju rok med dvoročnim učenjem in način nalog (eno- ali dvoročno). Razlike v protokolu med posameznimi skupinami so povzete v tabeli 1.

Tabela 1: Skupine glede na vrsto naloge. Razmerje rok med dvoročno vadbo: A – $50\% : 50\%$; B – $75\% : 25\%$.

Skupina	Uvod	Učenje	Test
A	dvoročno A	dvoročno A	dvoročno A
B	dvoročno B	dvoročno B	dvoročno B
C	enoročno	dvoročno A	enoročno
D	enoročno	dvoročno B	enoročno
E	enoročno	enoročno	enoročno

Merjenci so dobili navodila, da naj kolikor je le mogoče natančno sledijo gibanju reference, ob tem pa naj bodo gibi čim bolj gladki. Gibanje reference je namreč služilo bolj kot indikator hitrosti gibanja. Na ta način zagotovimo večjo ponovljivost gibov med celotno serijo. Osebe so izvedle 255 gibov v eni seriji, ki je trajala približno 30 minut. Celotno serijo lahko glede na tip izvedbe razdelimo na 3 dele:

1 **Uvod** – 85 gibov brez motilnih sil. Motilne sile se občasno vključijo in občasno spremenimo razmerje rok (približno vsako šesto ponovitev).

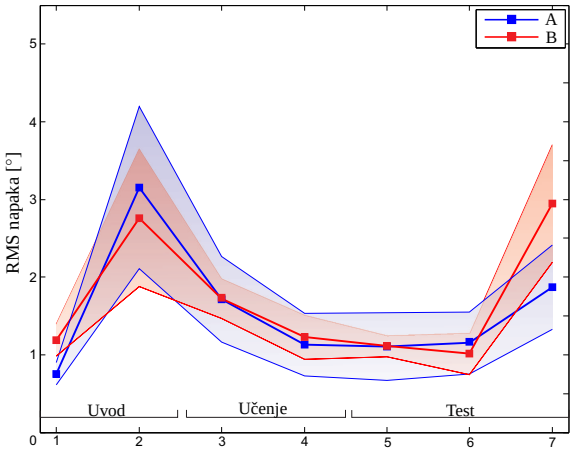
2 **Učenje** – 85 gibov s trajnim vplivom motilnih sil z izbranim razmerjem rok.

3 **Test** – 85 gibov z motilnimi silami. Motilne sile se občasno izključijo in občasno se spremeni razmerje rok (približno vsako šesto ponovitev).

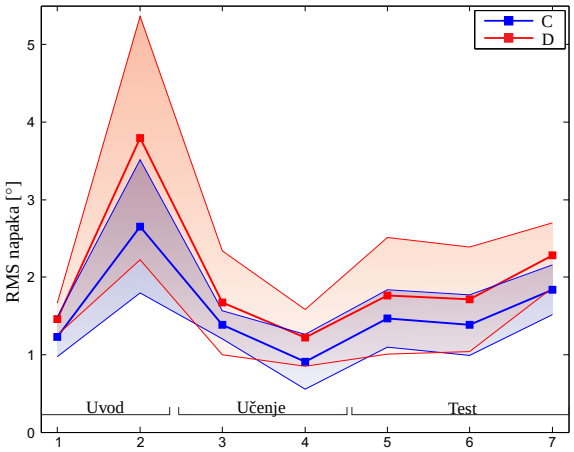
Med *uvodnim* delom se oseba seznami s sistemom in se vzpostavi stabilna začetna nemotena izvedba nalog. Skupini A in B izvajata uvodni del dvoročno, skupini C in D pa le z dominantno roko. Med fazo *učenja* na uporabnika neprestano delujejo motilne sile. Med tem časom se uporabnik prilagodi delovanju zunanjih sil med gibanjem. Skupine A do D izvajajo to fazo z obema rokama. Med *testno* fazo motilne sile še vedno delujejo, vendar jih občasno izklopimo, da lažje ocenimo nivo učenja. Ta del skupini A in B opravljata dvoročno, medtem ko skupini C in D gibe ponovno izvajata le z eno roko. Kontrolna skupina E vse gibe izvaja enoročno z dominantno roko.

3 REZULTATI

Razlika med načinoma vadbe A in B je le v razmerju rok, ki med dvoročno vadbo vodita prijemalo. RMS napaka



Slika 4: *Napake rotacije za skupini A in B med vadbo: uvodna seznanitev s sistemom (1); nenadna motnja med uvodom (2); začetek učenja (3); konec učenja (4); začetek testa (5), konec testa (6); nenadna odstranitev motnje med testom (7).*

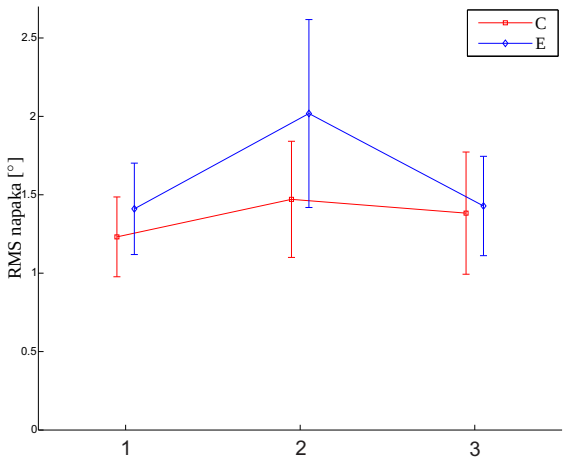


Slika 5: Napake rotacije za skupini C in D med vadbo: uvodna seznanitev s sistemom (1); nenadna motnja med uvodom (2); začetek učenja (3); konec učenja (4); začetek testa (5), konec testa (6); nenadna odstranitev motnje med testom (7).

zasuka tekom celotne vadbe je prikazana na sliki 4. Celotna vadba je sestavljena iz posameznih faz predstavljenih na sliki 4: uvodna seznanitev s sistemom (1); nenadna motnja med uvodom (2); začetek učenja (3); konec učenja (4); začetek testa (5), konec testa (6); nenadna odstranitev motnje med testom (7). Debelejši črt predstavljata srednje vrednosti RMS rotacijskih napak skupin A (modra) in B (rdeča), tanjši pa nakazujeta raztros vrednosti (standardna deviacija).

Podobno kot v prejšnjem primeru vidimo na sliki 5 srednje vrednosti RMS napak za skupini C (močnejša modra črta) in D (močnejša rdeča črta) s pripadajočimi vrednostmi standardne deviacije (tanjše črte).

Vpliv dvoročne vadbe na enoročno izvedbo gibov je predstavljen na sliki 6. Začetni napaki po uvajanju s sistemom (1) sta podobni. Pri skupini C, ki je učenje izvajala dvoročno, začetna napaka ob prehodu na enoročno vadbo sicer naraste, vendar precej manj kot naraste napaka ob nenadni izvedbi motilnega navora pri skupini E (2). Ob koncu vadbe sta napaki ponovno povsem primerljivi (3).



Slika 6: *Vpliv dvoročnega treninga na enoročno sposobnost. Napake so predstavljene med: uvodom (1), začetku enoročne vadbe (2) – za C je to začetek testa, za E pa začetek učenja in konec enoročnega testa (3).*

4 RAZPRAVA

Obe sliki 4 in 5 prikazujeta napake med različnimi fazami vadbe. Vidimo, da je napaka največja med uvodom, ko se nenadoma pojavi motilni navor (2). Ob začetku dvoročnega učenja (prvih 10 ponovitev faze učenja – 3) je napaka zaradi dvoročne izvedbe manjša, ki se pa zaradi uspešnega motoričnega učenja še zmanjša do konca faze učenja (zadnjih 10 ponovitev – 4). Ta vrednost se ustali in ohrani tudi med testiranjem za skupini A in B (slika 4; fazi (5) in (6)). Pri skupinah C in D pa se napaka ponovno poveča (5), vendar se do konca testne faze (6) bistveno ne spremeni. Naučene motorične vzorce nam potrjujejo tudi gibi med testno fazo, kjer nenadoma odstranimo motilne sile (7), saj je napaka med temi gibi bistveno večja kot med običajnim testom (6). Napaka je podobna tisti, ki nastane med uvodom in nenadnim pojavom motilne sile (2).

Vpliv dvoročnega treninga na enoročno gibanje nam prikazuje slika 6. Na sliki vidimo primerljivi napaki med fazo uvoda (1). Napaki pa nista več enaki, ko obe skupini začneta izvajati gibe le z dominantno roko pod vplivom zunanjih sil (2). Za skupino E je to takoj na začetku učenja. V skupini C to ustreza na začetku faze testa, saj je učenje dvoročno. Opazimo manjšo napako v skupini C, torej pozitiven vpliv dvoročne vadbe. Končni vrednosti napak na koncu testnih faz sta zopet podobni.

5 ZAKLJUČEK

Osebe so se sposobne prilagoditi spremenjenim dinamičnim razmeram med dvoročno vadbo. Nakazan je bil prenos naučenih vzorcev pridobljenih med dvoročno vadbo na enoročno izvedbo ekvivalentne naloge, kar potrjujejo smotrnost dvoročne vadbe kot metode rehabilitacije oseb z mišično-živčnimi poškodbami. Signifikantnih razlik med učenjem pri različnih stopnjah prispevkov posameznih rok ni. V različnih obdobjih rehabilitacije lahko tako manj prizadeta roka bolj ali manj asistira prizadeti, kar naj bi vse imelo pozitivne vplive na okrevanje.

6 ZAHVALA

Raziskavo je finančno omogočila Javna agencija za raziskovalno dejavnost Republike Slovenije. Avtorji bi se še posebej radi zahvalili vsem prostovoljcem, ki so sodelovali v raziskavi.

Literatura

[1] R. Shadmehr, F.A. Mussa-Ivaldi, “Adaptive representation of dynamics during learning of a motor task”, *The Journal of Neuroscience*, vol. 14, no. 5, pp. 3208–3224, 1994.

[2] J.L. Patton, F.A. Mussa-Ivaldi, “Robot-assisted adaptive training: custom force fields for teaching movement patterns”, *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 4, pp. 636–646, 2004.

[3] S.E. Criscimagna-Hemminger, O. Donchin, M.S. Gazzaniga, R. Shadmehr, “Learned dynamics of reaching movements generalize from dominant to nondominant arm”, *Journal of neurophysiology*, vol. 89, no. 1, pp. 168–176, 2003.

[4] J.K. Burgess, R. Bareither, J.L. Patton, “Single Limb Performance Following Contralateral Bimanual Limb Training”, *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 15, no. 4, pp. 347–355, 2007.

[5] M.E. Stoykov, D.M. Corcos, “A review of bilateral training for upper extremity hemiparesis”, *Occupational therapy international*, vol. 16, no. 3–4, pp. 190–203, 2007.

[6] A. Shumway-Cook, M.H. Woollacott, “Motor control: translating research into clinical practice”, *Lippincott Williams & Wilkins*, 2006.

[7] J.J. Summers, F.A. Kagerer, M.I. Garry, C.Y. Hiraga, A. Loftus, J.H. Cauraugh, “Bilateral and unilateral movement training on upper limb function in chronic stroke patients: a TMS study”, *Journal of the Neurological Sciences*, vol. 252, no. 1, pp. 76–82, 2007.

[8] G. Prange, M. Jannink, C. Groothuis-Oudshoorn, H. Hermens, M. IJzerman, “Systematic review of the effect of robot-aided therapy on recovery of the hemiparetic arm after stroke”, *Journal of Rehabilitation Research and Development*, vol. 43, no. 2, pp. 171–184, 2006.

[9] P.S. Lum, C.G. Burgar, C.G., H.F.M. Van der Loos, P. Shor, M. Majmundar, R. Yap, “MIME robotic device for upper-limb neurorehabilitation in subacute stroke subjects: A follow-up study”, *Journal of Rehabilitation Research and Development*, vol. 43, no. 5, pp. 631–642, 2006.

[10] M.J. Johnson, H.F.M. Van der Loos, C.G. Burgar, P. Shor, L.J. Leifer, “Experimental results using force-feedback cueing in robot-assisted stroke therapy”, *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 13, no. 3, pp. 335–348, 2005.

[11] M. Trlep, M. Mihelj, U. Puh, M. Munih, “Rehabilitation Robot with Patient-Cooperative Control for Bimanual Training of Hemiparetic Subjects ”, *Advanced Robotics*, vol. 25, no. 15, pp. 1949–1968, 2011.

RAZVOJ PRENOSLJIVEGA INERCIJSKEGA SISTEMA ZA MERJENJE ČLOVEŠKE KINEMATIKE

Tadej Beravs, Peter Reberšek, Domen Novak, Janez Podobnik, Marko Munih
Laboratorij za robotiko
Fakulteta za elektrotehniko
Tržaška cesta 25, 1000 Ljubljana, Slovenia
Tel: +386 1 4768465; fax: +386 1 4768239
e-mail: tadej.beravs@robo.fe.uni-lj.si

POVZETEK

V prispevku je predstavljen sistem inercijskih merilnih enot, ki jih odlikujejo majhna velikost, brezžična komunikacija in odprta arhitektura. Sistem je lahko uporabljen za splošno analizo človeškega gibanja ali pa integriran v eksoskelete za spodnje okončine. Vsaka merilna enota sestoji iz pospeškometra, žiroskopa in magnetometra. Iz izhodov večih senzorjev so izračunani koti v sklepih človeškega telesa. Po kalibraciji je bil sistem preizkušen na lesenem modelu noge in na dejanski človeški nogi. Kot referenca je bil uporabljen optični merilni sistem Optotrak. Razlike med inercijskim merilnim sistemom in Optotrakom so bile manj kot 2 stopinji za leseno nogo in manj kot 5 stopinj za človeško nogo. Sistem tako predstavlja učinkovito metodo sledenja in analize človeškega gibanja.

ABSTRACT

This paper presents a system of inertial measurement units characterized by a small size, wireless transmission, and open architecture. They can be used for general human movement analysis or integrated into lower limb exoskeletons. Each unit comprises an accelerometer, gyroscope and magnetometer. Human joint angles are calculated from multiple units. After calibration, the system was tested with a wooden leg mockup and an actual human. The Optotrak optical measurement system was used as a reference. Differences between the inertial measurement system and the Optotrak were less than 2 degrees for the wooden leg and less than 5 degrees for the human leg, suggesting that the system represents a promising possibility for wearable movement tracking and analysis.

1 UVOD

Eksoskeleti za spodnje okončine so običajno namenjeni bodisi ojačanju sposobnosti zdravih oseb ali pa nudenju pomoči fizično oslabljenim osebam. Eksoskeleti, namenjeni klinični terapiji, so trenutno neprenosljive naprave, pritrjene na tekoči trak (npr. Lokomat (Hocoma A. G.) in LOPES [1]). Obstaja pa tudi precejšnje zanimanje za razvoj prenosljivega robota, pritrjenega na človeško telo, ki bi uporabnika podpiral pri hoji v vsakdanjem življenju. Primer takega robota je eksoskelet Hybrid Assistive Leg (HAL) [2], namenjen podpori in rehabilitaciji fizično oslabljenih oseb. Še en primer je eLEGS (Berkeley Bionics), ki je namenjen predvsem paraplegikom.

S senzoričnega vidika lahko eksoskelet obravnavamo kot inteligentnega robota, ki za uspešno interakcijo s človeškim telesom in okoljem potrebuje več različnih senzorjev. Tak robot ima več zahtevnih nalog. Najprej mora združiti informacije večih senzorjev in tako pridobiti celostno sliko okolja. Poleg tega se mora znati odzivati na dogodke, kot so nevarnost padcev ali pa želja človeka po spremembi hitrosti hoje. Za uspešno delovanje torej eksoskelet potrebuje kompleksno mrežo senzorjev. Med njimi so tudi senzorji, ki merijo kinematiko človeškega telesa. Eksoskelet za noge, razvit na univerzi MIT [3], naprimer meri kote v kolku in kolenu s pomočjo rotacijskih potenciometrov. Podobno HAL [2] meri kote v sklepih s pomočjo pospeškometrov, žiroskopov in potenciometrov. Ker pri hoji sodelujeta tako spodnji kot zgornji del telesa [4], je zaželjeno, da lahko eksoskelet z dodatnimi senzorji, ki niso vgrajeni v sam eksoskelet, meri tudi gibanje zgornjega dela telesa.

V naših raziskavah potrebujemo prenosljiv sistem senzorjev za zgornji in spodnji del telesa, ki ga lahko integriramo z eksoskeletom ali pa samostojno uporabimo za analizo človeškega gibanja. Optimalno rešitev predstavljajo takoimenovane inercijske merilne enote (IME), ki združujejo pospeškometer, žiroskop in magnetometer. IME so že bile uporabljene v komercialno dostopnih sistemih, kot je Xsens MVN (Xsens Technologies B. V.). Ker pa nobena tržna rešitev še ne izpolnjuje naših potreb, smo se odločili za razvoj lastnega inercijskega merilnega sistema, ki bi imel prednosti pred obstoječimi sistemi.

Najprej želimo z uporabo triosnih integriranih elektronskih komponent zagotoviti majhno velikost in maso senzorja. Razviti senzor ima širino le 30 mm in višino nekaj milimetrov (za primerjavo ima Xsens MTX dimenzije 58x50x23 mm). Nadalje morajo biti senzorji s centralno enoto povezani brezžično, da ne motijo merjenca med hojo. Za vključitev v kompleksnejše sisteme mora imeti merilni sistem modularno strukturo, odprto arhitekturo in zmožnost delovanja v realnem času z visoko vzorčno frekvenco.

2 MATERIALI IN METODE

2.1 Strojna oprema

2.1.1 Inercijske merilne enote

Posamezna inercijska merilna enota (IME) vključuje tri senzorje: triosni žiroskop Invensense, triosni akcelerometer STmicroelectronics in triosni magnetometer Honeywell. Maksimalne frekvence vzorčenja so 1000 Hz za žiroskop in pospeškometer ter 160 Hz za magnetometer. Senzorji so priključeni na I^2C-vodilo.
Vsaka IME ima vključen tudi modul Atmel ZigbBit za brezžično komunikacijo. Napajalna napetost za vse komponente je 3 V. Napajanje zagotavlja baterija Wellpower s kapaciteto 250 mAh, ki enoti omogoča neprekinjeno delovanje do 8 ur. IME ima brez baterije velikost 30x20x5 mm in je prikazana na sliki 1.

Slika 1: *Inercijska merilna enota brez baterije. Njena velikost je 30x20x5 mm.*

2.1.2 Prenos podatkov in osrednja enota

IME so brezžično povezane z osrednjo enoto. Na strani IME prenos podatkov zagotavlja brezžični oddajnik ZigBit. Na sprejemni strani uporabljamo močan sprejemnik Atmel, saj velikost in poraba energije tam ne predstavljata omejitve. Ta sprejemnik ima doseg nad 15 metrov. Preko SPI je povezan z osrednjo enoto (slika 2), ki lahko naenkrat sprejema podatke z 8 IME pri frekvenci 300 Hz.


Slika 2: Osrednja enota z ločeno anteno za vsako IME. Njena velikost je 200x137x55 mm.

2.1.3 Kalibracija

Pri idealni IME so vsi senzorji ortogonalni, nimajo odmika od ničelne vrednosti in imajo točno tako občutljivost, kot jo je podal proizvajalec. Ker to v praksi ne drži, smo pospeškometre in magnetometre kalibrirali po postopku, opisanem v [5]. Za odmike žiroskopov od ničelne vrednosti poskrbi Kalmanov filter.
IME so bile pritrjene na leseno kocko s stranico dolžine 15 cm. Obenem je bilo na kocko nameščenih tudi 8 infrardečih markerjev sistema Optotrak, s katerimi smo lahko optično spremljali orientacijo kocke. Kocka je bila vrtena okoli navpične osi v šestnajstih korakih po 22,5 stopinj. Nato je bila postavljena v ravnino, pravokotno prvi ravnini vrtenja in znova vrtena okoli navpične osi v šestnajstih korakih. Končno je bila postavljena v ravnino, pravokotno prejšnjima dvema ravninama vrtenja in znova vrtena okoli navpične osi. V vsaki orientaciji je bilo posnetih 5 s meritve. Vse meritve smo vnesli v optimizacijski algoritem [5], ki je določil kalibracijske parametre za vsak senzor.

2.2 Človeška kinematika

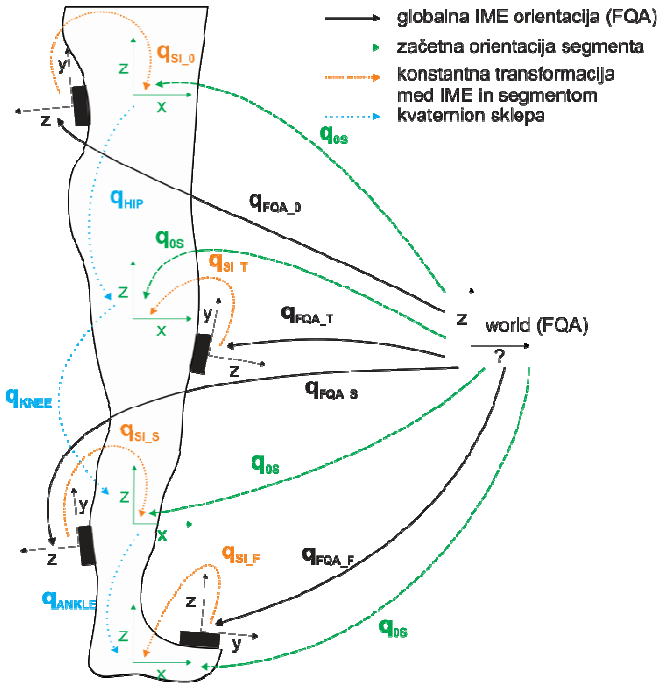
Surove meritve, ki jih dobimo z IME, moramo pretvoriti v obliko, bolj primerno za analizo človeškega gibanja: kote v različnih sklepih človeškega telesa. Tako moramo najprej izračunati orientacijo vsake IME, nato pa iz teh orientacij pridobiti orientacije segmentov telesa in kote v sklepih. V vseh korakih računanja uporabljamo kvaternione. Slika 3 tako prikazuje vse uporabljene kvaternione.

2.2.1 Orientacije IME

Začetno orientacijo vsake IME lahko pridobimo z algoritmom Factored Quaternion Algorithm (FQA) [6], ki je natančen le v mirovanju. Ob začetku meritve tako merjenec stoji pri miru, algoritem pa za vsako IME pridobi kvaternion $\boldsymbol{q}_{FQA}$, ki predstavlja absolutno orientacijo enote v koordinatnem sistemu, kjer osi določata gravitacija in zemeljsko magnetno polje.
Ko je začetna orientacija znana, lahko spremembam orientacije sledimo iz surovih meritev pospeškometra, žiroskopa in magnetometra s pomočjo t.i. "unscented" Kalmanovim filtrom (UKF), različico linearnega Kalmanovega filtra [7]. V primeru nelinearnih sistemov UKF običajno dosega boljše rezultate kot razširjeni Kalmanov filter (klasična razširitev Kalmanovega filtra za nelinearne sisteme).

2.2.2 Orientacije segmentov in koti v sklepih

Ko poznamo orientacijo vsake IME, moramo pridobiti začetno orientacijo vseh segmentov telesa $(\boldsymbol{q}_{0S})$. S tem lahko opišemo orientacijo segmentov neodvisno od načina postavitve IME na segmente.
Ker merjenec ob začetku meritve stoji zravnano, stopala pa so poravnana, takrat z-os koordinatnih sistemov vseh segmentov kaže navzgor (torej $z_{0S} = [0,0,1]$).



Slika 3: *Kvaternioni, ki opisujejo začetne orientacije IME in segmentov telesa.*

Poleg z-osi potrebujemo še en vektor, da lahko definiramo koordinatni sistem. Ta vektor mora imeti komponento, ki kaže bodisi lateralno bodisi naprej/nazaj. V našem primeru smo se poslužili IME na spodnjem trupu. Pri tej IME naj bi bila ena os usmerjena nazaj brez lateralne komponente. Os, ki kaže nazaj, lahko pridobimo iz FQA in uporabimo kot drugi os začetnega koordinatnega sistema segmentov. Tretja os je nato pravokotna prvima dvema.
Kvaternion q_{0S}, ki ga definirajo te tri osi, opisuje začetno orientacijo vseh segmentov v globalnem koordinatnem sistemu. Sedaj lahko izračunamo transformacijo od koordinatnega sistema IME do pripadajočega segmenta:

$$q_{SI} = q_{0S} \cdot q^{*}_{FQA} \quad (1)$$

Če orientacijo IME v globalnem koordinatnem sistemu (q_{IW}) pomnožimo s q_{SI}, dobimo orientacijo segmenta v globalnih koordinatah. Za sklep med poljubnima segmentoma 1 in 2 je kvaternion sklepa nato definiran kot

$$q_j = q_{seg2} \cdot q^{*}_{seg1} \quad (2)$$

Ta kvaternion lahko preprosto pretvorimo v kote v sagitalnih in koronalnih ravninah sklepa.

2.3 Validacija

Inercijski merilni sistem smo validirali s primerjavo kotov v sklepih, ki jih izmerijo IME, s koti v sklepih, ki jih izmeri brezkontaktni optični merilni sistem Optotrak Certus (Northern Digital Inc.). Vzorčna frekvenca je bila 100 Hz.

IME, Kalmanov filter in računanje kinematike telesa smo validirali v dveh korakih. Meritve smo najprej izvedli na preprostem lesenem modelu človeške noge, nato pa še na človeku med hojo. V obeh primerih smo razlike med IME in Optotrakom ovrednotili z računanjem absolutne razlike med koti v sklepih, ki jih izmerimo z IME, in koti, ki jih izmerimo z Optotrakom. Kote smo za kolk in gleženj računali v sagitalnih in koronalnih ravninah, za koleno pa le v sagitalni ravnini, saj lahko koleno lahko modeliramo z eno samo prostostno stopnjo.

2.3.1 Lesen model noge

Lesen model človeške noge je sestavljen iz medsebojno povezanega lesenega stopala, meča, stegna in trupa. Na vsakega od segmentov smo namestili po eno IME. Markerje sistema Optotrak smo namestili na lateralno stran noge.
Izvedli smo tri sklope meritev. V prvem sklopu smo kolk in koleno dvajsetkrat premaknili v sagitalni ravnini. Rotacija je bila približno 60 stopinj za vsak gib. V drugem sklopu smo nogo dvignili v vodoraven položaj, nato pa izvedli 50 lateralnih gibov po približno 7 stopinj rotacije. V tretjem sklepu smo kolk dvajsetkrat premaknili v sagitalni ravnini. Rotacija je bila približno 20 stopinj za vsak gib.

2.3.2 Človeška hoja

Drugi korak validacije smo izvedli na enem prostovoljcu. IME smo namestili na štiri segmente telesa: trup, stegno, meča in stopalo. Na vsak segment smo namestili tudi štiri markerje sistema Optotrak. Izvedli smo šest sklopov meritev: merjenec naredi tri korake iz mirovanja (15x), hoja na mestu (30 korakov) fleksija/ekstenzija kolka (10x), abdukcija/addukcija kolka (10x), fleksija/ekstenzija kolena (10x) in fleksija/ekstenzija gležnja (10x).

3 REZULTATI

3.1 Kalibracija

Tabela 1 prikazuje kalibracijske parametre za enega od pospeškometrov: občutljivosti in odmike od ničelne vrednosti za vsako os ter vse tri ortogonalizacijske kote. V idealnem primeru bi bile vse občutljivosti enake 1, vsi trije odmiki enaki 0 in vsi ortogonalizacijski koti enaki $\pi/2$. Rezultati za ostale senzorje so bili podobni.

Tabela 1: Primer kalibracijskih parametrov pospeškometra.

občutljivost [korekcijski faktor]			**odmik od ničelne vrednosti [m/s^2]**			**ortogonalizacijski kot [rad]**		
S_x	S_y	S_z	B_x	B_y	B_z	α	β	γ
1,011	1,004	0,995	0,084	0,050	0,107	1,522	1,571	1,593

3.2 Validacija na leseni nogi

Absolutna napaka (razlika med koti v sklepih, izmerjenimi z IME in Optotrakom) je prikazana v Tabeli 2 za kolk, koleno in gleženj.

Tabela 2: Absolutne napake za model lesene noge, predstavljene kot mediana, 25. percentil in 75. percentil.

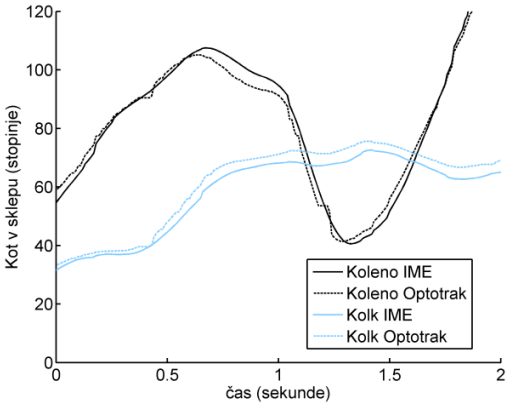
kot v sklepu	absolutna napaka [°]		
	mediana	*25. %*	*75. %*
kolk, sagitalno	0,38	0,03	0,97
kolk, koronalno	0,43	0,06	1,60
koleno, sagitalno	0,53	0,09	1,31
gleženj, sagitalno	0,38	0,14	0,73
gleženj, koronalno	1,11	0,16	3,72

3.3 Validacija na človeku

Absolutna napaka (razlika med koti v sklepih, izmerjenimi z IME in Optotrakom) je prikazana v Tabeli 3 za kolk, koleno in gleženj. Primer kotov v sklepih je prikazan na sliki 4 tako za IME kot Optotrak.

Tabela 3: Absolutne napake za človeško nogo, predstavljene kot mediana, 25. percentil in 75. percentil.

kot v sklepu	absolutna napaka [°]		
	mediana	*25. %*	*75. %*
kolk, sagitalno	2,29	1,13	4,01
kolk, koronalno	2,86	0,99	6,65
koleno, sagitalno	1,75	0,88	3,09
gleženj, sagitalno	4,04	1,68	7,64
gleženj, koronalno	5,00	1,99	8,43



Slika 4: *Koti v sklepih človeške noge, izmerjeni z IME in Optotrakom.*

4 RAZPRAVA

Rezultati validacije na leseni nogi kažejo, da naše IME lahko merijo orientacije in kote s približno eno stopinjo odstopanja od meritev Optotraka. Rezultati meritev na človeku so malce slabši, vendar je mediana napake še vedno nižja od 4 do 5 stopinj. Razlike med IME in Optotrakom so manjše v sagitalni ravnini, kar je spodbudno, saj za zaznavanje mnogih dogodkov med hojo potrebujemo le kote v sagitalni ravnini (npr. [8]). Vseeno pa so gibi v koronalni ravnini pomembni med obračanjem in v nestabilnih razmerah (npr. potencialni padci), zato jih ne smemo zanemariti.

Razlike med rezultati IME in Optotraka seveda niso le posledica slabosti IME. Do napak pride tudi zaradi vplivov, kot je naprimer premikanje enot in markerjev na koži. Ti vplivi bi morali biti manj opazni pri meritvah na leseni nogi, ki je zelo toga, kar deloma razloži manjše razlike med IME in Optotrakom v primerjavi z meritvami na človeku.

V eksoskeletu bi naš sistem lahko uporabili za zaznavanje dogodkov in sprejemanje odločitev, vendar pa bi feromagnetne kovine in elektronske komponente lahko povzročile motnje na izhodu magnetometra. V tem primeru lahko Kalmanov filter nastavimo tako, da magnetometer obravnava kot manj zanesljivega. Ta sprememba bi bila preprosta, vendar pa bi morda zmanjšala točnost sistema.

5 ZAKLJUČEK

Uspešno smo razvili sistem inercijskih merilnih enot, sestavljenih iz pospeškometrov, žiroskopov in magnetometrov. Sistem meri kote v sklepih z mediano napake (glede na referenčni sistem) manj kot 2 stopinji za gibe lesene noge in manj kot 5 stopinj za gibe dejanske človeške noge. Je tudi zelo majhen, brezžičen in modularen, zato ga lahko samostojno uporabimo za analizo človeškega gibanja ali pa vključimo v kompleksnejše sisteme.

Zahvala

Raziskavo sta podprla Javna agencija za raziskovalno dejavnost Republike Slovenije ter Sedmi okvirni program Evropske skupnosti (projekt EVRYON, grant 231451).

Viri

[1] J. F. Veneman, R. Kruidhof, E. E. Hekman, R. Ekkelenkamp, E. H. van Asseldonk in H. van der Kooij, "Design and evaluation of the LOPES exoskeleton robot for interactive gait rehabilitation," IEEE Trans. Neural Syst. Rehabil. Eng., vol. 15, pp. 379-386, 2007.

[2] H. Kawamoto in Y. Sankai, "Power assist method based on phase sequence and muscle force coordination for HAL," Adv. Robotics, vol. 19, pp. 717-734, 2005.

[3] C. J. Walsh, K. Endo in H. Herr, "A quasi-passive leg exoskeleton for load-carrying augmentation," Int. J. Hum. Robot., vol. 4, pp. 487-506, 2007.

[4] T. Wannier, C. Bastiaanse, G. Colombo in V. Dietz, "Arm to leg coordination in humans during walking, creeping and swimming activities," Exp. Brain Res., vol. 141, pp. 375-379, 2001.

[5] D. Jurman, M. Jankovec, R. Kamnik in M. Topič, "Calibration and data fusion solution for the miniature attitude and heading reference system," Sens. Actuat. A-Phys., vol. 138, pp. 411-420, 2007.

[6] X. Yun, E. R. Bachmann in R. B. McGhee, "A simplified quaternion-based algorithm for orientation estimation from Earth gravity and magnetic field measurements," IEEE Trans. Instrum. Meas., vol. 57, pp. 638-650, 2008.

[7] R. van der Merwe in E. A. Wan, "Sigma-Point Kalman Filters for Integrated Navigation," Proceedings of the 60th Annual Meeting of The Institute of Navigation, Dayton, OH, June 2004, pp. 641-654.

[8] J. M. Jasiewicz, J. H. Allum, J. W. Middleton, A. Barriskill, P. Condie, B. Purcell et al., "Gait event detection using linear accelerometers or angular velocity transducers in able-bodied and spinal-cord injured individuals," Gait Posture, vol. 24, pp. 502-509, 2006.

IMPROVEMENTS OF THE ELECTRONIC DRIVER DESIGN FOR DIELECTRIC ELASTOMER ACTUATORS

Mitja Babič, Jadran Lenarčič
Department of Automation, Biocybernetics and Robotics, Jožef Stefan Institute, Ljubljana, Slovenia
e-mail: mitja.babic@ijs.si

ABSTRACT

Actuators based on Dielectric Elastomers (DE) are a promising technology in robotic and mechatronic applications. Up to now, the practical application of DE-based devices is also hampered by inappropriateness of the electrical driving circuit, which often degrades the actuator response and energy efficiency. In this paper design improvements to a novel electronic driver for DEs are presented. The driver is based on a custom electronic driver derived from the flyback transformer topology. First, a two transistor flyback converter is described, which guarantee an improved energy efficiency and overvoltage protection in comparison to the classical design. Second, the input current measurement approach is shown, which is necessary to prevent damage to the switch under transient conditions and to measure the system efficiency. Finally, the improved driver performance is tested on a fixed capacitor load, whose value represents the actual DE-actuator capacitance.

1 INTRODUCTION

Dielectric elastomers (DEs) are solid elastic dielectrics which generate forces and undergo large deformations in response to large applied electric fields (see Fig. 1). Such an electrically induced deformation can be used to produce useful mechanical work. In the past decade, numerous DEs have been retrieved and discovered. Due to their inherent multi-physics interactions, researchers have identified DEs as promising materials for building solid-state sensors and actuators. However, many experimental devices exhibit effective performances that are below expectations [1, 2].

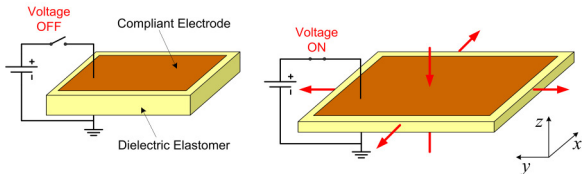


Fig. 1: *Electrically induced isochoric area expansion of an electro-active film.*

Up to now, the practical application of DE-based devices is hampered by poor material reliability, non-linear and non-repeatable electro-mechanical behavior, non-trivial control, and the inappropriateness of the electrical driving circuit [3–5], which often degrades the actuator response and energy efficiency.

From the electrical standpoint DE-actuators are treated as capacitive loads. Voltage-regulated power supplies, such as highvoltage, bench-top power supplies or high-voltage, portable DC-DC converters, are inappropriate for capacitor-charging applications since they consume at least double the energy that is being stored in the capacitor and furthermore provide limited peak currents that slow down the capacitor-charging process [6–8]. In addition, all these phenomena also reduce the DE actuator efficiency and limit its dynamic response.

In order to circumvent the aforementioned control issues a novel activation strategy for controlling DE actuators was proposed by the authors in a previous research work [9]. The presented electronic driver in [9] is derived from the flyback converter topology and it is able of delivering to the DE actuator middle-frequency, current-pulse trains with a duty-cycle dependent on the actuator-position error.

In this study some design improvements of the electronic driver presented in [9] are shown. The driver's design is based on an estimation of the DE actuator's electrical parameters (DE-actuator capacitance) and a dynamic response (settling time). Beside the design of a transformer also the optimal driver's switching and protection circuit components needs to be chosen in order to perform the appropriate driver's functioning and safety operation. The two transistor flyback converter is presented, which guarantee improved energy efficiency and overvoltage protection in comparison to the single transistor design. The input current measurement through the power switch with a serial current-sensing resistor is described. The current measurement is necessary to prevent damage to the switch under transient conditions for voltage-mode control and is an input to the control loop for current-mode control. Finally, the improved driver performance is tested on a fixed capacitor load, whose value represents the actual DE-actuator capacitance.

FLYBACK CONVERTER

The flyback converter basic scheme is shown in Fig. 2. The primary advantages of the presented circuit are the low parts count and simplicity, making it useful for relatively low power levels, up to approximately 150 W. The flyback converter is also widely used for high output voltages, thanks to the isolation between the input and output, which is achieved using the transformer [10,11]. Note that high voltages and medium currents are also the requirements of DE-actuator technology.

Regarding to Fig. 2 the working principle can be defined as follows. When the power transistor is turned on, the energy is stored in the transformer; meanwhile the load current is supplied from an output filter capacitor. When the power transistor switched off, the energy stored in the transformer recharges the filter capacitor with the current lost while delivering the load current.

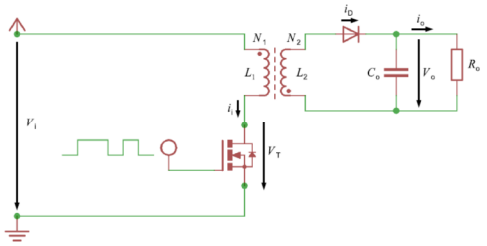


Fig. 2: *Flyback converter schematic [10, 11].*

From the electrical standpoint DE-actuators behave as capacitive loads. The DE-actuator's equivalent electrical *RC* circuit model depicted in Fig. 3 [9] is composed of the DE-actuators variable capacitance C_E, the serial variable electrode resistance R_{ES} and the parallel variable resistance R_{EL}, which represent the leakage current flowing through the DE-actuator's capacitance C_E. When the flyback converter filter capacitor C_o, shown in Fig. 2, is replaced with the DE actuator's electrical model represented in Fig. 3 [9], the DE-actuator's electrical driver shown in Fig. 4 is obtained [9]. The aim of the electronic driver is to charge the DE capacitor and provide additional energy to supply the drain resistance current $i_o = V_o/R_o$ and the DE actuator's losses.

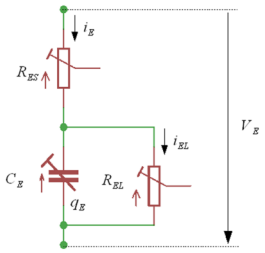


Fig. 3: *DE actuator electrical model [9].*

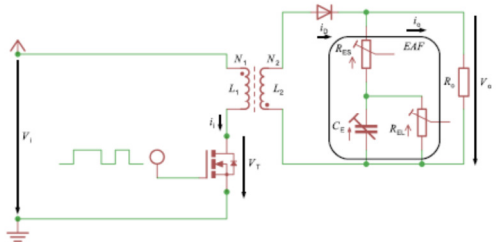
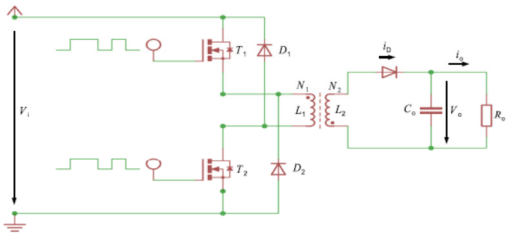


Fig. 4. *Proposed electrical driver schematic for DE actuators[9].*

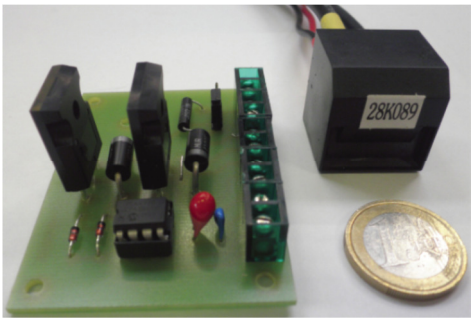
Two transistor discontinous-mode flyback

The main drawback of the electronic driver represented in Fig. 2 is that the off-voltage of the primary side power switch is inherently unconstrained. The excessive voltage on the drain pin may lead to an avalanche breakdown and eventually damage the MOSFET. Therefore, it is necessary to limit the peak voltage by additional snubbers and/or clamps that generally dissipate power and reduce efficiency [10, 11].

This issue could be overcome by the two transistor flyback converter schematic shown in Fig. 5a. The basics of the circuits operation are similar to the single transistor design. The difference is that the leakage inductances energy, and therefore, the power transistor drain voltage, is no longer unconstrained. The power transistors in the OFF state are subject to only the maximum DC input voltage plus a couple of diode drops, therefore no snubbers or clamps are required. In the OFF state the transformers primary becomes a current source that forward biases diodes D_1 and D_2, which return the energy stored in the transformers primary to the input source. The primary current is the same as the single transistor design. The ON resistance is still much lower than the single high-voltage transistor, so a higher power level can be obtained using the two transistor design. Efficiency will be higher than the single transistor design since the energy that would be dissipated in snubbers and clamps is recovered to the input source in the case of the two transistor design [11, 12].



(a) Circuit diagram [10, 11].



(b) PCB circuit board prototype.

Fig. 5. Two transistors flyback converter.

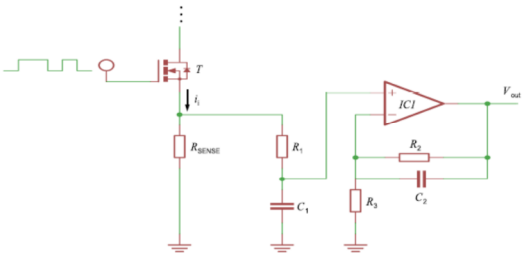
The design of the two transistor converter is basically the same as for the single transistor, special care must be taken in order to maximize the transfer of the magnetizing current to the reflected load and to avoid a delay in the transfer of current to the load. The slope of the leakage inductance current decay should be maximized, thats mean the leakage inductance current should be rapidly reset to zero [11].

The used two transistors flyback converter PCB (Printed Circuit Board) circuit board prototype is shown in Fig. 5b on

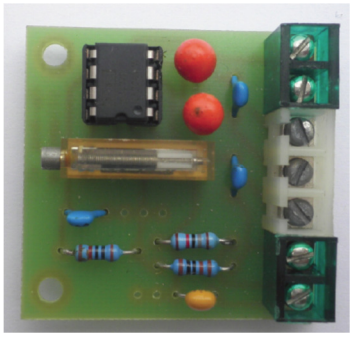
the left part, meanwhile on the right part is shown the employed miniature ferrite high-frequency high-voltage transformers model 28K089 by Information Unlimited.

Measuring input current of a flyback converter

In most converters there is a circuit that monitors the current through the power switch. The current measurement is used to prevent damage to the power switch under transient conditions for voltage-mode control and as an input to the control loop for current-mode control [10,13]. The DC current measurement into the converter is especially suitable in high-voltage application, where the secondary current can be calculated based on the primary current measurement. In the flyback converter topologies the input current through the power switch can be measured with a serial current-sensing resistor.



(a) Circuit diagram [13].



(b) PCB circuit board prototype.

Fig. 6. Measuring the flyback converter input DC current.

Fig. 6a shows the circuit schematic used for measuring the flyback converter input DC current, meanwhile the utilized PCB circuit board prototype is shown in Fig. 6b. The circuit use a current-sense resistor (R_{SENSE}) to measure the instantaneous current through the power switch. The voltage developed across the current-sensing resistor is than integrated by an active RC filter over a single converter cycle. The average value of the input current can be determined by dividing the RC filter integrated value with the current-sensing resistance value. In order to obtain an accurate measurement the filter corner frequency (f_{RC}) must be much lover in comparison to the converter switching frequency or $f_{RC} << (f /100)$. The filter corner frequency is defined as

$$f_{RC} = \frac{1}{2\pi C_1 R_1}. \quad (1)$$

The current-sensing resistor has a small value in order to obtain an accurate measurement and to minimize the voltage drop those currents create across the resistance. Therefore the sensing voltage must be amplified by the amplifier gain

$$A_{RC} = 1 + \frac{R_2}{R_3}. \quad (2)$$

By selecting the values of R_2 and R_3, it is possible to achieve any scaling factor desired. Finally the capacitor C_2 provides additional filtering if needed, with the filter corner frequency

$$f_{RC2} = \frac{1}{2\pi C_2 R_2}. \quad (3)$$

EXPERIMENTAL RESULTS

To test the validity of the proposed driving circuits, an electronic driver prototype was assembled and tested on simple electrical load in order to compare its response with the proposed design parameters. The adopted experimental set-up comprise an input signal generator (low-voltage), the tested electronic driver, an electric load and an output voltmeter. The electric load is an RC load (C_o = 1 nF, R_o = 100 MΩ), which values are chosen within the ranges of typical DE electrical impedances.

The experimented electronic driver prototype is based on a high voltage flyback transformer with nominal parameters L_1 = 19 μH, R_1 = 0.12 Ω, N_1 = 2500, L_2 = 1 μH, R_2 = 8.1 kΩ, N_2 = 9, a maximum output voltage V_0 = 7 kV and operating power P_T =10 W. In order to allow the flyback converter to operate in discontinuous mode, the maximum duty cycle is set to D_{max} = 45 % and a switching frequency of f = 10 kHz is selected. Based on the selected RC load and limiting the output voltage to V_{omax} = 5 kV, the time response is τ_{el} = R_oC_o = 100 MΩ · 1 nF =0.1 s, the drain current is i_{omax} = V_{max}/R_o= 5 kV / 100 MΩ = 50 μA and the power dissipation is P_{omax} = V^2_{omax} / R_o = (5 kV)^2 / 100 MΩ = 0.25 W. The required input voltage value is V_{imax} = 1V [9].

Step response

During the open-loop step-response measurement the electronic driver is activated with a step PWM (Pulse-Width Modulation) pulse-wave signal and the output-voltage response is measured. Fig. 7 reports the system's open-loop response to a pulse wave. The rise time is measured from 10 % to 90 % of the steady-state values and the fall time is measured from 90 % to 10 % of the steady-state values. The measured rise time $t_r \approx$ 9 ms and fall time $t_f \approx$ 230 ms are in accordance with the previous calculated values [9]. It is worth highlighting that the step response depends on the proposed electronic driver's design parameters, in particular on the charging time and on the time constant of the applied RC load (τ_{el}). The converter output ripple is far below the position sensitivity of the DE actuators $\Delta V \approx$ 50 V.

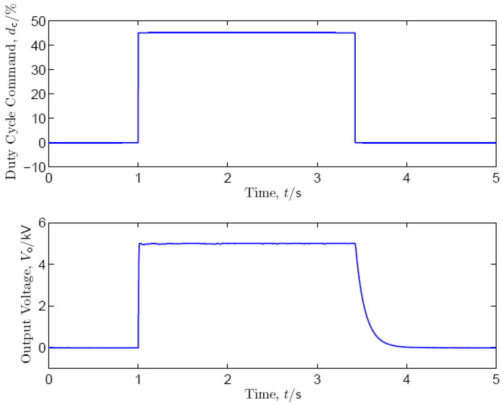


Fig. 7. Open-loop step response of the proposed electronic driver.

Input current measurements

The converter input current measurement using the proposed current-monitor circuit is shown in Fig. 8. The test is performed on a constant RC load and at constant input voltage for different duty cycle values (Fig. 8 upper figure - Duty Cycle). There is some noise on the error amplifier output present (Fig. 8 lower figure - Input Current), which can be reduced by additional filtering or creating a better circuit layout. As can be seen in the lower waveform in Fig. 8, good results were obtained at high-current and low-current levels during the dynamic tracking. Note that in order to increase the monitor circuits frequency response, the RC values of the low-pass filter should be changed. However, the obtained results are in accordance with the results obtained with a precision ammeter.

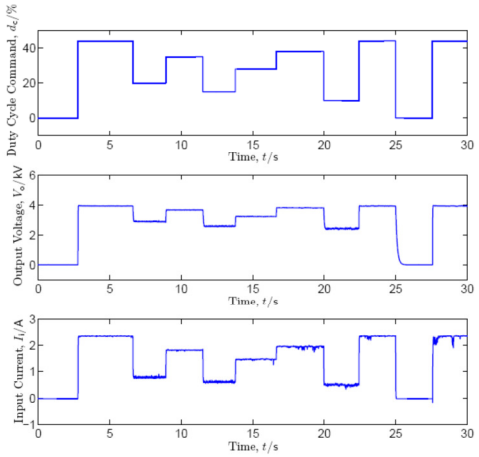


Fig. 8. Current-monitor of the proposed electronic driver.

5 CONCLUSIONS

In this paper improvements of the electronic driver's design for compliant actuators based on dielectric elastomers has been presented. The motivation for this work is based on the observation that DE-based actuators intrinsically offer great potential for development of novel solid-state mechatronic devices. However, the practical application of DE-based actuators up to now is hampered by the inappropriateness of the electrical driving circuit which often degrades the actuator's response and its energy efficiency.

The presented novel electronic driver is more performing, has a simpler architecture and is composed of a smaller number of components. Additionally, it should be more compact and less expensive than traditional driving circuits, which are based on high-voltage voltage-regulated DC/DC converters or DC power supplies. The driver's switching and protection circuit components were presented, which are required to perform the appropriate driver's functioning and safety operation. The performance of the driver was tested on a fixed capacitor load, whose value represents the actual DE-actuator capacitance. The experimental results are in accordance with the obtained calculations of the driver's design process.

REFERENCES

[1] Bar-Cohen, Y., 2004. *Electroactive Polymer (EAP) Actuators as Artificial Muscles: Reality, Potential and Chal lenges*, Vol. PM136 of *2*. SPIE Press, March.

[2] Kim, K. J., and Tadokoro, S., 2007. *Electroactive Polymers for Robotic Applications: Artificial Muscles and Sensors*. Eds. Springer.

[3] Choi, H. R., Jung, K. M., and Nam, J. D., 2007. "Robotic applications of artificial muscle actuators". *Electroactive Polymers for Robotic Applications Artificial Muscles and Sensors*, March, pp. 49–90.

[4] Vertechy, R., Castelli, V. P., and Waldron, K. J., 2006. "On the driving circuits of electro-sensitive elastomers". pp. 14–16.

[5] Xie, S. Q., Ramson, P. F., Graaf, D. D., Calius, E. P., and Anderson, I. A., 2005. "An adaptive control system fordielectric elastomers". *IEEE International Conference on Industrial Technology, 14*(17), December, pp. 335 – 340.

[6] Mita, K., and Boufaida, M., 1999. "Ideal capacitor circuits and energy conservation". *Americal Journal of Physics, 67*, pp. 737–739.

[7] Hayworth, B., and Warrilow, D., 1978. "Constant powercharging supplies". *Electro-Optical Systems Design, 10*(3).

[8] Elwell, R., Cherry, J., Fagan, S., and Fish, S., 1995. "Current and voltage controlled capacitor charging schemes". *Magnetics, IEEE Transactions on, 31*(1), pp. 38–42.

[9] Babic, M., Vertechy, R., Berselli, G., Lenarcic, J., Castelli, V. P., and Vassura, G., 2010. "An electronic driver for improving the open and closed loop electro-mechanical response of dielectric elastomer actuators". *Mechatronics, 20*(2), pp. 201 – 212.

[10] Brown, M., 1990. *Practical switching power supply design*. Academic press.

[11] Pressman, A. I., 1998. *Switching Power Supply Design*. 2 edition. McGraw-Hill Professional.

[12] Varga, C., 2001. "Power converter topology and mosfet selection for 48-v telecom". *Firechild Semiconductor Corporation*, June, pp. 1–11.

[13] Bottrill, J., 2005. "Measuring input current of a buck or fly-back converter". *Power Electronics Technology*, November, pp. 34–38.

DESIGN OF A MANIPULATOR AND RELATED TOOLS FOR DEDICATED OPERATIONS

Anton Ružič
Department of Automation, Biocybernetics and Robotics
Jozef Stefan Institute
Jamova 39, 1000 Ljubljana, Slovenia
Tel: +386 1 4773407; fax: +386 1 2519385
e-mail: anton.ruzic@ijs.si

ABSTRACT

In this paper we present our design and experimental activities aimed to the development of a special purpose manipulator that is adequate for automated cleaning of small to medium sized appliances. We describe the specific requirements that originate from the geometry and surface materials of the objects to be cleaned, from the nature of the impurities to be removed and finally from the target manipulator environment. We then present the experiments performed to develop a cleaning technology appropriate for robot use and application. Finally we illustrate the synthesis of a reconfigurable manipulator that carries the developed cleaning tools and is deployable in a fully automated production facility.

1 INTRODUCTION

After the initial decades of robotics, when the main application target areas were typical industrial operations, the effort has widened to novel robot systems and to other less conventional applications. One of such jobs is surface cleaning.

Cleaning service robots are already used in actual homes for dust cleaning. Most of domestic robots use vacuuming technology [1], much less sweeping and mopping [2] and rarely washing cleaning technology. New technologies for home cleaning are studied [3, 4]. For in-door cleaning tasks, benchmarks have been set to compare robots efficiency [5].

Research projects have investigated cleaning in other environments as airports, shopping centers, open urban areas and even in a radioactively polluted plant [6, 7, 8].

The goal of our work was to find a solution for automated cleaning of small to medium sized appliances, manufactured in an industrial environment. Examples of such products are computer printers, home vacuum cleaners and similar. The cleaning requirements are very tough, as these appliances' surfaces are much more delicate in comparison with floor areas. For delicate surface cleaning approaches with vision and force control have been studied but to our knowledge no such general purpose robot product is available [9].

We present some of the work we carried out in order to synthesize and develop a special purpose cleaning manipulator, effectively removing the present manual cleaning stages. This include the determination of process requirements and solution characteristics, the development of a dedicated cleaning technology [10] and the final synthesis of a special purpose cleaning manipulator [11].

2 SPECIFICATION OF PROCESS CHARACTERISTICS

Initially we identified all relevant target process characteristics and specific requirements. This is an outstandingly important task, as a satisfactory manipulator solution must conform to all requirements.

Figure 1: *Example of appliances with glossy and soft plastic cover surfaces: a computer printer and a vacuum cleaner.*

The manipulator should be able to clean small to medium size appliances like home computer printers, vacuum cleaners and similar devices, having the following specific surface and interior characteristics (see examples on Fig. 1):

- relatively soft and delicate plastic surfaces,
- glossy surface finishing, visually exalting matter on it,
- the glossy surface make also evident even light mechanical damages;
- the interior with electronic or electric components;

Additionally, the solution should be able to clean satisfactorily a class of products, e.g. printer models.

Finally, the solution should enable stand-alone operations as well as the possibility to insert and use the manipulator in a larger production environment.

In the latter case, the cleaning process is typically executed in the final stages of production, just before final packaging. The production environment requirements are

- the typical production cycle from 15 to 25 seconds;
- 24/7 manipulator operation;
- minimal down-time for short-term maintenance;
- operability in real production environments, possible integration with a typical continuous production line.

3 RESULTING DESIGN REQUIREMENTS

Basing on the specified requirements we first determined the necessary solution characteristics for the cleaning technology/cleaning tools and for the manipulator structure.

3.1 Cleaning technology characteristics

The necessary cleaning technology characteristics are:

1. the cleaning process should effectively to remove any dirt, dust or other impurities from the shiny surface;
2. for a typical appliance case, cleaning must remove:
   - plastic burrs;
   - dust, deposed on the plastic covers;
   - lighter grease stains;
3. the cleaning process should not create scratches and abrasions;
4. washing technology can't be used, only cleaner liquid:
   - liquid can be used in small quantities,
   - it must be neutral to the soft plastic material;
5. the cleaning operation should be fast, in line with the required cycle times of up to 15 to 20 seconds.

3.2 Manipulator construction characteristics

Necessary characteristics for the manipulator structure are as follows:

1. The construction of the manipulator carrying the cleaning tool should be able to adapt to a certain variety in appliance exterior shapes and dimensions;
2. cleaning technology maintenance should be fast;
3. to withstand 24/7 operation in real production, the manipulator and cleaning technology must be robust;
4. to enable possible insertion into a production line, the manipulator should allow objects to pass-through.

4 SOLUTION OUTLINE AND NECESSARY EXPERIMENTAL AND DESIGN STEPS

From the established design requirements we determined the overall solution framework.

In general terms, we decided to design a special manipulator, carrying some cleaning tools; together this system must be capable of applying the cleaning method and technology all over the appliance surface.

In contrast with the usual manual cleaning process, we decided to

- apply the cleaning process (whatever cleaning process we would develop) uniformly to the whole surface area of the appliance, without identifying more and less clean subareas.
- to avoid human-like firm contact wiping and to design a special purpose fast rotational cleaning tool, carrying some cleaning material to be defined later.

Cleaning coverage is achieved by a combination of the (fast) rotating cleaning tool and (slower) movement of the tool along the surface.

In order to achieve shorter cycle time, we choose that

- the manipulator will have a number of parallel kinematic chains, each carrying one cleaning tool;
- all parallel manipulator arm branches will apply cleaning concurrently, each on one surface area part.

As each chain covers simpler surface areas the manipulator can be kinematically simpler and thus more robust.

The cleaning material would eventually become saturated with impurities and has to be replaced. An appropriately designed tool exchange system ensures fast replacement.

Finally, the parallel axes manipulator should have a transport path through it according to the appliances' geometry.

5 DEVELOPMENT OF THE CLEANING TECHNOLOGY

We developed a cleaning technology including a suitable cleaning material, appropriate cleaning liquids, a cleaning liquid dispensing system, and a complete cleaning tool.

5.1 Determination of the cleaning material

To determine the cleaning material to be mounted on a rotational cleaning tool we tested a number of materials, some of which are illustrated in Fig. 2. We fixed them on a appropriate carrier actuated rotationally by an electromotor. This experimental axis was mounted on an industrial robot.

Figure 2: *Some of the evaluated cleaning materials.*

To attain comparable and repeatable results, we applied controlled amount of impurities to the test appliance surface. We swept the tested material on the experimental axis over the surface. We evaluated the cleaning performance at various rotational speeds, linear speeds and different distances between the axis to the test surface (see Fig. 3).

Some of quantitative and qualitative tests included:

- capability to a) blow away burrs and dust before contact b) absorb grease and other sticky impurities;
- time to achieve clean surface (if attainable);
- determination of impurity absorption capability;

- influence of cleaning liquids;
- cleaning material resistance to wearing.

From the tested materials we chose a special microfiber cloth and continued development with this material.

Figure 3: *The cleaning materials evaluation setup.*

Surface damage like scratches or obvious loss of surface shininess was determined optically and quantitatively using a surface profiler (see Fig. 4).

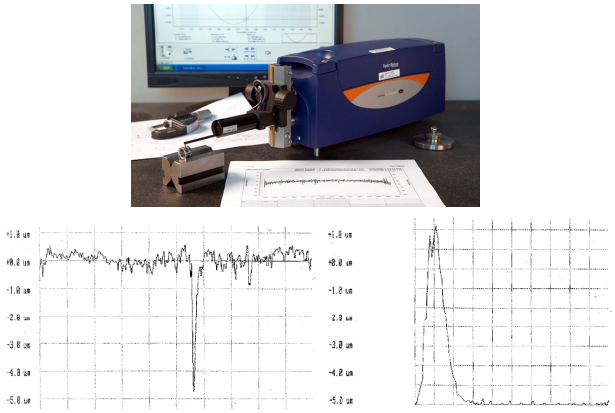


Figure 4: *Surface profiler sampling head; an example of a profile and its respective spectrum characteristics.*

5.2 Determination of optimal cleaning cloth disposition and cleaning tool parameters

In a previously described phase, we chose the cleaning material cloth and the basic disposition of the cloths on the cleaning tool in form of axially disposed sheets. We determined that tool should be moved along the cleaned surface perpendicular to the tool centerline (axis).

In the next step, we performed a number of experiments to determine the best characteristics for the final cleaning tool design:

- the configuration of each single cloth (e.g. thickness, single or double form and similar and dimensions);
- the number of cloth sheets, disposed uniformly around the cleaning tool;
- the radius of the cleaning tool at the cloth sheet attachment line and the cloth sheet extent (outer radius from the tool centerline);
- the tool rotational speed (revolutions per second)
- the range of acceptable cleaning tool movement speed along the cleaned surface.

We also quantified another important characteristic:

- the radial range of the cleaning tool's effective operation, given by the inner and outer radius.

This information was used for the determination of the number of cleaning tools and their disposition for a specific cleaning manipulator configuration for an appliance.

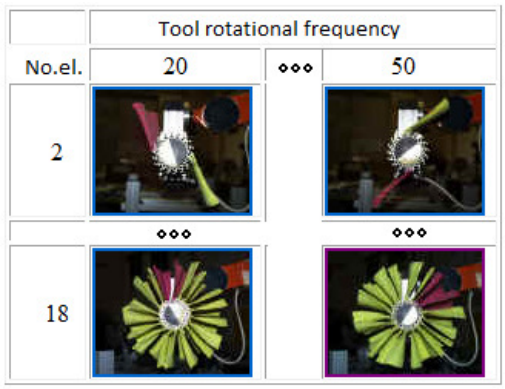


Figure 5: *One of the experiments for the determination of cleaning tools structure and for the determination of some operational parameters.*

In the Fig. 5 we illustrate just one in the series of experiments and measurements carried out in order to design the cleaning tool.

5.3 Mechanical design of the cleaning tool

After we defined the cleaning tool's cylinder diameter, cloth number and dimensions, we designed its precise mechanical structure.

To achieve fast manipulator's cleaning tools exchange, we designed an appropriate connection mechanism for the tool.

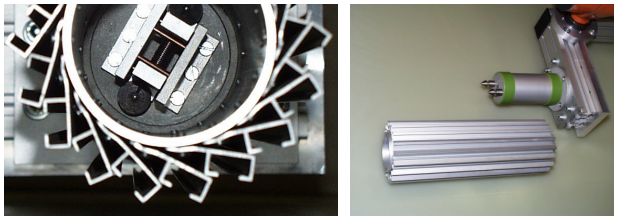


Figure 6: *Actual cleaning tool construction (with the cleaning cloths removed).*

A secure lock enables quick release of the tool with soaked cleaning cloths and engagement of a replacement tool with fresh cloths in just a couple of seconds. The basic structure of the actual designed tool (without the attached clothes) is shown in Fig. 6.

According to the robustness requirement, we designed the structure to withstand the forces resulting from a 3° misalignment between the manipulator carrier and tool axis.

5.4 Selection of cleaning liquids and design of a liquid dispensing system

We selected the cleaning material so that it is functional for dry cleaning. For further cleaning improvement, we also tested use of several liquids. We tested solvents, detergents and blends. We found that the solvents generally ensure better cleaning while having some negative aspects:

- they can have reactions with the plastic surface;
- explosion and flame hazard measures must be taken;
- exposure to some solvents can result in health hazards.

The detergents have less of these negative aspects. However, their accumulation in the cleaning cloths requires more frequent tool exchanges.
We designed an appropriate liquid dispensing subsystem.

6 DESIGN OF THE CLEANING MANIPULATOR

We outlined some manipulator design characteristics in section 4. Here we overview some additional design choices resulting from the cleaning tool structure.

- The manipulator is configurable; its configuration is determined and set up specifically for an appliance or a class of similar appliances to be cleaned
- the manipulator have a number of parallel kinematic chains, each carrying one cleaning tool;
- each parallel axis has one to two degrees of freedom; a simple axis is linear; in the another type the tool axis is moving perpendicular to a plane, along a curve lying on this plane and defined by a fixed mechanical template;
- during the cleaning process, all manipulator arms/chains will move concurrently, each cleaning one part of the whole surface area;

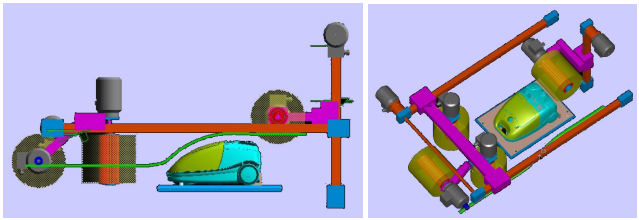


Figure 7: *A possible manipulator axis configuration, determined for cleaning of a particular class of appliances.*

This configurable manipulator has a frame base to which a number of axes can be fixed. Each axis origin and tool movement direction is set so that the rotating cleaning tools, moving along the axes, cover the entire appliance surface to be cleaned. To ensure the coverage we consider that

- a single linear axis has the cleaning coverage volume approximately equal to a block (to a rectangular parallelepiped) of dimension W x H x L, where W is the width of the cleaning tool, H is the height effective tool cleaning depth and L is the axis travel range;
- similarly, a single template axis has the volume coverage obtained by sweeping a rectangle W x H along the curve of the template.

If the manipulator is to be actually installed on a production line, the axes must be disposed accordingly.
In general, we can find more valid manipulator axis configurations for some appliance as (see Fig. 7). Most often, we can find a configuration having fewer axes, but requiring longer time to clean the appliance and vice versa.

7 CONCLUSION

In the paper we described the design of a special manipulator for automatic cleaning of small to medium sized appliances with enclosures from delicate glossy plastic material. We developed a cleaning technology and synthesized the cleaning manipulator configuration and structure for a class of appliances to be cleaned. Our tests showed that the developed technology is appropriate.
The following work should address an automatic procedure, possibly implemented in a form of a computer program, to determine the cleaning axes type, number and path for the reconfigurable industrial robot cleaner, given the geometrical model of target appliances to be cleaned.

References

[1] IRobot 2011. Roomba Cleaning Robots. http://www.irobot.com.

[2] Evolution Robotics 2011. Mint automatic floor cleaner. http://www.evolution.com/.

[3] Gizmodo 2009. Panasonic unveils Fukitorimushi: a worm-like robot for cleaning floors. May, 10. http://www.gizmag.com/.

[4] De Vogel M. 2011. Dust vacuum systems - time seems mature for robot vacuuming, Elektro Retail Magazine, Vol. 80, pp.20-23.

[5] Prassler E., Hägele M., Siegwart R. 2006: International Contest for Cleaning Robots: Fun Event or a First Step towards Benchmarking Service Robots, Field and Service Robotics, Springer Tracts in Advanced Robotics, Springer Berlin / Heidelberg, pp.447-456.

[6] Eureka 1996. Independent Floor Cleaning Machine for Hypermarkets, Eureka project: 1094 Clean. http://www.eurekanetwork.org.

[7] Dario, P. 2011. A Real Case of Deployment of Deployment of Service Robots for Urban Hygiene. In: Proc. 4th International Expert Days on Service Robotics, Hausen (Germany), February 23-24.

[8] Guizzo E. 2011. Robot Vacuum Sucks Up Radiation at Fukushima Plant, IEEE Spectrum Automation Blog. http://spectrum.ieee.org/blog/robotics/robotics-software/automaton.

[9] Prassler E., Ritter A., Schaeffer C. and Fiorini P. 2000. A Short History of Cleaning Robots, Autonomous Robots, Vol. 9, No. 3, pp.211-226.

[10] Ruzic, A. 2011. Development of a cleaning technology suitable for use on automatic devices, Jozef Stefan Institute report IJS DP 10787.

[11] Ruzic, A. 2011. Design of a special purpose robot manipulator, Jozef Stefan Institute report IJS DP 10788.

Zbornik 14. mednarodne multikonference

INFORMACIJSKA DRUŽBA – IS 2011

Proceedings of the 14th International Multiconference

INFORMATION SOCIETY – IS 2011

Kognitonika

Cognitonics

Uredila / Edited by

Vladimir A. Fomichov, Olga S. Fomichova

http://is.ijs.si

10.–11. oktober 2011 / October 10th–11th, 2011
Ljubljana, Slovenia

Preface / Predgovor

Second International Conference on Cognitonics (Cognit 2011)

During the last decade many scholars at various international and national conferences, in the books and papers have said and written about serious distortions in the development of the personality and national cultures caused by stormy development of informational technologies and globalization processes.

The conviction that it is not only necessary but also POSSIBLE to make something constructive and significant for compensating these distortions underlay the elaboration of the foundations of a new scientific discipline called Cognitonics (see Informatica. An International Journal of Computing and Informatics (Slovenia), 2006, Vol. 30, No. 4, pp. 387-398, www.informatica.si/vol30.htm#No4 and the Proceedings of the First International Workshop on Cognitonics (Slovenia, Ljubljana, 12-13 October 2009) - a subconference of the 12th International Multiconference Information Society, Jozef Stefan Institute, Ljubljana, Slovenia; http://is.ijs.si/is/is2009/zborniki.asp?lang=eng; Proceedings A, Section "Cognitonics").

Cognitonics aims (a) at explicating the distortions in the perception of the world caused by the information society and globalization and (b) at coping with these distortions in different fields by means of elaborating systemic solutions for compensating the negative implications of the kind for the personality and society, in particular, for creating cognitive-cultural preconditions of the harmonic development of the personality in the information society and for ensuring the successful development of national cultures and national languages.

The birth of Cognitonics was stimulated by the ideas of Philosophy, Cognitive Linguistics, Artificial Intelligence theory, Web Science, Applied Linguistics, Art theory, Cognitive Psychology, and Cognitive Biology.

Two factors seem to be especially important from the standpoint of achieving the goals of Cognitonics:
- informational technologies have been developing extremely quickly and have been expanding unusually broadly, they penetrate not only into every office and laboratory but also into every school class and every family;
- it is necessary and promising to use the power of modern informational technologies in order to very quickly and broadly disseminate the found effective methods of compensating the negative distortions in the development of the personality and of national cultures in information society.

The goal of the conference is to combine the efforts of the scholars from numerous scientific fields and educators in order to establish a new synergy aimed at ensuring the harmonic, well-balanced development of the personality, national cultures, and national languages in the modern information society and, as a consequence, to compensate a number of broadly observed negative distortions.

From the standpoint of educational practice, Cognitonics proposes an answer to the following question: what precious ideas and images accumulated by the mankind, at what age, and in what a way are to be inscribed into the conceptual picture of the world of a person in order to harmonize his/her intellectual and spiritually-coloured emotional development and to contribute to the successful development of national cultures and national languages.

Cognitonics formulates a new, large-scale goal for the software industry and Web science: to develop a new generation of culture-oriented computer programs and online courses (in the collaboration with educators, linguists, art historians, psychologists) - the computer programs and online courses intended for supporting and developing positively-oriented creativity, cognitive-emotional sphere, the appreciation of the roots of the national cultures, the awareness of the integrity of the cultural space in the information society, and for supporting and developing symbolic information processing and linguistic skills, associative and reasoning abilities of children and university students.

The Program Committee has accepted for the conference 19 papers from four continents: Asia (Japan), Europe (Croatia, Finland, Germany, Greece, Italy, Poland, Romania, Russia, Slovenia, Sweden), North America (Mexico, USA) and South America (Brasilia).

The editors would like to thank the authors of the papers for their contributions and the members of the Program Committee for their precious comments ensuring the high quality of the accepted papers and making the reading as well the editing of this section a rewarding activity.

Vladimir Fomichov, Olga Fomichova

The Second International Conference on Cognitonics – COGNIT 2011 (October 10 - 11, 2011; http://is.ijs.si/is/is2011/konference/kognitonika/cognitonics.html#prog) – **a subconference of the Multiconference "Information Society 2011" (October 10-14, 2011, Jozef Stefan Institute, Ljubljana, Slovenia, Central Europe**, http://is.ijs.si/vabilo_eng.asp?lang=eng)

Co-Chairs: Prof. Dr. Vladimir A. Fomichov (*) and Prof. Dr. Olga S. Fomichova (**)

PROGRAMSKI ODBOR / PROGRAMME COMMITTEE

Dr. Rosa Maria Bottino (Consiglio Nazionale delle Ricerche Technologie Didattiche, Genova, Italy)

Dr. Dan Dumitru Burdescu (Department of Software Engineering, University of Craiova, Dolj, Romania)

Dr. Bridget Cooper (The David Goldman Informatics Centre, University of Sunderland, United Kingdom)

Dr. Vladimir Fomichov (Faculty of Business Informatics, National Research University “Higher School of Economics”, Moscow, Russia)

Dr. Olga Fomichova (State Educational Centre "Dialogue of Sciences", Moscow, Russia)
Dr. Toshio Okamoto (University of Electro-Communications, Japan)

Dr. Luciano Seta (Institute for Educational Technologies, Italian National Research Council, Palermo, Italy)

Dr. Gaba Tsayang (Faculty of Education, University of Botswana, Gaborone, Botswana)

Dr. Linda Turner (School of Health, Faculty of the Professions, University of New England, Australia)

Mag. Hirohiko Yasuda (Shimonoseki Technical High School, Shimonoseki City, Japan).

A MAP OF COGNITIVE TRANSFORMATIONS REALIZED FOR EARLY SOCIALIZATION OF CHILDREN IN THE INTERNET AGE

Vladimir A. Fomichov (*) and Olga S. Fomichova (**)

(*) Faculty of Business Informatics, National Research University "Higher School of Economics"
Kirpichnaya str. 33, 105187 Moscow, Russia; (**) State Educational Centre "Dialogue of Sciences",
Universitetsky prospect 5, 119296 Moscow, Russia
e-mail: vfomichov@hse.ru and v.fomichov@snhu.edu

ABSTRACT

The paper grounds the necessity of much earlier socialization of children in the Internet age. The main goal is to make children (including teenagers) be aware of possible social consequences of their misuse of informational technologies, in particular, of the cell telephones and the Internet. An original method of early forming the cognitive subspace of moral values and social responsibility is stated. It is a part of the System of Emotional-Imaginative Teaching (the EIT-system) developed and tested by the authors during 1990s – 2000s. For describing this method, a new formal notation for representing transformations of the learners' cognitive-emotional sphere and the spectrum of information processing skills is proposed, it is called the notation of the maps of cognitive transformations. The described method of early socialization and the EIT-system as a whole are interpreted as an important component of cognitonics - a new scientific discipline aiming at explicating the distortions in the development of the personality, national cultures and languages caused by informational technologies and globalization and looking for the ways of finding systemic solutions for coping with these distortions.

1 INTRODUCTION

The stormy progress of the Internet since the 1990s has tremendously expanded the sphere of using the informational technologies (IT). In the developed countries, IT have reached practically every home. One of the consequences is that IT have considerably speeded-up the globalization processes. Every coin has two sides. During two decades the scholars in various countries have observed a number of negative shifts in the development of the personality, national cultures and languages caused by IT and globalization processes. These were the principal reasons for the birth of a new scientific discipline – cognitonics [5 - 7, 9 - 10]. It aims (a) at explicating the distortions in the development of the personality and national cultures caused by the peculiarities of information society and globalization and (b) at coping with these distortions in different fields by means of elaborating systemic solutions for compensating the negative implications for the personality and society of the stormy development of informational technologies and globalization processes, in particular, for creating cognitive-cultural preconditions of the harmonic development of the personality in the information society and for ensuring the successive development of national cultures and national languages.

From the standpoint of educational practice, cognitonics proposes an answer to the following question: what precious ideas and images accumulated by the mankind, at what age, and in what a way are to be inscribed into the conceptual picture of the world of a person in order to harmonize his/her intellectual and spiritually-coloured emotional development and to contribute to the successful development of national cultures and national languages.

One of the serious, large-scale negative phenomena observed mainly during last decade is that the teenagers (and some times younger children) have received the possibility to distribute any information about their peers and adults with the help of the cell telephones and the Internet. Unfortunately, a small part of children have used this possibility for bullying, in particular, for distributing discreditable photographs and texts [12].

That is why we believe that a possible way out is to elaborate the methods of much earlier socialization of the child than it is usually done in order, on the one hand, to eliminate or considerably diminish children's aggressiveness. On the other hand, for contributing to the birth in children of the feeling of social responsibility and to the understanding by children of the severe consequences suffered by them, their peers, and adults.

The arguments of the kind form the content of Section 2 of this paper. Section 3 describes the objectives and composition of the System of Emotional-Imaginative Teaching (the EIT-system) elaborated by the authors and indicating a new way of solving the problem of much earlier socialization of children in the computer age. Section 5 outlines the principal ideas of our approach to solving the problem of much earlier socialization of children. For this, a

new graph notation is introduced in Section 4 – the notation of the maps of cognitive transformations (MCT). Each MCT describes the transformations of the cognitive-emotional sphere of the learner and of the spectrum of his/her cognitive skills achieved as a result of employing pedagogical methods.

2 THE NEED OF MUCH EARLER SOZIALIZATION OF CHILDREN IN INFORMATION SOCIETY

Let's consider a number of phenomena caused by IT that can be interpreted as negative shifts in the development of the personality in information society. Twenty or more years ago the typical consequences of a conflict between a child and his/her classmates were the use of insulting nicknames or (in some schools and some classes) the fights between the classmates. It was bad for a child involved in a conflict of the kind, but nowadays the consequences of such conflicts may be even tragic. The reason is that now the school students of even middle grades possess the well-developed skills of using e-mail, Internet, cell telephones. However, they are not socially mature and are not ready to suffer all the consequences of their deeds.

The paper [12] describes a new, tragic phenomenon of information society: cyberbullying. According to this paper, cell telephones, profile home pages and social networking service are a breeding-ground for cyber-bullying in modern Japan. The moral pressure imposed due to these technical means may become unbearable for the teenager, and, unfortunately, there are known the cases when the child decides to leave the life. The other negative deeds performed with the help of modern information communication technologies are the attacks of young hackers against computer systems of socially very significant objects and military objects.

The common reason for all considered negative deeds is that children possess very high skills of using IT but are very far from being socially mature and, as a consequence, from being aware of own social responsibility. An analogy can help to grasp the essence of such situations. A tiger-cub, while playing with a chicken, hit the chicken with his paw. The tiger-cub was not aggressive, he didn't attack the chicken, didn't want to hurt the chicken on purpose. It was just playing, the tiger-cub was not aware of the power of his paw in comparison with the weakness of the chicken, but, as a consequence, the chicken couldn't move any more.

The same happens to young children and teenagers, they can't clearly understand the severity of their actions many times multiplied by the power of modern IT. A simple negative intention is turning into a powerful tool of destroying the human's Self. And, in return, the Self of the one who becomes aware of his/her power is in danger too: he/she is not ready to suffer the consequences.

3 SHORTLY ABOUT THE SYSTEM OF EMOTIONAL-IMAGINATIVE TEACHING

The method of early forming (in the child's conceptual picture of the world) the cognitive-emotional subspace of moral values and social responsibility to be introduced in Section 5 is a part of an original system of educational methods and teaching/learning materials called the System of Emotional-Imaginative Teaching (the EIT-system).

The core of the EIT-system was elaborated by O.S. Fomichova in the first half of the 1990s and has been expanded in the second half of the 1990s and in the 2000s. The EIT-system is underpinned by the Theory of Dynamic Conceptual Mappings (the DCM-theory) developed by V.A. Fomichov and O.S. Fomichova [2 – 4, 7 - 8].

3.1 The objectives of the EIT-system

The principal goal of the EIT-system is to develop in young children and teenagers:

- the skills of processing symbolic information, the reasoning abilities;
- a mature, rich cognitive-emotional sphere, the ability to perceive and appreciate the beauty in all its manifestations, in particular, in the deeds of people [9];
- the ability to understand the peculiarities of the conceptual picture of the world of the communication partner;
- the understanding of the complex system of social agreements;
- the skill of grounding the own point of view, of participating in a dialogue;
- the feeling of belonging to a very long chain of previous and future generations as the principal cognitive precondition of sustainable development;
- proud as concerns the own connection with great national culture, the openess to the achievements both of national and world culture,

3.2 The educational program of the EIT-system

.For achieving the indicated objectives, a collection of the interrelated educational methods and an original cross-disciplinary educational program have been developed by O.S. Fomichova. The elaborated program is intended for teaching children during twelve years, where the starting age is five to six years. The program has been personally tested in Moscow with great success by O.S. Fomichova over a period of 21 years. It includes the following series of lessons: (1) a two-year course (the age of learners is 5 to 7 or 6 to 8 years) of studying foundations of reading and speaking English as a foreign language (FL), including learning basic elements of English grammar (Present Simple and Past Simple Tenses); (2) a course on understanding the language

(a part of FL) of describing the nature and feelings evoked by nature; (3) a course on understanding the symbolic language of painting; (4) a course on understanding the language of poetry (with the accent on understanding metaphors and descriptions of nature); (5) a course aimed at (a) first acquaintance with sciences and (b) developing the abilities to argument the own opinion, to raise objections, etc.; (6) a course on improving the knowledge of English grammar (during mainly the fifth year of studies); (7) the course "Foundations of secure living activity in information society". In fact, the lessons of courses (2) to (7) may interchange.

The developed program can be interpreted as a model of a system of disciplines traditionally learned in school (elementary, middle, high) and in the courses forming the humanitarian component of university education. The main thing is that this program incorporates the interacting elements belonging to different disciplines and providing the possibility to achieve in practice the goals of cognitonics as concerns the well-balanced development of the personality in information society.

4 THE MAPS OF COGNITIVE TRANSFORMATIONS

During last two decades, at least two graph notations have been introduced and employed for explicating the essence of educational processes. The vertices (or nodes) of a concept map represent the basic concepts of a studied discipline, and the edges (or arcs) correspond to the relationships between these concepts [1, 11]. The conceptual-visual dynamic schemes (CVD-schemes) are the marked oriented graphs introduced by V.A. Fomichov and O.S. Fomichova, in particular, in [2, 5] for inventing effective teaching analogies. Such graphs establish a correspondence between the components of a piece of theoretical material to be studied and the components of a well-known or just created by the teacher but bright fragment of the inner world's picture of the learner.

One of the principal goals of cognitonics is to create the cognitive-emotional preconditions of the well-balanced, harmonic development of the personality in information society. That is why we propose a new graph notation allowing for reflecting the initial and achieved states of the cognitive-emotional sphere of the learner.

By definition, *a map of cognitive transformations (MCT)* is an oriented graph with the marks of the vertices of three classes (or types). *The A-vertices* are represented by the rectangles (or blocks) with *single contour*; the texts inside these blocks describe the theoretical materials underpinning the teaching methods. The rectangles corresponding to *B-vertices* have the *double contour;* the marks inside these rectangles are the texts describing the activity at a lesson (or lessons) either of a teacher or of the students. The *C-vertices* are represented not by the rectangles but *by the ovals*; the texts inside these ovals describe the initial or achieved state of the cognitive-emotional sphere of the learner or the state of the spectrum of the learner's cognitive skills.

The Figures 1 and 2 commented in the next section are the examples of the maps of cognitive transformations.

5 A NEW METHOD OF FORMING THE COGNITIVE SUBSPACE OF MORAL VALUES AND SOCIAL RESPONSIBILITY

Taking into account the existence of the mentioned negative phenomena, the following new objectives for influencing the development of the personality in modern information society can be formulated.

It is necessary to start earlier than it is traditionally done to acquaint children with the very complex system of social agreements. Since this system is based on numerous symbols, the scholars need to pay more attention to developing symbolic information processing skills of young children and teenagers. In addition, it is necessary to early acquaint children with the idea that different people may have considerably different inner world's pictures (i.e. conceptual systems), and it is very important to take this into account while interacting with people. It is important to explain to young children and teenagers that practically every person has various connections with many other people. That is why a suffer of a classmate, etc., in fact, causes the suffer of many other people: mother, father, brothers, sisters, grand mother, grand father, etc.

The Figures 1 and 2 represent the basic components of our method. The central idea of the method is as follows. Rather often a child tries to distinguish himself/herself among his/her peers by means of emulating a bad pattern of the adults' behaviour: smoking, aggressiveness, following the formula "Might is right", etc. This applies not only to the teenagers but also to children at the age of 10 – 11. It is important to underline that the bad patterns of the adults' behaviour (drinking, etc.) most often are the consequence of misfortune, despair. Normally, children have no despair, they simply emulate the adults.

As a result of employing the stated method at lessons under the framework of the EIT-system, a child acquires (by the end of the second year of studies, at the age of 6 – 7 years) the possibility to distinguish himself/herself not by means of a deviant behavior but with the help of mature thoughts and thoughtful behavior.

The figure 2 shows that one of the important preconditinos of employing our method during the second year of studies is the well developed figurative (or metaphoric) thinking. The scheme of creating this precondition is as follows:

Reading and discussing complex texts iin English as a foreign language (FL) at the age of 5 –6

→ mastering a rich sublanguage of FL for expressing the beauty of nature and the feelings evoked by nature

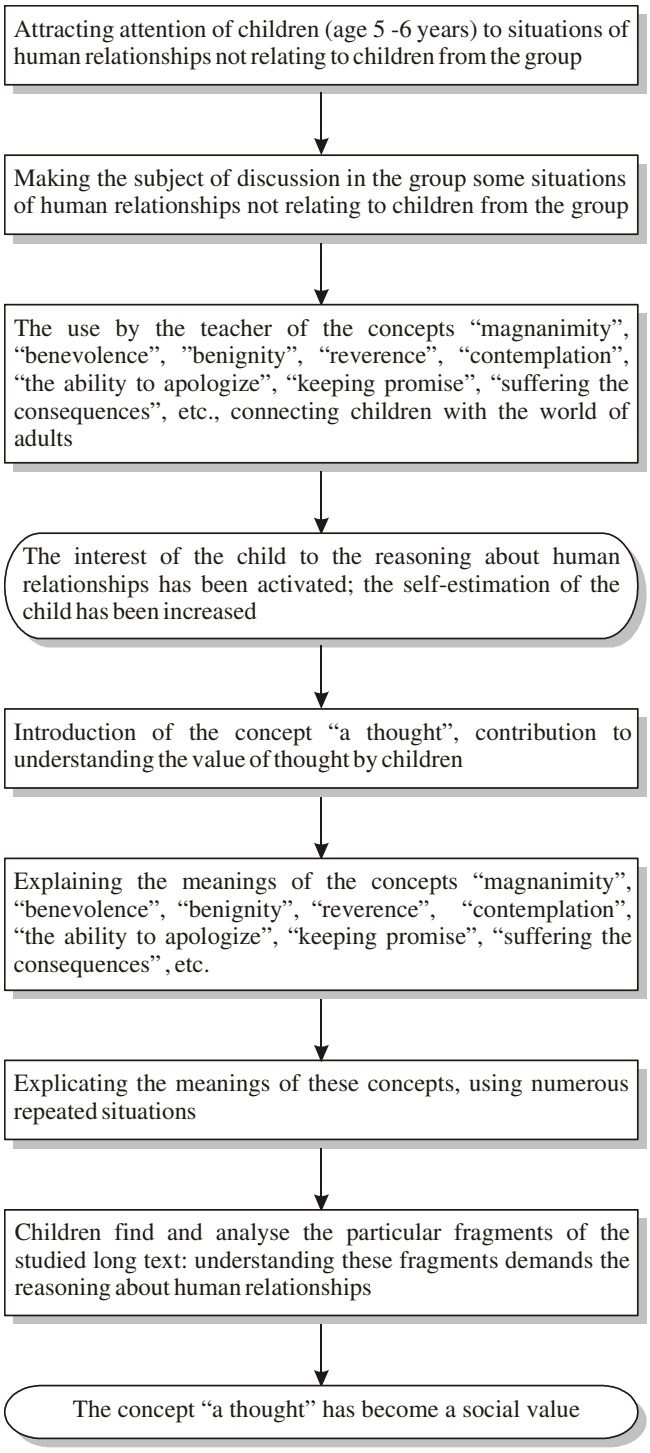


Figure 1: *First year of forming a cognitive-emotional subspace of moral values and social responsibility.*

→ development of figurative reasoning + development of the awareness of the social role of Natural Language

→ understanding poetical metaphors → creating metaphors

→ understanding the symbolic language of painting

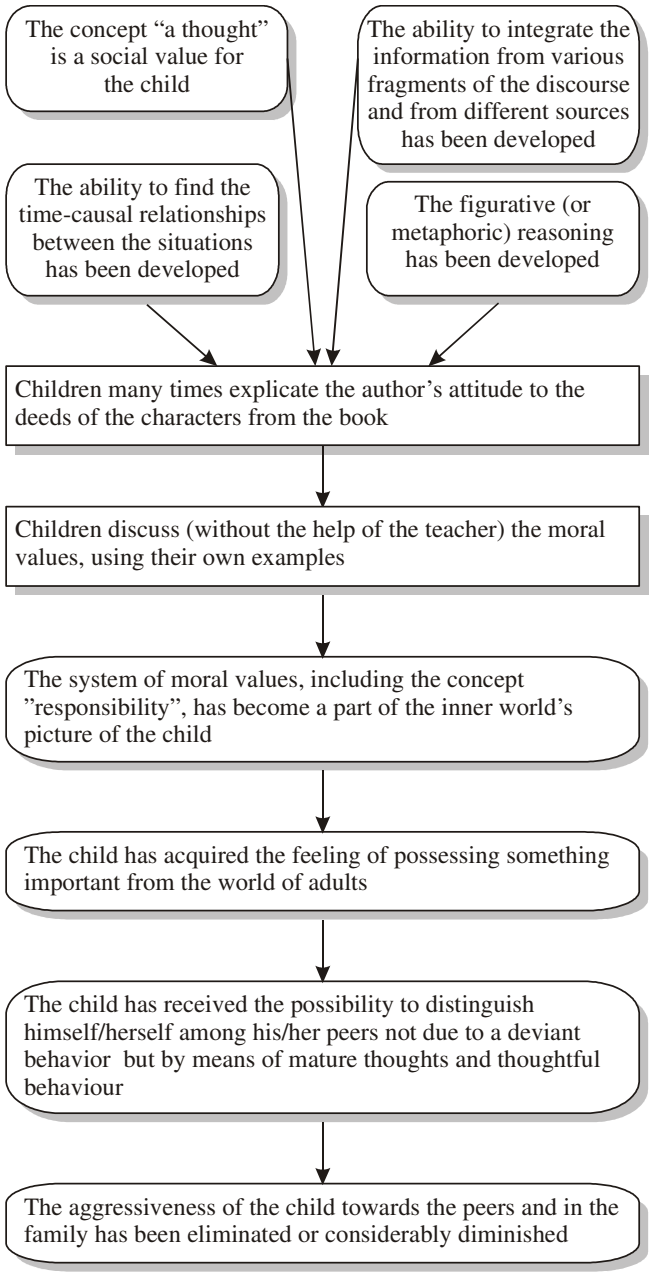


Figure 2: *Second year of forming a cognitive-emotional subspace of moral values and social responsibility.*

→ development of the ability of decoding the messages conveyed by the masterpieces

→ realization of the "Thought-Producing Self" (see [4, 5, 7, 8]) and the improvement of the feeling that a person is a link in the long chain of previous and future generations.

6 CONCLUSION

Our study started in early 1990s and including both theoretical and practical aspects has shown that the objectives of cognitonics concerning educational practice are realistic. In particular, taking into account the dangerous sides of acquainting socially unripe children (including teenagers) with modern informational technologies, we argued in this paper the necessity of looking from new positions at the problem of early children's socialization in the computer age. Our vast experience has shown that it is possible to do a lot for eliminating or considerably diminishing the aggressiveness of children, using traditional teaching materials (the fairy-tales, thrilling stories, etc.) but in a new way: extracting from these sources the attitude of the author to the deeds of the characters from the text and forming step by step the cognitive subspace of moral values and social responsibility. The essence of our approach to early socialization of the child was explicated with the help of two maps of cognitive transformations (MCT). It seems that the notation of MCT will be of help for the scholars elaborating the methods of contributing to the well-balanced, harmonic development of the personality in information society.

References

[1] D.D. Burdescu, M.C. Mihaescu, M.C. Ionascu, I. Buligiu, B. Logofatu. Expanding mental outlook with the help of concept maps. *Informatica (Slovenia),* 2010, 34 (4), p. 535-540.

[2] V. Fomichov and O. Fomichova. The Theory of Dynamic Conceptual Mappings and its significance for Education, Cognitive Science, and Artificial Intelligence. *Informatica. An Intern. Journal of Computing and Informatics (Slovenia),* 1994, 18 (2).

[3] V. Fomichov and O. Fomichova. A new theoretical and practical approach to early positive developing child's consciousness. In *R. Trappl (Editor), Cybernetics and Systems'98. Proceedings of the 14th European Meeting on Cybernetics and Systems Research. Vol. 1*, Austrian Society for Cybernetic Studies, Vienna, 1998.

[4] V. Fomichov and O. Fomichova. The social responsibility of computer science specialists for the creative potential of the young generation. *International Journal of Artificial Intelligence in Education. Special Issue on AIED 2010,* 2000, 11 (2).

[5] V. Fomichov and O. Fomichova. Cognitonics as a New Science and Its Significance for Informatics and Information Society. *Special Issue on Developing Creativity and Broad Mental Outlook in the Information Society (Guest Editor Vladimir Fomichov), Informatica. (Slovenia)*, 2006, 30 (4).

[6] V. Fomichov and O. Fomichova (Eds). Proceedings of the First International Workshop on Cognitonics, Slovenia, Ljubljana, 12 – 13 October 2009. In *Proceedings of the International Multiconference Information Society – IS 2009, Slovenia, Ljubljana, 12 – 16 October 2009, Vol. A*. Jozef Stefan Institute, 2009, pp. 427-470; available online at http://is.ijs.si/is/is2009/zborniki.asp?lang=eng.

[7] O. Fomichova. *Humanitarian Education – an Answer to the Challenge of Time*. Moscow State University Press, Moscow, 2009. In Russian.

[8] O. Fomichova and V. Fomichov. Computers and. the Thought-Producing Self of the Young Child. *The British Journal of Educational Technology,* 2000, 31 (3).

[9] O. Fomichova and V. Fomichov. Cognitonics as a New Science and Its Social Significance In the Age of Computers and Globalization. *IIAS-Transactions on Systems Research and Cybernetics*. Vol. VII, No. 2. Intern. Journal of The International Institute for Advanced Studies in Systems Research and Cybernetics. Published by The International Institute for Advanced Studies in Systems Research and Cybernetics (IIAS), Tecumseh, Ontario, Canada, 2007.

[10] O. Fomichova and V. Fomichov. Cognitonics as an Answer to the Challenge of Time. *Proceedings of the International Multiconference Information Society- IS 2009, Slovenia, Ljubljana, 12 – 16 October 2009. The Conference Kognitonika/Cognitonics*. Jozef Stefan Institute, 2009, pp. 431-434; available online at http://is.ijs.si/is/is2009/zborniki.asp?lang=eng.

[11] J.D. Novak. Learning, Creating, and Using Knowledge: Concept Maps as Facilitative Tools in Schools and Corporations. Lawrence Erlbaum Ass., Mahwah, NJ, 1998.

[12] H. Yasuda. An Information System in School for Risk Management of the Internet: Preventing Cyberbullying Without Prohibitions. *Proceedings of the International Multiconference Information Society – IS 2009, Slovenia, Ljubljana, 12 – 16 October 2009. The Conference Kognitonika/Cognitonics*. Jozef Stefan Institute, 2009, pp. 435-439; available online at http://is.ijs.si/is/is2009/zborniki.asp?lang=eng.

THE PECULIARITIES OF DICE ENGLISH COURSE METHODOLOGY FOR THE ACQUISITION OF SECOND LANGUAGE

Eloisa Le Maitre de Oliveira Lima
Graduated in Language Arts
Specialized in Early Childhood Education
Founder and Director of DICE E.C.
Master Degree in Neurolinguistics
Federal University of Rio de Janeiro UFRJ
Fonte da Saudade St., 193 – Lagoa-Rio-BR
Tel: + 5521-3215-2380
E-mail: eloisa@dice.com.br
www.dice.com.br

ABSTRACT

This paper presents DICE English Course, which stands for: DEVELOPING INTELLIGENT AND CREATIVE ENGLISH, a school in Brazil established in 1986 which proposes the acquisition of languages under the Piagetian Ontogenetic Theory of Intelligence. It attends children since 6 months of age up to teenagers (15 years old). Through the sequential evolutionist constructivism proposition, the pedagogical activities are proposed to the children during the classes. DICE is a private institution and a part time study, not a regular school for the study of other subjects. The students can enroll for 2 or 4 classes a week, the classes last 1 hour and 30 minutes each. The native tongue(s) of the students is (are) not used during the classes, since our project deals with the acquisition of English Language. It is also considered as a language immersion site where spontaneous and natural English is spoken since the student arrives at DICE. Working under tailored syllabus for each level of intelligence and language proficiency possibilities, the themes are developed and always under contexts, there is no focus on isolated words. The DICE Methodology is continuously adjusted and improved, as well as approved since the results are effective due to the comprehension and the proficiency of oral and written English which is close to native speakers of English.

1 INTRODUCTION

One of the main goals of DICE ENGLISH COURSE is to deconstruct methodological strategies used for teaching adults other languages and construct strategies appropriate for a language to be acquired by children [1].

As children are the subject to acquire other languages, it should be taken into consideration that the individuals bring from birth the brain architecture equipped with the Faculty of Language and more, the language cognition has a Critical Period to be developed. These two fundamental characteristics of the human beings constitute the preconditions for language(s) to develop in children's mind. The adults, on the contrary, are done in regard to the critical period [1, 2 3].

Didactic issues concerning the methodological background are based on the ontogenetic theory of intelligence, and from this theory we assume the stages or levels of intelligence and take them into consideration to structure our pedagogy [4].

Consequently, it is relevant to organize the groups of students, although the age is not neglected. The dosage of the English language grammatical structures, the language which is in progress to be acquired at DICE, is tailored into Syllabuses which not only consider the student's level of the comprehension of the world but also to determine the English language structures to be explored in class [4]. And more, we satisfactory plan our classes by thematic units for a pleasant context for children. The classes are planned on data based on the following syllabus and tailored to each group.

The language instructors are conducted to develop and update the method through Teachers Training Courses and weekly group of studies when the relevant matters are proposed for turning theory into practice what constitutes Dice's praxis [6 – 9].

The pedagogical materials used during the classes are properly selected according to the thematic unit which imposes a context bearing the challenge to each level of mental capacity of the configuration of each group. The technological resources are also well used although selected under same categories mentioned above and considering their proficiency in English.

The school environment is designed to input an atmosphere of English culture and real life situations. Objects of several contexts integrate the school and are disposed to be used at any need as if they were created a live dictionary [10].

2 PRECONDITIONS FOR LANGUAGE TO BE DEVELOPED

2.1 **The Critical Period**

The **critical period** is a limited time in which an event can occur, usually to result in some kind of transformation. It is a time in the early stages of an organism's life during which it displays a heightened sensitivity to certain environmental stimuli, and develops in particular ways due to experiences at this time. If the organism does not receive the appropriate stimulus during this "critical period", it may be difficult, ultimately less successful, or even impossible.These observations have led some to hypotheses about a critical period for certain areas of human learning, particularly the language acquisition [3].

2.2 The Faculty of Language

As Chomsky [2] puts it, "Evidently, development of language in the individual must involve three factors: (1) genetic endowment, which sets limits on the attainable languages, thereby making language acquisition possible; (2) external data, converted to the experience that selects one or another language within a narrow range; (3) principles not specific to FL." [FL is the faculty of language, whatever properties of the brain cause it to learn language.]

2.3 First Language acquisition

The Critical Period Hypothesis states that the first few years of life constitute the time during which language develops readily and after which (sometime between age 5 and puberty). Language acquisition is much more difficult and ultimately less successful. The critical period hypothesis was proposed by Eric Lenneberg [3].

2.4 Second language acquisition

The theory has often been extended to a critical period for second language acquisition. Older learners of a second language rarely achieve the native-like fluency that younger learners display, despite often progressing faster than children in the initial stages. David Singleton [11] states that in learning a second language, "younger = better in the long run". Younger people learning a second language typically achieve fluency more often than older learners. Older learners may be able to speak the language but will lack the native fluidity of younger learners. So (1) is Universal Grammar in the first theoretical sense, (2) is the linguistic data which the child is exposed to [11, 12].

3 THE DICE METHOD AND THE CONSTRUCTIVIST PEDAGOGY

The framework of Dice methodology merges with the Piagetian proposition for how the individuals develop the intelligence, how the knowledge is retained, enlarged and subordinates a more complex knowledge. We consider that inventions and discoveries are possible in the language acquisition process when the student is provoked into a creative insight. The conservations, which give support for changing paradigms in mind, consequently, upgrade the intelligence to face other mental maturations in life development. The endowment, the social conditions stimuli, and interaction all together trace a continuous process along life supplying new paradigms for inferring over the society. A new language in progress feeds a wide range of cultural aspects and, consequently, suggests that more intellectual possibilities enlarge the individual's developmental process [6, 8, 9].

"Education, for most people, means trying to lead the child to resemble the typical adult of his society . . . but for me and no one else, education means making creators. . . You have to make inventors, innovators—not conformists" [13].

Children develop best in a class with interaction, what is an issue of great value when language is the subject of interest in an educational process.

By contrast, in cooperative relations, power is more evenly distributed between participants so that a more symmetrical relationship emerges. The game possibilities, for instance, constitute one of the most important means for this distributive action to occur since "the game is to a child what the work is for the adult" [8]. Thus, authentic forms of intellectual exchange become possible. Children's thinking is not limited by a dominant influence, "the teacher does not teach he helps the students to learn" [6]. The reconstruction of knowledge is a favorable condition for the emergence of constructive solutions to problems. The knowledge triggers unlimited, flexible and regulated degrees of behavior by the logic of argument rather than being determined by an external authority. In short, cooperative relations provide the arena for the emergence of operations. Methods for the learning of other languages, when admitting repetition of ready statements does not support the development of intelligence, depriving and above all boring the young learners [4, 6, 9,14, 15].

When an active problem situation is proposed to students, it promotes a challenge that leads to a more complex activity coming. This novelty to come should be carefully graded since it can't be less simple than the preceding nor overcoming sequential steps on making it so complex that enables the students to solve it to succeed.

The child performs an action which has an effect on or organizes objects, and the child is able to note the characteristics of the action and its effects. "A child can only execute a task for what he is competent to" [7-9]. For different contexts and different kinds of objects, the child

is able to differentiate and integrate their elements and effects, actions that suggest a progressive possibility to a more complex attitude in a near future. As the development of cognitions continues to occur, randomly variations bring him novelties which promote, in different contexts, individual and unique situations for the usage of the language. Since very young age, acting over different kinds of objects, the child is able to differentiate and integrate its elements and effects. The needs, the desires, the willing under the interactive socialized relationship during and at any time of the immersion in the other language bring a wild range of actions and consequently, other various objects associated into reality participation by the part of the group of children. The child establishes a new level of knowledge and insight along with unexpected and striking contextualization possibilities.

The achievement of the formal thought (starting point for the adulthood) then, opens possibilities and starts to evolve a distance, by the part of the individual, from the object of studies. For this reason, according to J. Piaget [4] proposition of stages, it is not anymore mandatory to promote hands on activities for the learning of a language except the native. The grammatical issues and rules of the language are now technically useful. A comprehensive awareness of the language is necessary to generalize and combine the possible arrangements accepted by the society who masters the following idiom as its native language. All languages set defaults for instructional and academical status [4, 14, 15].

ADAPTATION, the dual process of assimilation and accommodation, susceptible, interfacing several areas of the brain, through continuous and imbalanced events experienced, at disposal, by the individuals during suggestive new stimuli forwarding new knowledge, coming to merge parts of whole greater process of the evolutional development of intelligence and, consequently any other cognition. The ADAPTATION process is mentioned to occur trigged by necessity, by curiosity, when spontaneous activity is in progress, by intentional proposition - Educational purposes [4, 16].

According to the sequential procedure in which the intelligence supports the learning and vice versa, there is evidence that any matter of interest is a possibility to motivate the students a new activation towards an accommodation to-be. And, consequently, generalized by the usage and effectiveness of what is learnt. In a school for the acquisition of a second language, as its code is an arbitarian statement it should be imposed from social interaction. To avoid repetition under same patterns constitutes a *sine qua non* condition for the learning to be internalized and in continued process. By means of ASSIMILATIVE attitudes subordinating ACCOMODATIVE ones, the language to be acquired is supported by an intelligent act to an effective learning process to-be. It should be carefully dosed and interesting to the learners' interest, not at all far from the subjects/learners. The learning is a personalized act, and each individual has individualized differences to be recognized by the teacher while teaching [4, 14-16].

As a result, the child starts to recognize still more complex patterns and to construct still more complex objects. Thus a new stage begins, which will only be completed when all the child's activity and experience have been re-organized for higher levels. This process takes a dialectical form in which each new stage is created through the further differentiation, integration, and synthesis of new structures out of the old, that the sequence of cognitive stages are logically necessary rather than simply empirically correct. Each new stage emerges only because the child can take for granted the achievements of its predecessors, and yet there are still more sophisticated forms of knowledge and action that are capable of being developed [4, 14-16].

4 TEACHING TEAM COMPETENCE FOR THE APPLICATION OF THE METHOD

The teachers at DICE participate in continuous studies of the theory about the Piagetian propositions, on linguistics, English language issues to carefully understand not only the mental stages of thought which students experience from birth to adolescence but also to be able to comprehend the possible expectations of the group being taught: their interest maintaining up high conditions. Secondly, they should be totally fluent as native or near native in all the language abilities. The theoretical support is the unique and valuable strategy to tailor class plans with a Piagetian attitude. The adequate planning sequence is absolutely necessary for the structure of contents so that the activity proposed becomes a challenge to keep up a pleasant learning. The activities should be under a perfect coherence and should not be so difficult that the child cannot resolve it, nor too easy that is foolish (repetitive), nor a charade (guessing). It must, therefore, be useful, real, contextualized (not through isolated words!) and functional, making the learning of a foreign language an instrument for a global communication. The plans are elaborated by the teacher and verified by a coordinator, the immediate professional to control the applicability of the method 6].

5 CLASS PLANNING, MATERIALS AND ACTIVITIES PROPOSED

To mind Education through intelligence implies a special labor toward the team in charge of the students. The multidisciplinary approach is another issue of importance. The structures of a language itself are not enough for CONSTRUCTIVIST educational process, thematic units are equally necessary to the method dynamic. The activities proposed should face needs, language structure

and matter. Once decided the thematic highlighted for a project development, studies of other implied disciplines should be accomplished in order to feed the context for creative class plans' challenges forwarding the mental level of capacity as well as responding for their ultimate interests. The activities are proposed by means of GAME - the play is a typical and spontaneous activity of the children. The game constitutes the most important strategies onto this methodology.

They enable to work not only the language contents in progress as the interaction among the students. While playing a game, moral implications arise; cooperation among the members of same group occurs and competitiveness with the group(s) against increases the interest to learn to win. The instructions to understand the goals of the game, the applicable rules, and the contents (vocabulary and or sentence forming, for instance) should be explored, determine and controlled by the teacher. The teacher, such as a referee, is expected to be fair, promoting an enthusiastic atmosphere and never losing the focus of the learning process to be accomplished.

The time duration of the activities can't be overlooked due to the spam of concentration at each level.

Considering the language abilities to be acquired (oral, reading and writing), the oral is privileged always and, even when the group is already able to read and later to write, the two latest ones never exceed 30% of the time class lasts.

The organization of the student, the internal balance to make awareness possible, should occur. It should be suggested and demanded by the teacher during the entire class. Note well: the conceptualization of organization in our context is not as a synonym for behaving well.

For the students to learn in a pleasant mood, in satisfactory level of attention and engagement, the dose of affectivity, the tonus sustained during the activities, must be a concern and matter of attention by the teacher during the time of the class [17].

Other activities proposed at DICE and suggested in the syllabuses we design and the didactic material implied:

the singing of songs supported by audio CDS, DVDS, I-pods, musical instruments; the listening of stories - through materials such as books, films, puppets (hand or fingers) and other sort of concrete ones: toys and real objects - story books are used not only to be told by the teachers but to practice and improve reading and writing skills. Dramatization, plays, situational stories, spontaneous conversation, debates, discussions are activities proposed to evaluate their comprehension and to motivate the student's fluency. We suggest costumes, ordinary clothes; masks or other accessories to better personalization of the character. Dice building has a multi-use salon where we share a movie theater and a theater where we also present the plays. The scenarios are designed and produced inside the school. We carefully decorate the entire school about the thematic being explored along the unit in progress to express the language with more directivity to vocabulary in study. The height of the children is considered when placing the decoration pieces on walls and other places in order that the symbolic students can take advantage as if a story happens around. These chart board elements can be handmade at DICE or, external materials can be provided for a more realistic visual impression on in three dimensional possibilities. Activities with hands-on are very much explored at DICE for live experience. These kinds of activities need a great variety of materials such as for art graphic ones. Different types of inks, special papers, wood, fabric, glues, foam, scrap, scissors for specific purposes, beads, thread, needles, pliers, seeds, cleaning things, special foods, stove, kitchen equipment. Audio and visual materials are very much used such as CDS, DVDS and other ones.

The graphic representation activities (matrixes, murals, booklets, special books) to verify their understanding of the language (drawn, read and written) are produced in our art department, also tailored according to each group needs. These materials demand a tremendous flexibility and the teachers respond for their selection and adjustments whenever necessary.

A wide range of combinations generated to variations of activities for each different materials are part of the creativity of the team while developing projects.

The computer aid is used as a complementary tool for research, preparation of tasks and self-studies. Our media center is equipped with some of recent generation, and our library gathers around 3000 books.

Integrated events such as: video club, book club, sleeping overs, garage sales, thematic parties, traditions and celebrations; cinema sessions, program in American Schools for immersion purposes are planned systematically for the contact with public in general to improve fluency, spontaneity and adequacy and polish the use of the idiom [1, 10].

6 ATTESTING THE ENGLISH LANGUAGE PROFICIENCY

As a private institution we refer the Cambridge International Exams to certify the students' competence and degree of the proficiency in the language acquired at DICE. DICE students are the best and youngest qualified students in Rio de Janeiro – Brazil. Our students achieve 100% average of approval on the three of the four exams – Key English Test, Preliminary English Test and First Certificate in English those they apply each year (see www.cie.org.uk for more details).

As a matter of fact, we do not change the appropriateness of DICE methodology to focus on the following exams. On the contrary, the efficacy of the method suits

conditions for high standing results and consequently well successful performance of the students.

7 CONCLUSION

Confine students into a classroom is not anymore acceptable at any school at any subject at any time. The families are not the specialists and they are not to dictate the rules at school what reduce their view of Education at their homes. On the other hand the reality outside the school is not conformist; it demands "*multitaskers*" and practical professionals.

Education is an open wide system, the contemporary life speed and the welfare not neglected and must come up together when children are the perspectives. That is why schools are there for and must go forward. We should consider a child born today as a part of a new generation from the ones born yesterday. Schools should considered renovation part of their system, creating the new to conquer the students' happiness and enthusiasm to come improved the next class day.

As humans are the ones gifted with speech, language is just available to us. This cognition radically differs us from other animals. The obvious is hard to be noticed. To teach languages for the ones who don´t know how to write yet is a challenge!

We understand that the students should be exposed to language usage to nurture the spontaneity and confidence for communicating with anyone at any time and circumstance. The more they speak the better they express themselves, commit fewer mistakes and are able to comprehend natives from any part of English speaking countries. Parents who think English to be fluent as first language already realized that the world is not reduced to the children's surroundings also that English is not just for leisure, or to enjoy a game on line but as an instrument for global communication.

"The baby brain is perfectly designed to what he needs to do: to learn the world" [18].

References

[1] E. Lima O. Second Phase during the Pre-Verbal Language Development Period: Sensing the Consonants, UFRJ, Br. 2009.

[2] N. Chomsky. Language and Mind (extended edition). USA. 1972.

[3] E. Lenneberg. Biological Foundations of Languages, USA. 1967.

[4] J. Piaget, B. Inhelder. The Psychology of the Child. Basic, Inc. 1969.

[5] P. Ur. A Course Language Teaching: Practice and Theory, Cambridge. 1996.

[6] L. Lima O. A Escola Secundaria Moderna. Piaget Aplicado ao Ensino Brasileiro. Br.1976.

[7] L. Lima O. Piaget Para Principiantes. Br.1980.

[8] L. Lima O., A. Lima O. Uma Escola Piagetiana. Br. 1983.

[9] L. Lima O. Piaget Sugestão aos educadores. Br. 1999.

[10] E. O. Lima. As a quisições linguísticas. In: A. O. Lima. Conversas. Rio de Janeiro, 2005.

[11] D. Singleton. Laguage Acquisition: The Age Factor, Clevedon, Philadelphia: Multilingual Matters. 1989.

[12] S. Krashen, R. Scarcella, R.,M. Long Child-Adult Differences in Second Language Acquisition. USA. 1982.

[13] J. C. Bringuier. Conversations with Jean Piaget. Chicago: University of Chicago Press. 1980.

[14] J. Piaget. Intelligence and Affectivity Annual Reviews Inc, California, USA. 1981.

[15] J. Piaget. Play, Dreams and Imitation in Childhood. W.W. Norton & Company, Inc. California, USA. 1962.

[16] M. Imhoff. Piagetian Ontogenetic Theory of Intelligence, California State College. 1962.

[17] M. Guasti T. Language Acquisition the Growth of Grammar, USA. 2004.

[18] A. Gopnik. The Philosophical Baby, USA. 2009.

PRACTICAL REASON AND PRACTICAL HUMANITIES WORKING WITH MEDICINE

Ronald Schleifer
George Lynn Cross Research Professor of English and Adjunct Professor of Medicine
University of Oklahoma
Department of English
Norman, OK 73019
USA
Tel: 011 405 325 6219; Fax 011 405 325 0831
Email: schleifer@ou.edu

ABSTRACT

This paper describes the work of Aristotle's practical reason (or *phronesis*) in the work of physicians in the patient-physician encounter. It argues that "humanistic under- standing" needs to complement biomedical science. Such understanding attends to the situated engagement and practical ethics of the work of physicians in order to conceive of both healthcare and cognitonics as pragmatic and performative worldly activities.

1 INTRODUCTION

In their essay entitled "Cognitonics as a New Science and Its Significance for Informatics and Information Society," Vladimir Fomichov and Olga Fomichova argue that

> one of the principal tasks of Cognitonics is to contribute to the progress of the scholarship of teaching and learning by means of delivering effective methods of preparing not only future teachers working with children and teenagers but also specialists in various fields aimed at enhancing the ethical component in their every-day professional activity, the responsibility for the future generations. [1: 393]

Such a task is particularly apposite in the training of physicians, particularly insofar as, for all its scientific formalism, cognitonics seeks above all to close the "increasing gap between spiritual and intellectual development of the person" [1: 388] in the information age. In many ways, the juxtaposition of "spiritual"—or what I and my colleagues have called "humanistic" [2]—understanding and data-based information is at the heart of the practice of medicine. Indeed, the use of the humanities in the training of physicians pursues the practical engagement of physicians striving to develop with their patients both concepts of and therapies to promote health, well-being, or coping with illness, even as they are trained in seeming disinterested "pure" scientific objectivism, best exemplified under the concept of "evidence-based medicine." Such evidence-based medicine pursues disinterested science in controlled blind studies. In this paper —as in my recent work—I pursue the concept of "paradigm-based medicine" in order to focus on the "medical humanities" as it has developed in recent years in the United States that delineates the "practical humanities" of *situated engagement*. The "practical reason" of my title is the standard translation of Aristotle's concept of *phronesis*, his description of the work of ethics in everyday life that strives toward situated interested action rather than abstract understanding. In the *Nicomachean Ethics* his chief examples of the phronetic skills of practical reason are navigation and medicine.

Here, then, are three things I focus upon in this paper: (1) *The situated engagement* of medical practice as it is understood through the categories of humanistic understanding. This is best understood under the category of "the chief concern" of medicine that my colleague, Dr. Jerry Vannatta, and I have been working out in relation to the practicalities of clinical practices [3]. (2) *The practical ethics* and *practical reasoning* of medicine, as it is understood through Aristotle's category of *phronesis*. It is my contention that *phronesis* itself is best understood in relation to narrative cognition—what I have called "narrative knowledge" [4]—insofar as both narrative and the practical reasoning of *phronesis* focus on the *ends* of action. And (3) *by way of conclusion, how knowledge—and, indeed, cognitonics itself—can be conceived* as *pragmatic* and *performative* insofar as knowledge and the combination of "spiritual" and "intellectual" development, at least as it functions in the physician, is always situated in the moment of the encounter of patient and physician. Knowledge so conceived and cognitonics so conceived ask us to rethink the Kantian intellectual and aesthetic categories of disinterestedness and to supplement the necessary and sufficient truths of evidence-based medicine with the necessary but not sufficient conditions of paradigm-based medicine.

2 THE SITUATED ENGAGEMENT OF MEDICINE

The situated engagement of medicine is closely related to the sense of the wholeness of the patient as a person—which

Fomichov and Fomichova call "the concept of the Thought-Producing Self" [1: 389]—and the wholeness of the story that patients necessarily bring to their physicians. By far the most common task of the physician is eliciting stories, and it has been estimated that most physicians will perform this task around 200,000 times in a career [5: vii]. That is, when a patient first meets his/her physician, the doctor creates a record that is normally referred to as the "History and Physical Exam." The record begins with the patient's "chief complaint" and then creates the "History of Present Illness," which is a narrative account of what brings the patient to the physician (which includes the chief complaint). Lately, Dr. Vannatta and I have argued [3] that along with the chief complaint, a *formal* aspect of the History and Physical Exam, should explicitly include the patient's "chief concern." That is, we have suggested that in the process of facilitating the story of the (medical) events which brought the patient to the doctor, the physician should explicitly attempt to ascertain the patient's chief concern, and record it along with the chief complaint. This process we believe will demonstrate to the patient that the doctor knows the difference between symptoms and concerns. It will also serve early in the interview to aid in the physician's engagement with the patient. Finally, it will serve as the basis for the patient-physician *negotiation*—what I later describe as their *deliberation*—concerning the ends of the medical practice, the conception and goal of what would be "good health" or whatever other outcomes that the situation allows

The chief concern Dr. Vannatta and I describe pertains to the *ends* of healthcare, what Martha Nussbaum describes as "a vague end, health," and the issue in medicine, as she says, "will be to get a more precise specification of the end itself" [6: 98], namely the meaning of "health" in relation to this particular patient. Such a concern, articulated by the patient with the facilitation of the physician, promotes the development of a practical framework for the everyday practices of medicine, particularly in relation to the patient's awareness of what his illness *means* in the context of the ongoing story of his/her life. Thus, we argue that the situated *action* of medicine *begins* with the patient's chief concern. The chief concern is ultimately practical: the patient's fear that this ailment will keep him from performing his job, her concern that her medication will deprive her family of other necessities (a particular problem in the United States), a fear of losing close relationships, an overwhelming belief that quality of life is more important than length of life, or the contrary sense that everything possible, regardless of the cost in wealth and pain, should be pursued in the face of life- or experience-destroying illness. We believe that the *formal* inclusion of the chief concern within the History and Physical Exam would make the self-conscious awareness of narrative and the mutual patient-physician *deliberation* over the end and goal of the patient's narrative—namely, a functional definition of health that grows out of the patient's situation—will make these important aspects of medical diagnosis and treatment *routine*.

With such a routine, another way of understanding the profession of medicine arises alongside the monolithic conception of the profession. This new sense understands the profession as a collection of situated practices, the very *action* that is for Aristotle the heart of ethics in its attempt to realize *eudaimonia*, in his general sense of a well-lived life and, in medicine, the realization of the supreme ends of healthcare, health understood as the alleviation of suffering, as the restoration of well-being, or as the profound gift of possibilities of simply carrying on. These definitions of health, like *eudaimonia* itself, are always entwined with successful narrative—successful both in its execution and its reception in the relationship between teller and listener—and, as I am arguing here by alluding to the *Nicomachean Ethics* (in which Aristotle develops his conception of *phronesis*), they entail ethical action in the world.

3 *PHRONESIS*: THE PRACTICAL REASONING OF MEDICINE

The situated engagement of medicine, I am suggesting, is the scene of narration, which is why the humanities—studying what takes place at scenes of narration and how they can be inflected by understanding and paradigmatic protocols—should be an essential part of medical education and medical practice. This is particularly true in the contemporary world, where—at least in the so-called developed countries of Euro-America and elsewhere—chronic conditions have significantly replaced acute diseases. Chronic conditions—diabetes, heart disease, arthritis, etc.—unlike acute conditions (such as a broken leg or dysentery) call for the institution of new norms in new situations, and such institutions are the work of situated action rather than timeless transcendental truths. It is also the work of Aristotle's virtue-based ethics, which focus on the practical *situated engagements* with experience in a manner very different from the transcendental pronouncements of normative ethics ("Do no harm") or even the transcendental measures of utilitarian ethics ("pleasure is always better than pain"). At the heart of Aristotelian ethics is his concept of *phronesis* or "practical reason." In an important and sustained engagement with Aristotle, Martha Nussbaum has noted the ways that his understanding of ethics and value emphasizes: action over logic, deliberation over pronouncement, and, with these, the performative moment over essential transcendental truths.

At the heart of Nussbaum's analysis is an argument against a mistaken translation from the *Nicomachean Ethics*. This mistake occurs in an important claim of the *Nicomachean Ethics*, Aristotle's contention in the standard translation that "we deliberate not about the end, but about the means to the end" [cited in 7: 61 n.15]. The correct translation, she contends, is "we deliberate about what pertains to the end," which "includes, as well, the further specification of what is to

count as the end" [6: 61 n. 15; in this, she is following Wiggins, who translates this passage "we deliberate (*bouleuometha*) not about ends but what is towards ends" [8: 225]). In other words, practical reason—unlike the seeming once-and-for all transcendental truths of scientific reason—doesn't assume its ends are determinately self-evident in the manner of the deontological ethics of Kant or of the utilitarian ethics of Bentham: it is precisely the work of practical reason to deliberate on the constituent features of the end or goal of action altogether—deliberation that is, by definition, essentially *performative*.

What practical reason most shares with narrative—and, I am contending, with "the performative" as well—is experience. Aristotle insists, Nussbaum argues, "that practical wisdom . . . . must be grasped with insight through experience" [7: 68]. In this account, *phronesis* functions like perception itself and it is no accident that Nussbaum entitles the chapter of *Love's Knowledge* dealing with Aristotle's practical reason, "The Discernment of Perception" and later describes "an Aristotelian perception-based morality" [7: 165]: "practical insight," she argues, "is like perceiving in the sense that it is noninferential, nondeductive; it is an ability to recognize the salient features of a complex situation" [7: 74]. The aim of perception is "that of focusing on some concrete particular, either present or absent, in such a way as to see (or otherwise perceive) it *as* something, picking out its salient features, discerning its content" [7: 77]. Ludwig Wittgenstein's related example is the "perception" of sadness: "I can only see, not hear, red or green,—but sadness I can hear as much as I can see it. Think of the expression, 'I heard a plaintive melody'. And the question is: 'Does he hear the plaint?'" [9: II,xi, p. 178]. For Wittgenstein the salient feature is at the heart of "perception" in such a way that perception itself, however immediate it feels, must be more than simple sensation. (In this example, even though the attribution of sadness might perhaps be associated with the isolatable "feature" of a minor modality of music, this attribution still leaves open the question of what makes this modality of music "sad.") This is of the utmost importance because it allows for *paradigms* of experience even when the felt experience of immediacy suggests unique and non-paradigmatic events. Such paradigms are at the heart of the humanities as a discipline—and, perhaps, at the heart of cognitonics as well—and they allow, I believe, for the humanities to teach practical arts like medicine. That is, there are "salient features" in the practices of medicine that explicitly tie together perception and practical reason even when the seasoned physician (what Aristotle calls the *phronimos*) seems simply and immediately "to perceive" what is important.

Nussbaum argues that "experience is concrete and not exhaustively summarizable in a system of rules" [7: 75], yet another way that practical reason is like narrative is the fact that it attends to rules even while it is not wholly governed by them—such rules are necessary but not sufficient—just as narrative organizes itself in relation to narrative genres and, some argue (I do as well), narrative grammars, that are necessary but not sufficient so that while narratives are not wholly governed by them, these rules allow us to understand why narratives are so universally acknowledged as such. (Even very young children recognize ill-formed narratives.) Moreover, the narratives patients bring to physicians are particularly not wholly governed by systems of rules insofar as they remain open-ended, what Alasdair MacIntyre calls "not-yet-completed" narratives [10: 223]. Frank Kermode's argument in *The Sense of an Ending* is that it is the "end" of narrative that creates its power, even when endings necessarily remain provisional [11], and it is precisely the nature of the stories that patients bring to physicians that their "end" is not clear—either as a diagnosis or as a life-episode insofar as patients bring to physicians, almost of necessity, stories without an "end": a story with a beginning, and middle, but, at best, only a hoped-for conclusion. And it is precisely the not-yet-completed narratives of patients that require the practical reason of *phronesis*. In this way, I am arguing, rules for practical reason, like rules and structures that aid us in analyzing and, indeed, apprehending, narratives, are necessary but not sufficient. For this reason, I believe, Nussbaum argues that "a general account of the techniques and procedures of good deliberation, . . . may give us necessary conditions for choosing well; it cannot by itself give sufficient conditions" [7: 93].

4 THE PRAGMATIC AND PERFORMATIVE KNOWLEDGE OF MEDICINE

Let me offer a short description of two doctors encountering a patient that I present in a recent essay to demonstrate the contrast I am suggesting between humanistic and data-based understanding (based upon Dr. Vannatta's first-person narrative).

> Mrs. Jones, an elderly woman with serious bed sores, accompanied in the hospital by her daughter, was faced with the necessity of surgery. But hearing the plan from the attending physician, she refused surgery even after the doctor informed her that, without surgery, these sores would not heal and she would die of infection. Both mother and daughter listened carefully, and the mother stated, "I can't have surgery today, because the moon is over my chest and I will die of a heart attack in surgery." Her daughter agreed and explained that she had her mother's durable power of attorney and that they both make all major decisions based on the Farmer's Almanac and the major zodiac signs. Frustrated, the physician asked them to think it over and let him know when they could schedule surgery. Every day Mrs. Jones and her daughter rejected surgery because of some problem with the moon and its phase. In the face of this seeming stubbornness, the physician confronted the patient, and

the following morning the daughter informed him that they had hired a different doctor, one whom they trusted.

The attending physician contacted the new physician by telephone, and he [agreed to] see Mrs. Jones, but only after her illness was taken care of and she left the hospital. Two days later, Mrs. Jones refused surgery again; the moon was over her abdomen, and surgery would ruin her bowels. The new doctor was at the nursing station, and the attending physician explained the situation and asked him if he would go into the room with him and talk to them. He reluctantly agreed. Upon entering the room the patient and her daughter smiled, the room immediately warmed.

"Hello, Mrs. Jones," he said and shook her hand. He then turned to the daughter and introduced himself. They were all aglow.

"I hear you need some surgery," he said enthusiastically.

"Not so sure," Mrs. Jones replied.

"Let me look," showing concern. "Yep, you sure will need surgery on this. When do you suppose we can do this?"

The patient looked nervously at the daughter. The daughter shifted in her chair, looked briefly at both doctors.

"Well, my daughter makes all my decisions, and the moon is over my abdomen now, so . . . ."

"Oh!" he responded, "you use the almanac?"

"Yes," the patient said.

"Do you plant your garden by it too? How was your garden this year?" This was followed by a three-minute colloquy of tomatoes, corn and turnip greens.

"So, the moon is over your abdomen?" The new doctor moved closer to the bed and touched her belly. "And this would mean . . . ."

"Bowel trouble," Mrs. Jones replied.

"And if it's over your head?"

"A stroke."

"Oh my!"

"And your chest?"

"Heart attack."

"Boy, then we can't do that!" the doctor replied.

He turned to the daughter and asked, "Do you have an almanac?"

"Yes," she said slowly.

Then back to the patient, "where would the moon need to be to do surgery?"

The patient looked shocked, glancing quickly around the room. "Well, I guess, uh, oh, well . . . ."

He moved back to the bed, kindly touched her lower leg.

"How about your lower leg here?"

"I guess so."

Looking at the daughter, the doctor inquired, "When is the moon over the lower legs?" Immediately he moved to the chair where she sat and helped her look it up. They studied and discussed, changed their minds a few times. It was a negotiation to behold.

"The 17th that's it," the daughter said emphatically.

"Yes, that would be a safe day. Three days from now. We will get it scheduled, and Mrs. Jones, you are going to do so well." With the patient looking surprised at her daughter, the new doctor left the room looking enthusiastically over his right shoulder and said, "I'll drop by every day and make sure of it." [4]

In this narrative, the patient presents her doctors with a not-yet-completed narrative. While the second physician understands the degree to which this narrative is *about* the patient herself, her "Thought-Producing Self" and her chief concern in relation to her condition, the first physician simply attends to what he takes to be a clear-cut medical problem and medical solution. Such solutions are sometimes complete in themselves and need neither negotiation nor consideration in larger contexts of understanding: after all, broken bones—but also bedsores—do not seem to lend themselves to elaborate negotiation. Yet because the features of the Self are necessarily not-yet-completed, the second physician seems to understand that any ailment can be the site of practical deliberation and negotiation: in medicine deliberation about "health"; more generally, deliberation about, Aristotle's *eudaimonia*. That is, human ailments are the site of the work of *phronesis*. That work, as I am suggesting and as this narrative suggests as well, is the deliberation concerning the (sufficient) constituent elements of the "ends" of the stories patients bring to physicians, stories whose shape take up and articulate the chief concern that a patient always brings with his ailment.

5 CONCLUSION

Let me conclude then with some practical suggestions for both medical pedagogy and also for a wider sense of cognitonics beyond its focus on the spiritual and intellectual development of young people. Medical education should include the use of humanistic understanding to achieve training in:

1. **listening to patients**, which entails the articulation of the "salient features" of narrative in order to develop systematic strategies that allow physicians to recognize efficiently and, hopefully, automatically the information patients bring to the patient-physician relationship;
2. **interviewing skills**, which entail systematic strategies for obtaining such information; and finally
3. **actions or virtues**, which entail systematic strategies for behaviors appropriate to the patient-physician interaction.

Such skills are necessary but not sufficient aspects of medicine—one certainly wants physicians in full possession of all the "data" of biomedical science—but their achievement would allow physicians to routinely recognize the concern and "Thought-Producing Self" of each patient they encounter, and it would help them join with their patients in negotiating the meaning of health for each particular engagement. Moreover, such features could be developed into a "checklist" [12] to become everyday part of medical practice.

Let me conclude with an observation from my recent study of popular music to give a sense of the situated knowledges of medical *phronesis* in terms of jazz improvisation and offer a working model of pedagogical cognitonics beyond school-age children. In *Modernism and Popular Music*, I cite Richard Middleton's discussion of what he calls a "theory of gesture" in relation to jazz improvisation. Following the argument of Charles Keil, Middleton notes that "Western musical analysis," by virtue of "its dependence on notation," focuses on the hierarchic formal organization of music, while, in Keil's argument, "non-notated musics . . . foreground *process*. They are much more concerned with gesture, physical feel, the immediate moment, improvisation" [cited in 13: 21]: "when a jazz saxophonist comes up with a triple forte screech," Keil writes, "is he having reed trouble or is it the climax of his solo? Only the gesture's place in the overall process can determine the answer" [14: 345]. If it were the "accidental" screech of reed trouble, it is not a gesture at all, but a simple accident. A gesture, one might say, is a complex accident, purposeful and aleatory at the same time (insofar as a screech might be unintentional and yet still purposeful), and essentially *subject* to interpretation. I am quoting this discussion of music because it suggests the performative and paradigmatic nature of understanding more generally and, perhaps, the performative and paradigmatic nature of cognitonic science. The paradigms we traffic in the humanities, I am suggesting, grow out of the relationship of gesture and process.

That is, it is no accident that Nussbaum describes the "good deliberation" achieved by *phronesis* in the action of seasoned physicians as "like theatrical or musical improvisation, where what counts is flexibility, responsiveness, and openness to the external" [7: 74] because the encounter with experience—whether it be a jazz performance, the performed narratives of anxious patients, or the performances of cognitonic science—calls for a theory of gesture that does not "label," as Keil says, pre-existing action but grasps the event of the moment as something that is part and parcel of an "overall process." Such grasping is facilitated by a sense of paradigm, as Dr. Vannatta and I use the term in describing paradigm-based medicine. The chief paradigms we have described gather the agents and genres of narrative into a system by which one can *hear* and *recognize* what is "unsaid" in the patient's story; and they also conceive of the patient-physician encounter as a *dramatic* situation, the formal drama of which can be facilitated by a sense of narrative agency and value-based ethics.

References

[1] Vladimir Fomichov and Olga Fomichova, "Cognitonics as a New Science and Its Significance for Informatics and Information Society," Informatica 30 (2006), 387-398.

[2] Jerry Vannatta, Ronald Schleifer, and Sheila Crow, *Medicine and Humanistic Understanding: The Significance of Narrative in Medical Practices*. Philadelphia: University of Pennsylvania Press, 2005. A DVD-ROM publication.

[3] Ronald Schleifer and Jerry Vannatta, "The Chief Concern of Medicine: Narrative, *Phronesis*, and the History of Present Illness," *Genre* 44 (2011): forthcoming.

[4] Ronald Schleifer, "Narrative Knowledge, *Phronesis*, and Paradigm-Based Medicine," *Narrative* 20 (2012): forthcoming.

[5] Mack Lipkin, Samuel Putnam, Aaron Lezare, *The Medical Interview: Clinical Care, Education, and Research*. New York: Springer-Verlag, 1995.

[6] Martha Nussbaum, *The Fragility of Goodness: Luck and Ethics in Greek Tragedy and Philosophy*. Cambridge University Press, 2001.

[7] Martha Nussbaum, *Love's Knowledge: Essays on Philosophy and Literature*. New York: Oxford University Press, 1990.

[8] David Wiggins, "Deliberation and Practical Reason," in Amélie Oksenberg Rorty, ed., *Essays on Aristotle's Ethics*. Berkeley: University of California Press, 1980, pp. 221-240.

[9] LudwigWittgenstein, *Philosophical Investigations*, trans. G. E. M. Anscombe. Oxford: Blackwood, 2001.

[10] Alasdair MacIntyre, *After Virtue: A Study in Moral Theory*. University of Notre Dame Press, 1984.

[11] Frank Kermode, *The Sense of an Ending: Studies in the Theory of Fiction*. New York: Oxford University Press, 1967.

[12] Atul Gawande, *The Checklist Manifesto: How to Get Things Right*. New York: Picador, 2011.

[13] Ronald Schleifer, *Modernism and Popular Music*. Cambridge University Press, 2011.

[14] Charles Keil,"Motion and Feeling through Music," *Journal of Aesthetics and Art Criticism*, 24 (1966) 337-349.

THE SOCIAL PERCEPTION OF LANDSCAPE AS A MEANS OF NATURAL, EVOLUTIONARY AND EDUCATIONAL COMMUNICATION

Rita Micarelli, Atelier of Mediterranean Landscapes – Tuscany and International Institute for Advanced in System Research and Cybernetics (IIAS)
and Giorgio Pizziolo, IIAS and Atelier of Mediterranean Landscapes, Tuscany,
Department of Urban Planning, Faculty of Architecture University of Florence,
rita.micarelli@libero.it
pizziolo@unifi.it

ABSTRACT

A progressive exasperation affects the relation *men/society/environment* and puts at risk its very evolutionary potentialities. At the same time the increasing influence on *natural world* by the *virtual globalized powers* accents and exasperates their separation and put at risk their reciprocal communications. A *new way of learning* is needed, able to re-constitute experiential modalities and encourage the reciprocal integration between these two conditions. A new idea of the *life environment* –virtual and natural- can be suggested towards a concrete and socially practicable *learning*, based on a multiplicity of landscaping research-action, able to bring to light unexpected *evolutionary learning contexts,* where Landscape becomes the interconnection's hinge between *territory/society Entities* and *virtual /experimental Minds*.

1 LANDSCAPE AND SOCIAL LEARNING: TWO HOLISTIC PHENOMENA

The official signature of the European Landscape Convention (ELC- Firenze, 2000) signaled a switch in the traditional cultural, scientific, and social common idea of the landscape.
About this innovative idea we could emphasize the social perception of landscape as a fundamental step to re-open new relationships between individuals/societies and their *own landscapes* conceivable as *Life Environments*.
In this new conception the Life Environments become the places of the social experience, and consequently the places to which everyone -in the course of his daily life paths– refers his perceptions, emotions, latent desires, and instinctive critical or proposal valuations.
This spontaneous *inaccessible phenomenon* of perception is widely experienced from the individual to the social level and –for this reason – it become a fundamental *fount of information* autonomously and continuously produced among the people. Such an 'information fountain' could constitute a very social opportunity to challenge the pervading authoritarian given information, hierarchally imposed by the globalized world.
To play this concrete challenge it is fundamental to transform these instinctive inaccessible activities in a very social *welcome learning* towards adult and child people. We called this kind of understanding a 'friendly reciprocal learning procedure'. Such a procedure can be practiced among different social groups throughout lasting experiential processes referred to different life environments.
These modalities become effective only by a very 'equal full immersion' in a living context where expert, adult people and children interact to re-discover their life environments as the places where they can re-cognize themselves as an unexpected Holistic Subject able to play a role of contemporary *Genius Loci.*

2 TENSIONS, EXASPERATION AND SOCIAL REACTIONS TOWARDS THE GENERAL AUTHORITARIAN TRENDS

The informational phenomena which influence the contemporary conditions become more and more uncontrollable by the populations, while the enormous gaps between them and the dominant powers are going up and up every day.
Despite this pervasive tendency a growing awareness of this degeneration is raising from the *social humus* which is still -unexpectedly- fertile and creative. Two kinds of *spontaneous creative ways to react* emerge from this *humus* :

- the renewed *social awareness* of the living contexts (in terms of cultural roots, sense of belonging, emotion, knowledge, economy, and so on)
- the *appropriation of the net work information system,* free, independent and self manageable by people, extendable at the planetary scale beyond all territorial constraints.

Actually these two ways can become exasperated without reaching their common desires of independence, freedom and autonomy, which should become practicable in terms of evolutionary perspectives.

About the first way we notice that the local territorial awareness manifested within specific contexts (human historic settlements, landscapes, traditional managements of local resources, local economies) could emphasize some latent tendency toward new isolation which -even if

positively practiced – could lead them to new forms of exasperate egoistic 'social identity'.

About the second way we could notice that the creative reaction has been propagated at the planetary scale by a *virtual information and by a free inter exchange system*, already used by a number of individuals and social groups. This reaction, even if extremely positive, could propagate a further illusion among the *people of the users*, bringing their ephemeral attention toward a superficial knowledge of the contemporary phenomena. Actually a lot of the users can be attracted and suddenly involved in this system, without reaching an adequate awareness of their real subordination towards the dominant powers and their imminent deprivations in terms of social, economical and cultural autonomy. In other words, this way is not able to stimulate processes of cyclical reflections and appropriated reactions which could enable this people to go beyond the mere indignation or the violent protests.

In other words, *if* the *first way s*hould tend to a wider relational perspective (without losing its deepness), the *second* should reach a deeper relational dimension, keeping its unlimited horizons. A different interactive communication system could be helpful to involve the users, the local communities, the virtual environments and the real life environments in an wider interexchange system.

In both cases adequate relationships between all of the subjects above mentioned are needed to go beyond their respective intrinsic limitations. In this interactive relational perspective,
- *on one side* the informational content of internet- exchange could generate learning processes -based on cyclical reflections- and promote concrete experiences among the users related to a variety of life environments and reciprocally linked by the planetary network;
- *on the other side* the local groups could open up their cyclical activities towards wider external life environment contexts to implement their relationships without losing their original characters and cultural roots.
The two modalities could be reciprocally influenced and integrated, through the creation of a new *common landscape* – a wider relational entity, *virtual and real* at the same time. They could move like rhizomes (dynamic and rooted in the soil, free but relationally linked at the same time) to approach a *new learning dynamics*, like the one showed by the evolutionary phenomena.

How to promote this *learning dynamics* in our contemporary disintegrated conditions?
To answer these urgent questions we have been promoting different experiential learning processes, based on the relationships *Individual/social/ life environment*, the fundamental *evolutionary structure* which could lead us toward a renewed *Ecology of Mind and Nature* (as defined by G. Bateson).
Such extraordinary epistemological suggestions led us to experience various modalities of this learning, developed in a relational way between life environments and social contexts, and recently extended to new virtual/real landscapes referred to the wider contemporary realities.

How to address these learning experiences to the *social /life environment contexts*?
How to consider the involvement of the child in these activities?
All of the above described conditions have been progressively accented in the recent years. Actually the young contemporary generations risk of losing the perceptive experiences towards their life environments together with their spontaneous social relationships. Hence, they tend to emphasize only the virtual
-instantaneous but superficial- interactions. The traditional educational procedures can't cope with this unstoppable, frenzy phenomenon. Nevertheless we could think that some attracting landscape experience would be welcome even among these *absent placeless groups*.
Same study- examples, already texted and in course, can be helpful to understand the meaning and the *methodological effectiveness* towards the adult and child groups, the living contexts and institutions progressively involved. Different social groups and living contexts in North/Central Italy participated in such activities along the last twenty years. Their processes can be valued under multiple points of view (procedures, cyclical balances, evolutionary trends) referred to the re-conquered self consciousness by the groups through the re-discovery of their life environments.

These researches, selected among a lot of study cases are distinguishable in two groups, listed as following :

1. **Landscape- art, friendly participation, social awareness, didactic experiences**
2. **Explorations among users and inhabitants of contemporary virtual landscapes and real life environments, encouraging their participative meetings**

3 THE FIRST GROUP

1-Landscape- art, friendly participation, social awareness, didactic experiences

This first group includes three experiences developed on various territorial contexts with several different social groups (adults, children) invited and spontaneously disposed to these activities, i*nvolved in several creative processes* concerning *social perception of landscape*. A variety of friendly meetings, spontaneous conversations, and promenades, guided by people throughout their life spaces, have been experienced. In general the knowledge and the perception have ranged from the widest landscapes to the smallest public spaces selected in loco. Such activities have been developed as *process- experiences*, progressively and chorally implemented by experts and people (meeting, discussing and learning together) in a pleasant atmosphere, in a very 'friendly learning'. All of these experiences allowed experts and people to achieve a new appreciation of their own life environment, and some time they opened up

also new opportunities of social creative participation in its transformations (beyond any traditional political-planning or project methods).

THE EXPERIENCES

1a - ***A first methodological approach of "Social Perception of the Landscape*** **in the experience of Suvereto. Province of Livorno, South Tuscany (1998)**[1]

The place: Suvereto Village, south Tuscany, Piombino - Livorno

The promoters: University of Florence, :National Research Council

The research aims: Relationships between the population and the territory, working through the Landscape in a participative planning perspective, beyond the landscape classic mosaics to a Landscape holistic social perception and evaluation

The participants: scientific consultants, social, scholar and institutional groups, all directly involved to obtain answers to our questions.

The procedure we have followed:

- we involved every subject in the knowledge process; accepted as a very intriguing play by the participants
- we focussed the discussions, on "landscape ambits", with the explicit intent of recognise those ambits on the base of the differences among them, in terms of living experiences.
- then we asked a personal evaluation to the subjects that participated in the experiment, through different modalities (drawing and description for the elementary school children, proper and more specific questions for the junior high school students, and articulated questions for adults who had been provided with direct answers in "classroom meetings").
- we examined the collected material, working out different hypothesis of interpretation.

Finally we have promoted a new meeting to discuss with the people who had been interviewed the results of our research and the validity or the lack of validity of the suggested hypothesis, and we accomplish a balance of the whole operation and its possible developments.

The quality of the perception and the valuations that everybody was showing out, and expressed by different languages:

The **child** expressed their critical opinions or positive appreciations throughout surveys on the places, informal colloquies, sketches, coloured drawings, discussed and valued by the scholar groups together with the experts and their teachers, as a very didactic activity.

The **eldest group** pointed out some landscape marks invisible to others, which they perceived as their own, yet socially valuable and shared.

The **young people**, studying or working elsewhere, remained deeply rooted and strongly related to their original life environment, recognised as a sort of privilege, to respect and save jealously, as a firm point in their lives.

In general **all of the participants** manifested their positive attitudes towards the common life environment by a concrete sense of responsibility.

1b – The Vezzano's participative landscape experiences :villages, agriculture, waters and… ***enjoyed panoramas*** **!(1998-2002). La Spezia province, North Italy**

A synthetic description

Vezzano is a small village of the Liguria's Rising Sun Coast, at the confluence of Vara and Magra, and overlooking the more distant gulf of La Spezia, the Apennine roads, and the Apuane Alps. In ancient times this village, exchange and sentinel crossroad for goods, abundantly rich in fruits, wines, oil, gardens, was inhabited by noble families and by farmers. The village is a blaze of colors, blue, green, and grays among waters, stones and the sky. But nowadays Vezzano is risking to be choked with highway junctions, the container storage areas of La Spezia harbor, the dried areas of the polluted rivers, just like the outskirts of La Spezia, city that in the last thirty years has been piling up poisons, traffic jams, and pollution, without getting going toward the wonderful and great destinies of the Gulf.

A very intelligent reaction towards these increasing diseases has been manifested by the populations throughout *creative landscape research-actions*, expressed by various combined activities (scholar groups, –pro loco- associations, experts and public bodies). A choral participated rediscover of their life environment led all the groups to appreciate their "landscape privilege" in a perspective of maintenance and promotion of *their* more significant places, itineraries, local histories, and traditions. Specific activities have been promoted and developed in this sense and also new interactive collaborations involved different groups, while interesting proposals in terms of self promotion and management of this Life Environments came to light.

1c -Art, Sociality, and Science walk in a Mutual and Friendly Learning process -Pescia, North Tuscany, a participative research-action

This activity became an opportunity to experiment a new method of learning, through a cognitive and creative process focussed on the relationships among inhabitants and their life environment

The authentic friendly learning process has been progressively built up by progressive friendly meetings between experts and motivated groups of people to develop *new relations with their life environment.*

In this process each group expressed its own valuations, desires and options (social problems, urban renovations, a

[1] The coordinator of this research is Giorgio Pizziolo. Rita Micarelli and Paola Paradisi have taken part to the project experience as well, the first one for the social ecology, and the latter for the psychosocial investigation.

developing program, political disagreements, special needs.) and discussed about (new schedule, suggestions, proposals, further reflections, and theme choices for the creative participation to the transformation of one's own context) towards further developments for the process, in cyclical verifications and creative shared perspectives.
In particular the relationships between School and Territory encouraged *new participative landscape interactions*, experimented and developed through participative procedure, appropriate to the anthropological, social, and territorial condition of Pescia, a small town so much differentiated in its varied territorial and social "zones". Such a procedure will allow us, together with the "vertical" experiences of the social/place relations, to get to a program of **re-qualification of the life environment of relational kind**, by developing the great proposal qualities contained in the "**relational systems**", keeping in mind that the program experimented in Pescia will certainly become the first stimulating pilot experience.
N. B.: in the course of the presentation the slides will be showed

4 THE SECOND GROUP

2 Explorations among users and inhabitants of contemporary virtual landscapes and real life environments, encouraging their participative meetings

This second group of experiences began only *recently* and at the present is in progress. We are trying to involve some group of young *web professional and users* of the informatics networks, often (even if latently) conscious of the intrinsic limits of such so potent instrument towards their personal concrete experiential potentialities.
The hypothesis is to create new concrete opportunities of reciprocal exchanges between *widespread virtual communications* and *cultural life environment phenomena*.
According to this initial hypothesis we are proposing '*new contacts' modalities* which affect at the same time individuals and groups, together with their landscape references (both in concrete life environment terms and in *virtual landscape* terms). All contacts must be focussed on *thematic meetings* and concrete *lasting process perspectives*.
These meetings can produce different interesting results both towards *young and child people*, wide-ranging from didactic experiential level to real opportunities in terms of social knowledge and sustainable economies based on the initially characterized landscapes.

A perspective for the second group: procedural and methodological implementation hypotheses

2a -The integrated economies of the Commons
2b- The micro innovative or restored activities towards mountain economies
2c- The creation of new social managements of urban life environments as cultural concrete opportunities to go beyond the social exasperate protests and legal claims
2d- The U.F.O. at Portovenere, a child unexpected Landscaping vision

Virtual and real contexts: how to experience their evolutionary meetings ?
The contexts where such meetings could be promoted by a relational extended idea of Common Good (from the classic Commons to the Informatics' Networks (as suggested by E. Ostrom). In this way reciprocal wider exchanges, not generically approached but focussed in *thematic terms* on information, didactics, scholar experiences, professional collaborations, experiential procedures, comparability, mutual specific supports, autonomous economies. They could become the new structure to involve persons and groups coming from different conditions (cultures, lifestyles, economies, life environments) brought in contact by such integrated (virtual and concrete) holistic modalities.
The European Convention remains a basic reference for all these new activities and for the promotion of a new way to deal with the life environment for the involved populations. Hence no more classic methods no more models to apply, but rather different criteria and wider approaches, sensitive and participated observation of the phenomena, informational network's interexchange, could be helpful to constitute unusual *real and non hierarchic virtual groups*-experts, citizens and users who work together in a new, *extended life environment* **.**
In such environment experiments and researches could be constantly and mutually compared, in an open cyclic progression. In this perspective the virtual and the living contexts (with their users and inhabitants) could be widely referable to the relational *Mind and Nature Principles (*G. Bateson) in a new evolutionary wholeness, as a new unexpected **landscape,** where the Nature, the Mind and the non hierarchic network information can equally interact.

Acknowledgements

Our epistemological basic references and acknowledgements have to be addressed to G.Bateson's (Mind and nature, A sacred Unity); E. Ostrom (Governing the Commons); V. Mancini and W. Bandler (social perception and Mathematics of Relations).

Special acknowledgements are dedicated to the friend, members and professors of IIAS, The International Institute for Advanced Studies in System Research and Cybernetics (G. Lasker, A. Aydin K. Hivaki, K. Boullart,, H. Schinzel, J.R.L. Chandler. M. Malatesta, A .Grappone, S. Santoli, V. Fomichov and O. Fomichova). With them we have been discussing our hypotheses during the IIAS InterSymp international scientific conferences which are annually held in Baden-Baden (Germany).

COMPUTER-AIDED EDUCATIONAL INTERVENTION IN TEENAGERS THROUGH THEIR PEER'S SUBCULTURE

Maria Guadalupe Velázquez-Guzman & Felipe Lara-Rosano***
*Universidad Pedagógica Nacional **Universidad Nacional Autónoma de México
México
e-mail: gvelazg@gmail.com

ABSTRACT

In this paper we propose educational strategies based on computer-aided communication-oriented Internet social networking sites as Facebook and MySpace and the Complexity Approach to Human Systems for enhancing the social development of the students, taking into account their peer's subculture.

Keywords: Peer's groups, educational intervention, school practices

1 INTRODUCTION

The school, as educational institution, is the essential nucleus for the formation of future citizens and social actors [21]. In fact, it is the place of meeting of diverse individualities and social activities where values are affirmed, as well as the forms of being and coexisting.

However in our contemporary societies, from the least developed to the so called "first world" countries, we find now a social environment of great violence and brutality that influences in a negative way the development of the young people, driving them toward an aggressive cruel behavior.

Therefore the educational institutions have the great challenge and the commitment not only to teach mathematics, science and language. They also must contribute to the reconstruction of the social fabric by forming the new citizens [11]. These new citizens should be critical and creative social actors that participate in the construction of better ways of coexistence. According to Stacey [17], when people communicate with each other to accomplish the joint action of living and acting together, they are continuously relating to each other in a responsive manner. Marion [7] asks what catalyses social behavior and concludes that humans are said to cooperate because that is the best way of achieving individual goals. In addition, individual human action is catalyzed by symbols: the ideas, concepts, opinions, beliefs, emotions, projections, and values like the understanding of the personal differences, the cooperation for the well of the collective whole, the search for justness and equity and the practice of tolerance.

In this paper we identify in the actual school practice the obstacles and facilitators for the individual and social development of the students for living together and propose educational strategies based on computer-aided communication-oriented Internet social networking sites as Facebook and MySpace and the Complexity Approach to Human Systems for improving the effectiveness of the implemented school practices.

2. THE SCHOOL SOCIAL DEVELOPMENT OBJECTIVES FOR LIVING TOGETHER

Let us consider a school group community as a complex system, composed by the teacher and the students peer group, defined as a collective of human agents that interact with one another in a common internal environment defined by the school rules and the group interactions. According to Mead [9] it is in the detailed interaction between people, their ongoing choices and actions in their relating to each other, that their minds and selves arise.

However, the schools are located in neighborhoods where are seated families that belong to very diverse socioeconomic strata. The socio-cultural differences of individuals and families present a diversity of values, beliefs, lifestyles, and behavior rules that are part of a plural society. This diversity is reflected in the students peer group. In accordance with this cultural and economic social heterogeneity, the objective of forming citizens that have the sensibility of participating with the other ones in the construction of a living together community, it is a work that is one of the main challenges of the education. In fact, the school must form citizens to be able:

a) To understand and accept the principles and forms to live together in a heterogeneous community.
b) To understand the common benefits of collaboration and mutual help.
c) To balance the respect to the individuality with the collective good.
d) To recognize and respect the personal differences.

Evaluating in the actual school practice the accomplishment of these objectives, we found [22] the following conflicts in the practice implementation that hinder the educational targets:

- The prevalence in the school of the control, surveillance and punishment measures. McGregor [8] distinguishes between autocratic and participative styles of group management. He argues that autocratic styles are based on the view that people dislike work and avoid it if they could. They need to be coerced into working, being punished if they do not work and rewarded if they do.

This is the "stick and the carrot" approach that is still so prevalent in schools.

- The absence of dialogue between teachers and students. It is important that the rules are the result of a practical ethical knowledge. Therefore, they should contemplate the rights of everybody [21].

The previous results let us pose the following questions:
Which educational strategies could be implemented in the school for developing effectively the future citizens?
Is it possible a different organization from the current conventional forms?
Is it possible that the teachers as educational agents go so far as to modify their behaviors in order to become better builders of future citizens?

3 PROPOSALS FOR A NEW EDUCATIONAL INTERVENTION PARADIGM

Taking into account that the maintenance of a relationship is dependent upon the members each striving to reach an acceptable balance between their own desires and needs and that of others, according to the approach of Relational Dialectics [1], [13] we developed the following proposals [22]:
1) That the dialogue and the agreements be the means to solve novel situations. Following the perspective of Gadamer [4] about the community sense, the practical knowledge that arises through the contrasting experience is a phrónesis, because it is knowledge to distinguish what is good and what is bad for the community. This moral knowledge subsumes the individual preferences to the communality correct, making familiar to the students a community sense that before was strange. These dialogue and agreements may be easily implemented through Internet mediated social networks discussions.
2) That teachers and pupils, by means of the dialogue and the argumentation, participate in the making and upgrading of the rules, defining collectively the principles, rights and responsibilities of the members of the community to make possible the coexistence for the common good, nevertheless the plurality and the personal differences. In fact, 59% of students with access to the Internet [19] report that they use social networking sites to discuss educational topics including career and college planning, and 50% use the sites to talk about school assignments. Some parents and teachers say that using these sites helps students improve their reading, writing, and conflict resolution skills and learn to express themselves more clearly.
3) That the values that strengthen the community sense like: friendship, respect, participation, solidarity, common good and the feeling of belonging to a community are encouraged. Under the surface of the social, individual resonances harmonize in the sense that people develop a shared view. Each individual is like an attractor and when individuals come together to form a group, they resonate with each other, producing through their communication a social attractor [7]. Prigogine [12] casts his theories in terms of entities resonating with each other and evolving as collective ensembles. Therefore a strategic policy enhancing the above values through the use of Internet based social networks may result in the development of social attractors that reproduce themselves changing the social pattern.
4) That the rules take into account the diverse needs for the members and their interactions. Langton [6] talks about the inability to provide a global rule, for changes in a complex system's global state, making it necessary to concentrate on the interactions occurring at a local level between agents in the system, like those of social networks. He states that it is the logical structure of the interactions, rather than the properties of the agents themselves, which is important, thus potentially elevating interaction to primacy.
These proposals are further supported taking into account the organizational power of adolescent peer's groups.

4 ORGANIZATIONAL POWER OF ADOLESCENT PEER'S GROUPS

In the school adolescents, nevertheless the differences, have a strong tendency to constitute peer's groups where adults are left out. Handel [5], affirms that the peer's groups are complex organizations that perform a socialization function among their members. In these organizations the members establish goals and make agreements directed to certain matters, interests and concerns. These agreements must follow certain procedures.
The communication-oriented Internet sites such as social networking as Facebook and MySpace and photos and video sharing sites such as YouTube have emerged as new media very popular among adolescents and have shown a big impact on the organization of adolescents peer's groups. The challenge for parents and schools alike is to eliminate the negative uses of electronic media while preserving their significant contributions to education and social connection [18].
Richard Savin [14] sustains that from a perspective of the human psychological development, it is important that inside a net of interactions (like Facebook and MySpace) the individual finds a place for itself, a status. In this net there are necessarily a pattern of power relationships based on the different capacities and personal abilities, being this pattern a constant feature in the structure of human relationships.
The peers derive their standards to think and to judge, their individual references, behaviors and values from their group. According to Corsaro, many activities of peer's interaction look for a sense of self and opposition to the adult world [2]. The peer's group defines types of relationship that are qualitatively different from the relationships with the figures of authority. Some of these new types of relationships are reciprocity (not possible in the relationships with the adults), identity, belonging, solidarity and support. The peers feel that they participate at the same level in decision-making and the establishment of

norms for the group. Certainly the group rewards its members, but it also punishes them with distancing, ostracism and other disapproval expressions.

5 THE ADOLESCENTS GOALS AND THEIR WORLD'S REFERENCE FRAME

It is important to recognize that the adolescents possess natural organizational powers that it is important to guide with the educational intervention for their own social development. Each individual is like an attractor and when individuals come together to form a group, they resonate with each other, producing through their communication a social attractor [22], [7].

The word potency refers to the characteristic quality of the human being to advance toward objectives when the person is endowed with self-reliance and is able to accomplish her/his objective [3]. However, potency is not enough because she/he needs a reference frame about the natural and social world, that should be structured and have an internal coherence, where she/he should find her/his place. Without this framework she/he will get confused and will be unable to act upon in a proper way. Savin's reflection is similar [14], since he refers to the importance that the individual finds a suitable place inside a net of interactions.

However, a reference frame is not enough to guide the adolescent action, because the adolescent also needs a set of objectives as the focal point of her/his actions and the expression of her/his effective values. This necessity to have defined objectives responds to an existential primary need that demands satisfaction, without caring the means.

The contributions of Fromm [3] and Savin [14] are the base to understand the adolescents' auto-organization, the individual necessity to belong to a peer's group and the origin of the leadership that some of them impose over the organization. However, when these necessities are ignored then the adolescents in their social diversity are not properly guided. They end up forming contesting groups mainly with tendencies of untamed dominating power.

Sherif and Sherif [15] suggest that the individuals must be guided toward humanist and not destructive goals, because their nature is malleable, notwithstanding that by nature they differ in capacity, abilities, resources and experiences. It is starting from this differentiation that they can contribute to the solution of their problems. The uniformity is not a source of strength in a group.

What we want to emphasize is that in the adolescent peer's groups we can find the organizational potency, the search of goals, the values of reciprocity, solidarity and belonging. However without an educational guide from the school this emerging social organization can take any form.

6 ORGANIZATIONAL ELEMENTS FOR AN INTERVENTION PROPOSAL IN THE SCHOOL

The school has a functional structure that could be used to guide the suggested peer's group policy [23]. The school has teaching and learning objectives, collective projects and practices, implicit socialization actions in the teacher-student relationships, behavior rules and sanctions. All these organizational elements should be used to implement an integral adolescent socialization program through communication-oriented Internet social networks sites like Facebook, MySpace and Twetter. Examples of these collective projects and practices to implement an integrative socialization of the students through the social network media are the following.

1) The individual contribution of the student to the work in class (presentations, discussions, questioning, inquiring)
2) The organization of educational games and sport teams.
3) The organization of the working environment in the classroom.

These community activities could be the means for the students to become responsible social actors able to contribute for the social coexistence and collective good [23]. Individuals join groups, motivated by internal values attractors, creating a collective whole that is more than the sum of their individualities. If some of these internal values are related with living together in a heterogeneous community with a balance between the respect to the individuality and the collective good and participating in the definition of the collective principles, rights and responsibilities of the members of the community then this peer's group will affect the way of how their members evolve [22] improving the coexistance with others peer's groups.

7 CONCLUSION

We think that the values that strengthen the community sense like: friendship, respect, participation, solidarity, common good and the feeling of belonging to a community must be encouraged. The peer's community is the first kind of community encountered by the teenager when she/he tries to be independent from her/his family, where she/he shares values, objectives, and interests. When individuals come together to form a group, they resonate with each other, producing through their communication a social attractor and a social hierarchy characterized by a leadership structure. In fact, under the surface of the social, individual resonances harmonize in the sense that people develop a shared view, notwithstanding that in this organization there is a hierarchy, rules to be followed and sanctions to be assumed provided that the members are accepted in the group. Moreover, in the same school coexist different opposing subgroups that try to differentiate because of their strength, age, experience, abilities, and daring. Therefore a strategic policy for educational intervention should aim to transform the rivalry among these subgroups into a cooperative force with common interests at a higher level through computer-aided communication. This would enhance the values in the peer's subculture toward the development of social attractors that would change the social pattern toward a cooperative one at a school level. This transformation could be implemented

through computer-aided intervention strategies designed and implemented to make use of communication-oriented Internet social networks sites like Facebook, MySpace and Twitter.

Acknowledgement: This work was supported in part by the project PAPIIT IN105909, Universidad Nacional Autonoma de Mexico.

References

[1] Baxter LA & Montgomery BM. (1996). *Relating: Dialogues and Dialectics* New York: Guilford Press.

[2] Corsaro, W. (1985) *Friendship and Peer Culture in the Early Years*. Norwood, NJ: Ablex Publishing Corporation

[3] Fromm E. (2006) *Anatomia de la destructividad humana. México, Siglo XXI* **.** 20 ed.

[4] Gadamer, H. G. (1990). *Wahrheit und Methode I.* Tübingen: J.C.B. Mohr (Paul Siebeck).

[5] Handel, G. et al. (2008) *Children and Society*. Los Angeles, California: Roxbury Publishing Co.

[6] Langton C. (1989). "Artificial Life" in Nadel L & Stein D. (eds) *Lectures in Complex Systems.* Reading, Mass: Addison-Wesley.

[7] Marion M. (1995). *The Edge of Organization: Chaos and Complexity Theories of Formal Social Systems.* Thousand Oaks, CA: Sage Publications Inc.

[8] McGregor D. (1960). *The Human Side of Management.* New York: McGraw-Hill.

[9] Mead GH. (1934). *Mind, Self and Society*. Chicago: University of Chicago Press.

[10] National School Boards Association (2007). "Creating and Connecting: Research and Guidelines on Online Social –and Educational- Networking", NSBA.org, July.

[11] Ortega P. (2006). "Educar para convivir" in *Educación, Valores y Desarrollo Moral Vol II*. A. Hirsch-Adler (ed) Mexico:Gernika pp 89-117.

[12] Prigogine I (1997). *The End of Certainty: Time, Chaos and the New Laws of Nature*. New York: The Free Press.

[13] Rawlins WK & Holl M. (1988). "Adolescents' Interactions with Parents and Friends: Dialectics of Temporal Perspective and Evaluation," *Journal of Social and Personal Relationships,* Vol 5, pp 27-46.

[14] Savin, R. (1976) "An Ethological Study of Dominance Formation and Maintenance in a Group of Human Adolescents" in Muuss, R *Adolescent Behavior and Society.* New York: Random Hansom 1980.

[15] Sherif, M & Sherif, C. (1964) *Problems of Youth*, Chicago: Aldine Pub. Co.

[16] Stacey RD, Griffin D & Shaw P. (2000). Complexity and Management. London: Routledge.

[17] Stacey RD. (2001). *Complex Responsive Processes in Organizations*. London: Routledge.

[18] Subrahmanyam K & Greenfield P. (2008). "Online Communication and Adolescent Relationships". *The Future of Children,* Vol 18 No 1, Spring, pp 119 - 146

[19] Velázquez-Guzmán MG & Escobedo CI. (2008). *Agresores, Agredidos y Mediadores.* México: Universidad Pedagógica Nacional.

[20] Velázquez-Guzmán MG & Lara-Rosano F. (2007). "The Complexity of Building up Communitarian Coexistence in the School: A Hermeneutical Approach", Acta Systemica, vol 7, no 2, pp 17-24.

[21] Velázquez-Guzmán MG & Lara-Rosano F. (2009). "Social Development for Living Together: School Practice Diagnosis and Systemic Proposals", ", in *Personal and Spiritual Development in the World of Cultural Diversity Vol VI,* GE Lasker & K Hiwaki (eds). Windsor, CANADA: IIAS, pp 31-36.

[22] Velázquez-Guzmán MG & Lara-Rosano F. (2010) "Teenagers Educational Intervention through the Peer's Subculture", in *Personal and Spiritual Development in the World of Cultural Diversity Vol VII,* G.E. Lasker & K. Hiwaki (eds). Windsor, CANADA: IIAS, pp 27-32.

LOST IN CYBERSPACE(S)

Vassilios Dagdilelis
Department of Educational and Social Policy
University of Macedonia,
54352, Thessaloniki, GREECE,
e-mail: dagdil@uom.gr

Maria Bontila
Philologist,
Ph. D. History of Education,
Artistotle University of Thesaloniki
e-mail: bontila@edlit.auth.gr

ABSTRACT

Blogs and other platforms of Web 2.0 comprise digital social forums, where everything can be discussed and analysed. The new communication media are not associated with political developments only but have become part and parcel of everyday life. These new communication media are examined, and some results are also presented.

In March of 2011, a few days before the meeting of the Greek government with the representatives of the IMF, an anonymous e-mail was circulated in the Greek cyberspace which proclaimed an impending economic disaster: "Bankruptcy is a matter of days and has been programmed for the 25th of the month" was written, amongst other details. The content of this email was reproduced by dozens of Greek blogs and caused a form of national panic. The Greek government, of course, denied it, denouncing it as provocation, and enlisted the Cyber Crime Unit of Greece (C.C.U.) to find the instigator. Eventually, the culprit was located on the island of Crete: a blacksmith by profession, without any knowledge of economics and with no links whatsoever to the establishment. He quite simply played a joke: a joke that even rocked the already weak Greek Stock Market. This event was not the first of its kind -either in Greece or abroad.

At the same period, the gatherings in Tahrir Square in Egypt were not only organised, as stated through Facebook and the other social network spaces, but also had a domino effect on the other social spaces in the Arab world, as well as wider afield, from Winsconsin to the '*indignados'* in Spain and the '*aganktismenoi*' in Greece (the indignants). Blogs comprise a social digital forum, where everything can be discussed and analysed. They seem to constitute a new form of the *Ecclesia of Demos* of ancient Athens, with the only difference being that the physical presence of the participants is in this case not required. As the majority of bloggers are on the 'periphery' of power and do not have direct access to new information, their subject matter is mainly based on an analysis and evaluation of the issues which appear in the official Mass Media. In this way, they too project and create propaganda for their own positions, which in turn, become an object of criticism and further analysis by others.

Three years ago, Barak Obama's presidential campaign was the first in the 21st century in which all the means of new-age political campaigning where made full use of to achieve the ultimate goal. For instance, Obama used Twitter not only to send general messages, as did Hillary Clinton, but he went one step further where he read his voters' comments and responded to them. This might seem like a minor detail, however, it contributed greatly in granting him the lead, because he applied this new medium of communication to directly connect with his voters at a personal level. In addition, the website he created, FighttheSmears.com, gave the opportunity to his supporters who belonged to the digital generation to counter anti-black propaganda openly [4] with what appears to have been positive results.

The new communication media are, of course, not associated only with political developments but have become part and parcel of everyday life. They have influenced the way of life of many people who for various reasons are forced to live away from their families. For instance, today, thousands of migrants are able to connect on a daily basis with their families back in the homeland through the use of the new communication media. The so-called *technology of love*, allows mothers who have had to leave their children behind to communicate with them easily and cheaply and in this way to shorten the distance that separates them and to limit the repercussions of her absence [7].

We, thus, find ourselves before a new communication reality, owing to a variety of reasons: firstly, the rate of technological advancement and the increasing possibilities this offers; secondly, the decreasing cost of broadband

connections; thirdly, the children of the digital generation, who growing up with technology, have assimilated it to such a degree that it now '*forms an integral part of them. It is something like the air they breathe'* [4]. And fourthly, the loss of supremacy of the traditional mass media, which are more often than not attached to and defend the positions of political parties. This has resulted in them having lost their credibility with most people. Blogs, then, have emerged as the new digital means of communication. Young people, and not only, have pinned their hopes on blogs to supply a contact medium that is livelier, more stimulating and free from the politico-economic interests of the large media organisations and party politics.

The increased influence of blogs and the social network spaces has been established for quite some time now. Their impact is not confined to state boundaries but extends across nations, abolishing borders and the power of the system. In February of 2010, the number of blogs worldwide came to 156 million, of which each one individually offers a forum for public speech, and collectively, have contributed to securing political and cultural pluralism, the dissemination of information, and the biggest potential for the expression of ideas. On the other hand, the number of Facebook users came to approximately 550 million. Nevertheless, the two figures should probably not be compared as they do not relate to the same factors. Meaning that the most likely explanation for the increase in the number of Facebook members is due to the rise in the number of groups ('*friends*' on Facebook or *sig:* special interest groups, etc), in contrast to blogs, which play a role that is closer to and reminiscent of traditional communication media. Blogs are, to a large extent, means of *personal journalism*.

It is often supposed that the new communication media function in a liberating capacity and that they will help change the world. However, the questions that need to be posed are: Do they, in fact, have this power? Are there meaningful changes to the dissemination of information? How does the average citizen deal with the deluge of information that they receive daily? How vulnerable might they be to crises and how susceptible to the opinions of others? Are citizens given essential information and dialogue or do they, paradoxically, find themselves today more trapped than ever? Can they cope with the new data and are they able to make appropriate choices? Ultimately, do blogs comprise an alternative information medium? Have they perhaps in the last five years of their rapid growth gained so much power that they have in essence become problematical for citizens? Has the new communication medium actually changed the message or has the situation remained where the medium is still the message? [2].

And in the final analysis, what proportion of the population have direct access to these new communication media? According to Morozov [3] there were 19, 235 Twitter accounts in Iran at the time of the recent unheavals, that is to say, a mere 0.027 % of the population. Could these people, perhaps, have played a determining role in the Iranian uprising and we thus find ourselves before a new social phenomenon -that of the new social media as intermediary agitator?

On a daily basis, we become receivers of messages and communications whose validity we can neither control nor test. In the case of printed matter in the mass media, journalists can be held legally and morally accountable for what they pen, and the reader has the possibility of taking libel action against the writer of the article. Moreover, knowing the political line of the paper, one can elect to buy it or not, to trust the news items presented therein or not, or based on the criterion of the author and his/her opinions, to choose which articles and stories to read and which not. Whereas in the case of the new digital networks, not only do we most times find ourselves defenseless before the maze of information (which sometimes is even a farrago of useless information), but also false reports and deliberate misinformation can be spread at lightning speed and with relative ease. An individual is helpless before anyone and everyone who wants to sully one's name and reputation. In Greece, each time, for example, that the ex-Minister of Infrastructure, Transport and Networks , D. Reppas, attempts to pass a new bill through parliament, the news item appears, almost without fail, in various blogs that when he was a university student, he had connections with the Colonels' junta of 1967. In this way, certain circles or rackets attempt to damage the Minister's image in casting doubt on his ethics and democratic convictions, by adopting Goebbels' infamous motto "Sling mud, something always sticks".

Unverified news reports and rumours from no-one knows where, always anonymous, often false and conflicting are rapidly and uncritically relayed by dozens of blogs, thus causing a constant-swelling in the number of people who have access to this communication. Anonymity is the biggest problem in this new communication environment, because it permits unreliable information to be easily and thoughtlessly propagated to the entire world. In order to prevent the problem becoming more acute than it already is, various codes of conduct for bloggers have been proposed and manifestos are in circulation, which aim to limit this arbitrary behaviour and to ensure ethical standards of operation [6]. Tim O'Reilly, who since 2007 has played a leading part in this campaign in has proposed the so-called 'Blogger's Code of Conduct' for acceptable conduct for bloggers. One example is that nothing should be written that cannot be said to one's face or that one would not put their name to [11]. In an interview on BBC Radio Five Live he said 'I do think we need some code of conduct around what is acceptable behaviour, I would hope that it doesn't come through any kind of [legal/government] regulation it would come through self-regulation' [5]. There is the fear of censorship or something similar being imposed, as happened in China where the Establishment is very rigid. For instance, accounts that use a pseudonym are barred from operating. And despite the fact that the communications media in that country are

claimed to have been democratised, passing into the hands of many, they have gained so much power that they have become the establishment and pose a threat to citizens [1]. Many claim that western culture too is under threat on account of the amateurism that characterises the users of the new communication media. On the Internet, we can find answers to practically everything that we are looking for, but we are never sure of the validity of what we have found. In his article, *Twitter Can't Save You,* Lee Siegel [9, 10] makes the comment that although a few years ago the talk was all about how the Internet would become freer, today we are discussing how we will control it more. In a speech in 2005, Hillary Clinton extols the liberating possibilities of the new communication media, claiming that people all over the world will be able to use them as tools to promote democracy and human rights. In his criticism of this speech, Morozov [3] states that Hillary is obviously not aware that these new communication media in Russia are indeed being used in a liberating capacity, so to speak. People use them, he says with a note of sarcasm, not to discover what democracy is but to find ways to be free of their extra kilos.
Other questions which arise in connection to this are: the new communication media may very well have been democratised in the sense they provide unimpeded access to all, but how are they actually used? And has their use changed? There is widespread recognition that the Internet and the new technologies played a determining role in the recent uprisings in the Middle East. Analytic reports suppose that it was through these communication media that citizens rapidly mobilised, provoking serious ruptures in the autocratic establishments. But is this really the case or has their contribution perhaps been overestimated? Morozov [3] claims that in regards to new technologies, those who are in possession of power are just as informed as the citizens might be. They, too, use all the possibilities that technology offers, in order to keep up with the intrigues and the scheming, or to intervene in creating a climate that they want to prevail, in order to direct and shape circumstances. The extent of the intervention by those in power is not merely keeping tabs on bloggers and being informed about what is being discussed and what is being organised, but they are in a position to physically obstruct communication, as happened in Iran and Egypt, when the Internet was prohibited [2011]. Governments and police authorities know in advance the feelings and the movements of dissidents. They even proceed to analysing the messages and texts that circulate on the Internet, tracking those people that tend to move in nonconformist circles. At the same time, they propagate their own positions, imposing them on, or in the best case, submitting them to various bloggers.

Morozov also claims [3] that even when there is no disorder or civil unrest, governments still spy on the Internet space, propagate their views, and shape consciences. The metaphor 'Bread and Games' of ancient Rome is apt today; the thought remains the same, only the means have changed. Citizens of today, instead of viewing the gladiators in the arena or the first Christians being eaten alive by the ferocious lions, sit on their couches watching the latest American movie, which the government of their country permits them to download free from the Internet, with the ulterior object of distracting them. Rather than become involved in the current political and social issues, the people are diverted to focusing their attention on their modes of entertainment. Technology has no ethics: it is used by all to achieve their goals and to propagate their views -from the rebels across the Arab world who want a revolution to the neofascists all around the globe who want a pogrom.

Finally, do the new communication networks assist the development of collectivities and solidarity, and do they help make citizens more active? Are Facebook users transforming the ties of friendship (Facebook-friendship) into a political stance, joining groups which clearly assert a socio-politico-demanding character? Do they then progress to the next stage of action, i.e., participation in social and political movements and mass demonstrations or taking individual initiatives with a political character? Or are Facebook users perhaps making their own brand of revolution from their homes, using up their 'participation' in all manner of social movements by writing their opinions on their personal pages, adding their names to lists, and signing petitions? Is the Internet taking users to the streets or does it merely cover them emotionally (in other words, that it is not only them who are concerned with this issue and that they are not alone in this struggle, in whatever way they wage it)? Are we simply caught up in network-activism -a digital deluge of ideas and only that? Or is the digital generation going to create a genuine '*agora*' of ideas and a new culture of public dialogue, where they will activate critical thought and lead to the active participation of citizens in the scheming and plotting ? Would all that has taken place in the Arab world lately have happened anyway, even without the existence of the new communication media or did they in fact have a vital catalytic role [8]?
As this new way of communication has only been in existence for a few years, the scene has not yet been fully formed. We are still in a transitional stage and it is too soon to have a clear account of the situation. It is our view that the discussion should not be centred around the democratization of the information and communication media but rather it should focus on their facilitation of citizen participation.
The new communication reality which has taken form in the last few years requires skills of a different kind. What individuals need is to have a level of critical digital literacy which will allow them to move easily in the world of the Internet and to be discerning when selecting the channels of information and communication. Much cogitation and re-cogitation is needed in order to help us, on the one hand, recognise the processes in which we are trapped and on the other, become aware of our practices and choices. A characteristic example is the analysis of the events in Lybia. While the majority, if not all of us, based on the information in the mass media reports, believed the image of the "bad"

Gaddafi exploiting the Lybians, the analysis of one blogger claiming that the uprising was set up and organised by financial centres abroad whose demands Gaddafi did not succumb to, came to turn the tables. In this situation, what is the truth and how can the average citizen obtain it? Which version of events is true and how can the average citizen verify it? Or should we just accept that we do not need any more truths and that the breach of our (what up till now we considered) '*certainties*' that the new communication media have caused is more than sufficient? In this situation, too, information is sacred; the difference being the agonies of doubt suffered, which can, however, lead to totally new ways of reading the world around us.

These new forms of communications and relations create a new societal reality, comprising at the same time a challenge in the way individuals, social groups and their interconnections can be represented. We anticipate that these reproductions will help to show the properties of the new communication relationships, which although being not visible at a small level, become apparent when reproduced on a mass scale. Graphics are a typical way of representing social relations. Today, the advances in digital technology have given us more powerful possibilities of reproduction than were possible in the past. In this way, instead of individuals we can have reproductions of entire 'clouds'. In other words, interconnected clique-like formations of people – in graph-theory terms. At the same time, not all inter-connections and relationships are of equal weight (because a Facebook-friendship, for instance, is not the same as an intimate relationship) and as Google's RankPage algorithms have shown, not all the nodes are of equal value. It becomes apparent, then, that the new concepts have to be created on graph theory and new processing algorithms have to be worked out. Lastly, it is our conviction that the pioneering work by R. Tamassia et al [12], on Automatic Graph Drawing and readability of diagrams will play a major role in understanding these graphics, as well as the relationships they establish.

References

[1] Keen A.(2007). The cult of the amateur. How today's internet is killing our culture. USA: The Doubleday Broadway Publishing Group

[2] McLuhan M. (1989). The medium is the message in Livieratos, K., Fragkoulis, T. (eds) To minyma tou mesou, I ekrixi tis mazikis epikinonias (The message of the medium, the explosion of mass communications) Athens: Alexandrea

[3] Morozov E. (2011). The Net Delusion: The Dark Side of Internet Freedom, PublicAffairs

[4] Tapscott D. (2009). Grown Up Digital: how the net generation is changing your world. New York: McGraw-Hill

[5] BBC News (2007). Available from: http://news.bbc.co.uk/2/hi/6502643.stm [Accessed 2/5/2011]

[6] Cyberjournalist Net (2003). Available from: http://www.cyberjournalist.net/news/000215.php [Accessed 2/5/2011]

[7] Digital Diasporas: Migration, ICTs and transnationalism (2011). Available from: *http://www.crassh.cam.ac.uk/events/1328/* [Accessed 2/6/2011]

[8] Rich Fr. (2011). Wallflowers at the Revolution. Available from: http://www.nytimes.com/2011/02/06/opinion/06rich.html?_r=2 [Accessed 4/6/2011]

[9] Siegel L. (2011). Twitter Can't Save You. Available from: http://www.nytimes.com/2011/02/06/books/review/Siegel-t.html?_r=1 [Accessed 20/5/2011]

[10] Siegel L. (2011). The Net Delusion and the Egypt Crisis. Available from: http://artsbeat.blogs.nytimes.com/2011/02/04/the-net-delusion-and-the-egypt-crisis/?ref=review [Accessed 25/5/2011]

[11] Wikipedia, 2010. Available from: http://en.wikipedia.org/wiki/Blogger%27s_Code_of_Conduct [Accessed 27/5/2011]

[12] Tamassia R. , Batista G., Batini C. (1988) Automatic Graph Drawing and Readability of Diagrams IEEE Transactions on Systems, Man, and Cybernetics, Vol. 18, No. 1, January/February 1988.

IMPRESSIONISM IN THE MIRROR OF COGNITONICS

*Olga S. Fomichova * and Vladimir A. Fomichov* **
* State Educational Centre "Dialogue of Sciences",
Universitetsky prospect 5, 119296 Moscow, Russia
** Faculty of Business Informatics
National Research University "Higher School of Economics"
Kirpichnaya str. 33, 105187 Moscow, Russia
e-mail: vfomichov@hse.ru and vfomichov@gmail.com

ABSTRACT

The paper aims at representing a new way of considering impressionism under the frame of cognitonics. It helps teenagers and university students to better understand the world of social networks, the real world around them, and to be ready to resist the so called emotional attacks via social networks. An original algorithm of transforming the negative emotions (caused by the messages received from social networks) into the positive ones is proposed. This algorithm considers the possible reactions of a human (including the recommended reactions) to the emotional attacks via social networks. It is proposed to include an analysis of the kind into the program of the course "Foundations of secure living activity in information society".

1. INTRODUCTION

According to the dictionary, an illusion is a false idea, belief, or impression. One of the greatest illusions created by the Internet is the illusion of true life in the Cyberspace. One can earn money, spend free time, enjoy communication, go shopping, and even enjoy evil delight making other people harm. But Cyberspace lets teenagers pretend, lead an imaginary life under another imaginary name (nick).

Impression, the first impression in particular, is an ability of human intelligence. It is an idea, a feeling, or an opinion about something/somebody, especially one that is formed without conscious thought. The first impression is very often far from being true. It is caused by a detail and trifle.

Though the second impression is very often much more correct, the first one is much more bright, because it is coloured by a strong emotional response. The distortions in the perception caused by the first impression as well as the illusion that the reality has created the Cyberspace are rooted in the emotional sphere of the person.

The balance between the intellectual sphere and emotional sphere of the person is destroyed, and it leads to either overestimation of the problem or underestimation of possible emotional or intellectual reaction of a person to a challenge of any kind.

Let us consider impressionism under the frame of cognitonics [1 - 4]; more exactly, the way it helps to turn a visual feast into a splendid opportunity to understand how strong the first impression is:

- How it provides an opportunity to make children and teenagers see the possibility to view the world in a different way;
- To notice the transient appearances of one and the same object (communicative situation) depending on the season, weather, the background of the viewer, his/her mood, etc.

2. "SECRETS" OF THE IMPRESSIONISTS

One of the "secrets" of the impressionists is hidden in their manner of painting. To enrich the colour of their canvases, the impressionists made use of what is known as division of color and optical blending. For example, to represent a green meadow, they put little dabs of blue and yellow on the canvas which were supposed to combine to form green in the eye of the spectator.

The impressionists, for example, Claude Monet, devoted themselves to capturing in paint the fugitive effects of light falling on objects and the play of reflections.

Due to this style of painting, the impressionists manage to render the effects of sunlight, vibrations of water and air, the thousand and one reflections on water, etc.

Studying impressionism is a great pleasure. The emotional response of children can't be overestimated. In case children are asked to come very closely to the pictures, they become deeply impressed by the fact that they can't see anything except for the mixtures of dabs, a kind of colourful chaos. Watching the pictures from some distance, children come to understanding the beauty of canvases, in fact, they make their discovery of transfiguration. An apparent transformation from the colourful chaos to the visual feast produces a deep lasting effect on the mind and feelings of the child. It is a discovery of an illusion (first impression which is false, though bright).

Take “Water lilies. Green harmony” by C. Monet. In fact, there are no white lilies there, though the water surface is dotted with the common white water lilies.
Children suppose that it is one more example of illusion: we are sure that water lilies are white, but in fact it is impossible to point at and list one white lily. They come to understanding the effect of transparency of the water, on the one hand, and the reflections of the sky, clouds, plants edging the pond, and the whipping willows, on the other hand.

3. UNEXPECTED CONCLUSIONS DERIVED FROM STUDYING IMPRESSIONISM

The acquired experience of perception can be applied to various communicative situations at school with teacher, classmates or any kind of misunderstanding.
When a child or a teenager is at sea or in a fix, he/she is sure that everything is “black”. But the idea that he or she hasn’t see pure white lilies on the canvas makes him/her think that it may be an illusion, and everything is not so bad, and it is necessary to step aside and look at this situation from some distance.
It is the process of establishing a link between the constructed mental representation of the seen pictures and the constructed inner visual image of the life situation. The suggested conclusion: no panic, no chaos, no black situations. It is not passive refection, it is an active transfiguration that makes the life brighter, stimulates a creative response to it.
Another important discovery done by children is connected with the possibility to view the world in a different way and to notice the transient appearance of one and the same object depending on various things.
The ability to see ordinary things in a new way stimulates curiosity and desire to reveal personal perception. It reveals the ability of the child to view the world actively, creating his/her own images, metaphors, corresponding to his/her conceptual picture of the world.
For instance, children are asked to describe the lilies in the pond or the pond itself. The following descriptions were given by children.
Example 1 (a girl, 6 years old): “Near the castle, there was an old park with wide spreading trees and a pond with white lilies. At dusk, lilies are falling asleep, and it seemed that someone had eaten whipped cream from the blue cap of the pond”.
Example 2 (a boy, 7 years old): “At night the pond looked like a mirror reflecting bright sparkling stars, and it seemed that one could scoop out a bucket full of stars”.
Example 3 (a girl, 7 years old): “At dusk, the pond is fringed with silvery light and lookes like an ancient looking glass of the moon being lost by carefree crescent”.
The images of the pond are rooted in the own life experience of eating whipped cream or noticing the reflections of the starlet sky in bucket, barrel, or well in the country or playing with a grandmothers looking glass in the silver frame.
Such kind of work proves the importance of the impression as a bright flash evoking emotions, making clear various links between different domains.

4. HOW TO MAKE FEELINGS BECOME THE SUBJECT OF THE THOUGHT

Though impression is formed without conscious thought or specific knowledge, the child should be taught to analyze his/her impressions, to appreciate them. Impression is a kind of impulse sent by the outer world and accepted by the child unconsciously. Impressionism aimed at revealing impression. It provides the splendid opportunity to see the world with the eyes of the painters who were deeply impressed by it and elaborated a special language to express the admiration or just a way of viewing. Such kind of approach to studying impressionism leads to analyzing the accepted impulse, it leads not only to emotional response but also to intellectual response. Feelings are becoming the subject of the thought. It may prevent children from impulsive decisions caused by the first impression that looks like an emotional attack.
In the information society, the possibility of such unexpected emotional attacks by social nets is increasing. The way impressionism is taught under the frame of cognitonics is one of the keys to the solution to this problem, because of constructing the mental representation of what is called “the illusion of white lilies”.

5. THE CROSS-DISCIPLINARY COURSE “FOUNDATIONS OF SECURE VITAL ACTIVITY IN INFORMATION SOCIETY”

The information society we live in has its peculiarities, advantages, and disadvantages, as any other society. In order to speak about successful socialization of children and teenagers, it is necessary to make children understand the ways of living, participating in the social networks, communicating via e-mail, taking on-line courses, receiving information, etc. They should know how to avoid negative “digital” situations or overcome them, how to distinguish virtual reality and emotions caused by that virtual reality from the real life and emotions caused by that life.
Children should be aware that in both cases they should be ready to suffer the consequences of their careless behaviour or ignorance. Children should be taught the rules of acting in the digital space, paying special attention to the moral standards, lest they should hurt somebody’s feelings while communicating with the help of informational technologies (IT). They should understand the power of IT and the responsibility of the users. The teachers should find a correspondence between situations taking place in the real life and similar situation from the virtual reality.

For example, if someone reveals aggression in any way, he/she can suffer the same aggressive feedback, can be hurt. In case of the digital space, not a child or a teenager him/herself is hurt and experiences pain, but his/her feelings are hurt and his/her reputation is in danger because of the quickly spread negative information.

The core idea of the course "Foundations of secure vital activity in information society" is the same for real and virtual life: treat others the way they want to be treated, show compassion and consideration, learn from success and failures, appreciate learning opportunities, etc.

But children should be aware of the difference between person-to-person communication and person – digital environment – person communication, because the digital environment has its own power that can enhance the communicative (and any other) situation.

The goal of the course is to contribute to successful socialization of children and teenagers in the information society. The subject of the course is to introduce students into the digital space, paying special attention to ethics, to the rules of interaction and communication.

Students acquire knowledge of what is strongly prohibited and what kind of behaviour is expected. The course shows the clear difference between virtual reality and reality and establishes correspondence between the ways of perception by people of various situations happened in digital space and real life, on the one hand, and the possible consequences caused by that difference.

The idea of this cross-disciplinary course and ethical approach to its development are suggested under the frame of cognitonics.

6 AN ALGORITHM OF TRANSFORMING NEGATIVE EMOTIONS INTO POSITIVES ONES

Let us consider the possible reactions of a human (including the recommended reactions) to the so called emotional attacks via social networks. An analysis of the kind could be introduced into the program of the course "Foundations of secure living activity in information society".

Head Module of the Algorithm
"Processing Messages from Social Networks"

begin
If the first impression (a strong one, it can be either true or false) is POSITIVE then
Procedure 1
else { the first impression is NEGATIVE}
Procedure 2
end

Description of the Procedure 1
begin
Make conclusion 1: no harm at the moment of getting an impression;
Make conclusion 2: the feelings are not hurt;
Make conclusion 3: The situation is over
End

Description of the Procedure 2
{The condition of calling this procedure: the first impression is NEGATIVE, that means that the impression causes panic, confusion at the moment of getting it}
begin
FIRST AID: a reminder of the white lilies on the canvas by Claude Monet (if the mental representation is strong and clear);
Make conclusion 1: It is a situation of uncertainty, not apparently a bad one;
Make conclusion 2: The situation needs reflecting, reasoning;
Start some kinds of intellectual activity, thinking over the situation and diminishing emotional activity;
A little later make conclusion 3: The situation is getting much more balanced, less harmful;
Make conclusion 4: The situation is turning into an intellectual riddle;
Make conclusion 5: Now the situation causes another kind of emotions, they are based on the feeling of curiosity;
Make conclusion 6: A situation of another kind has emerged, it aims at solving the riddle
{the transformation of the emotions into the positive ones is over}
End

7 CONCLUSIONS

The development of civilization is the endless process of challenges and answers. Internet, new informational technologies (or Cybermind, if we use a metaphor) are a challenge. It is not only an intellectual challenge but a spiritual challenge as well.

The answer to the intellectual challenge of the digital time is clear and successful. The applications of informational technologies in all spheres of the human activity changes the standard of living, expands the communicative possibilities. A breakthrough in teaching, training, working, medical care, finance, etc. creates what is called Information Society.

On the other hand, the illusion of the true existence of the Cyberspace gives birth to new kinds of emotional attacks via e-mail, social nets, and cell telephones (see, in particular, [5]). It is difficult to resist these attacks, because lots of teenagers and grownups become aware of the information together with the attacked child.

Impressionism as a manner of painting, rooted in the idea of the first impression, being taught under the frame of cognitonics helps to construct the vivid mental representations of the illusive situations in the minds of children and teenagers. The example considered in this paper is positive, impressive, and it

is based on children's life experience (they acquired it while watching the painting).

On the other hand, they have already thought about their own examples taken from the real life: one child is sure that the dog is angry, because it barks; another child says that his particular dog is kind, because it wags its tail every time the child sees it.

To make children and teenagers understand how the illusion and first impression work, explaining to them emotional constituent of these notions, providing them with thrilling and clear examples is very important, especially in the information society when they deal with one partner of communication but with many partners from the nets.

People communicating via social networks don't take into account the child's mood, character, the events of the day, child – parents relationships at the moment, background of the child. New possibilities of IT demand much more developed ability of the child to resist to any reply or replies, much stronger confidence in oneself, and clear understanding how the illusion works.

We know that the lilies on the pond are white, but, in fact, there are no white lilies in the pond: everything depends on the maturity of the viewer.

References

[1] V.A. Fomichov, O.S. Fomichova. Cognitonics as a New Science and Its Significance for Informatics and Information Society. *Special Issue on Developing Creativity and Broad Mental Outlook in the Information Society, Informatica. An Intern. Journal of Computing and Informatics (Slovenia)*, 2006, Vol. 30, No. 4, pp. 387-398.

[2] O.S. Fomichova, V.A. Fomichov. Cognitonics as a New Science and Its Social Significance In the Age of Computers and Globalization. *IIAS-Transactions on Systems Research and Cybernetics. Vol. VII, No. 2. Intern. Journal of The International Institute for Advanced Studies in Systems Research and Cybernetics.* Published by The Intern. Institute for Advanced Studies in Systems Research and Cybernetics (IIAS), Tecumseh, Ontario, Canada, 2007, pp. 13-21.

[3] O.S. Fomichova, V.A. Fomichov. Cognitonics as an Answer to the Challenge of Time. *Proceedings of the International Multiconference Information Society- IS 2009, Slovenia, Ljubljana, 12 – 16 October 2009. The Conference Kognitonika/Cognitonics*. Jozef Stefan Institute, 2009, pp. 431-434; available online at http://is.ijs.si/is/is2009/zborniki.asp?lang=eng.

[4] O. Fomichova. Humanitarian Education – an Answer to the Challenge of Time. Moscow State University Press, Moscow, 2009. In Russian.

[5] H. Yasuda. A Risk Management System to Oppose Cyber Bullying in High School: Warning System with Leaflets and Emergency Staffs. *Informatica. An International Journal of Computing and Informatics (Slovenia)*, 2010, Vol. 34, No. 2, pp. 255-259.

RESTRUCTURING OF KNOWLEDGE SYSTEM ON THE INFORMATION NETWORK

Joonho KIM
School of International Relations
Tokyo International University
049-232-1111, fax: 049-232-7477
toko9a@gmail.com

ABSTRACT

Network is global commons, and general intellectual structure of the contemporary world is composed of this globalized network of information on the cyber space. If network fails, the world collapses. So the urgent task of network information society lies in the change of negative information flow to the knowledge structure, with active and critical commitment and constructive engagement. This paper tries to provide a consideration to overcome the negativity of information system.

1 INTERNET AS AN INFORMATION SYSTEM

In the 21st century, a new world is not discovered but created. Human reinvents his/her world, by digitalization, on the network, which is a space abstracted from the concreteness of life. Digital is a magic of 0 and 1, and this fantastic contrast of light and darkness formed contemporary world with sound, image and letters. …With digitations all of the media become translatable into each other [9, p.3].
Digital is networked through communication technique and produced in the electronic public sphere that links and socializes individuals.
General intellectual structure of the contemporary world is made by this globalized network of information formed on the cyber space. But the indigestion of information has played havoc with the knowledge base of global society. Here the flow of information often represents negativity of globalization. Therefore, nothing is as important and essential to this age as structuring humanism out of this flow, which constitutes a theme of globalism.

2 INFORMATION FLOW IN THE PUBLIC SPHERE OF INTERNET

2.1 Strata of intellect in the internet sphere

There exist sharp conflicts between reality and knowledge in human society.
How do we know what we believe we know [12, p.9]? Ernst von Glasersfeld, a radical constructivist, poses the provocative proposition that all we can ever know about the real world is what the world is not [12, p.14].
In contemporary world, various patterns of intellectual substance are mixed, on the internet plane, which gives rise to tensions between virtual reality and human knowledge.
A similar structuration of scientific knowledge and technological innovation is taking place under the new paradigm of informationalism [2, p.8].
Human intellectual contents are classified into 4 levels: data, information, knowledge, and wisdom. Each level represents the step of intellectual advancement corresponding to the development of human cognitive ability.
Information is a part of knowledge, and wisdom is the integrated expression of knowledge.
Information can be interpreted from various points of view: for example, philosophical view, mathematical view, economic view, legal view [9, pp.12-19].
And, the component of information is easily recognized. It is the explicit ingredient of knowledge, where what we want to know may be itemized. Information consists of facts, specific intellectual artifacts (often arranged in sets or bunches). It is impersonal (not a matter of opinion) [5, p.19].
The atomic data develop into the informational molecule.
The distinction between information and data could be equated with that between a raw material and the manufactured product [9, p.11].
And then, organic integration of information into knowledge system, and furthermore, up to the transcendency of wisdom.
Information is based upon objectivity apart from human views, while knowledge has existential feature as it contains value concept, when human confronts with reality.

Here we can classify the information as technē, and knowledge as epistēmē.

> …the mark of a deep separation between epistēmē and technē as kinds or modes of knowing is well evidenced… Epistēmē is the etimology of epistemology, and technē, technology [4, p.27].

Knowledge as epistēmē is the synthesis of information as technē and human value judgment in society, because knowledge is socially generated from the activities and relations of an interpretative community [10, p.96].

And knowledge is usually understood as making some claim on the general, usually with some normative or mainstreaming connotations [3, p.98].

2.2 Flow of information energy

Today human cannot exist outside of network, which is society that is real as well as virtual. The relationship of human to network defines the characteristic of contemporary society.

Digitalization = networking = globalization, as synonym and tautology, represents the flow and structure of global society. Digital networks are global, as they know no boundaries in their capacity to reconfigure themselves. So, a social structure whose infrastructure is based on digital networks is, by definition, global. Thus, the network society is a global society.

A network society is a society whose social structure is made of networks powered by microelectronics-based information and communication technologies [2, p.3]

Network is channel, globalization is movement. Globalization is a vast and rapid flow of information [2, p.3]. Networks process flows. Flows are the streams of information between node circulating through the channels of connection between nodes [2, p.22].

Information is energy. The flow of this energy produces power. And we witnessed the magnitude of this power in Algeria, Tunisia, Libya, Syria and in London and other cities of UK, in the year 2011.

Information energy, flowing through the cellular phone, gave rise to people's movement. Freeing of this energy stimulate revolutions. And this seemingly soft revolution has led to hard revolution, which is social and political changes, like above.

Flow of molecular particle of technē, which is not filtered by human concern can be harmful. Information seems to be objective, but is not value-free.

Value is what is processed in every dominant network at every moment in every space according to the hierarchy programmed into the network by the actors acting upon the network [2, p.26].

Information attacks human vulnerability and inclines toward unethical direction. This negative character of information is abused by market power and misleads the global society.

The development of new telecommunications infrastructures is not some value-neutral, technologically pure process, but an asymmetric social struggle to gain and maintain social power, the power to control space social processes over distance [13, p. 21].

Human is often dominated by material interest and physical instinct; money and pleasure. The network, full of financial derivatives and Viagra information, attacks the global society.

Residual of information produced superfluously, like spam mail, becomes a torrent and violates the realm of atomized individuals.

Polluted information breaks the equilibrium of production and consumption of information

Now economic crises are familiar to our daily lives, and the world is united by pornography, which governs the mental geography of humankind.

Thus flow of negative information, general phenomenon of globalization, rules the world. If network fails, the world collapses, if network is polluted, then human society will be deteriorated, because network is the world. The sustainability of network environment has become vital condition for human life.

3 TRANSFORMATION OF NETWORK INFORMATION INTO HUMAN KNOWLEDGE

3.1 Network as the global commons

We are living in the network civilization. Network externalities have become a most feared word. Network is defined as global commons, which needs the internet governance. Human participates in the network space, as media citizen.

The commons of society are made of electronic networks, be it the media inherited from the mass media age, but deeply transformed by digitalization, or the new communication systems built in and around the internet [2, p.30].

The character of public space network as global commons is open, common and equal.

A whole philosophy of life.is embedded in this rather simple piece of code. A whole philosophy of equality among people, of equal access of people to other people. A philosophy that anything that everyone has to say is equally worthwhile [7, p. 149].

3.2 The possibility of global knowledge system

Structuring the flow of information on the public sphere of internet needs the communal engagement of human being, and it presupposed the peculiarities below, as its subjective condition.

A. Criticism

Network as pubic space needs critical value judgment. Without critical rationality the publicness of network is damaged. Information has no function of self-examination, which should be provided exogenously. But people are mostly sunken in the network, with non-critical one-dimensional thought. Active and

critical commitment in the public space of global network diminishes this one-dimensionality. And humanism as the criterion of criticism becomes the central concept in the sustainable development of network.

Criticism presupposes the objectification of mixed flow. Self-objectification is needed of network. That means that people are needed to get out of the flow lest they be drowned during their enjoying the net surfing.

Gnōthi seauton- know thyself- means the objectification of self. To know is to objectify something. Thus gnōsis as the general intellect of network era requests the objectification of seemingly objective information. This means the subject-object confrontation on the internet sphere should be justified and fortified. Some authors assert it is already happening.

> Our societies are increasingly structured around a bipolar opposition between the Net and the Self [1, p.3].

B. Construction

Information is flow, and knowledge is structure. Flow and function confront conceptually with the structure. The urgent task of network information society lies in the change of information flow to the knowledge structure. In other words, it means the construction of knowledge system based upon the structuralization of information.

Construction is theorized as a cognitive operation, knowledge is the product of construction.

Objective fact, data or information is a part of knowledge and therefore they are apart from human judgment or ideas, as knowledge is based on humanity as a whole.

Restructuring of network information is critical and constructive.

Constructive theories (implicitly) embrace meta-theoretical assumptions that centre a singular real-world reality (the assumption of ontology) and sharply distinguish this real world reality from knowledge about it (the assumption of epistemology) [6, p.4].

Wikipedia, as a typical example of constructed knowledge, has various character of digitalized knowledge structure: common objectification, constructive, developmental, critical, intersubjective, and it contains global rule, criteria, common engagement, objectified theme….This is a global forum, around which social knowledge structure is formed, and a kind of dialectical commitment into the human intellectual concern is successfully achieved.

3.3 Intellectual existentialism and the meaning of education in the network era

In “inform-ation”, we are informed of something, whereas we try to know something, in “know-ledge”. We are passive in the former concept, and active in the latter. But in contemporary world, “inform-ation” as active flow subjectifies itself, while passive human intellect becomes objectified. This subversion must be reversed. Human should recover own subjectivity (“know-ledge”), which means the structuralization of the flow, with critical attitude on the pseudo-gnōsis.

The ability to integrate information into the existence of human being and construct the organic system of knowledge is needed.

Information fragmentality should be transformed into holism and structurality of knowledge system based on humanity, human wholeness.

Contemporary network system is a mechanism of distorted information flow, which is related to the reproduction of market capital.

How to recreate information as knowledge, how to overcome information capitalism by the intellectual humanism, pose the contents of network existentialism.

Information with humanity develops into knowledge and makes existential transformation occur. Transformation of information into the system of knowledge means ideation of information system, which endows human value on the information.

This value endowment on the objectivity of information is best accomplished through direct education. Human contact or Oriental guru-disciple relationship in the network era is reformative as well as traditional.

The most appropriate and efficient methodology of global system of education lies in the effective distribution of direct education in school and indirect education on network. Not the share of information but that of knowledge through the human contact shall be more and more highly appreciated, as the digital technology as an educational method develops.

Direct education should be used as the supplemental mechanism to correct the indirect network information, which will become the main engine of contemporary education system, so that the meaning of education should be revaluated in the internet society. And reform of education system is required globally.

Network knowledge will and should be the universalized method of education, because it gives cheap and equal opportunity of access to educational materials. People will teach and learn among themselves via network. How the directness of school education can be successful as supplementary educational method will be the most essential topic of the new globalized information age. And the building of network which has the self-cleansing function, with critical and constructive mechanism, is essential to the intellectual future of human being.

4 CONCULSION

Three decades ago, the term “Global Brain (by Peter Russell, 1982)” had been introduced to describe the developed mass information flow of then contemporary society, even long before the internet system was known to people.

Now every individual computer device forms an informational node, like a neuron. Suddenly the earth has become a real huge brain, since two decades ago, by the technology which connects

the nodes. And this system produced the dark side as well as bright side.

It is required that humankind should reconstruct the new system to survive the risk and instability, described above. Transforming the information into knowledge is vital to human future. And the network system should be cleansed with critical commitment and constructive engagement. So the construction of network system with self-cleansing mechanism has become the basic theme and question of new era.

One of the answers lies in the possibility of efficient and effective composition of direct education and indirect learning through network system. It can be said that education is what structuralize the information.

Direct education system, including schooling, as person to person teaching and learning system, should be reconsidered its significance in this IT era. We should rediscover and reconfirm the meaning of education.

The key to survive this information era is educational system based on the harmony of necessary flow of information and the share of human value judgment through the direct contact between people, that is, the restructuring of knowledge system on the internet network.

Dam, which is constructed and structuralized in the river, controls the flow of water to produce electricity. If we can control the torrent of network information, we can produce and provide great value for human intellectual and physical welfare.

The structuralization of knowledge system is the ideation of information, through the global engagement, which is constructive and critical. In other words, it means the endowment of human value to the information.

Now nobody would deny the global world is one brain. But no brain exists without heart.

Global Brain must be humanized by "Global Heart".

References

[1]Castells, Manuel (1996) *The rise of the network society*, Blackwell Publishers.

[2]Castells, Manuel (2004) *The network society: a cross-cultural perspective*, Edward Elgar Publishing.

[3]Gerbner, G., L. Gross, M. Morgan and N. Signorielli (1982) 'Charting the mainstream: television's contributions to political orientations', *Journal of Communication*, Vol. 32, No. 2.

[4]Hazelrigg, Lawrence (1989) *Social science and the challenge of relativism*, University Presses of Florida.

[5]Hill, Michael (1999) *The impact of information on society*, Bowker-Saur.

[6]Hosking, Dian Marie (2007) 'Can constructionism be critical?', in Holstein, J. and J. Gubrium, (2007) *Handbook of constructionist Research*, Guilford Publications.

[7]Lanier, Jaron (1997) 'Lecture delivered at the New Media Minds Forum Center for the Arts', San Francisco, September 11, 1997.

[8]Livingstone, Sonia (1999) 'Mediated knowledge', in Gripsrud, Jostein *Television and common knowledge*, Routledge.

[9]Saxby, Stephen (1990) *The information society*, The Macmillan Press.

[10]Schroeder, K. C. (1994) 'Audience semiotics, interpretive communities and the ethnographic turn in media research', *Media, Culture and Society*, Vol.16, No.2.

[11]Simon, Craig (1998) 'Internet governance goes global', in Kubálková, Vendulka, Nicholas Onuf and Paul Kowert *International relations in a constructed world*, M.E.Sharpe.

[12]Watzlawick, Paul (1984) *The invented reality*, W.W.Norton and Company.

[13]Wilson, Mark and Kenneth Corey (2000) *Information tectonics: space, place and technology in an electronic age*, Jonh Wiley & Sons.

AND THEY SPOKE WITH MANY TONGUES (SOME BASIC ISSUES OF LEARNING, LANGUAGE AND SOCIETY)

Peter Mbaeyi
Postfach 101219
63012 Offenbach
e-mail: sdeschle@csc.com

ABSTRACT

The caption of this contribution is the summary of the phenomena described as having been attendant to the quest that manifested itself in the drive to erect the Babylonian Tower. Through a phenomenological approach, a number of hidden issues and problems are unraveled and are seen to epitomize a number of salient questions that repeatedly reverberate through society, and all of which have the following phenomena in common: cognition, language, knowledge and learning in their manifold interactions. The approach focuses on the interplays between cognition and language – the generation of many languages (many tongues) – and will be described by looking at the yearnings for (1) the ultimate informer / educator – internet; and (2) for ultimate science. In view of the wide span of themes involved only a skeleton delineation can be presented on the basis of issues of perception-cognition complexes, and their role in the crystallization of faculties and languages. Unexpectedly, this also reveals relationship to another legend.

1 INTRODUCTION

The legend / myth of the Babylonian Tower in the Pentateuch phenomenologically seen is a description in imagery form of issues and questions that haunt the problems of intention / intentionality. Babylon, as a civilization obviously attracted people from far and wide, i.e. to say people of diverse backgrounds – of diverse cultures and languages. So one should expect that it was not unusual to encounter people of diverse languages at activity points and sites, e.g. construction sites. Today's globalization offers ample examples of this at nearly all workplaces and sites around the globe. Historical accounts about the composition of the residents of Rome and other major population centers of the Roman Empire abound with references to the presence of peoples of diverse languages. Therefore, from pure physical perspectives, the existence and presence of different languages side by side as documented by human population of different backgrounds may be said to be accompanying phenomena of civilizations. Consequently, admixtures of human misunderstandings and the conflicts that follow in their wakes are no abnormal phenomena. So, what is to be understood as special about the legend of the Babylonian Tower? Or, alternatively stated, what is its message?

2 FROM ONE LANGUAGE TO MANY

According to the legend / myth, people developed appetite to know where 'God' as the ultimate ruler resides; the translation of this into firm intent / intentionality led to the resolve to build an edifice high enough to reach Gods place of abode. In the course of executing this resolve, people developed different tongues (languages), and with time, no longer understood one another. In short, the realization quest generated misunderstandings. Implicit in this narrative is the phenomenon that people embodied initially as a whole separated into sub peoples (parts with different languages) as a result of the pursuit of the realization of the quest described above.

The issue to be looked at is not whether the narrative describes something that can be taken as a canon of belief; rather attention is directed at the timeless phenomena embedded in the narrative, thus giving the narrative an overarching role. The intent-intentionality complex is one such phenomenon; it is too involved for description within the limited space offered by a short paper of this kind. It spans an entangled set of themes that includes taking account of Libet Experiment – a complete theme for itself. Consequently attention is paid to the following chain of observations:

According to the narrative, there was something whose existence and eventual location – a place to be frequently sought out and found – were undisputed, implying that there were unifying understandings about these.

For purposes here, the following definition will be taken as given:

DEFINITION D1: Language is the medium and conveyer for and of organized-ordered anthropic articulations and expressions of anthropic cognitive processes as well as their representations in forms accessible to anthropic perceptual processes.

QUESTION Q1: Within what language were such unifying understandings achieved, reached?

Therefore, the following assumption will be taken for granted:
ASSUMPTION A1: There existed a language X.
It is also essential to attach a semantic to the word 'understanding'.
DEFINITION D2: Understanding implies coincidences, collusions, confluences of articulations, expressions and representations generated through contributory perception-cognition complexes inherent to all participants in such a manner that discernments, decisions, judgments over things jointly shared by all participants are uncontested and undisputed.
In the absence of definition D2, misunderstandings will be said to exist.
Further, according to the legend: In the course of designing and erecting the desired edifice for seeking out the assumed and presupposed location of the one cherished thing, misunderstandings arose and these expressed themselves in the emergence of different, i.e. diverse tongues (languages), which generated misunderstandings.
In real world, on a smaller scale, the misunderstandings arising from dissentions at diverse levels between designers (e.g. architects), engineers, artisans, resources enablers (e.g. money givers) and the conceiver / commissioner of a project are ample illustrations of the messages conveyed through definition D2 in relation to the narrative in use here.
Quintessentially therefore, one obtains the following observation:
On the basis of the language X, languages Y1, Y2... emerge. In short, we go from a single language to many.
The following assumption is needed as an aid:
ASSUMPTION A2: Languages are interwoven with the description of a phenomenon designated as 'perspective'.
The differentiations achieved up to those points are summarized as follows:
LEMMA L1: An entity – abstract, imaginary or real – observed, i.e. viewed from different platforms (vantage positions) stimulates the emergence of conflicting coincidental contradictory perspectives of one and same thing by observers of similar physical composition.
Taking the perception-cognition complexes as attributes into account, this leads to the following question:
QUESTION Q2: In the 'interactions and interplays' of entities which are materially of similar compositions but are parameterized by non-material constitutivities which are merely comparable without necessarily being similar, are there standardizable patterns that provide gauges for insights into the behaviors of the individual entities and to that of the whole obtained by taking all entities together.
REMARK R1: Interactions and interplays have been put in inverted commas because a wider description that incorporates more designations requires a recourse to the German word 'Wechselwirkung', whose semiotic-semantic equivalents in English constitute a theme for itself to be presented in the near future (year 2012) in conjunction with conceptualization of the notion of 'linguistic turn'.
In order to apply the foregoing to society, the following definition is needed:
DEFINITION D3: An anthropic society is an ensemble of entities called populations in which the entities are humans.
REMARK R2: The adjectival 'anthropic' in definition D3 is essential because there are phenomenologically different categories of populations giving rise to diverse societies.
In view of the existence of psychic phenomena, anthropic societies satisfy the proviso (conditions) set out in question Q2.
REMARK R3: Except for the case of hyper-intelligent houses, if they were to come into being, ensembles, i.e. societies of classical houses, do not satisfy the proviso of question Q2.
One of the characteristics of anthropic society is the quest for knowledge which, if extended to an abstract thing designated here as ultimate knowledge and substituting this for 'God' in the narrative in use here the fraying phenomena sketched out here apply to most admissible contexts that can be considered, and experiences provide evidences that do not contravene the differentiation described so far.
The travel companions of knowledge are the phenomena with the collective designation learning.

3 LEARNING

This will be explained through attention to natural learning; this arises from the fact that all human beings begin life within the initial phase of post-birth developmental biology during which acquaintances and acclimatization with different environments are carried out employing the perception-cognition complexes with which human beings are equipped from the onset. The opening observation is therefore:
QUESTION Q3: Does the legend in use here have any relevance whatsoever for natural learning (which will be defined later below)?
Affirmative answer to question Q3 requires the following assumption:
ASSUMPTION A3: There is a possible converse of the legend in use here, i.e. that of going from many languages to one.
That is:
QUESTION Q4: Given a set of languages L1, L2..., is there one single language L which is common to all, signifying the unification of all? In short: Can the following ensemble
{ [L], [L1,L2,...] } exist?
REMARK R4: To pose question Q4 fully and stringently requires a detour that eventually involves the Russellian Paradox which cannot be accommodated within the available space.
The relevance of question Q4 arises naturally through the perception-cognition complexes that are at the core of the themes of integrated perception.
REMARK R5: It is essential to differentiate integrated perception from holistic perception, which will be explained in other writings and publications.

Human beings are endowed with sense organs (sensors) that continuously deliver streams of promptings (ticklings, tinglings etc.) to an assembly of variable centers (domains) in the brain; these centers (domains) are in continuous interactions, interplays, intercourses etc. with one another. One set of senses are oriented to exospace – the space outside of the human – and deliver promptings that give the human feelings and ideas about the status in and of exospace. Another set is located internally in the human, i.e. in situ, and deliver promptings about the status of the endospace of the human.

DEFINITION D4: For simplicity, the centers (domains) described above will be designated 'cognitators'.

REMARK R6: Measurable, quantifiable promptings define things called data.

ASSUMPTION A4: The senses are unique, in that they have distinct ways of articulation and expression – some sorts of mini-languages.

On basis of assumption A4, 'cognitators' are continuously being fed with diverse languages.

Observations and experience provide evidence that the brain has the aptitude and ability etc. to integrate the diverse languages from the senses into a wholesome thing that will carry the designation cognition. In other words, the interactions and interplays of 'cognitators' generate a new language since cognition has its own class of articulations and expressions.

Consequently, there is a natural phenomenon of a single language being the fusion of several languages. The situation anyhow is not as simple as stated here, but does provide a guiding model for looking deeper into the dynamic pattern formation in the brain. Because further dynamical developments of cognition lead to other phenomena like the crystallization of self as basis of 'I', and so forth. Deeper thrusts into these phenomena require the invocation of things like mapping and transformation theories provided by logic and mathematics which cannot be stated ad hoc within the space available. Anyhow, the salient relevant points have been described above and allow for a definition of natural learning:

DEFINITION D5: Natural learning constitutes capability and capacity of an entity to generate centers of focus within itself for survival within its contexts of existence, where these centers of focus are the result of synthesis of inputs from the contexts of its existence.

REMARK R7: Not yet fully encompassed is the question of what role pre-birth developmental biology plays in post-birth developmental biology.

REMARK R8: The focus on language for elucidating these connections arise from the fact that in neurological disorders that distort cognition phenomena one of the prominent symptoms appearing in most cases is that of language in very general sense. Some of the distortions of cognition that can arise are for example disintegration of same, 'decoherentization' of same, indicating that the unifying language embodied by cognition can again fray up into parts, e.g. in case of stroke, and other diverse types of neurological injuries. In such cases we have the phenomena described in the narrative used in this paper showing that the converse can fall back into the original. More compact designations for the aforesaid are integration/disintegration, aggregation/disaggregation.

4 EDUCATOR AND GENERAL LEARNING

The following definition guides the observations to be made here:

DEFINITION D6: In an ensemble made up of entities that satisfy the proviso of question Q2, a participant entity that serves as center of focus as defined in definition D5 will be called an educator if it can moderate and modulate the inputs reaching it and in addition influence the entities from which inputs come.

REMARK R9: Cognition can also be an educator with respect to the senses provided certain developments (deeper inner differentiations) have taken place. These are philosophical issues far removed from the present context.

REMARK R10: In the real world of computer systems the issue raised in question Q4 has a partial answer which brings it closer to assumption A1 of this paper. There is a language – the machine language – to which all the other programming languages dock. But it does not provide a unification of all programming languages. To date, the furthest development of computer systems is the internet and therefore raises the following question: Can the internet be an educator in terms of definition D6?

DEFINITION D7: Cognition which differentiates into further cognitive dimensions will be designated an ultimate educator because of its capability to regenerate and reorganize its own cognition that subsequently influences the senses to which cognition is linked.

REMARK R11: Can the internet function as an ultimate educator?

General learning is to be considered as expansive abstractions of natural learning and differs somewhat from the classical understanding of learning – the acquisition of repeatable specific skills and techniques. The latter corresponds somewhat more to modularization of cognition complexes that lead to crystallizations of faculties. The former, general learning is tuned more to the requirements of change phenomena. Because change involves modifications and transformations of languages in context and manner closer to the processes whereby cognition emerges as described above.

5 CONCLUSION

The tour de horizon given here through the stimulus of the legend of the Babylonian Tower and its converse as stated in assumption A3 unraveled a number of phenomena concealed in the legend but with significance for core operations and systems of human life. As noted in remark R8, these confront human existence with another phenomenon that has similar characteristics under a different name – the legend of Sisyphus. In this legend it

was the case of repeatedly rolling a heavy stone up a hill, and thereafter experiencing that the stone subsequently rolles down back to the foothill. The legend used here along with its converse show that the quest for wholes is attended by the disintegration of wholes to parts. The generation and disintegration of cognition is a very important issue with which society has to contend with in many circumstances, for example in the understanding of health processes. It induces the need to consider the possibilities of reflecting on alternatives that take human life away from the search for ultimate quests. An example for this need for alternatives is raised by the quest for the ultimate science. Will an ultimate science that unifies all sciences, if achieved, be saved from the destiny and fate described in the aforesaid observations made from the legend in use here? What overarching cognition of the most general type will be required in order to produce a wholesomeness that represents an ultimate science in a manner that all other sciences will be represented and contained in it? The difficulty of such an endeavor is illustrated by the project of Russell and Whitehead that produced the monograph called the 'Principia Mathematicae' which was an attempt to generate a unifying language for all mathematics. How will the ultimate language that unifies all languages look like? Even though science was not described separately in the exposition above, it is implicitly taken to be connected with the problem of cognition and the emergence of faculties and also to be related to learning.

CONSCIOUSNESS TRANSFORMATION: FROM ANALYTICAL TO SYSTEMS THINKING

Tadeja Jere Lazanski
University of Primorska, Faculty of Tourism Studies
Obala 11a, 6320 Portorož, Slovenia
Tel: +386 5 6177059; fax: +386 5 6177020
e-mail: tadeja.lazanski@turistica.si

ABSTRACT

Analytical thinking has been a dominant mode in science for centuries. Nowadays, the majority of society still fall into the trap of analytical thinking, which is short-term thinking with insufficient feedback information and knowing the deeper meaning of a challenge. Consequently, people remain unsatisfied, sad and generally in lower emotional conditions. Systems thinking appeared in the 1950s, when systems philosophers and engineers started to think from the perspective of a whole and used this approach, firstly in the industrial area. Systems thinking searches for industrial solutions, but it also is an important part in the conscious transformation of analytical thinking. It represents human awareness of the situation as a whole and it causes a shift of consciousness, in which long term solutions are of greater importance than short-term ones. This paperpresents the process of the growth of systems thinking awareness, which represents the shift of consciousness, its benefits and its deficits.

1 INTRODUCTION

According to Ackoff [1], analytical thinking has been the main mode of thinking in Western science for the last 400 years. It brought technological development to the world but also split consciousness. Consciousness is an awareness [2], awareness of being aware [3], wakefulness or control system of the mind [4], strictly analyzed matter and environment, judged and graded in order to explain what matter actually was. As the Mayan calendar represents the evolution of consciousness [5] and thus explains the reason of analytical thinking as a main mode in the last 400 years [6], we can claim that the time has come when complimentary thinking to analytical thinking comes into power in order to maintain balance in natural, organizational and evolutionary systems. The awareness that consciousness influences thinking, consequently talking and acting, brings contemporary human beings to the results where analytical thinking is not an answer to the challenges and questions of modern life. In this paper, we present the process of analytical thinking and its transformation to systems thinking.

2 ANALYTICAL THINKING CONSCIOUSNESS

Analytical thinking is thinking without considering the feedback effect. It represents the "iceberg trap", which short-sighted and conventional solutions, when making decisions. The big picture or the picture of wholeness represents ones' understanding the depth of a challenge and taking into concern all points of view.

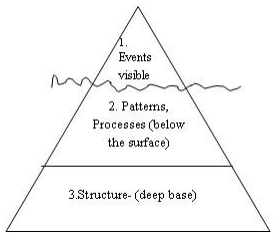


Figure 1: *The Top of the Iceberg as a short-sighted trap*

As it is shown in Figure 1, the trap of short-sightedness is the analytical thinker's failure to concentrate on the 2nd and 3rd, levels of Processes and Structures. What "sinks" the strategy is the same thing that sinks ships: the invisible part below the surface. Following Haines [7], 87% of an iceberg is below the waterline. Decision-makers should consider the two levels below the surface as those that can sink their strategies and efforts. Analytical consciousness is defined by the observer as: place, time, and observer's reality (Fig.2) **.** If the observer thinks of place, time and observed object, then he depends on calendar and location if he wants to obtain the results of the observed object. If the observer thinks of inputs, process and outputs, he depends on outputs and a wish to achieve the outputs unconditionally by influencing the inputs and the process.

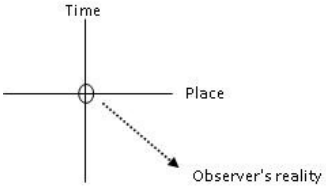


Figure 2*: Two-dimensional analytical consciousness [8]*

Time is the very foundation of conscious experience [9]. It is a consideration, and an agreement between people. Place is also an agreement between people. Every time and place has its own consciousness [8], as well as the observer, whose consciousness reality is the intersection between time and place as shown in Fig.2. The observer's consciousness reality was different in 19th century than today, e.g. the car was rare but today is a normal transportation means. When we take several individuals with their own analytical viewpoints, we get many separated perceptions, which have something in common: they represent separated, (none synthesized) thinking or points of view. They represent separate entities, without any interconnections. Each of them has its own reality, its own consciousness. We can see the linear process of bringing the observer from inputs to the outputs.

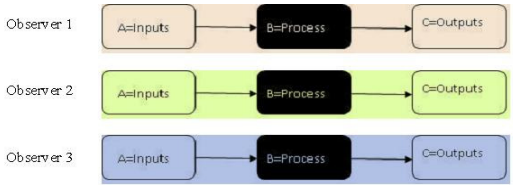


Figure 3: *Analytical thinking: from the left to the right*

An example of the three observers with analytical consciousness can be seen in Fig.3. Each of the observers wants to reach the outputs C; they are parallel in time but in different places and want to achieve different outputs, since they have given different inputs (marked by different colors). They have analytical consciousness, without thinking of interconnectedness and acceptance each other's point of view. None of them thinks of the environment, just about the outputs and the processes. A typical example of this analytical consciousness can be encountered every day in the real estate business, when someone wants to sell the house with a help of real estate agency. The house value will differ due to many different point of views: the owner's, the buyer's, the real estate agents', the tax officer's, the evaluator's and the neighbor's. Analytical thinking can bring compromises but seldom consensus among observers.

3 SYSTEMS THINKING

One of the definitions [10] of systems thinking says that system thinking emphasizes looking at wholes rather than parts, and stresses the role of interconnections. It is a circular and focuses on closed interdependencies, where Observer 1 influences Observer 2, Observer 2 influences Observer 3, and Observer 3 influences Observer 1. It opens a window on our mental models, translating our individual perceptions into explicit pictures that can reveal subtle yet meaningful differences in viewpoints. System thinking shows a big picture perspective, one of wholeness, which is connected to patterns and finally to the structure. Patterns are trends, or changes in events over time. Whenever one sees a pattern of events, one comes closer to grasping the systemic structure driving that pattern. Structure always gives answers to the questions why a pattern is happening or what is causing an event. Thinking at the structural level means thinking in terms of feedback connections. Systems thinking means understanding the patterns and structure that lie below the surface of the "iceberg".

	Action Mode	Time orientation	Way of perceiving	Questions you would ask
Events	React!	Present	Witness event	What's the fastest way to react to this event now?
Patterns	Adapt!	↓	Measure or track patterns of events	What kind of patterns of events seem to be recurring?
Structure	Create change!	Future	Causal loop diagrams and other system thinking tools	What structures are in the place that is causing these patterns?

Figure 4: Levels of Understanding [10]

Figure 4 shows the richness of the three levels of understanding: reaction to an event, adaptation of patterns and creating change going to the structure.
The real power of structural-level thinking comes in the fact that actions taken at the level of event are creative, because they help the observer to shape a different future, the future that he wants. His ability to influence the future process increases as he moves from event-level to pattern-level and to structural-level thinking. The art of thinking at the systemic structure level comes with knowing when to address a problem at the event, pattern or structural level, and when to use an approach that combines the three. [10]

4 SYSTEMS VS. ANALYTICAL THINKING

In the beginning human beings experienced themselves as one with the nature. To survive they needed to understand and control the world. This kind of thinking soon become predominant and the experience of one with the nature ("oneness, wholeness") were lost. Breaking things down into parts, analytical thinking became the way how people thought. Mass production is an example of analytical thinking. As people left farms and went to work to the factory, they learned to do isolated tasks the way engineers wanted them done [11]. The essence of systems thinking is to focus on the whole. The parts are no longer primary focus. The parts are essential but what is more important is the interrelationship between the parts as they work together to fulfil the purpose of the whole system. According to Checkland [12], "scientific thinking is almost synonymous for analytical thinking. There are three senses in which science is "reductionist". The real world is so rich in variety and messy that it is necessary to simplify it, to select some items to examine out of those which could be looked of. Secondly, William of Ockham emphasized there is much to be gained by being reductionist in explanation –Ockham's razor. Thirdly scientific outlook absorbed Descartes's advice to break down problems and to analyse component by component. Meadows[14] says that ever since the Industrial

Revolution, Western society has benefited from science, logic, and reductionism over intuition and holism. Systems thinking is intended for people who may be wary of the word "systems" and the field of systems analysis, even though they may have been doing systems thinking all their lives. On the other side, handling independent variables is the essence of analytical thinking [13]. Understanding interdependency requires a way of thinking different from analysis; it requires systems thinking. Analytical thinking and systems thinking are quite distinct. Analysis is a three step thought process. It takes apart that which it seeks to understand, then attempts to explain the behaviour of the parts taken separately, and finally it tries to aggregate understanding of the parts in to an explanation of the whole. Systems thinking uses a different process. It puts the system in the context of the larger environment of it is a part and studies the role it plays in the larger whole. Systems thinking and analytical thinking will come to be thought of as twin components of scientific thinking [12].

5 TRANSFORMATION OF CONSCIOUSNESS: FROM ANALYTICAL TO SYSTEMS THINKING

According to Hitchcock, consciousness arises as a man develops more accurate images of the world and of himself as a part of that world. That accuracy includes the depths of reality, both inner and outer, and the depths of psyche as well [15]. Transformation of analytical consciousness to systems consciousness takes a certain time. Calleman [5] describes transformation of consciousness in a shape of the Mayan calendar, where he explains nine levels of the Earth consciousness, presented by Mayan pyramids as nine step models. For development of analytical thinking are the most important, 7th and 8th levels, which represented the strong development of analytical thinking in science and society, whereas the 9th level represents the rise of systems thinking (conscious co-creation).

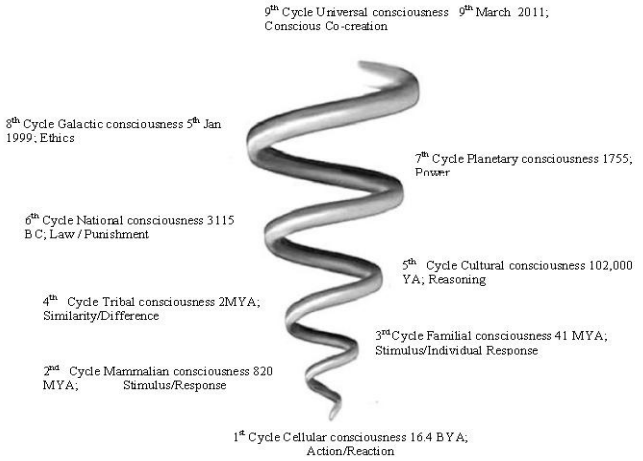


Figure 5: Cycles of *Consciounesss Evolution [5], [6]*

Figure 5 shows the Mayan nine-level spiral (originally as a pyramid *[5]*), system, which represents the evolution of consciousness. The two dates for the big bang are alternative estimates of the age of the universe based on different assumptions. Calleman's assumption for the first level started with the Big Bang, 16.4 billion years ago, and ended 820 million years ago when the first living cells appeared. In a recent study based on WMAP[1] measurements, the authors settled the age of the universe as 13.7 billion years, but include a range of estimates up to 16.5. billion years, based on different assumptions as to the nature of the universe [5, 16]. On the other four steps, consciousness evolved from stimulus/response, to stimulus/individual response, to similarity/difference, to reasoning. This was started by the shamans of the tribal groups when they developed reasons as to why things are/were the way they are. They have since then evolved into priests and religions through the cultural cycle and formed a single pantheon with the religion of Egypt. The last four steps of consciousness as an evolutionary system are the most important ones, since we can find parallels with events that happened and happen today if we carefully observe this system. The sixth step of consciousness, which began in 3115 BC, was law and punishment and lasted until 1755 AD. The concept of good and evil developed (Adam, Eve and the apple – the idea of retribution) but also the Laws of Nature and Science were discovered. The seventh step of consciousness from 1755 to 1999 was a consciousness of power, in which there was no place for integration but analyzing, separation [6]. Leading minority, industrialists anticipated but linearly: Now I am here and this will be my future, no matter what. From A to C and in between was the B, where they tried to influence the process. The same was with leading minority: royalty, aristocracy. No one thought of the environment (people, nature ...). Their consciousness was not able to anticipate in synthesis, where they could see themselves as a whole with the environment and its concern regarding their visions for the future. As a note there is no good or bad power as consciousness has no good or evil, it is the individual viewpoint of the application of power that makes the use of power good or evil. To the universal consciousness everything is just an experience – there is no good or bad experience) The evolution of human consciousness did not allow the systems thinking to appear as common thinking before the time. This time came around fifties with Wiener, Bertalanffy and many others. The consciousness that evolved at the 8th level during 1999 and 2011 was a consciousness of ethics, in which all the towers of manipulation of negative power started to collapse. We have reached the 9th level of consciousness, which is a consciousness of co-creation, a consciousness of thinking in systems. For strictly analytical minds, the present time of

[1] The WMAP (Wilkinson Microwave Anisotropy Probe) mission is designed to determine the geometry, content, and evolution of the universe via a 13 arcminute FWHM resolution full sky map of the temperature anisotropy of the cosmic microwave background radiation.

systems and co-creative consciousness will be very demanding, since systems lead to, interdependent positions among elements, including people; they influence each other and are aware of co-creative, responsible and complementary roles among each other. The transformation from analytical to systems consciousness brings natural thinking in systems, which always takes into concern the environment and the feedback information, Fig. 6.

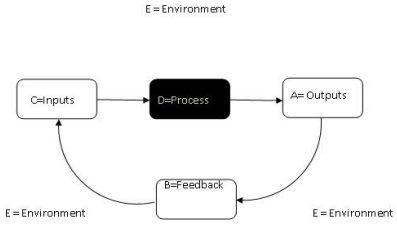


Fig. 6: *Systems thinking: from the right to the left*

Transformation from analytical thinking to systems thinking is visible, since the observer uses as his primary questions the questions about the influence of his vision or (A-outputs) to the environment (E-other people, nature, society), uses feedback information (B-what will my vision bring to the E) and asks himself what will my vision (A-outputs) bring to the environment (E) and what is the current situation (C-inputs, ideas, teams, co-creation) for achieving the (A) and how can I help in the process (B), either with help or without any worries if he cannot influence the process. The essential differences between the systems and analytical diagrams lie in the beginning of observer's thinking; the analytical observer starts to think from left to the right and this thinking leads him in a linear direction to the outputs, regardless of the consideration of environment or the questions of what the outputs bring or what the observer wants to achieve. The observer with systems consciousness starts to think at the right side, anticipating the outputs, their impact on the environment and their feedback for the present situation and inputs as well as the process. The complementarity of both types of thinking lies in the fact that analytical thinking mostly concerns visible events and consequently short-term solutions; systems thinking considers both: short- and long-term solutions, which as a necessity uses the results of the analysis and analytical thinking.

5 CONCLUSION

This paper has presented facts about analytical and systems thinking, the differences between them and the consciousnesses that are present at both. We showed analytical and systems thinking as phases in observers' growth and as periods in time of the earth evolution of consciousness as researched in *[5]* and *[6]*. We presented the differences and complementarities of analytical and systems consciousness. Consciousness is a natural path, which can be accelerated with the knowledge and the awareness of existence of systems. Systems thinking awareness leads towards systems consciousness, which is necessary in everyday life, when searching for solutions in industrial, human or societal realms. Analytical thinking can be of great help when used in order to newly synthesize analyzed elements. Thus, we come to the conclusion that if the observer with an evolved systems consciousness starts to connect, synthesize, interconnect, he must have co-creative ideas and solutions, in which he understands all points of view, also analytical ones, which in this case become the parts of systems solution and this way the complementary ones. The results achieved with conscious co-creation are optimal for the environment of today and of the future.

REFERENCES

[1] Ackoff, R. (1999) *Ackoff's Best: His Classic Writings on Management.* New York: Wiley& Sons.

[2] De Mello, A. (1992) *Awareness;*Doubleday. New York.

[3] Lungold, I. (2004) The Condor Flies to Eagle, lecture in White Horse, Canada. June 22nd 2004

[4] Farthing, G. W. (1992). *The Psychology of Consciousness.* Upper Saddle River, NJ: Prentice Hall.

[5] Calleman, C.J. (2004) *The Mayan Calendar and the Transformation of Consciousness*; Rochester Vermont: Bear and Company.

[6] Jere Lazanski, T. (2009) Systems thinking : ancient Maya's evolution of consciousness and contemporary systems thinking. V: DUBOIS, Daniel M. (ed.). *Computing anticipatory systems : CASYS '09 - ninth International Conference, Liège, Belgium, 3-8 August 2009*, (AIP conference proceedings, vol. 1303). Melville, N.Y.: American Institute of Physics, 2010, p. 289–296.

[7] Haines, S. (2005) *Pearls of Wisdom*; Haines Centre International, pp.2–5.

[8] Lungold, I. (2004) The Condor Flies to Eagle, lecture in White Horse, Canada. June 22nd 2004.

[9] Falk, D. (2008) *In Search of Time: The Science of a Curious Dimension*; Thomas Dunn Books. New York.

[10] Anderson, V.,Johnson L. (1997*) Systems Thinking Basics: From Concepts to Causal Loops;*. Pegasus communication. Williston.

[11] Papageorge, A. (2004) *GoInnovate!* The 13 Essential elements for generating Innovation in Your Workplace; GoInnovate Publishing. San Diego. USA

[12] Checkland, P .(1999) *Systems Thinking, Systems practice* ; John Wiley & Sons. Chichester. UK

[13] Gharajedaghi, J. (2006) *Systems Thinking: Managing Chaos and Complexity*; Butterworth-Heinemann. Burlington. USA.

[14] Meadows, D. (2001) *Thinking in Systems: A Primer*; Chelsea Green Publishing. Vermont. USA

[15] Hitchcock, J. (1999*) Healing our Worldview: The Unity of Science and Spirituality;* Chrysalis Books. West Chester. USA

[16] Tegmark, M., et al., Cosmological Parameters from SDSS and WMAP, Phys.Rev.D69:103501, 2004.

INFLUENCE OF COPYRIGHTS OVER THE RELATIONSHIPS IN eLEARNING

Ivan Pogarcic, Marko Pogarcic**, Matej Pogarcic****
*Polytechnic of Rijeka, **University of Rijeka, Faculty of Law,
***University of Ljubljana, Faculty of Civil Engineering and Geodesy
Trpimirova 2, 51000 Rijeka, Croatia
Tel: +385 51 35 37 53; fax: +385 51 67 35 21
e-mail: pogarcic@veleri.hr, knjo@net.hr, tkoma@net.hr

ABSTRACT

Teaching is a concrete realisation of two combined processes: learning and teaching. These processes are supplementary, they create unity and their synergic bonds cannot be broken. Teaching can integrally be observed as meaningful and intentionally organised system or as a concrete realisation of relationships between its basic elements: students and teachers who are gathered around context with clearly defined goal. Relationships are complex and multilevel starting from a human level in relationship student-teacher up to modern frames where these relationships attain virtual characteristics. The common goal – adoption of required knowledge and acquirement of specific skill – corresponds with all three basic elements in different manners. Student and teacher approach to goal within specific context from different points of view but with the same motivation: the fastest possible and the most efficient adoption of contents. Usage of ICT implies new problems in teaching process. The most evident problem is ownership over educational content and consequently influence on relationships between elements that comprise teaching. This paper aims at indicating several, primary legal, but also ethical and moral aspects of teaching, mostly of its form presently familiar as eLearning or mLearning. To answer the question "Does teaching have an owner?" is not simple, because it implies the question of which part of teaching should be owned.

Keywords: eTeaching, eLearner, eLearning, ICT, intellectual property, authority

1 INTRODUCTION

Teaching and learning are supplementary processes. They are tightly connected into one complexity named class. Class as a whole, that is meaningfully and purposely organised system, comprise basic elements: student/attendant and teacher/tutor who are gathered around learning content. The goal is to adopt needed knowledge and to acquire certain skills. Goal is common but it corresponds with all three mentioned elements in different ways. Attendant and teacher approach goal from different standpoints but with the same motivation: the fastest and the most efficient acquisition of content. Realisation of necessary activities is being done through processes of teaching (a teacher) and learning (a student). Context in which the class is being executed is the fifth important process element [1]. Attendant and teacher are, on the other hand, elements that determine the other elements in different ways. Put in the interrelationship by articulating their goals, they primarily decide upon learning content and indirectly the context of class realisation. The both processes have during the history, by influence of different factors, changed and qualitatively modified. That is actually special dimension of the processes' dynamics. The both processes, individually or together, make a continuum in which an access to their organisation, form of realisation and applied technology and tools influence the acceptance or dismissal of certain paradigms in learning/teaching. Attendant and teacher are basically main initiators of learning as process, so they are put in a relationship that defines their specific rights and obligations. Clearly, such process should be, due to obvious reasons, legally regulated or a proper realisation of obligations and protection of rights derived from a process or emerged as its result/product should be ensured.

Aim of this study is to implicate some, primarily legal, but also ethical and moral aspects of learning, especially its form presently known as eLearning.

2 IS CLASS A BUSINESS?

Increase of knowledge in general, and technology improvement specifically, influence all aspects of human activities, the class included, with different dynamics and different intensity. Consequences of this trend are: increase of knowledge quantity combined with the requirement to shorten time needed for acquiring the same. Especially underlined is the need for continuant adjustment to new situation through permanent updating of the learned and completing it with new facts. The latter is today known by its popularly used name – the lifetime learning. On the other hand, increase of knowledge directly reflects learning

content by depriving its static, characteristic to traditional forms of class. Not to misguide one, learning content has never been completely static, but has also never had such dynamic provided by the information communication technology (ICT) usage today. For instance, acceptance of allegedly object-oriented paradigms has the most drastic influence over the learning content and mode of its preparation. Implementation of ICT into a specific process determines a form of its realisation. Accordingly, the application of technology defined by methodology names the efficiency and completion of the process realisation. Need for a constant application of new facts has influenced traditional forms of class. Every society, for the purpose of its sustainability, defines and controls obliged forms of class that help the individual to grow, maturates and acquires certain skills that ensure existential and some other needs. However, an urge for lifetime education, aside modern technological possibilities demands also the forms of class presently often called informal education [2]. In such a manner, eLearning as form of class in lifetime education can also emerge in different sorts of education as compulsive, vocative, supplementary and additional/hobby education. Clearly, such form of class demands, but also enables, a different approach. Education has, with enormous help of ICT, become a part of business that can have its own marketing and own market. Still, in such way a legal aspect of learning has been emphasized that no longer has to be a part of a competent ministry, but it certainly requires a legal regulation.

3 WHO ARE THE LEGAL SUBJECTS IN eLEARNING?

If eLearning is treated as business or, better yet, as product that can in specific way be consumed, then such a product should be commercialised. By commercialisation eLearning becomes a product that requires its buyers/consumers. Naturally, all characteristics that will ensure a qualitative market niche for such a product should be provided. This indirectly means that specific form of class supported by ICT should have checkable qualities, commercial protection etc. Clearly the mentioned primarily applies to learning contents but also to a form of realisation that can vary from one case to another. The main question here is: Can knowledge be owned? An answer to that question demands proper description of main participants on learning and their rights. For instance[3], the authors have a long time ago recognized a difference between learning and teaching. Luckily, eLearning has, with the help of ICT, largely eliminated that confusion. In formal education, comprehensive and vocational, legal aspect have been solved in a higher level, mostly state, since that form of education has a vital importance for development and sustainability of a society. Informal education requires a special regulation of participants' rights and obligations. Service Level Agreement or abbreviated SLA is form of legal act that regulates such relationships. Such form of relations' regulation is especially important in ICT environment. eLearning requires such regulation precisely because of the proclaimed characteristics of time and location independence. Proper definition of relations ensures more qualitative results to all participants [4].

Authors believe class can be treated as information system, at least from a standpoint of students, simply because student during a class acquires new knowledge, skills and competencies. This simultaneously means that student isn't acquainted with information that are offered and that quality of these information should be ensured by a person who prepares learning contents and learning process. Quite contrary, this quality should be guaranteed, since it will influence the realisation of class's goal – an important element of learning. Learning contents have to fulfil several conditions that ensure their quality. From the aspect of this study, important is to mention a fact that learning content is treated as applicable, in several stages, on the computers. That means such forms can appear in forms conditionally named as computer program – though it is much more then that. Such modes can be found as LMS or CLSM (Learning Management System, Content Learning Management System). Goals and context in which class is being held will be used as a frame in which legal regulation of broader social context will be applied.

4 eLEARNING, COPYRIGHTS AND INTELLECTUAL OWNERSHIP

Hence, does knowledge has an owner or can cognition be held exclusively by an individual and can he hold rights over certain knowledge? Here the concepts of knowledge and cognition are purposely equalled by giving them certain material characteristics that can define them as a property. In a cognitive process a common start refers to idea. Idea can have an owner, but not the exclusive one. The same idea can be developed by more individuals. Still, the realisation of idea has commonly a material ground, so it can have the first and exclusive owner. If eLearning is now observed as realisation of someone's idea, clearly that realisation has an owner and his rights over the property should be protected. However, this indirectly means that ownership can be obtained over the "package" but not over its content. That is the point of eLearning.

Many different authors will develop certain course – CLMS – in different ways, some of them maybe even in the same way, but the final goals will be identical if required by context. The specificity developed by authors as computer form or some hybrid form is their ownership that can be commercialised. Therefore terms-categories: intellectual property and copyright are globally protected. Protection is demanded primarily because of material usage that can be obtained by using a product or because of any kind of misuse.

The Republic of Croatia has, by Croatian Copyright Act (National Gazette, Nr 01-081-03-1344/2, Zagreb, and 7th October 2003) [5] determined as follows:

1. Author's work is an original intellectual entity from literature, science and art that has an individual character, regardless to form and way of expression, sort, and value or purpose if not differently specified by this Act (Article 5, Paragraph 1) and
2. Author of work is a physical person who created a work, and
3. Author owns a copyright over his work by simply the act of creating that work (Article 9)
4. Copyright comprises moral rights of author, property rights and also his other rights (Article 13, Paragraph 1).

Computer program is protected as linguistic work by this Act if it is original in sense that it represents own intellectual property of his author. Term computer program comprises of computer program's expression in any form, including the preparatory designer material. However, ideas and principles that are basis for any computer program's element, including the ones used for creation of its interfaces, is not protected by copyright (Article 109). While intellectual property is a common name for possession of non-material goods that is products of a human mind.

In eEnvironemnt, the learning content is shaped and packaged in forms that are convenient for the ICT usage and realisation of eLearning so legal protection is required due to several reasons. Learning content is in a way "atomised" in forms popularly called Learning Objects (LO). These are in different ways prepared segments of learning contents that can be used by computers. Every eCourse is a combination of these elements. Accordingly, property can be over individual element but also over the complete eCourse.

The other kind of legal regulation applicable in this case is treatment of computer criminal [6]. In this area, Europe acknowledges different approaches that are, mainly based upon Convention of Cyber-crime (http://www.poslovniforum.hr/info/konvencija.asp, 06-09-2011) that represents international legal instrument that has originally regulated problems connected to usage and download of information and data, but also the biggest, the most detailed and the most qualitative European document of its kind. The following table compares several states of the EU, Japan and the Republic of Croatia.

Table 1: Comparison of handling the cyber criminal

State	Specificity of regulation
Federal Republic of Germany	The first to adopt Act of Data Protection The second Act of preventing the economic criminal Act of Copyright 30 years of experience in solving cases connected to cyber criminal
Great Britain	Computer Misuse Act Regulation The basic act of hacking is still a work of a hacker Unauthorised modification of computer content
Japan (shaped on the influence of the USA)	The basis is Japanese Criminal Law Act on unauthorised approach that regulates: - criminal activity of falsification - criminal activity of misconduct - criminal activity of hacking the business transaction over the computer
The Republic of Slovenia	Criminal Act that regulates: - criminal activity of unauthorised approach to a computer system - criminal activity of unauthorised modification of content, destruction or damaging of data
The Republic of Croatia	1997. Croatian Criminal Act was originally introduced into the computer criminal 2002. the Republic of Croatia has passed Act of adopting the Convention of Cyber-crime

What sort of criminal is possible within eLearning? If the CLMS applications can be used by computers, then they can easily be misused. Naturally, protection of copyrights over such products demands treatment provided to other product of similar structure and composition. Again, one should underline the fact that eCourse is a product that will enable eLearning as package of cognitions that shouldn't have owner, so the ownership is possible just over the package. There is one more important circumstance in this area. That is alleged Open Source products or products which authors dispose without insisting on protection of copyrights [7]. Naturally, this also requires specific protection that is imposed only in case of further commercialisation and profit gaining by the same products.

5 RIGHTS AND OBLIGATIONS OF eSTUDENT AND eTEACHER

Improvement in development of ICT technology has caused "a rearrangement of cards" within education as a human need and activity. Influence of technology has mainly reflected on structure of certain educational contents with unpleasant result of changing the competencies in specific professions. The consequence reflects in a problem of competencies of all persons who use ICT technology as a tool or methodical asset. This primarily refers to eStudent and eTeacher. They absolutely have to acquire certain competencies in this field.

For eTeacher this situation can be frustrating de to several reasons. Numerous assignments, even if they are familiar to him/her, couldn't be done individually but in association with others – members of a team for preparation of eClass. Nevertheless, eTeacher still has an obligation to shape a pedagogical-didactical-methodical form of class. Legally,

this scenario predefines situation that is clear on the institution's level. Another option is more dangerous for teacher's creativity. If there is an adequate form of SCORM, teacher will arrange his/her puzzles with a danger of loosing his/her own creativity. This case requires a legal regulation that covers copyrights of those who created SCORM and obligations of its users.

eStudent is also not completely free of rights and responsibilities. However, the student is, at least in comprehensive and vocative education, exempted from certain obligations according to his/her age. Responsibility of student is conferred, up to a certain age, to another person – parents or teachers. Situation slightly differs within higher education and informal education. eStudent or informal ePupil can, same as eTeacher, use some educational objects out of SCORM. Naturally, in this case, student has to consider copyrights and proper use of authorised educational objects in comparison to unauthorised trespass.

eClass, through collaborative and cooperative form of work, can put eStudent into a position of idea initiator of some qualitative and later on, generally accepted form of class. In that case, eStudent becomes an author of educational object, and his/her rights should be protected. Though, it can be concluded this opportunity is rare, it still exists, both in theory and practice. In most faculties students through their obligations towards class create different "products", right paper works, undertake researches and create projects. In most technical faculties students use different CAD and similar computers that help them to create valuable solutions. This certainly requires protection of their rights.

6 CONCLUSION

eClass as a form of class will probably prevail du to several reasons, however, these won't be elaborated in this study. Authors believe that education as process, in organisational sense, within these circumstances significantly moderates so it requires radical changes of educational and realisation determinants. One can almost say "nothing is the same anymore". Naturally, that is not the case.

It should be emphasized that legal regulation of education can have a basic skeleton regardless to level or purpose of education, but, precisely because of its level and purpose it requires certain supplements and additions. Comprehensive education, starting from kindergarten to high education is regulated completely by competent ministry. Still, by a gradual penetration of private sector in this area, regulation should be adjusted to new circumstances. Aside from diversification to state and private sector, which is important in this study, it should be mentioned another classification of education: one required by profession of its participants and another out of hobby. Rights of participants are specific, especially in the latter.

Cyber criminal acts can be hard to prove: due to specificity of technological system, short period provided for proving the criminal act, international components, etc. Its comprehension demands appropriate profession knowledge. This, on the other hand, demands a help of experts, aside the appropriate expertise of the police, state law, and courts.

References

[1] D.D. Pratt, Good Teaching: One Size Fits All? In Issue *New Directions for Adult and Continuing Education*, Vol. 2002, Issue 93, pp. 5–16, Spring 2002

[2] D. J. Trump, M. McIver: Think Like a Champion: An Informal Education In Business and Life. Vanguard Press, 2009, ISBN-10: 1593155301, ASIN: B002QGSY1S

[3] R. Ackoff , D. Greenberg: Turning Learning Right Side Up: Putting Education Back on Track , Pearson Prentice Hall, 2011, ISBN-13: 978-0132887632

[4] S. Emmett, B. Crocker: Excellence In Procurement: How To Optimise Costs And Add Value, Publisher: Liverpool Academic Press, 2008, ISBN-13: 978-1903499405

[5] National Gazette, Nr: 01-081-03-1344/2, Zagreb, 7th October 2003

[6] C. Easttom, J. Taylor: Computer Crime, Investigation, and the Law, Course Technology PTR, 2010, ISBN-13: 978-1435455320

[7] B. Ö. Czerkawski (editor): Free and Open Source Software for E-Learning: Issues, Successes and Challenges, Information Science Reference, 2010, ISBN-13: 978-1615209170.

SAVE THE CULTURE AND THE CULTURE SAVES YOU. THE ROLE OF CULTURE AND ART IN MODERN SOCIETY

Monika Jadzinska
Academy of Fine Arts in Warsaw, Poland
Faculty of Conservation and Restoration of Works of Art
Wybrzeze Kosciuszkowskie 37, 00-379 Warsaw, Poland
Tel: +48226251251; fax:+48226229649
e-mail: monika.jadzinska@gmail.com

ABSTRACT

This paper presents the way of changing the role of culture and art in creating the identity in societies. The new forms and range of participation of people in the sphere of culture create new challenges and difficulties. At the same time the protection of the cultural heritage is to be enforced. Some solutions and tools are proposed. A few interdisciplinary, long-term projects are mentioned as the examples.

1 INTRODUCTION

At the early 21st century societies from all over the world are facing an overwhelming number of challenges: demography changes, climate changes, economical changes, technological changes, globalization. Alain Touraine, the Nestor of French sociology, writes in his latest book *Apres la crise* (*after the crisis*) that the public no longer exists in the meaning of being a whole, homogeneous organism, linked together in the name of interests of classes or groups [1]. The only platform that determines the individual and collective identity today is culture and art, and they are of strategic importance.

Culture and art can be considered to be the most significant forms of human creativity. At the same time they are just like a barometer of its time, the most complete document of spiritual, social and material civilization, the mirror of transformation of the human being, nature, science, religion etc. and a component of individual and collective identity. We need to support, develop but also to protect them. Therefore, they should be protected and conserved to survive heritage of our and our ancestor's time to the future generations. In order to prevent them against combined risks, the concerted actions are needed.

2 EVOLUTION AND CULTURE

Up until the Neolithic period, the development of culture and cultural characteristics was taking place at the same slow pace as genetic evolution. After that, the cultural evolution, as a part of development of civilization, started to accelerate. But the pace of change was still slow. In paintings, for example, in the Italian Trecento, even after phenomena of Giotto, all the rest were very uniform. Quattrocento was already more varied, and Cinquecento distinguished several schools. The pace of change became faster and faster. In 17th century we have got a few big strands (realism, neoclassicism, baroque), in the 18th century – a large number of strands, and in the 19th century – an increasing number of strands evoked and supported by new technological and material possibilities [2]. And we reach 20th century, when artistic programs seem to be like a huge kaleidoscope of ideas, forms, matter, and meanings, inspired by a variety of life, disciplines, media and many other factors. All that changes evoke different approaches to the problem of preservation of the work of art.

The role of art also has changed. From the early Renaissance to the 19th century art, according to the concept of Vasari, grand historiographer of Italian art, painting was "a window open to the nature". Visually basing on the nature, it was still an isolated form that has got its own space, light, shadow, and narration. The function of an artwork was strictly connected with religious, ideological, documentation, hedonistic or decorative features. We can observe the transition of the role of art from taking it only as an object executed according to the technological and material rules for specific function, through the cult of the artist's individual character, until the dominant of the context (social, motion, historical, artistic and others). This change, by all means, the methods of cognitonics lead us behind hard dogmas, to the acceptance of the entire range of the values and cultural diversity.

Modern and contemporary art began to be an interpretation of reality through the use of a system of characters to represent the real and metaphysical world. It means that it is a reflection of transformation and diversity. New opportunities in technology, the process of globalization but also heterogeneous system of values create a new object of art as a complex system of features and meanings. We can say that new division of art is not due to the type of material or technique but due to the manner of receipt, the participation and creative activity of the viewer [3].

3 NEW FORM OF PARTICIPATION

3.1 New type of perception

"Creativity is rooted in the combination of contemplation and action" wrote Vladimir Fomichov and Olga Fomichova [4]. In art, through the ages, activity was, we can say: "one-sided". The only active side was its creator – an artist. The viewer could only contemplate the artwork. Improvement of techniques was carried out in parallel to the innovation that impact on the visual perception of the viewer. More or less since the mid-19th century technical and material innovations, as well as tools for the collection and exchange of information, communication and visualization, has created new forms of artistic expression of vision, ideas and thoughts. The existing model of perception of art based on the contemplation became inadequate for understanding the condition of the cultural creation and its reception in today's world. In the 1930s American philosopher, John Dewey, in his concept of esthetic pragmatism, merged the idea of art with direct experience of life [5]. A work of art no longer is considered to be a matter of aesthetic, but by means of a close relationship with daily, ordinary forms of life in terms of "art as experience". Collection of works of art by the audience took another dimension. The artist began working in a particular area, a single situation in which he had to penetrate a viewer's mind and constitute themselves on the basis of his experience. There has been a transfer of the task from the object to the viewer. A further step was to establish the significance of the presence of the audience, which was already referred to. Robert Morris was the first, who described and identified a completely innovative contribution to minimalism in art. In the contrary to the traditional art, which requires only observation and contemplation (passive stance), the modern and contemporary art forced a personal involvement of the recipient, who is standing between an "active observation" and "active co-creation".

3.2 Commercialization of cultural heritage

The process of commercialization of culture constitutes a threat, while at the same time many opportunities. Tourists established the chance for cultural property, being the source of economical support, but at the same time a crowd of people determine a great risk for fragile, susceptible to deterioration, cultural property.

3.3 Art in the digital age

The Internet is an ubiquitous medium demolishing all divisions (geographical, economic, classes etc.), giving the possibility of the existence of everything and everyone. The difficulty is quantifying – anyone can issue their tracks, from day to day to become a star. However, this is a fascinating opportunity to easily reach a common contact with culture. Hence the organizations involved in the development of culture, just like galleries and museums, have increasingly richer offer on their website, not limited only to the information about the collection, description of historical material, portraits, but entering the educational departments in the form of an intelligible and an easily comprehensible specially for young generation (short movies, the games, fun activities).

4. THE NECESSITY TO PROTECT CULTURAL HERITAGE

We need to enforce our efforts on the protection of our common (for example, European) and the national cultural heritage. There is a huge responsibility to take care of our own heritage, also in terms of local, national, small communities. Future generations will estimate our time basing on the heritage that we have left them.

4.1 What do we want to preserve?

Care of cultural heritage is based on preservation of its authenticity, which is one of the most important factors conditioning the value of an artwork and the manner of proceeding in the scope of conservation theory and practice [6]. Studying the evolution of art we can discover that the concept of authenticity appeared to be an agent that determined the form in which a work would be preserved [3]. It has influence on all research, decision-making models, as well as methods of preservation, display and active conservation-restoration of a work of art. The twentieth century brought about a set of rules, according to which the original matter possesses the absolute value. But very soon it proved to be not sufficient and the criteria of authenticity changed, enriched by cultural diversity and taking into consideration elements of intangible cultural heritage [7].

Contemporary art being beyond the limits of traditional media, giving the concept of a priority over the materiality, challenging new issues arises for conservators used to focus on the material degradation, following the huge tradition from Alois Riegl to the Venice Charter. In contemporary art the message established in the course of this creative process gained much more significance than the final product. What took place was the transfer of focus from the artists' creation to the artists themselves and cultural, social, political and geographical background [8].

5. PROBLEMS AND EXAMPLE OF SOLUTIONS

We still observe too small interest of the culture in society. It started in the process of education of children and young people. It can be noticed, for example, as a disastrous reduction in the level of readership or lack of interest of what has happened to the cultural property in global and local scale.

5.1 Solutions

- changing the role of culture as a fundamental platform for social development and integrity
- more care of cultural heritage (national and international programs for taking care and conservation of cultural properties), systematical protection with regard to climate change, risk assessment and other changes
- use the potential of culture and art as a real chance for the region and people living there - note that cultural institution creates new quality in the place (city) that was not associated with art so far, but now is blooming just because of it. Tate Modern – 100 million pound a year, 3000 new workplaces, 23% growth of turnover in hotels and gastronomy; encouraging better collaboration between public and private sectors, as well as open innovation between different research activities and business sectors
- awareness of the benefits and risk connected with tourism - cultural patrimony attracts millions of visitors every year to monuments, historical places, museums and galleries, archeological sites
- culture-oriented education of children and youth
- science on active duty of culture and art
- note the changes within the same art - the objective is to call the active participation of people in a work of art in culture. Art in the urban area (using the commercial opportunities and tools); art as a space of experiment, which can lead to the creation of new forms of social and economic life; art as a cognitive-emotional sphere of life
- new role of culture institution – for example some of museums in Poland (Museum of Modern Art in Warsaw and Wroclaw), in spite of the fact they have got no building or collections (or rather small), are very active, also in a public space reconstructing social sphere by promotion the culture and art
- mobilization of social energy in restoring the culture proper place in the life of the country – for example, the Pact for Culture in Poland, the platform which is a social movement that set up the citizens of the culture during the Congress of Polish Culture in 2009, in the name of the idea that one percent of the state budget should be located into the culture initiatives. And they persuaded the government to do it in 2011.
- Activation of small communities into their own tradition, the tradition of the village, city, region. Giving the social dimension, values, commitment to the protection of cultural heritage

5.2 Tools

We need to create a long-term research needs and objectives in the area of preservation and use of cultural heritage in the context of global change. Some international and national projects are planned to establish priorities, timelines and specifying instruments and resources required for its implementation. We must be aware that cultural heritage needs protection and conservation responses executed not only in the traditional, active manner but also in an extensive form of sparring knowledge, experience, information, resources, best practices, methodologies and guidelines.

6. EXAMPLES OF PLATFORMS, PROJECTS OR SMALL INITIATIVES IN THE SPACE OF PROTECTION OF CULTURAL HERITAGE - IN THE GLOBAL, EUROPEAN OR LOCAL SCALE

JPI

"The JPI on Cultural Heritage and Global Change: a new Challenge for Europe (scientific actions for next 3, 10, 25 years) has recently defined a common vision which needs to be translated into a strategic research agenda (...) A Coordination Action is needed to allow Member States and Associated Countries to implement the Joint Programming for providing jointly areas where public research programs can respond to major societal challenges". The aim of creating such platform is to implement the action programs for the area of research on cultural heritage in a changing world, protect the World Cultural and Natural Heritage and the Safeguarding of the intangible Cultural Heritage. Climate change, other environmental changes, human activities and security risks threaten Europe's cultural heritage. Cultural heritage contains assets which are unique and irreplaceable in their tangible forms of historic buildings, collections, sites and movable objects as well as in their intangible value, which includes history, collective memory and identity.

Identity Card-CHIC

The aim of the EU project is to set CHIC - a system of the <Cultural Heritage Identity Card>, which will introduce a systematic collection and storage of data on immovable heritage objects across European and neighbouring countries (www.eu-chic.eu). The main objective is to develop and test guidelines, needed for efficient compilation of data, pertinent to each monument under observation. The system of CHIC will support sustainable maintenance, note conservation and rehabilitation of historic sites and monuments in the space of time-varying alteration, caused by human interventions and environmental impacts. Partners of the project will promote and assist an introduction of the CHIC system in their countries and further facilitate its use in the neighbouring countries through their links with governmental authorities, responsible for their cultural heritage protection and preservation.

International Network for Conservation of Contemporary Art INCCA - sharing knowledge, experiences and ideas
The aim of a few UE projects (started in 1999) was to create a networks of professionals connected to the conservation of modern and contemporary art (conservators, curators, artists, scientists, researchers, archivists, art historians and many others professions) and continue the actions to protect modern and contemporary art (Inside Installations - Preservation and Presentation of Contemporary Art; PRACTICS, ACCESS2CA). INCCA is a digital platform for information, knowledge, experience exchange. Members allow access to each others unpublished information (artist interviews, condition reports, installation instructions etc) through the INCCA Database for Artists' Archives. Till February 2011 the network grew from 23 to almost 600 members (including independents and students) from 400 organizations in around 50 countries. "This INCCA website www.incca.org contains information on all kinds of activities in the field; projects, seminars and conferences via the news section, educational possibilities as well as links to numerous websites. The resource section provides direct access to good practice documents such as theoretical articles, case study descriptions and practical guidelines and formats".

NET-Heritage
European Network on Research Programme applied to the Protection of Tangible Cultural Heritage. Net–Heritage is a significant initiative to coordinate national research activities (Research and Technological Development (RTD) programs of European countries in the field of protection of tangible cultural heritage (www.netheritage.eu)

LOCAL, NATIONAL SCALE
Children and young people educated in the field of cultural heritage and history
The example: The Warsaw Rising Museum which has prepared an extensive educational offer. Apart from offering museum lessons for schools, encourages participation in numerous certified factual and methodical training courses, open lectures and seminars as well as projects surpassing traditional teaching methodology and in-service training courses but also several different competitions. A unique initiative is the Little Insurgent's Room, it is a portion of the exhibition dedicated especially to children. The youngest visitors can start their "history adventure" in an area adjusted to their physical and psychological needs, among toys and games related to the historical epoch, cared for by well trained instructors.

"SAFE IT FOR THE FUTURE – INTERVIEWS WITH WARSAW'S ARTISTS"
National Polish project: *Save it for the future - interviews with modern Warsaw's artists* (ed. I. Szmelter and M. Jadzinska) consists of ten interviews with Warsaw's most respected artists (Magdalena Abakanowicz, Miroslaw Bałka, Stefan Gierowski, Aleksandra Jachtoma, wife and pupils of Aleksander Kobzdej, Teresa Pągowska, Jacek Sempoliński, Wieslaw Sadley, Leon Tarasewicz, Mariusz Woszczyński) in order to gather the knowledge on their artistic legacy, technique, materials and approach to problems of existing and destroying. The aim of the project was to protect now-a-days heritage by recognition conceptual and material sphere of works of art of the group of one of the most honorable artists in Poland. It was published in form of album with three CDs.

7 CONCLUSION

Impact of global change forces new awareness and activities in the sphere of culture. We should find and use the possibility to engage people into the culture and art. Parallel to it we should be aware of multiple risks due to ageing, adverse environmental conditions, and human pressure. That inclines to initiate interdisciplinary, integral, long-term projects dealing with the protection the cultural heritage. Treating culture and art as significant forms of human creativity, barometer of its time, the most complete document of spiritual, social, material civilization and its transformation, and also as a component of individual and collective identity, we have to gather our efforts to preserve it for future generations.

References

[1] A. Touraine. *Apres la crise* (*After the crisis*). Seuil. 2010.
[2] M. Rzepinska. *Siedem wieków malarstwa europejskiego* (Seven Centuries of European Paintings). Wraclaw. 1991.
[3] M. Jadzinska. Authenticity in Installation Art, *Ph.D. Thesis.* Institute of Art, Polish Academy of Science, Poland (accepted to print) 2010. In Polish.
[4] V. Fomichov, O. Fomichova. Cognitonics as a New Science and Its Significance for Informatics and Information Society. *Informatica* 30, pp. 387–398, 2006.
[5] J. Dewey. *Art as Experience*. London. 1934.
[6] J. Jokilehto. *Questions about 'authenticity'*. In K. E. Larsen, N. Marstein (eds.), *Conference on Authenticity in Relation to the World Heritage Convention, Preparatory Workshop, Bergen, Norway, 31 January-2 February 1994,* Bergen, pp. 10-21. 1994.
[7] *Convention for the Safeguarding of the Intangible Cultural Heritage*. UNESCO 2003. Paris. http://portal.unesco.org
[8] *Inside Installations. Theory and Practice in the Care of Complex Artworks*. Eds. T. Scholte, G. Wharton, Amsterdam. 2011.

SUSTAINABLE DEVELOPMENT OF HERITAGE TOURISM IN THE INFORMATION SOCIETY

Marija Valčić, Lana Domšić
College of Business and Management "B.A. Krčelić"
V. Novaka 23, 10240 Zaprešić, Croatia
e-mail: marija.valcic@vspu.hr; lana.domsic@vspu.hr

ABSTRACT

This paper suggests that many of the negative effects of globalization and inadequate tourism growth can be compensate by the use of intelligent ICT solutions in the development of heritage tourism. The encounter between cultural tourism and information and communication technologies represents an opportunity to preserve national culture, create partnership and enhance destinations value in information society. To this aim, we propose a model for local Destination Management Systems of heritage destinations (DMS) which could help to achieve a more globally responsible paradigm for the tourism industry and facilitate the management of destinations and the coordination of the local suppliers. DMSs provide interactive demonstrations of local amenities and attractions and enable consumers to build their own itinerary based on their interests and requirements. All the stakeholders within the destination are linked with each other in order to create collaborative action and a genuine, sustained growth in heritage tourism. An example is given of the Croatian World Heritages Sites and their status on the Web.

1 INTRODUCTION

The globalization and changes in modern society represent both opportunity and threat for national cultures, heritage and identity. Cities and regions are loosing their traditional heritage and keeping up with global trends and fashions rather than reviving local traditions, history and values. By doing so they become less attractive places to live, work, visit, and invest in. Therefore, economic development driven by international tourism, especially in smaller communities, may be short lived and have many negative effects.

The tourism industry, facing its everlasting dilemma of growth versus sustainability, can be regarded as contributing to critical trends in world development rather than an instrument of world peace and tolerance (as praised by international organizations like UNESCO) [1]. It is clear that the traditional approach pursued by the tourism business is in need of fundamental revision. The increased cultural diversity of today's world, together with the growing access to the heritage through the Internet, is likely to result in strong pressure for its fundamental restructuring. A technological link should be established between heritage and a sustainable tourism economy based on the cultural richness of places. Heritage should be considered as a bridge between the past and the future of a community, a reflection of founding values, history, and identity [2]. The cultural capital embodied in buildings, artifacts, sights, songs, and rites permits the transmission of the culture of a people through time and space. The heritage industry, small in size, information-intensive and creative provides a means to interact and to learn about host communities.

ICTs and Internet can favor the reconciliation of heritage and tourism, supporting a process of the empowerment of local stakeholders and of creative encounter between host and guest communities. This progress should support creativity, collaboration and appreciation of national and world heritage, taking into consideration the cultural and social capital that is necessary for sustainable community economic development.

2 GLOBALIZATION AND ITS IMPACTS ON THE TOURISM

Globalization has profound implications on competitiveness, trade and tourism policy. In an increasingly global and competitive market, characterized by standardization, cultural crisis and pollution, tourism is today more than ever regarded as a "problem area", something to constrain and regulate, rather that a strategy to pursue cross-cultural integration across the world. Technical progress in transportation has enhanced the physical accessibility of destinations, but this has not been matched by an equal increase in cultural access, the subjective capacity to recognize and attribute a value to the cultural features of the visited places. There has been shift toward commercialization and short-term decision making at the expense of conserving cultural heritage and, in the generality of cases, tourism development has resulted in increased strain from tourist pressure on host communities. It fails to contribute to the elevation of their economic status and shows to be a short-lived option for development. For the area of heritage tourism, a new model is necessary that joins contributions from local institutions and private enterprises that are looking for ways to build on their existing potential, organizing them developing and enhancing local cultural assets, emphasizing the unique character of the place [3].

The transition to an information society creates new opportunities and threats. Information and communication technologies are developing and expanding extremely quickly

and have a huge impact on tourism because they deeply affect its organization and governance and travel value chains and thus the economics of the industry [4]. The long-term success depends largely on its ability to bring about integration at the host community level, enabling the public and private sectors to cooperate and use local resources for development efficiently.

3 ICT SOLUTIONS IN ACHIVIENG SUSTAINABLE HERITAGE TOURISM

A key challenge for the enhancement of cultural heritage is to bridge the different perspectives of how tourism organizations and service providers can present their cultural heritage in a way that appeals to the interests of the international tourism audience. The goal is to generate value from local knowledge and information, and make it available to consumers worldwide. The use of ICT is necesarry and it involves various stages of the operation of the heritage industry like content creation and communication, value enhancement, and market strategy. A basic prerequisite for sustainable tourism is allowing individuals and communities an opportunity to be included and connected. There is a need to develop the aspects concerning the use and development of tools, technologies, and methodologies to facilitate the efficient networking of information and communication systems in tourism. Utilizing the ICT infrastructure, communities are helped to become more strategic and entrepreneurial in managing their heritage [3].

Cultural heritage tourism is increasingly depending on ICTs for purposes of promotion, distribution and delivery of products and services. It also provides a tool for communication between tourism suppliers, intermediaries and consumers. Web-based visitation is becoming commonplace as the tourism industry adopts networked interactive multimedia technologies. It is affecting the ways that tourists become aware of destinations, the ways they select and experience destinations [5]. Tourism information needs an extensive representation of photos and graphics in order to provide a tangible image or experience to travel planners [6]. E-Tourism is maturing fast as a mainstream distribution mechanism and establishing Internet presence and e-commerce strategies will become critical for destinations to remain competitive.

ICT enables heritage destinations to expand their activities in the geographical, marketing and operational sense and play a particularly important role in managing relationship with customers [7]. Much information about cultural heritage already exists in publicly available resources, but can also be incorporated into more structured promotional and educational programs and platforms. The objective is to provide e-content for travel planning and education. Technical partners need to be brought together with local tourism organizations to design and implement Internet sites and portals that will aggregate information and services related to the cultures of each host city in terms of historical perspective, architecture, landscape, fashion, cuisine, and public culture. This joint-venturing process needs to be actively stimulated and supported by the local and regional governments and other nongovernmental organizations, in particular in regions that are now “disconnected” [3].

Web page of the site can be used in various ways and for different purposes including preservation, education and site management. It educates visitors about the site and need for its protection. It enables market segmentation and gives a costumer relevant information, contextualized and supplemented with history facts, stories and related objects and sites. Hypertext links can channel the different parts of the audience to different places on the page (for example, parts specifically designed for children. The sites often have limited space for exhibition and must choose the artifacts that will be exposed while the rest remain in storage. Unlike the real-space sites, web site virtual space is infinite and can be used to display objects that are not exposed at the site. Virtual visit allow access to the site for those audiences who have no possibility to travel to the destination and thus the sites better meet its mission of enabling public access. It is especially appropriate in the case of heritage sites in which physical visitation is discouraged in order to conserve the resource, or is not possible for financial or other reasons. Furthermore, Web sites play an important role in the creation of sustainable tourism. An essential element of sustainable development of cultural tourism is the behavior of visitors at the site. There are different channels for raising awareness. Quality of information about heritage encourages visitors to understand the characteristics of heritage and the need for its protection and helping visitors to enjoy the site appropriately. On the other hand, the promotion can significantly help in achieving financial and educational goals. Good presentation on the Internet can attract more visitors, if the site carrying capacity allows it. Then it increases the profit that can be used to fund educational activities, solve management problems and reach the goals and objectives of the site.

Destinations web sites and electronic reservation systems used in today's tourism industry are being transformed into integrated Destination Management System (DMS). DMS is a set of available interactive digital information about the destination and its related products and services which provide "total tourism product" or "travel experience" [8]. It gives users an access to a comprehensive picture of a destination touristic product, through different channels and platforms, providing extensive information on destinations and attractions, as well as the ability to perform search and booking in real time. DMS also serves as a destination management tool, a tool for marketing and promotion, and support for small and independent providers of tourist services. Furthermore, technological developments can be expected to lead to multichannel multimedia DMS serving purposes not only of travel information distribution, planning and fulfillment, but also of travel-related education and entertainment. Typical heritage destination DMS should include interactive maps, 3D applications, virtual tours, online exhibitions, interactive learning resources, games and fun tools, online collections and databases, user communication, community aspects, personalization and online shops. It should

be developed in the collaboration with educators, art historians, historians, artists, museum specialists, etc. It also provides accommodation reservation and events bookings, personalized navigation, interactive maps, travel journey planner about local weather and public transport. These identified e-Services can vary in their technological development, from some simple interactive structure to more complex interconnected e-services, not only using different media, but also linking and displaying the different e-services together. ITC solutions are not necessarily expensive for a heritage destination. Joining with other destinations in a common web page can significantly reduce costs and also better respond to customer needs. These portals perform two functions. Firstly, they can be seen as a marketing tool that projects the region in perspective to the rest of the world. Secondly, by educating tourists about the prevailing culture of the region, a more satisfactory travel experience can be offered. The value enhancement process is expected to produce a number of outcomes, which can be categorized as "guest satisfaction", "profitability" and "sustainability." These three categories result from the encounter of tourism demand with the supply of local culture. Guest satisfaction results from the intensity and quality of the cultural experience. Satisfaction is also clearly connected with profitability and therefore the cooperation is important at different levels in destination marketing [3].

4 EXAMPLE OF WHS IN CROATIA

For the purposes of this paper we conducted a research of the Web presence of Croatian cultural heritage tourism based on UNESCO World Heritage sites, showing the heterogeneity of this segment of the travel industry and the diverse origins and forms of its presence on the Internet. The analysis showed that most of the destinations are largely invisible on the Web and few provide web-based access to travel planning services. The visibility of a World Heritage Sites on the Internet is key factor in their emergence as a virtual heritage destination that influences the actual physical visitation. The cultural heritage tourism sector in Croatia seems slow to adopt new technologies. There are a number of barriers such as the low level of cooperation between stakeholders, the lack of the strategic vision and business planning and the limited levels of understanding of the eTourism potential. This is affecting the presentation of the country on-line as there is currently no Destination Management System or a comprehensive portal that promotes cultural heritage destinations in Croatia.

World Heritage Sites are considered to be the centerpiece of the global heritage tourism industry. The original purpose of designation as a WHS under UNESCO's Convention was to assist with management and preservation of the cultural heritage site and to encourage the development of management plans. However it is believed to increase tourist visitation [9] and many WHS are becoming major cultural tourism attractions.

First of all, we wanted to see who puts the information about the Croatian World Heritage sites on the Internet, that is, who are the participants in creating images of these heritage destinations. We used Google, currently the most common search engine on the Web, and the term "World Heritage Sites in Croatia" in both English and Croatian. We analyzed the data collected from the first hundred pages that have appeared in Croatian and in English, and we classified these pages in several categories of heritage tourism stakeholders. The result of the analysis is shown in percentages in the table below.

WEB PAGE	SEARCH RESULTS IN CROATIAN	SEARCH RESULTS IN ENGLISH
Tour operators and agencies	16%	17%
Regional and local destinations	12%	2%
National portals, tourism associations and organizations	23%	15%
Media and publications	23%	9%
Events and conferences	2%	0%
Academic and educational sites (Universities, libraries, encyclopedias)	4%	10%
Social web (blogs, wikis, social networks)	4%	17%
Commercial service providers (hotels, restaurants, renting)	14%	10%
Individual attractions (sites, museums, monuments)	2%	0%
International tourism portals and on-line guides	0%	20%

Figure 1. *Composition of heritage tourism web pages*

The analysis shows that intermediaries have a predominate role in the dissemination of information on world heritage destinations in Croatia. Individual web sites offering information on WHS are set up by a wide variety of organizations with overlapping jurisdictions which are promoting destinations for different reasons with the shared interest to increase awareness of the WHS, although not necessarily to maximize physical visitation. There are a major number of national tourism organizations web pages which are often devoted to their development plans, programs, and projects. There are also a large number of web pages of tourism and cultural web portals and various media. In second place we found, in almost equal proportion, local and regional tourist offices, tourist agencies and service providers that help promote the destination with the main goal to increase the number of visitors. The biggest promoters of our heritage sites in English are different international portals and guides. A significantly large role has a social web where users themselves public information via blogs and social networks.

An important fact is that only 2% of the web pages in Croatian and 0% of pages in English are dedicated to individual heritage sites, which means that the majority of destinations don't have their own websites. Also there is no major common web site devoted exclusively to World Heritage Sites in Croatia. The only two destinations that have their own web sites are Plitvice Lakes National Park and the Stari Grad Plain in Hvar. These two pages contain options in several languages, reservations systems, information on local weather and pictures or videos of the site. But they still lack a number of elements to become the real DMS. The quality destination management system will depend, of course, on the maturity of a particular locality, the general awareness among the passengers of the heritage location brand and the degree of development of heritage tourism industry and tourism products and services around the site. For example, Dubrovnik as a historic city already possesses a well developed tourism infrastructure, specialized service providers and additional cultural attractions. Only the technology infrastructure lacks to form a DMS. It is important that persons responsible for site management embrace ICT as an important and necessary element in fulfilling the tasks of protecting, the public presentation and communication of heritage sites.

6 CONCLUSIONS

Cultural tourism development currently presents some very definite unbalances. On one side, it depends on localized and hardly reproducible resources. On the other, it is governed by an industry that is increasingly both global in nature and disconnected from the sources of cultural capital. Like any industry, tourism needs profit and investment incentives to grow, but both commercial interests and government entities should work to achieve a reconciliation of the inherent conflict between heritage and tourism. The potential from tourist growth can only be fully exploited if both policy makers and businesses remove unnecessary structural barriers to growth, by capitalizing on the opportunities that are based on cultural heritage and identity [3]. In other words, a new strategy is required for developing cultural heritage as a viable economic sector. With time, there should be a change from the administered industrial economy to an entrepreneurial economy accompanying and institutionalizing the information society.

Heritage tourism is currently in the process of systematization of information, communication and multimedia as means to achieve a competitive advantage and sustainibility. Awareness and knowledge about a destination and perceptions about its quality and value are factors that influence the motivation of visitors and the selection of a destination [10]. Destinations will gradually have to incorporate the online Destination Management System (DMS) that can improve promotion and management using integrated e-services. New Destination Management Systems can convey diverse, comprehensive and multimedia information on heritage destinations and surrounding products and services, and thus contribute to the destinations long-term competitiveness and sustainability.

In Croatia there are no DMSs at the national level or for the particular destinations, and cultural tourism in Croatia has not yet started to widely use information, communication and multimedia technologies. UNESCO sites are the main points of heritage tourism in Croatia. They don't have their own web pages, but are present at other web sites of special interests that provide limited information. The best solution would be a portal that would unite all destinations in one place and offer a complete information and experience as a base of heritage tourism. Collaborative networking should be established between all stakeholders and both the policy makers and the entrepreneurs should work together to raise awareness of e-tourism through training and education.

References

[1] UNESCO (2001) Message from the Director-General of UNESCO, Mr. K. Matsuura
http://www.unesco.org/dialogue2001/en/dgtourism.htm

[2] Graham, B., Ashworth, G. J., & Tunbridge, J. E. (1998) A geography of heritage: Power, culture and economy. London.

[3] Go, F.M, Lee, R.M., Russo,A.P (2003) e-Heritage in the Globalizing Society: Enabling Cross-Cultural Engagement Through ICT. J. of IT & Tourism 6(1), pp. 55-68.

[4] Buhalis, D. (1998) Strategic Use of Information Technologies in the Tourism Industry, Tourism Management 19(5), pp. 409-421.

[5] Davis, A.,Prentice, R. (1995) Conceptualizing the Latent Visitor to Heritage Attractions, Tourism Management 16(7), pp. 491 – 500.

[6] Buhalis, D. (2003) eTourism: Information technology for strategic tourism management, Pearson

[7] Buhalis, D., Spada A. (2000) Destination Management Systems: Criteria for Success – an Exploratory Study. Information Technology and Tourism 3, pp. 41-58.

[8] Buhalis, D., Owen, R., Pletinckx, R. (2006) ICT applications for World Heritage Site Management, in Leask, A., & Fyall, A. Heritage Site Management. Butterworth-Heinemann, London, pp.125-144.

[9] Shackley, M. (1998) Visitor Management: Cases from World Heritage Sites. Oxford: Butterworth Heinmann.

[10] Drost, A. (1996) Developing Sustainable Tourism for World Heritage Sites, Annals of Tourism Research 23(2), pp. 479-492.

COGNITONICS STUDYING THE SKILLS OF MIND

Tarkko Oksala,

IIAS and Aalto University, Department of Architecture

Otakaari 1, FIN-02150 Espoo

Tel: +358 445 444 873

e-mail: Tarkko.Oksala@gmail.com

Aino Oksala

University of Stockholm

Vanadislunden 1A 1101, 11347 Stockholm

ABSTRACT

The aim of this paper is to connect ICT to the classical ideas of skills and to the missing parts of them as well. Thus the duality of science and art is touched with real grasp. Firstly, the seeking of right way, rationality and potentiality of "techne" are discussed as applied to mentality. Secondly, the successful working under guiding emotion is considered as the lead-way to sustained mental abilities. This all is done in order to separate good and bad actions and developments stemming from them in cognitive technologies.

1 INTRODUCTION

Cognitonics [1] has its theoretical part or root (cognitology) but is also by name a practical and applied branch of science. Mind is divided into rational and emotional part. That is why we can notice two types of skills. Aristotle [2], [3] opens the rational part by defining:"Skills are right and ratio-based potentialities to act and fabricate".

In addition to that we have to notice, that some skills are also "working (acting) and feeling-based abilities". Then we take steps toward "tacit skills", which are difficult to rationalize [4]. In addition we may sum, that skills consist concepto-architecturally of:

Rightness, rationality, potentiality

Action, feeling/ emotion, ability

2 RATIONAL TECHNE OF MIND

2.1 Rightness of mind

Our mind can act in a right way or make mistakes or function even in a disordered way. The problem is then, what is right and under who's conditions. In cognition right thinking is assimilated often with truth, as in the case of knowledge-processes. In general, idealists think that the realization of the basic values is the right goal of action. In art we have canonized beauty to be one central value. Analogically some have used expressions like "beautiful mind". In addition, the meta goals of mind include:

Holy, true, right (itself)

Good, beautiful, skillful

2.2 Rationality of mind

Rational mind follows rules and notifies laws in seeking the truth. In pure mathematics following of rules is important, and in applied work semantic networks play their role. As exception to standard cases part of aesthetics is based on rules and part of mathematics is intuitionist. Thus the demarcation problem of rationality is challenging. Aristotle saw man as "animal rationem particeps", but also "a being capable of laugh" (or cry) [5]. Pure rationality of mind is correlated with knowledge and rational skills and is based on know-how consisting of factors:

Truth, argument, belief (conception)

Rightness, rationale, potential to act (rationally)

2.3 Potentials of mind

Potentials of mind are open. We do not know enough about mind and cognition to characterize their limits. In the theory of computation the concept of Turing machine is quite clear and gives artificial limits to mechanical thinking [6]. Also logic has its own limits. Intuition may continue working of mind where formal systems are not sufficient tools. Emotion, feeling and inspiration are open sources of innovation. Knowledge is the way to reflect our mind but in general we have to face factors of:

Belief, knowledge, volition

Desire, emotion, skill

3 EMOTIONAL TECHNE OF MIND

3. 1 Action of mind

We have too main approaches to study the action of mind. Some study the function of the brain some study the abstract function of mind. Action of mind is most naturally described via the transformations of mental states, which is a map about mental evenio. This all is abstraction, thinking as the target. At the same time thinking itself is based on abstraction. In a more concrete picture we may divide the action of mind as cognition consisting (architecturally) of:

Perception, memory, thinking

Abstraction, knowing, doubting

All these actions are subjects for the predicates of skills (techne). - Already Aristotle connected perception to presence, memory to past and thinking potentially to the future. Abstraction [7] was central factor in the theory of universals. Plato gave an example, how to define knowledge and Socrat was a master of critical doubting.

3.2 Emotionality of mind

Mind works partly in a rational way among the action types of cognition (above). A more accurate analysis, however, notices, the rest class we call here the emotional part of actions.

Perception may be modeled as a scientific experiment [8] but also a field of strong emotion manifested in arts [9/ 1946]. In science and technology we expect memory to work as an accurate storage of facts, but in art memories are creative [10]. Thinking is formalized using the laws of logic of mathematics, but natural thinking is emotion-driven.

3.3 Abilities of mind

Abilities of mind depend on how well we sum up the factors described above mainly from the positive and constructive point of view. Negative and destructive forces are, however, always potentially at hand in cognition, as we know from knowledge (good and bad knowledge [11]).

Disabilities of mind consist of wrong choices, antirational steps, powerlessness; miss-work, killing of conscience and lost abilities itself. They are mental but manifest in all bodily actions. Abilities of mind on the other hand are manifest in right action and in skillful action containing merits from all paragraphs mentioned in the analysis done.

Mind has an open variety of abilities as directed intentionally to the world. It also has its special ability to “gnoti s(e)auton” [5] (compare, [1, 395]. This makes the improvement of culture possible. ICT is one concrete product of that, but we can go much further due to the rapid development of cognitonics of today and potential developments of tomorrow.

Cognitonics saves the cumulative self-consciousness of mind from generation to generation. This transition of knowledge stays alive when we can transfer also partly the tacit dimension of right emotion. The world is, however open, and new generations should also find new values of their emerging cultures.

4 DISCUSSION: A Case study and our general challenge

The idea of the Techne of Mind has far reaching cyclic connections to education and didactics. First all skills are learned naturally or artificially. Secondly the improvement of learning and teaching is bound to certain skills of their own. These ideas have been noted in antiquity (see [12]) and later in a more exact form in the birth of experimental psychology (Wundt/ [13]) and experimental education (von Neumann [14], Meumann, [15]). It is natural that we give also answers to those questions in the time of ICT.

Typical examples of arts of cognition include the Art of memory. Ideas of how to make remembering and learning more easy and effective are studied in mnemonics, psychology, cognitive science and ICT (Majurinen, Oksala, 2009). The question concerns the formation of complexes (Mueller/ [15]) in a dialogue of external and internal realms

as recently discussed largely in information aesthetics [16], [6].

A study [17] concerning the teaching of information personal to use ICT tools with developing service-performance was carried in 2008 for Finnish Railways (Ratahallintokeskus). In it skills analyzed were divided in four groups: Information skills, Railroad skills, Customer skills, Situation skills. These targets were mapped with the so called "Theaters of memory" known in mnemonics and using the theory of effective complex formation. Also an ICT-based and electro-mediative simulator to test and improve these skills (above) was constructed.

In addition to special experiments and cases we may note the general vision: This large field of studying technics of cognition may be refreshed and updated in ICT-time. One catalyst to this process may be the updated notion of techne promoted in this paper.

5 CONCLUSIONS

We have shown how human cognition is at its best knowledge-based. At the same time knowledge is restricted, and human mind works under the guideline to be skillful. That is why on the other end of the evaluation of thinking we have to notice the emotional pole called conscience ((eu-) daimonion).

Acknowledgements

This paper is based on discussions with Joel Majurinen.

References

[1] Fomichov, V. A., Fomichova, O. S. (2006): Cognitonics as a New Science and Its Significance for Informatics and Information Society. Informatica (Slvenia), Vol 30, pp. 387-398.

[2] Aristotle (1984): Nikomakhoksen etiikka, suom. S. Knuuttila, Gaudeamus.

[3] Kotila, H., Mutanen, A., Volanen, V.-M. (2007): Taidon tieto (The Knowledge of Skill), Edita.

[4] Polanyi, M. (1966): The Tacit Dimension, Chigaco.

[5] Oksala, T. K. (1986): Homeroksesta Alvar Aaltoon, W+G.

[6] Oksala, T. (1981): Logical Aspects of Architectural Experience and Planning, Otaniemi.

[7] Mikkola, E. (1964): Die Abstraction Begriff und Struktur, Helsinki.

[8] Niiniluoto, I. (1975): Todennäköisyyden lajeista, in R. Tuomela (ed.), Yhteiskuntatieteiden eksakti metodologia, Gaudeamus.

[9] Aalto, A. (1946/1972): Luonnoksia, toim. G. Schildt, Otava

[10] Radford, T., Oksala,T. (1996): Creative Memories and the Finnish Sauna, in T. Oksala, G. L. Farre and G. E. Lasker (eds.), Design: Emergence, Content, APS Ci106

[11] von Wright, G.-H. (1961): Ajatus ja julistus, tt, WSOY

[12] Yates, F. A. (1958): The Art of Memory, Peregrine

[13] Schmit, P. (1979): Architecture and the Human Dimensions, Godwin

[14] Neumann, G. von (1907): Experimentelle Beiträge zur Lehre von der Ökonomie und Technik des Lernend, Zeitschr. f. experim. Pädag. IV

[15] Meumann, E. (1914): Experimentelle Pädagokik, III, Leipzig und Berlin

[16] Elfving, G. (1965): Information and Esthetic Evaluation, in Acta Philosophica Fennica, Fasc. XVIII pp. 7-11

[17] Majurinen, J., Oksala, T. (2009): Junaliikenteen informaatiokeskuksen toimintatapa, Helsinki (Ratahallintokeskus)

CULTURAL INHERITANCE AS PREREQUISITE TO ELEARNING

*Ivan Pogarcic *, Ida Hitrec*, Marko Pogarcic***
*Polytechnic of Rijeka, **University of Rijeka, Faculty of Law
Trpimirova 2, 51000 Rijeka, Croatia
Tel: +385 51 35 37 53; fax: +385 51 67 35 21
e-mail: pogarcic@veleri.hr, ida.hihrec@veleri.hr, knjo@net.hr

ABSTRACT

Education as process is comprised of two synergy-related segments: learning and teaching. Within each of these segments there are active participants mutually connected to other elements that create education as a complex. These include: education or educational contents, goals planned to be reached and context in which, and because of which, education is being realized. Postulates and mutual dependence of mentioned elements can be observed on several levels. Basic relationship can be observed on a primary level that can be determined by different circumstances. On the other hand, education has in time gone through changes caused by other activities in a real-time and scientific environment. Discovery of computers and their more intensive usage in education opened up huge possibilities but has also removed some limitation factors such as place and time of performing education. Possibility of virtual implementation of the usual educational unit, a classroom, has made space for other questions and emphasized need for careful treatment of the same. Virtual student and virtual teacher are only apparently virtual while realistically they do exist in space and time. Still, these do posses their own personal culture, customs and habits. Such circumstances can lead to possible misunderstandings and even conflicts that can be anticipated and avoided. Paper analyses students' and teachers' attitudes and possible cultural, moral and ethical relationships during the realization of eLearning as process.

Keywords: e-Learning, cultural inheritance, ICT, moral inheritance, learning style

1 INTRODUCTION

Presence of ICT in learning today is, more or less, indisputable. Specific forms of ICT tools and technologies depend upon several different factors. Naturally, financial aspect of learning's organization and providing finances are the most important in process. Considering importance of learning, regardless to social environment, possibilities provided by ICT are important. If one observes the highest level of learning supported by ICT – online learning – better known as the highest level of eLearning, then detachment from traditional forms of learning becomes more relevant. eLearning has been declared as form of learning where participants are independent from location and time. Not just teacher but also the other elements of learning should in these circumstances consider eventual differences. In continuance, the paper examines more important attitudes and definitions of culture and cultural relationships in general. Finally, an interpretation of students' attitudes at Polytechnic of Rijeka towards the subject has been attached.

2 eLEARNING IN THE VIRTUAL CLASSROOM WITH ECULTURE

Virtualization of forms of learning's organization provides additional possibilities, but it also initiates problems that must be treated carefully. Basic elements or factors of learning: student, teacher, learning contents, learning goals and context in which learning is performed, regardless to changes brought by ICT, remain constants. Form of learning is being realized through interaction of these elements. An intensity of ICT implementation during the process itself will move centre of model towards some of the elements. Traditionally, centre of learning has been located by student-teacher relationship with movements according to this rudimentary axis, whether towards student or teacher. ICT in learning determines a primal centre according to content, or form of content, while in the process of learning it shifts towards the student. Executive form of learning – section or classroom, when being transferred into the virtual forms, is mostly performing a virtual dehumanization of relationships that were used in traditional learning. Virtual students are real persons with all belonging characteristics, determined among others, by cultural and traditional norms of environment they originate from. Within frames of the Internet it is possible to form sections which success will depend upon considering such contributions of individual students. In any activity, cultural differences can be the main initiator and mode in achieving quality but can also be an obstacle in realizing the goals. In this way [1] not

considering the cultural differences in marketing can cause legal consequences and damage to promoter's reputation. This example has been mentioned in purpose since eLearning, especially in informal learning, is actually trade that requires qualitative marketing. Still, such mode of making a business retains its specificities in all, even in cultural needs. Here a culture can be observed from different aspects. One can accept a definition [2] of culture as a collective formation of mind that differentiates members of various groups. However, in eLearning the students will bring their cultural inheritance that must be mutually lined up. That way, during the learning they will form new mode of culture – culture of eClass. Hence, for purpose of this study a pragmatic definition is acceptable [3] according to which a culture is defined as scientific, transferable, interdependent group of symbols which meaning enables orientation for members of category and an option for solving the category's problems. Learning, regardless to its form of realization, is a process that has duration, goals and resources that are dispersed during a realization. Virtual environment will demand a bit differently assorted "cards". Since learning is process based upon cooperativeness of participants, and eLearning additionally emphasizes collaboration during a realization, importance of culture and cultural inheritance of individuals is more expressed. Especially it is this important when eClass is heterogeneous by its age and environment from which individual originates.

Which determinants of culture are important in eLearning? Maybe it can be said which parameter is more important, but certainly language, background in learning, or learning' goals are the most important. Material status and social relationships among religion and sense of aesthetic will at highest rate influence attitudes of individual and his/her definition of values. When ICT is strongly applied, term of language can be treated differently. It is important to differentiate living – talkative language from computer languages that can be required during the learning. eLearning misses one, in communication important form of language – body language. However, during the process of self-disclosure in learning its virtual realization can be obtained.

Social relationships of environment from which eStudent originates can become an obstacle to his/her socialization. Different treatment of certain groups of society and definition of status in certain society are just some indicators of this parameter's quality. Attitude towards aesthetical values is maybe not that disputable but it doesn't mean in some situation it cannot be crucial to success of eClass. If eStudent is part of eClass that has many different eStudents with different religious believes, eLearning should be organized carefully. In such case it is assumed that contents are formed in such manner that they don't insult certain religion. Still, the highest burden is on mentor or e-tutor that moderates the communication. Conflicts in this area are easy regardless of them being deliberate or not.

Prior education of individual is a potential that can be super-positioned within eLearning. Education eludes capability of individual to influence the environment and anticipate his/her attitudes in concrete situations.

For purposes of this study one should considerate some classifications of culture. For example [4], a culture of low context based upon written and spoken word (without any hidden meaning) and culture of high context where all facts derivable from mutual relationship can be differentiated. Within these frames one insists on construction of cooperation, trust and loyalty. It is obvious that eClass as heterogeneous group by education and prior experience, needs a culture of high level, since its realization strongly requires cooperativeness as well the collaboration that will help individuals to better attend the process. However, quoted definitions are too general and demand additional parameterization. Cultures of world [5] can be segmented according to the value of parameters determined by: power of alienation, power of avoiding the insecurity, level of individualism, domination of gender values and capability of long-term orientation. Each of these values can be elaborated in eEnvironment and are important for quality of relationships in eLearning. Equally important is the interdependence of individual parameters, but that goes beyond this study. However, in planning the eLearning' realization, mentioned parameters should be considered through means of valorization and their ranking.

3 RESEARCH AND SURVEY

For the purpose of this study some ethical issues that could accompany eLearning have been researched through this survey. The goal is to avoid ethically problematic design or behaviors. So we try to point out some examples.

Privacy issues. Online devices can invade privacy. Guidelines need to be set. They need to be clear and they should be enforced.

Uniformity of access. Ethical constructs that deal with justice and the administration of justice suggest that all individuals who participate in an activity should be able to do so with equal chances of success.

Non-biased, culturally equitable delivery and expectations. Signs and symbols can be subtle, and people may not be aware that a particular sign, symbol, or content item could be offensive to some groups. It is important to expand the rules of proper eLearning behaviors and to make sure that students are not posting or sending invasive or offensive items to fellow students.

Languages barriers. The e-tutors are ethically obligated to provide training, mentoring and support to learners who may not have the background or language skills to succeed in eLearning.

Posting and other concerns. "Netiquette" notwithstanding, impulse control is often lessened in an environment where one feels safe and fairly anonymous. One way to combat

rudeness in the discussion board is to attach a real identity and impose social control.

Cyber – bullying and cyber-stalking. "The Internet has been described as transforming society by providing person-to-person communication, similar to the telegraph and telephone as well as operating as a mass medium, like radio and television before it" [6].

This paper studies cultural features and behavior in virtual classroom with 274 examinees. The authors analyze the attitudes of students and teachers about issues concerning different cultural, moral and ethical relations in the realization of the teaching – learning process.

Survey has primarily enclosed questions regarding gender and financial security of students (ICT equipment and infrastructure), mode and frequency of using the Internet, safety and reliability in work, attitude towards appropriateness of learning contents in the Internet. The second part of questions (10-25) has applied to expressing opinions about communication in virtual environment: moral, cultural, ethical and other attitudes within which a communication is being realized. Explanation of survey follows in the continuance of paper. (See the Questionnaire in Appendix A at the end of paper)

4 EXPLANATION OF RESULTS INSTEAD OF CONCLUSION

Results of the analysis indicate almost all examinees, app 98%, posses their own computer respectively have a daily access to the Internet. The majority of students, app 95.41%, use the computer and the Internet always or regularly when accomplishing their faculty assignments. These results indicate importance of proper and timely education of students in accurate online behavior and usage of the Internet. Also, a great percentage of students use the Internet in learning. 83.28% of students use it always or frequently. All the above mentioned leads to a conclusion the Internet became an extremely important item of a higher education system. When questioned in what purpose do they usually use the Internet, 53.11% of examinees answered they use it generally for information purposes, followed by entertainment that was chosen by 16.72% of students. Considering the examinees' attitudes towards the Internet safety, most of them – 48.52% is neutral, with reference to they don't consider the Internet neither safe nor uncertain medium, 33.78% consider the Internet safe or fairly safe, while only 17.38% of them believe the Internet to be fairly or completely uncertain medium. The above stated indicate the examinees' attitude towards the ethics of other Internet users who influence its safety by their own actions. Though it has been noticed that less then half of students consider the Internet to be generally safe medium, 68.86% of them stated they always or frequently believe in contents browsed on the Internet for educational purposes. Besides, they also indicate their capability of making a difference between relevant and irrelevant, respectively relevant (accurate, true) and irrelevant (inaccurate, untrue) information. Further on, the results of the analysis indicate that precisely 60% of students have never used a false identity on the Internet that proves their ethical behavior. Even the higher percentage, 88.20%, answered they have never insulted others on the Internet by any kind of means. Still, out of 35 examinees who answered positively, 29 declared their insult was on purpose. The same percentage of students, as in false identity issue, 88.20% answered they have never hacked some web sites or interfered with other peoples' email that also proves students' ethical behavior. Nevertheless, this doesn't indicate a real situation, but imbalance between examinees' wishes and passiveness and/or technical ignorance for that sort of operation. The question posed about students' aspirations towards the following statement: "Never do something online you wouldn't do in a real world environment" the allocation of answers was as follows: only 5.57% examinees do not support statement; 6.56% cannot provide a concrete answer; 13.44% examinees partially disagree; 21.31% partially agree, while more then a half, 52.79% support the statement. This kind of answers' allocation should support positive behavior of examinees as persons who responsibly use medium such as Internet. When questioned about verbal confrontation over the Internet, the majority of examinees, 55.74% haven't provided a concrete answer, but instead have concluded that sort of behavior is a situation related. It can be presumed that were the question posed as "Would the examinees verbally confront someone face to face", the answers would have been different so there is imbalance between these answers and the ones given on the previous question, where majority students claim they would never do something online they wouldn't do in a real time environment as well. 86.56% stated they have never published a disinformation on the Internet. Similar percentage (90.82%) claims they have never published other author's work signed by their own name, nor have they published their own works signed by others (85.25). In the same context, 96.72% of examinees answered they have never published on the Internet some/someone's confidential data. When asked have they ever downloaded some contents or their segments from the Internet and presented them as their own work or for educational purposes, 60.33% of students answered negatively. It is obvious these answers support those given on a control question by which examinees should have stated how regularly they quote Internet sources used in their papers. Even 95.08% of students always or sometimes quote those sources proving they actually don't represent the Internet contents as their own papers. It was therefore pointed out the examinees actually sincerely answer the posed questions. Regarding the responsible and conscientious behavior when communicating on the Internet, most of examinees, 86.56% believe they actually behave that way. Only 7.87% of students stated they never or only occasionally behave responsible and conscientiously, while 5.25% has never considered their behavior. When asked about other Internet users and their

level of education in correct Internet communication, 68.52% of examinees consider the others undereducated, though the majority has previously declared they mark themselves positively. This indicates examinees consider themselves to be in a higher level then the rest of the Internet users, meaning they are not particularly self-critical when judging their own behavior. Moreover, 19.02% of examinees answered they cannot mark other Internet users' behavior, while only 12.13% of students answered they consider the others to be well trained for this sort of communication. Finally, the analysis results prove 39.34% of examinees believe the Internet usage increases the impression of their responsibility. Similar percentage, 37.38% cannot decide, while 22.95% believe the Internet usage does not make them look more responsible. To conclude, the answers are relatively balanced since the examinees have probably concluded that responsibility impression depends also upon the mode of using the Internet. When analyzing whether the students have prejudices towards the different ones and whether discrimination is a cause of their behavior, the question about their willingness to participate in online course together with students of different races, religious beliefs, national affiliation etc., 89.84% gave a positive answer. Only 5.57% answered they wouldn't participate, while 4.26% didn't want to express their opinion.

References

[1] M. Kotabe, K. Helsen: Global Marketing Management, Wiley, 2010, ISBN-13: 978-0470381113S.

[2] G. Hofstede, G. J. Hofstede, M. Minkov: Cultures and Organizations: Software for the Mind, McGraw-Hill, 2010, ISBN-13: 978-0071664189

[3] V. Terpstra, K. David: The Cultural Environment of International Business (Sb-Principles of Marketing), South-Western Educational Publishing, 1992, ISBN-13: 978-0538800037

[4] E. T. Hall: Beyond Culture, Anchor Books, 1976, ISBN-13: 978-0385124744

[5] G. Hofstede: Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations, Sage Publications, Inc, 2001, ISBN-13: 978-0803973244

[6] K.Y.A. McKenna, J.A. Bargh: Causes and consequences of social interaction on the Internet: A conceptual framework. Media Psychology, 1, 249–270, 1999

Appendix A: The questionnaire

1. What is your gender?
2. What year of study have you enrolled?
3. Do you have a daily access to the computer and the Internet?
4. Do you use the Internet in completing the faculty assignments (homework, seminars etc.)?
5. Do you use a computer and the Internet for learning (adopting a learning content)?
6. For what purpose do you use the Internet most frequently?
7. Do you feel the Internet is a safe medium?
8. Do you find confident the contents browsed on the Internet and used for educational purposes?
9. Have you ever used a false identity on the Internet?
10. Have you ever insulted anyone on the Internet by means of his religious, national, racial, gender or sexual affiliation?
11. If you have insulted someone, as stated above, was it on purpose?
12. Have you ever hacked the web sites or someone's email over the Internet?
13. How do you feel about saying: "Don't do anything online you wouldn't do in a real world environment"?
14. Are you willing to verbally confront someone on the Internet?
15. Have you ever placed some kind of disinformation over the Internet?
16. Have you ever uploaded on the Internet something that hasn't been created by you?
17. Have you ever published on the Internet your own paper signed by false/someone else's name?
18. Have you ever published on the Internet some/someone's confidential (secret) data?
19. Have you ever downloaded from the Internet some content or its segments and represented them as your own work?
20. Do you consider yourself sufficiently responsible and conscientious when communicating on the Internet?
21. Do you consider people in general to be sufficiently educated for proper communication over the Internet?
22. How often do you actually quote the Internet sources you use in seminar papers, homework, presentations etc?
23. Do you believe usage of the Internet in educational purposes strengthens the effect of your responsibility?
24. Would you, without any prejudices, participate in online course together with students of other races, religious persuasions, national affiliation etc?
25. You may, if you wish, make your own comment, notification or observation you feel is important for the research.

IMPROVING THE EFFECTIVENESS OF EDUCATIONAL PROCESS BY CLUSTERING

Dumitru Dan Burdescu[1], Octavian Neguletu[1], Marian Cristian Mihaescu[1], Bogdan Logofatu[2], Costel Marian Ionascu[3]

[1]Department of Software Engineering, University of Craiova
Bvd. Decebal, Nr. 107, 200440, Craiova, Dolj, Romania
Tel/fax: +40 251 438198;
e-mail: burdescu@software.ucv.ro, neguletuoctavian@yahoo.com

[2]CREDIS Department, University of Bucharest
Bulevardul Mihail Kogalniceanu 36-46, sector 5, Buharest, Romania
Tel: +4 021 315 80 95; Fax: +4 021 315 80 96;
e-mail: logofatu@credis.ro

[3]Analysis, Statistics and Mathematics Department, University of Craiova
Str. A.I. Cuza 13, 200585, Craiova
Tel/Fax: +4 0251 411 317
e-mail: icostelm@yahoo.com

ABSTRACT

Increasing the effectiveness of educational processes is one of the greatest challenges for information society. The paper presents the usage of M-Tree structure for classification of learners based on their final marks obtained in their respective courses. The classical building algorithm of M-Trees with our own accustomed clustering procedure was implemented. The data that is managed within M-Tree structure is represented by instances. The main goal of the structure is to provide information to students and course managers regarding the knowledge level reached by students. The proposed clustering procedure that is used for splitting full M-Tree nodes is designed to properly classify the learners.

1 INTRODUCTION

The ability to classify a student's performance is very important in internet-based educational environments. A very promising area to attain this objective is the use of special designed data structure. In fact, one of the most helpful usages of modern algorithms in e-Learning is classification. E-students are students that follow courses within an e-Learning platform. There are different educational objectives for using classification, such as: to discover potential student groups with similar characteristics and reactions to a particular learning strategy, to improve a student's capacity of learning, to group students who are failure-driven and help them improve their skills, to identify learners with low motivation and find remedial actions to lower drop-out rates etc. We have applied a classification method using unique algorithms which have a common base (tree classification).

Some possible outcomes of such analysis process are: predicting students' grades (to classify in three classes/clusters of low priority – week learners, medium priority – easy learners, high priority – competitive learners) from test scores.

M-tree [1,2] is a dynamic access method suitable to index generic "metric spaces", where the function used to compute the distance between any two objects satisfies the positivity, symmetry, and triangle inequality postulates. The M-tree design fulfills typical requirements of multimedia applications, where objects are indexed using complex features, and similarity queries can require application of time-consuming distance functions. In this paper we describe the basic search and management algorithms of M-tree, introduce several heuristic split policies, and experimentally evaluate them, considering both I/O and CPU costs. Results also show that M-tree performs better than R*-tree on high-dimensional vector spaces.

2 BUILDING THE M-TREE

Our implementation uses the following data structures:

```
struct Student{
      int IDStudent;
      float Lp;
      float Mp;
      float Maxp;
      int Where[3];
   };
```

The *IDStudent* is the identifier corresponding to each student entering the online distribution program. In order to evaluate these students, each chapter presents a final quiz containing notions belonging to the previously discused levels of priority. Thereby , each student will get a mark for each type of notion contained in the chapter:
- a *Lp* mark corresponding to *low priority notions*
- a *Mp* mark corresponding to *medium priority notions*
- a *MaxP* mark corresponding to *maximum priority notions*.

Each one of these marks is important, because they represent the guiding tool for a student. Example:
Example: Let the student identified by his/her *IDStudent=1002* take the quiz at the end of the chapter and gets the following results:
(Lp= 7.70, Mp=6.78, Maxp=5.00).
Right know, a teacher, or even an electronic program, is able to compute the minimum performance of this student, reaching the conclusion that the advanced notions of the chapter(indicated by Maxp) have not been covered properly by this student, and therefore , he/she needs to put more energy in this direction. These directions are given by *Where* vector, used for providing instructions regarding where should a student improve his/her level of knowledge. A graphical represantation is given below:

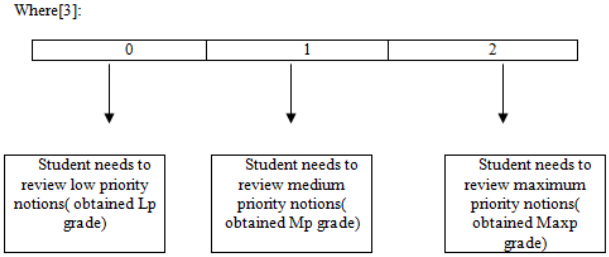


Fig. 1. *Where* vector, used for providing instructions regarding where should a student improve his/her level of knowledge.

As a start, Where[3]=(0,0,0). If any of this vector's component changes to 1 , this means it becomes active. For instance, if StudentA.Where[3]=(0,1,0), it means that he/she needs to review notions of medium priority level.
These students are then distributed and placed by our algorithm in a M-tree structure. Before moving on with the algorithm, let us present the structure of the M-tree. As presented in the earlier paragraphs, the M-tree is a spatial, metric tree, consisting of: *1 root* and *k leaves* containing students.(As we will see later, these leaves represent clusters of students). For now, let us stick to the structure of a M-tree, mentioning that this tree is actually a spatial one, where its leaves can be imagined as spheres, containing points, which are actually students. As far as the structure of a node is concerned, we have:

```
struct m_Node{
      int nrKeys;
      bool isLeaf;
      float radius[NMAX];
      m_Node *routes[NMAX];
      struct Student students[NMAX];
  };
```

The `nrKeys` represents the number of students contained in a node(cluster). As far as our M-tree is concerned, we pay extra attetion to the nodes, because it is very important wheather they are leaves(terminal nodes) or internal nodes, as we will see later in the algorithm. The boolean variable `isLeaf` points out our exact concern: weather a node is leaf or internal. Moving on, as previously said, we consider our nodes as spheres, and as any sphere, it is geometrically represented by its center *C(x,y,z)* and its radius *R*. However, in order to match these notions to our real implementation, we have constructed an abstract interpretation for this geometrical representation. The center *C* of a sphere(cluster) will be represented by the average student belonging to the set of students contained in that particular cluster. Instead of spatial coordinates *(x,y,z)* , our *centerStudent* will be represented by its elements *(IDStudent,Lp,Mp,Maxp)*, which were presented earlier. The *radius* of this abstract sphere will be represented by the distance between the *centerStudent* and the student (students) with the lowest results in that cluster. We will take a closer look to this abstract system later, when we will discuss the implemented algorithm. The *routes* represent the children of a particular node and of course, the `nrKeys` points inside the sphere, which are actually the *students*, as in our abstract system, a spatial point is represented by a student.
Our implementation is based on the idea of students distribution depending on their results to a quiz at the end of a chapter. For a better understanding of our implementation, let us consider a real situtation:
Let S be a finite set S of k students defined as $S=\{St_1,St_2,....St_k)$, $k>0$. Let us suppose that all these students have taken a quiz at the end of a chapter, in order to evaluate their level of knowledge. Each student is represented by his/her IDStudent, and his/her grades: Lp,Mp,Maxp (they were discussed earlier). Let us assume that, we want to create an hierarchy among these students, depending on their results. In order to do that, we need to group these students in *clusters*, each cluster having its own attribute. An attribute, for a cluster, represents the level of performance for that particular group of students. Moreover, these attributes are also used as indicators pointing out to the type of notion (low priority, medium priority, maximum priority) the student needs to review. After a group of students (cluster) is formed, a center is chosen, that is the average student in that group, and all the other students are distributed in a spherical manner arround him/her.
Computation of a radius for a cluster : the radius of a cluster represents the maximum distance possible, between the *centerStudent* and the rest. The bigger the distance is, the better, or the lower the results of that student are. Just as in real cases, when we say there is a big distance between this *average student* and *student A*, this means that Student A has either better results or worse results, we don't know for sure. Anyhow, should the distance between the *centerStudent* and any other student be greater than the cluster's radius, it means that the particular student does

not belong to that cluster, for the simple reason that he/she is smarter than all those students in that cluster, or his/her results are lower than any other's in that cluster.

The radius of a cluster is computed, depending on the results of each student, being the maximum distance between two students. We define the distance between *StudentA and StudentB*, as follows:

d_{AB}= max{(|Lp_A-Lp_B|, |Mp_A-Mp_B|, |$Maxp_A$-$Maxp_B$|)}. (1)

As you can see, when we measure the distance between two students, we are looking for the most marking difference between them. This also helps us in defining attributes of a cluster, depending on the type of notion students should focus on (low priority notions, medium priority notions, maximum priority notions). Moreover the above relation (1) guarantees that the radius of a cluster represents the biggest difference between the levels of knowledge for each student belonging to that cluster.

As example, let us consider student A with his/her results: (9.60, 8, and 7.50). Let us consider student B with his/her results: (7.60,8, 6.50). Following the relation at that computes the distance between two learners we get the biggest difference 2 (9.60-7.60). This is the biggest distance between them two. So they might have similar knowledge for medium priority notions (8,8), and maximum priority notions(7.50,6.50), but when it comes to low priority notions, we see a gap between them(StudentA -9.60 , StudentB -7.60). Let us suppose that StudentB has the lowest result in the cluster, and StudentA is the centerStudent. Then, as we have presented earlier ,they can be grouped in a cluster, with its radix 2 . Let us consider now, that StudentC gets the results: (5, 6.30, 5). We get:

d_{AC}= max {|9.60-5|,|8-6.30|,| 7.50-5|}=4.

d_{BC}= max {|7.60-5|,|8-6.30|,| 6.50-5|}=2.60.

As you can see, neigther of these distances is lower than our cluster's supposed radius, as the difference between StudentC and students StudentA, StudentB is huge, so there is no way, StudentC will become member of this cluster. Moreover, based on the present results of StudentA and StudentB , we can define an attribute for this cluster: all students belonging to this cluster will posses similar knowledge levels for medium and maximum priority notions, but the marking difference between them will be represented by the low priority notions, so all of them need to review this part of the chapter.

The main steps of the agorithm are:

Step1. We start from a simple representation of students identified by their elements:

IDStudent, Lp (score), Mp (score), Maxp (score).

Step2. We picture the set of the students who have taken the test as points in 3D space. Our algorithm involves two major operations: a clustering operation and a split operation.

We will first describe the *splitting* method. We have decided that these groups of students should have a maximum number of allowed members. Let us denote this number as the *filling factor* of a cluster (student group). Whenever the number of students in a particular cluster becomes greater that this filling factor, a cluster splitting is involved. This is how the M-tree extends its nodes. The splitting procedure works as follows:

At the beginning two random students from that cluster are chosen as centers for the new clusters resulting after splitting. Let us denote them *Student1* and *Student*. Next, we distribute the rest of the students arround the new centers. If, for instance, we have Student1 and Student2 as centers, the question is where should we attach Student3, to? We compute the distance between (Student1,Student3) and *(Student2,Student3),* using relation (1). *Student3* will go near that student which is more closed to him(that is from a level of knowledge point of view). After the distribution is completed, we start the *chooseCenter* method, which recalculates the new centers of the clusters, and if new centers are found, the entire discussed process happens again until new centers are found no more. This process is called the *Clustering Process*. After that the effective splitting happens, and the initial tree node is split into two nodes. When these clusters are formed, they are also assigned attributes(we have discussed about them earlier). What is interesting is that these atrributes suffer a constant evolution, depending of the students inserted in that specific cluster. As the clusters are subject to constant splittings and modifications whenever the number of students inside is greater than the filling factor, so do the attributes change, transforming a part of the old cluster in a better one(shelters students with higher levels of knowledge) or even a lower one (shelters students with lower levels of knowledge).

Clustering and splitting procedures in pseudocode are presented below.

Procedure Cluster_and_split

```
#Get the cluster which will be subjected to splitting procedure;
#Let StudentSet={Si| Si ϵ cluster, i=0,nrKeys, Si student};
#Choose
  S1 ϵ StudentSet(as new center1);
  S2 ϵ StudentSet(as new center2);
  oldCenterStudent1←S1;
  oldCenterStudent2←S2;
#While (!STOP_CONDITION) do {
For each other Si ϵ StudentSet, (Si ≠ S1 and Si≠ S2 ) do {
        Compute   d1i=max{(|Lp1-Lpi|,|Mp1-Mpi|,|Maxp1-Maxpi|);
        Compute   d2i=max{(|Lp2-Lpi|,|Mp2-Mpi|,|Maxp2-Maxpi|);
        If (d1i < d2i)
Attach Si to the cluster with center S 1
        else
Attach Si to the cluster with center S 2
        #Determine attributes
      }//end do
newCenterStudent1=chooseCenter(newFormed Cluster1, newClusterRadius1);
```

```
newCenterStudent2=chooseCenter(newFormed
Cluster2, newClusterRadius2);
   STOP_CONDITION←
(newCenterStudent1=oldCenterStudent1)
&&(newCenterStudent1=oldCenterStudent1))
}//end while
```

The *Determine attributes* sequence works in the following way. Every time a distance between two students is computed, after the marking difference between them is extracted, inside the *Where[3]* vector, the corresponding component of that type of level notion the maximum was extracted for, is incremented(we say a flag is raised for that component). Example: Let us consider that the marking difference between studentA and studentB regards the low priority notions, then studentA.Where[0]=1 and studentB.Where[0]=1.(See the *Where[3]* vector configuration in previous sections). Notice that this *Where[3]* vector changes for every new student in the cluster,because in the end of the clustering process, only the attributes of the average student will prevail, because he/she is the center of that particular cluster.

The *chooseCenter* sequence simply computes the biggest distance between all the students, it also sets the new studentCenter of the cluster, as we will see in the complete algotithm's pseudocode.

Now we will take a closer look to the *splitting procedure*:

Procedure Effective_split_of_initial_cluster

```
 m_Node = node for splitting
  m_node→isLeaf = false;
  # Alloc memory for leftChild and
rightChild of m_node;
  m_Node→leftChild = newCluster1;
  m_Node→rightChild = newCluster2;
  #Insert S = {
centerStudent1|radiusCluster1,
centerStudent2|radiusCluster2} in m_Node
  #Attach the rest of students Si
to leftChild or rightChild
  leftChild→isLeaf=true;
  rightChild→isLeaf=true;
```

After seeing these two important steps in the algortihm, it is time to present the entire process:

Procedure Build_M_Tree

```
#initializeTree;//create an empty root tree
#While (not endOfFile){
 #Get input data student(IDStudent, Lp,Mp,Maxp);
 #If  ( tree->root->nrKeys < filling factor )
  #make a simple insertion in the actual root
 #Else
  #apply Cluster_and_split;
}//end while.
```

6 EXPERIMENTAL RESULTS

Let us consider 6 students values for (Lp, Mp, Maxp) = {(8,7,6), (5,9,10), (7,7,7), (7,8,7), (6,6,6), (4,3,5)}. The value of filling factor is 4.

We insert students with IDStudent 1,2,3,4 in one cluster. Student 5 also needs distribution, but adding him/her to the same cluster is not possible, since a cluster can hold a maximum of 4 persons (filling factor is 4). So a split is mandatory.

After split the clusters are:

Cluster1={Student1,Student3,Student4}
Cluster2={Student2}

Then, student 5 may be inserted, and the clusters become:

Cluster1={Student1,Student3,Student4,Student5}
Cluster2={Student2}

A new split is necessary because cluster 1 is full. Thus, the following clusters are being obtained:

Cluster1={Student3, Student4, Student5}
Cluster2={Student2}
Cluster3={Student1}

Finaly, Sudent6 is inserted in cluster 3 and thus the following clusters are obtained:

Cluster1={Student3, Student4, Student5}
Cluster2={Student2}
Cluster3={Student1, Student6}

6 CONCLUSIONS AND FUTURE WORKS

The paper presents a study usage of an implementation of M-Trees building algorithm. The tree manages real data representing e-Students (students from an e-Learning environment). The instances (i.e. the students) are characterized by attributes representing the scores obtained when taking tests. The tests are classified in accordance with their level of difficulty: low, medium and high.

Within classical M-Tree building procedure a custom clustering procedure was used when splitting a node was necessary. The clustering procedure is designed such that produces an optimal grouping of students regarding the "distances" in knowledge among them.

The tests were performed with datasets representing 200 students, and the filling factor of a cluster, restricted to 18. As a result, we got 7 clusters with attributes, leading e-Students to notion reviews.

Another goal of our current implementation is to provide valid data distribution using specific data validation algorithms. Moreover, our algorithm is able, at the moment, to distribute e-Students which test their level of knowledge for one chapter only. We wish to extend this process for multiple chapters.

References

[1] Uhlmann, J.K.: Satisfying General Proximity/ Similarity Queries with Metric Trees. Information Processing Letters. vol 40, pp. 175-179 (1991)

[2] Ciaccia, P., Patella, M., Zezula, P.: M-tree: An Efficient Access Method for Similarity Search in Metric Spaces. In: VLDB'97, Proceedings of 23rd International Conference on Very Large Data Bases. pp. 426-435. Morgan Kaufmann (1997).

ADAPTATION OF LANGUAGE MODELS FOR POETRY RETRIEVAL IN WEB SEARCH

Magdalena Granos, Aleksander Zgrzywa
The Institute of Informatics
Wroclaw University of Technology
Wybrzeże Wyspianskiego 27, 50-370 WrocławPoland
Tel: +48 607379123
e-mail: magdalena@granos.pl

ABSTRACT

The paper investigates the language models for poetry retrieval purposes and raises the importance of access to advanced exploratory search engines, which, according to the Maslow's hierarchical pyramid of needs, meet the needs of higher order, associated with self-fulfillment and informational-cognitive aesthetic requirements of non-verbal transmission. Th context of work focuses on fulfilling the information needs of sublime attributes of poetry as a form of structurally categorized pattern, which is a crucial indicator to determine the search methods of poetry based on adaptation of language models. The issues addressed refer to probabilistic models that allow predict occurrences of words in a sentence on the basis of distance functions, relatively to similarity of words and phrases, as well as by k-nearest neighbors strategy, and frequency of words in relation to rankings. Moreover, the paper aims at showing the optimal performance of language models in search of the most effective methods of poetry retrieval. During these searches, the relevant information retrieval models reveal their role in terms of new opportunities for Internet search engines into the formative process of education and creation. Accordingly, the paper attempts to explore the ways of increasing web-search efficiency to make future research flexible, yet precise, in interpretation of queries.

1 INTRODUCTION

Amid escalating resources of complex data in website network, the adaptation of search tools for information from the extensive resources of knowledge, implies the inevitability of working towards increasing power of search systems. Globalization process of transformation in socio-economic context calls for smooth availability and flowing information, as well as an operative technology to provide efficient multi-contextual information services for efficiently informed society. Widely and easily accessible information is a common and inexhaustible wealth of human imagination and intellectual achievements, which are inalienable in consciousness for resources of life derivatives.

The information requirement is a space of human knowledge, experienced consciously as a query, leading to look for relevant reply. As far as a user of information system is considered, a measure of meeting the need for information involves the assessment of user knowledge about information systems and services. It also results from the size of a subset of user's population who positively evaluate a system. Eventually it concerns the extent to which people use different types of services, resources and information collections [1].

Along with the growing expectations of information, there is a pressing need for advanced exploration capabilities of Internet search engines which take into account not only the wide spectrum of phrases determined by semantics terms from surrounding reality in context of accompanying circumstances but also the aspirations of a higher order arise, associated with self-fulfillment and satisfaction of informational-cognitive aesthetic ambitions, according to the hierarchical pyramid of needs by Maslow as a consequence of once fulfilled expectations, owing from the vital functions [2]. Poetry, pervading the most subtle areas of human sensitivity, fills the most sophisticated information needs and becomes an area of particular desirability to find information about non-verbal lore.

2 STRUCTURAL PATTERNS OF POETIC WORKS

According to Chomsky, the ability to learn and use language is an innate property that generates the words in a hierarchical format of branched tree called the structure of the syntax [3]. Chomsky's revolution shows that the internal capacity of learning, processing and building syntax is universal to the species, since we are born with sensitivity to this type of structural system, functioning on the principle of tree [4].

Mehrabian communication model, where only 7% are the words of verbal communication, while 38% and 55%, respectively, constitute intonation and expression of non-verbal communication, shows the validity of new trends development of search mechanisms [5]. They have intuitively interpreted the meaning of messages, which are expressed in poetry through stylistic means, such as onomatopoeic words, accent, rhythm, rhymes, expression of feelings, mood, metaphor, the spatial organization of words and thematic motifs as well [6].

Although being non-verbal in communication, poetry, like musical notation, also appears as a structured form, that is operating according to the categorized models, such as the rhythmic pattern, metric foot, rhyming pattern, versification pattern, alliteration, which are effectively used to develop a search method, submitted in consequence of building a new linguistic model based on structural patterns of poetic works, and by adapting some linguistic models [7].

3 LANGUAGE MODEL FOR N-GRAM TEXT CATEGORIZATION

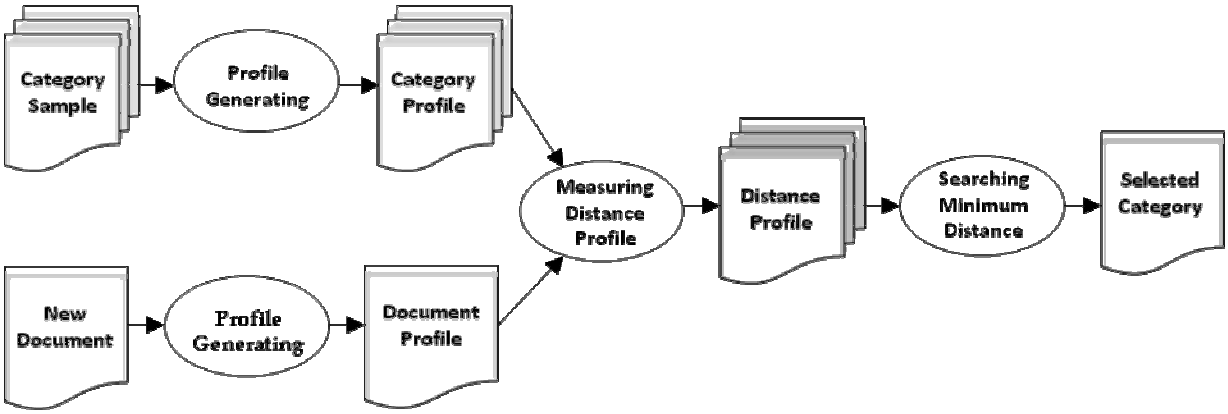


Figure 1: *Data flow in the N-gram model of text categorization*

N-gram model is a probabilistic model that allows to predict the next elements in sentence, while the probability P(Wn|Wn-1), where W is a word and n is a position in a sentence, of Wn word occurrence, which, as conditioned by the occurrence of Wn-1 preceding word, is equal to the ratio of the probability of co-occurrence of two words P(Wn-1, Wn) to the probability of P(Wn-1) preceding words (equation 1). N-gram is a subset of sentences, composed of n-elements of the sentence, which may be letters, characters, syllables, words or pairs of elements mostly of sizes from 1 to 5. The most elementary uni-gram model rejects the conditional context and defines each term separately, hence usually used to find information in case of structural complexity documents, whereas the second size bi-gram model determines the occurrence of the words preceding [8].

$$P(W_n \mid W_{n-1}) = \frac{P(W_{n-1}, W_n)}{P(W_{n-1})} \quad (1)$$

The functions of distance between the document and category profiles are determined according to the model, based on quantitative Zipf law, where word frequency is inversely proportional to the ranking, i.e. the higher the rank, the lower frequency (Figure 1).

The benefits of such approach is the ability to work both with short and long documents, also the minimum occupancy requirements of memory and computing, as well as a perfect fit to the texts of noisy sources.

4 LANGUAGE SEMANTIC MODEL OF STRUCTURAL SIMILARITIES

Searching of poetry from the sources of poetic corpuses is one by measuring the scale of the semantic similarities on structural paths of queries and lines of poems on the basis of

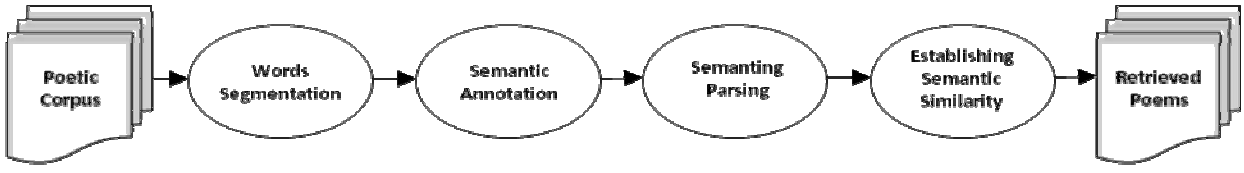


Figure 2: *The flow of data in the semantic model of structural similarities*

of semantic annotation and parsing (Figure 2).

When a semantic category for n_1 is different than for n_2, then $C(n_1,n_2) = 0$, where n_2, n_1 are the final nodes and have the same category, meanwhile $C(n_1,n_2) = sim(n_1, n2)$. $Sim(n_1,n_2)$ is the result of semantic similarity, while $sim(n_1, n_2) = 1/16$, what makes 2 branches between n_1, n_2, whereas $sim(n_1,n_2) = 1/256$ makes 4 branches, and $sim(n_1,n_2) = 1/4096 \sim 0$ tends to infinity. The T_1 tree consists of n_1 nodes. $C(n_1,n_2)$ is the number of subtrees with roots at the nodes n_1, n_2. $K(T_1,T_2)$ is a measure of similarity between trees T_1 and T_2, which is the sum of similarity measure $C(n_1,n_2)$ in each pair of nodes of trees T_1 and T_2 (equation 2) [9].

$$K(T_1,T_2) = \sum_{\substack{n1\in N_1 \\ n_2\in N_2}} C(n_1,n_2)$$

$$C(n_1,n_2) = sim(n_1,n_2) \qquad (2)$$

$$sim(n_1,n_2) = \frac{1}{4^N}$$

.

5 SEMI-SUPERVISED LINGUISTIC MODEL OF SARCASM IDENTIFICATION

Sarcasm means expressing the content which has an opposite meaning to the intentions. It is the activity of expressing the contrary meaning of what one has in mind to what has said in intention to transfer the message in a hidden manner. Peculiarity of sarcasm ambiguity makes it difficult to estimate if its sense is sarcastic or not.

In the model there are two stages of identification - a semi-supervised pattern and classification of sarcasm on the set of 66000 reviews for products in the selected shop website. Semi-supervised identification requires few data with annotation. The applied semi-supervised teaching is a learning process with little interference of human by providing the initial rules of marking the units, on the basis of which the next ones are collected. The initiator of the labeled sentences is subjected to annotation in scale of 1-5 of training set, where 5 indicates a purely sarcastic sentence, while 1 means a complete lack of sarcasm. The pattern is built on an underlying assessment concerning the star ranking of the product in a set of sentences, expressing the positive opinions issued by users of the shop service. The test data are enriched with the sentences with annotation, indicating the level of sarcasm in a set of 40 statements on a scale of 1-2 when there's no sarcasm, and 3-5 for sentences with sarcasm, where each sentence has received annotation from 3 testers. For each v-vector in the test set, Euclidean distance is calculated for each of the matched k-vectors of the extended set, where the matched k-vectors are defined as those that share at least one feature of pattern with v-vector.

Let L(v) be a level of sarcasm in a sentence for v-vector, while ti, i=1..k, where k is a vector with the smallest Euclidean distance to v-vector. Then v-vector is classified with the label l and the result is a weighted mean of k-nearest vectors from the set, according to the k-nearest neighbors strategy (equation 3) [10].

$$L(v) = \left[\frac{1}{k} \sum_i \frac{l_1 \cdot k_1 + l_2 \cdot k_2 + ... + l_i \cdot k_i}{l_1 + l_2 + ... + l_n} \right] \qquad (3)$$

For each feature v-vector in the test set, the Euclidean distance is calculated for each of the matched vectors in the extended test set, where the matched vectors share at least one pattern feature from v-vector. The result is a weighted average of the k-nearest test set of vectors. Although, not all sentences containing sarcasm are recognized, the result for a certain group of sarcastic sentences is satisfactory.

6 LANGUAGE SENCE MODEL

Sense Language Model (LSM) generates a probability question raised both from the term and the relevance of the document.

The quintessence of the model is semantic ambiguity of the word represented by polysemy, which occur while one word can have different meanings in different contexts and synonyms, when different words can have the same meaning in return.

Characteristics of the Language Sense Model are evaluation of the effectiveness on several corporas using a hierarchical method of smoothing as well as the feature of greatly exceeding the other models of linguistic action of a very high degree (for medium and long queries - 7.53% -16.90%) too.

The unigram linguistic model, where q_t and d_t mean the representation of query terms q and document d, q_{ti} is the i-th term in question q_t and n is the length in quiry q_t, which matches the terms and senses representations, that are possessed by each document in the model.

In this model, analogously to the model of terms representation, firstly the sense model d_s is generated for each document d, by using the representation of the importance and the probability of generating query $q_s q_{s1} q_{s2} ... q_{sn}$ is estimated, then the documents are arranged in ranked order, according to the likelihood that the model can generate, for example, for example: N9772277, N7758173 N7038963, N6066817, N5885165.

Summarizing, the language sense model generates a query probability of the representation of both terms, and meanings of the document, where $P(q_{ti}|d_t)$ and $P(q_{si}|d_s)$ denote the probability of generating the i-th query q_i, successively of terms and senses representation, and the representation of senses d_s can be further extended to represent the meanings of dh hyperonim as well as the meanings of d_r sources representation (equation 4) [11].

$$P(q_t \mid d_t) = \prod_{i=1}^{n} P(q_{ti} \mid d_t) \quad P(q_s \mid d_s) = \prod_{i=1}^{n} P(q_{si} \mid d_s) \qquad (4)$$

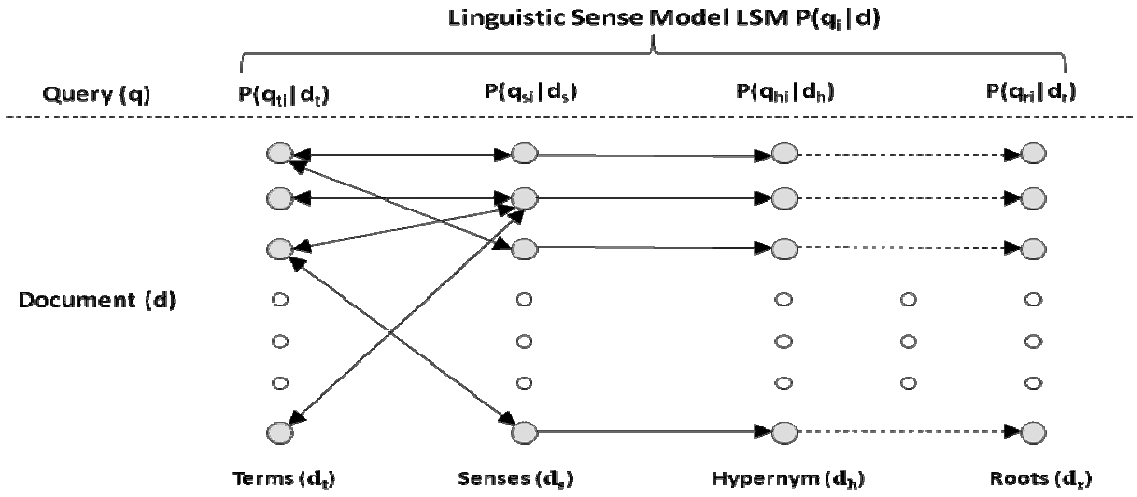


Figure 3: *Linguistic Sense Model P(qi|d)*

7 CONCLUSION

Usage of poetry retrieval methods, based on structural patterns of poetic works and adaptable linguistic models, including the distance measure of words and phrases similarities about non-verbal meaning as well as building new linguistic models and using structural patterns of poems, will allow the efficient acquisition of response to questions which concern poetry retrieval.

Exploring the websites databases of poetry works, based on linguistic models, shows that the present systems of poetry retrieval reveal weaknesses within identification of the contextual ambiguity and under accuracy of information retrieval.

Validity of information retrieval in context of non-verbal requires of users has being emphasized by this year's report by the International Telecommunications Union, which shows that yet 10 years in the whole world was less than 400 million Internet users, while today there are exactly 2.08 billion, which means that about 1/3 of humanity benefits from Internet resources, including 950 million users of mobile broadband and 555 million from fixed broadband with a steady upward trend of potential target groups and relative areas of application, carrying the indisputable benefits of thorough exploration to personal growth and flowering of modern education and information society.

References

[1] J. Reitz. Dictionary for Library and Information Science. Libraries Unlimited. pp. 357. 2004.

[2] A. H. Maslow. A Theory of Human Motivation. Psychological Review. Vol. 50, No. 4. 1943.

[3] N. Chomsky. New Horizons in the Study of Language and Mind. Cambridge University Press. pp. 1-195. 2000.

[4] J. McWhorter. Understanding Linguistic: The Science of Language. The Teaching Company. 2008.

[5] A. Mehrabian. Silent messages. Nonverbal communication. Reedition Publishing, Springer-Verlag, New York. 2002.

[6] B. Bojar. Słownik encyklopedyczny terminologii z zakresu języków i systemów informacyjno-wyszukiwawczych. 2002.

[7] R. B. Donald. Nonverbal Poetry: Family Life-Space Diagrams. Journal of Poetry Therapy. Vol. 14, N.3 pp. 159-167.2000.

[8] C. D. Manning, P. Raghavan, H. Schütze. An Introduction to Information Retrieval. Cambridge University Press. pp. 237 –252. 2008.

[9] W. Shu-Lei Chen. Semantic Structure Extraction and Retrieval of Chinese Poetry, National Tsing Hua University. pp. 1-49. 2004.

[10] D. Davidov, A. Rappoport, O. Tsur. Semi-Supervised Recognition of Sarcastic Sentences in Twitter and Amazon. Computational Natural Language Learning (CoNLL). 2010.

[11] Bao S., Zhang L., Chen E., Long M., Li R., Yu Y. LSM: Language Sense Model for Information Retrieval. Lecture Notes in Computer Science. pp. 97-108. 2006.

MEASURING THE DEGREE IN WHICH THE EFFECTIVENESS OF EDUCATIONAL PROCESS IS IMPROVED

Marian Cristian Mihaescu
Department of Software Engineering, University of Craiova
Bvd. Decebal, Nr. 107, 200440, Craiova, Dolj, Romania
Tel/fax: +40 251 438198;
e-mail: mihaescu@software.ucv.ro

ABSTRACT

A great challenge in e-Learning regards measuring the degree in which the effectiveness of educational process is improved. This paper presents a custom analysis process whose main task is to provide information regarding the effectiveness of switching to a different classification scheme for learners. A baseline classification scheme based on k-means clustering and a custom MTree clustering are presented. For comparisons, the classical characterization formulas are considered.

1 INTRODUCTION

One of the greatest challenges in e-Learning area is to continuously improve existing systems. In order to overcome the challenge, there are needed sound procedures whose task is to prove the challenger procedure creates a better system than existing one. The key issue regarding effectiveness of educational process is classification. The main goal of this paper is to obtain a better or at least acceptable classification scheme with less computation power.

Some possible outcomes of such analysis process are: predicting students' grades (to classify in three classes/clusters of low priority – week learners, medium priority – easy learners, high priority – competitive learners) from test scores.

Clustering algorithms are part of the unsupervised classification techniques. They try to group a set of items into subsets or clusters. The cluster algorithms' goal is to create clusters that are coherent internally, but clearly different from each other. In other words, items within a cluster should be as similar as possible; and items in one cluster should be as dissimilar as possible from items in other clusters. In this paper, learners represent items.

As baseline unsupervised classifier, the standard k-means algorithm is used [1]. K-means is the most important flat clustering algorithm. Its objective is to minimize the average squared Euclidean distance of items from their cluster center where a cluster center is defined as the mean or centroid of the items in a cluster.

M-tree [2, 3] is a dynamic access method suitable to index generic "metric spaces", where the function used to compute the distance between any two objects satisfies the positivity, symmetry, and triangle inequality postulates. The M-tree design fulfills typical requirements of multimedia applications, where objects are indexed using complex features, and similarity queries can require application of time-consuming distance functions. In this paper we describe the basic search and management algorithms of M-tree, introduce several heuristic split policies, and experimentally evaluate them, considering both I/O and CPU costs. Results also show that M-tree performs better than R*-tree on high-dimensional vector spaces.

2 BUILDING CLUSTERS WITH K-MEANS AND M-TREE

The classical standard k-means algorithm partitions a dataset on N instances into K clusters. The algorithm is:

procedure k-means (x_1, x_2, …, x_N; K)
{c_1, c_2, …, c_K} ← Select Random Centroids
for (k=1, k<K)
centroid$_k$ = c_k;//these are initial centroids
while (#centroids are not same){
for (k=1, k<K){
for (n=1, n<N){
j = index of corresponding cluster
#put x_n in corresponding cluster Cj
}//end for
}//end for
for (k=1, k<K)
# compute centroids for all clusters
}//end while

The most important discussion regards the computational complexity of k-means algorithms. Most of the time is spent on computing distances between items. This computing is performed when putting instance x_n in cluster C_j. One such operation costs *log(M)*. The reassignment step computes *KN* distances, so its overall complexity is *log(KNM)*. In the re-computation step, each vector gets added to a centroid once, so the complexity of this step is *log(NM)*. For a fixed number of iterations I, the overall complexity is therefore *log(IKNM)*.Thus, K-means is linear in all relevant factors: iterations, number of clusters, number of vectors, and dimensionality of the space. One of

the most important issues regards the number of iterations. In most cases, K-means quickly reaches either complete convergence or a clustering that is close to convergence. In the latter case, a few items would switch membership if further iterations are computed. This computation has a small effect on the overall quality of the clustering.
The classical M-Tree algorithm has been adapted such that its final structure has two levels. The procedure for building the structure takes into consideration both the desired number of clusters and the filling factor of a leaf node.

procedure MTree (x_1, x_2, …, x_N; K; F)
// K – the number of clusters
// F– filling factor
for (i=1, i<N){
  C_i = FindCentroid(centroids, x_i);
  if (#Leaf[C_i] has *F* instances)
    if (#we have k clusters)
      #put x_i in Leaf[C_i]
    else
      #split Leaf[C_i]
  else
    #put x_i in Leaf[C_i]
  RecomputeCentroids(Leaf[C_i])
}//end for

The computational complexity of this M-Tree procedure takes into consideration that each instance is considered only once. That is why the complexity of this operation is *O(N)*. Still, the number of clusters influences the complexity since the best corresponding cluster needs to be determined. The time taken for this operation id *O(K)*. Still, the recomputation of the centroid is not so costly as in the case of k-means algorithm due to the filling factor parameter. Thus, the overall complexity of employed MTree procedure is O(KNF).

3 CLUSTERING EVALUATION METRICS

A comparison of two distinct procedures defines the needed steps in order to obtain sound results. The presented analysis procedure compares two methods that are used for building clusters of items: k-means and MTree.
An input dataset is considered. Both k-means and MTree algorithms are than used for building clusters from the same dataset. For each set of clusters, there are computed specific indicators for characterizing obtained clusters. The indicators that are taken into considerations are:

Tightness Indicator:

$$Q = \sum_{i=1}^{k} \frac{1}{|C_i|} \sum_{x \epsilon C_i} d(x, \mu_i)$$

where $|C_i|$ is the number of points from cluster i. The value for Q will be small if the data points from the cluster are close. Thus, in the comparison analysis procedure the clusters with smaller computed values of Q have higher quality.

Homogeneity Indicator:

If the centroids of clusters are computed with formula: $r_k = \frac{1}{n_k} \sum_{x \epsilon C_k} x$, where x are the instances from cluster C_k than homogeneity indicator is:

$$H(C) = \sum_{k=1}^{K} \sum_{x \epsilon C_k} d(x, r_k)^2$$

The value for H will be small if a cluster has homogeneous structure. This, in the comparison analysis procedure the clusters with smaller computed values of H have higher quality.

Cluster Distance:

$$CD = \sum_{1 \le j \le k \le K} d(r_j, r_k)^2$$

Where j and k are indexes of clusters whose centroids r are taken into consideration. The value for CD will be big if the similarity among clusters themselves is low. Thus, in the comparison analysis procedure the method with bigger computed values of CD have higher quality.

Weakest Link between Points:
The weakest link for a cluster is the maximal value of all pairs of points belonging to the same cluster.

$$WL = \max (d(x_i, x_j)),$$

for all x_i and x_j belonging to the same cluster.

4 EXPERIMENTAL RESULTS

The experiment takes into consideration 15 students. For each student there are available two weighted parameters: the number of loggings and the time spent on-line. The real parameters scale such that all values are in the range 0 to 16.
The input dataset is:
A1(10.94 , 11.86); A6(11.02,2.28); A11(9.29 , 13.86);
A2(1.58 , 6.27); A7(11.23,9.37); A12(8.00 , 1.09);
A3(13.66 , 4.62); A8(7.35 , 3.99); A13(11.52 , 1.63);
A4(2.33 , 1.16); A9(9.4 , 11.84); A14(5.08 , 7.42);
A5(10.04 , 9.41); A10(13.43 , 8.97); A15(12.12 , 12.59).
This dataset is used for building three clusters with both k-means and MTree algorithms. The obtained distributions are presented in figures 1 and 2.

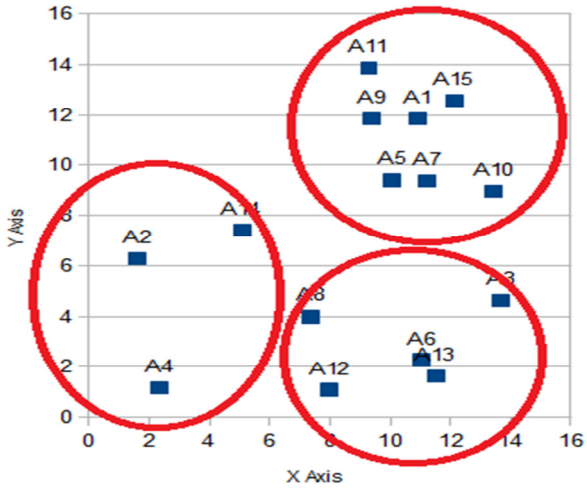


Fig. 1. Distribution of learners with k-Means

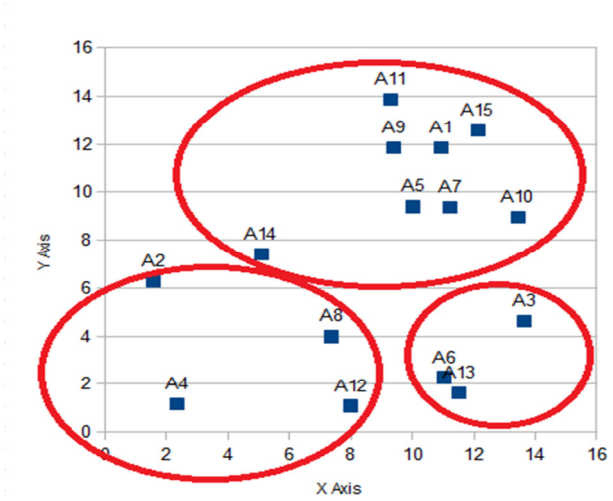


Fig. 2. Distribution of learners with MTree

The obtained clusters by k-Means clustering have the following centroids and composition:

C1 (2.99,4.95) *//Cluster 1's Centroid*
A2 (1.58, 6.27)
A4 (2.33, 1.16)
A14 (5.08, 7.42)

C2 (10.92, 11.13) *//Cluster 2's Centroid*
A1 (10.94, 11.86)
A2 (10.04, 9.41)
A7 (11.23, 9.37)
A9 (9.40, 11.84)
A10 (13.43, 8.97)
A11 (9.29, 13.86)
A15 (12.12, 12.59)

C3 (10.31, 2.72) – Cluster 3's Centroid
A3 (13.66, 4.62)
A6 (11.02, 2.28)
A8 (7.35, 3.99)
A12 (8.00, 1.09)
A13 (11.52, 1.63)

The obtained clusters by MTree clustering have the following centroids and composition:

C1 (10.04, 9.41) - *//Cluster 1's Centroid*
A1 (10.94, 11.86)
A2 (10.04, 9.41)
A7 (11.23, 9.37)
A9 (9.40, 11.84)
A10 (13.43, 8.97)
A11 (9.29, 13.86)
A14 (5.08, 7.42)
A15 (12.12, 12.59)

C2 (2.33, 1.16) *//Cluster 2's Centroid*
A2 (1.58, 6.27)
A4 (2.33, 1.16)
A8 (7.35, 3.99)
A12 (8.00, 1.09)

C3 (11.02, 2.28) *//Cluster 3's Centroid*
A3 (13.66, 4.62)
A6 (11.02, 2.28)
A13 (11.52, 1.63)

For each clustering procedure the evaluation metrics presented in section 3 were computed. The results are presented in the following table:

Indicator	Clustering Procedure	
	k-means	MTree
Tightness	7.55	8.52
Homogeneity	100.47	137.48
Clusters Distance	230.47	203.11

Table. 1. Tightness, homogeneity and cluster distance indicators for k-means and MTree distributions

The link analysis for both distributions is presented in the following table:

Indicator	Clustering Procedure	
	k-means	MTree
Weakest Link Cluster 1	0.9	1.21
Weakest Link Cluster 2	0.84	1.15
Weakest Link Cluster 3	0.87	0.51

Table. 1. Weakest link values obtained for k-means and MTree distributions

The k-means results are obtained using Weka [4]. Weka is a collection of machine learning algorithms for data mining tasks which has implemented the k-means clustering algorithm.

The MTree results are obtained using a custom Java implementation of the algorithm. The main differences of this implementation compared with classical MTree algorithm regards two aspects. One regards the general structure of the tree that is restricted to two levels. This means there is only one root node where centroids along with covered radius are placed. The second issue regards the way *k* (the number of clusters) and f (the filling factor) are managed. If the algorithm is required to produce a certain number of clusters, the instances are placed into appropriate clusters until a filling factor is reached. When this happens, a split is performed. Splitting is no longer performed when the desired number of clusters is reached. In this way, the clustering process is directly managed by the values *k* and *s*.

The comparison of the two obtained distributions reveals the fact that the MTree distribution clusters have lower quality than the ones obtained by usage of k-means. Still, the results obtained by MTree are very different from the ones obtained by k-means. All indicators presented in table 1 have better results for k-means than the ones obtained for MTree. It can be observed that the tightness and homogeneity are better (because they have smaller values) for k-means than for MTree.

Another comparison that may be done regards the mobility of centroids. Although the differences of computed indicator (Tightness, Homogeneity, Clusters Distance) values are not very small, the computed centroids are quite close.

6 CONCLUSIONS AND FUTURE WORKS

This paper presents a procedure that measures the degree in which the effectiveness of an e-learning process has improved. The analysis process is data centered. The data represents experiences provided by learners. In this study two features (attributes) characterize each learner: the number of loggings and the time spent on-line.

The goal of the procedure is to produce clusters of users using two different techniques: standard k-means algorithm implemented in weka and a custom flavor of MTree algorithm with a custom implementation.

The input dataset is restricted to a sample of 15 learners. This choice is because a manual inspection of the obtained clusters is desired. An automated analysis of the obtained clusters is performed by computing some basic clustering quality metrics: Tightness, Homogeneity, Clusters Distance and link analysis. The obtained results show an acceptable quality of the MTree clusters, although the computational complexity of the algorithm is much lower than complexity of k-means.

The main goal of the paper is to find an algorithm that produces acceptable results with complexity much smaller than a classical procedure.

The quality of the obtained clusters has a direct influence over the degree in which the e-learning process has been performed. Unsupervised classification (clustering) is one of the main methods for making evidence regarding the knowledge acquisition of learners. Once a high quality distribution has been discovered, a learner may by clustered at certain moments and progress may be evaluated. Of course, the process needs to be well defined and needs to be based on a high quality clustering procedure.

The future works regard different aspects. A first issue would be to replicate the procedure with more data. This may be accomplished on hundreds or even thousands of learners, if data is available. The clustering procedure is highly influenced by the initial centroids. In custom initialization is advisable. A good starting point may be obtained by using a k-means clustering on a sample dataset from the entire dataset. The quality of the clustering process is directly influenced by the choices made regarding k and f values. Thus, an initialization step may also refer to prior computation of the optimal number of clusters and optimal filling factor. The computation of these parameters may be delegated to other high quality clustering procedure that works on a data sample.

Finally, there may be defined procedures for assessing progress in time and even recommendations. The progress in time may be computed, classifying the learner from time to time. This may yield to a learning path that has been followed by the learner. More than this, there may be obtained recommendations for the learner. The recommendations may regard necessary actions to be taken by the learner in order to improve his/her learning curve.

Acknowledgement

This work was supported by the strategic grant POSDRU/89/1.5/S/61968, Project ID61968 (2009), co-financed by the European Social Fund within the Sectorial Operational Program Human Resources Development 2007 – 2013.

References

[1] MacQueen, J. B.: "Some Methods for Classification and Analysis of Multivariate Observations". 1. Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. University of California Press, pp. 281–297 (1967);

[2] Uhlmann, J.K.: Satisfying General Proximity/ Similarity Queries with Metric Trees. Information Processing Letters. vol 40, pp. 175-179 (1991)

[3] Ciaccia, P., Patella, M., Zezula, P.: M-tree: An Efficient Access Method for Similarity Search in Metric Spaces. In: VLDB'97, Proceedings of 23rd International Conference on Very Large Data Bases, pp. 426-435. Morgan Kaufmann (1997).

[4] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten: The WEKA Data Mining Software: An Update. SIGKDD Explorations, Volume 11, Issue 1 (2009).

LIMITS FOR THE SPREAD OF INNOVATIONS IN LANGUAGE NETS

Michael Schenke
Department of Computer Science and Communication
Merseburg University of Applied Sciences
Geusaer Straße 88,
D-06217 Merseburg, Germany
Tel: +49 3461 463018; fax: +49 3461 462900
e-mail: michael.schenke@hs-merseburg.de

ABSTRACT

The question how changes spread over networks is frequently discussed in literature. In a recent paper [11] a model was presented which does not only consider the spread of a change but also a possible extinction of the change. This is an aspect very rarely dealt with, but interesting not only in computer science. The result is an apparently complex model, a model of a dynamical graph with dynamic states. It can be shown rigorously that simpler models are not complex enough. That will be done in this paper.

1 INTRODUCTION AND PROBLEM STATEMENT

Growth phenomena, from the spread of viruses in computers or in medicine to new patterns of behaviour in society, are frequently considered in literature. A classical method is to model such phenomena by differential equations. The most classical tool is the logistic differential equation

$$p' = b\, p\, (1-p)\,.$$

Here p is the looked for function which describes the monitored quantity, e.g. the number of viruses in dependence of time, whereas p', the derivative, describes its growth. The basic version of the equation $p' = b\, p$ simply says that the growth is proportional to the already reached size. The well-known solution is the exponential function, which describes unbounded exponential growth. In this case b is a real numbered proportionality coefficient being necessary for mathematical reasons. It can be interpreted as a measure, how tightly size and growth are coupled. The third factor in the logistic differential equation calls for the fact that individuals eat up their resources. This slows down the growth. The bigger p is, the smaller will be $1-p$ and hence the whole product.

The solutions of the logistic equation initially also describe exponential growth, but later show various behaviours: bounded growth, chaotic behaviour, total destruction or else, each one being realistic in certain situations.This equation is also the starting point for the Piotrovskij law [9], which describes the expansion of a language change and which is the backbone of the field, e.g. [2]. It turns out, however, that the models are not precise enough in many aspects. One of the reasons seems that by equations only universal properties can be taken into account. During the last two decades hence a new approach has won more and more support, the use of dynamic nets, cf. [3], [15]. Here also local aspects can be considered. It makes a big difference for example in the spread of a new fashion, whether individuals are well connected or not. This can easily be built into net models, other than into differential equations. The model presented in [10,11] exhibits the whole life of the change, from its birth to its (possible) death.

The main objective of this paper is to explain the necessity for a certain complexity of the model.

2 THEORETICAL BACKGROUND: CELLULAR AUTOMATA AND DYNAMICAL NETS

A formalism which has become very popular for the description of growth phenomena during the last decades are cellular automata. Central ingredients of cellular automata are:

- a set of cells which can take one of finitely many states,
- all cells are identical in their capabilities,
- a geometry in form of a regular lattice (with dimensionality d) describing the relative

positions of the cells, very often Z^d, the set of all d-tuples of integers,
- fixed form of the cell-neighbourhood, identical for all cells,
- changes of state simultaneously for all cells at discrete points of time,
- state change depending only on the states of the cell and its neighbours, where the change function is identical for all cells.

An extremely popular example of cellular automata is J.H.Conway`s "Game of Life".
It is well-known that cellular automata are Turing-powerful and hence can model everything what can be modeled at all on computers. The practical capacity of such automata has been impressively demonstrated in [15], and they are a simple model indeed. But they are not practical for growth and spread phenomena. The point is the following: "Turing-powerful" means a principal capability. Nothing is said about speed.
As described above, the solutions of the classical differential equations exhibit exponential growth, at least in the beginning of the growth phase. In cellular automata, however, only polynomial growth can be modelled, as it is shown in the following

Theorem: Let A be a cellular automaton having a geometry with dimensionality d. Any state change of a cell c in A is impossible, if all neighbours of c are in the same state as c, then each state change propagates through the automaton with a speed v with $v(n) \in o(n^d)$.

The theorem is proven by induction on n, the elapsed time. Let O be the origin of the change, the first affected cell. It is easy to see that at time n each affected cell has a distance of at most n from O. Hence all affected cells are positioned in a d-dimensional cube of length $2n+1$ at most. This cube has a volume not bigger than $(2n+1)^d$. This proves the theorem.

What matters here is less the concrete statement, but more the interpretation: you cannot use cellular automata to model exponential growth. The number of cells where the state change has taken place at time n is more or less bounded by n^d, a polynomial function of the time. This contradicts not only classical mathematics but also the observations made in reality, when growth is observed.

Thus the demands concerning the model have to be relaxed. If one looks at the essentials of cellular automata, it is easy to see that the most prospective candidate for a relaxation is the geometry. In such a relaxed model there is no longer a regular geometry and hence no form of the neighbourhood identical for all cells. A consequence of that is a more complex function for the state change. Not longer just one form of the neighbourhood has to be dealt with, but the various ones which might appear. This new formalism, the dynamical net, has also been studied intensively in the literature, e.g. [3]. Formally, a dynamical net is a particular graph, which is mathematically a pair of a set of "nodes" and connections between the nodes. In the graph coming as a dynamical net the nodes have a state, and this state can change at discrete moments of time, the "steps". The dynamics of the net comes in two forms

1. The state of a node n can change in each step, dependent on the previous states of n and all its neighbours.

2. The topology of the net can change: New nodes and new edges can be created or removed at random.

It is not surprising that such a general model is fit for many areas of science. More surprisingly, it is detailed enough to give interesting insights. For this purpose, however, a deeper analysis of the field under inspection is necessary in order to find the appropriate parameters which govern the model, for example which function should describe the change of state in the nodes.

It is well-known that very little can be said about spread and growth phenomena in dynamical nets in general. Their speed typically can range between linear (as in a line) and instantaneous (in complete graphs). Typical examples where spread happens with exponential speed, as it is desirable for growth modelling, are trees. One typical property of nets should be mentioned:

Call the number of neighbours of a node its *degree*. Then the degrees of nodes in many natural nets (www, epidemic models, physics) obey to a certain distribution, the Pareto-distribution. A property of nets with a Pareto-distribution is crucial:

If such a net is sufficiently big, it usually has locally a treelike structure.

All notions in this statistical proposition have a clear mathematical meaning. The proof is known since 2002, cf. [3]. In our context this means that an innovation at first spreads with exponential speed, as it is usual in trees (and in models based on differential equations). Later the speed slows down due to the innovation`s success: There are no more unaffected neighbours. These considerations show, why in nets with a Pareto-distribution a pattern of spread is to be expected similar to the classical approaches. Much research has been done, how a Pareto-distribution can be assured, e.g.[1]. There are theorems which guarantee such a distribution, provided the above mentioned topology change of the net obeys to certain statistical rules. This can easily be done in simulations.

3 APPLICATION TO LANGUAGE CHANGE

The classical approaches, in particular those using differential equations, like the Piotrovskij law, did only model the spread phase and did not really say anything about the reverse process, mostly nothing at all. Our model, however, is capable to model the complete life of a change in a language, including its possible death. This way many open questions concerning irregularities in the behaviour of real language change can be answered, like the ones raised in [5].

The model developed in [10] has the following ingredients: Its nodes represent individuals or groups of them (for example, the staff of a small office or the neighbours' community of a village), which come in two states (for simplicity, but other models are feasable). These indicate the frequency of the new form: complete use of the new form (for brevity: infected) or no use at all (not infected). Each edge connecting two nodes introduces the interaction of people or groups of people (for example, two persons often talk to each other or two groups of people either have common members or closely communicate). Naturally, the human network is not stable in two ways:

1. The existing nodes change: New nodes may appear, other nodes die or cease to exist for other reasons.

2. The communication behaviour can change. New links may be established, old ones may be broken off. Mathematically: New vertices between the nodes may be introduced or removed.

Given these facts, it is reasonable to make the state transition function for a node n dependent on some parameters:

1. The state of n.
2. The number of neighbours of n, called *the degree of* n.
3. The share of infected neighbours of n as a subset of all neighbours of n.
4. A certain hysteresis time which prevents too frequent state changes.

The change in the net topology is gouverned by the following parameters:

1. In each step a fixed number of new nodes is created.

2. The end of a node is modelled by aging: There is a fixed probability for all nodes according to which a node is removed in some step.

3. New edges are created by a procedure called preferential linking. This means that the probability of some node n getting a new connection depends (linearly) on the degree of n at that moment. This way of creation leads to a very realistic distribution of the possible degrees. The number of nodes $no(d)$ in a network with degree d shrinks polynomially with d, cf. [BA99]. This ensures the Pareto-distribution of the degrees.

This model is not fully realistic. The simplifying abstractions are harmless, widely accepted, and they make the model simple enough to derive some facts by pure mathematics. But the approach is far too complex to analise it completely by mathematical means. Computer simulations became necessary. These are described in [11]. Like the model itself, the usefulness of the simulations is not restricted to languages. The computer model has undergone several tests as described in [10].

For that paper the example of the spread of German loan words in the Polish language has been chosen. For linguistic applications the general problem is that there are in general not enough data available to obtain statistically valid results for language change in general. Our data come from a database at the foreign languages department of Oldenburg University, [16]. In this data base for each known German loanword in Polish, among much other information, the year of its first appearance and the year of its last citation can be found. This alone does not say anything about the spread of the single words. There are also linguistic worries about the difference between the first and last citations on one side and its adaptation and disappearance on the other side. These reservations can be solved, however. In any case at least two quantities can be measured in our simulations and compared with the data from the database:

1. Is the ratio of words which are extinct today (as a subset of all words ever borrowed by the Polish language) comparable to the ratio predicted by the model?

2. We consider the duration lengths of the extinct words. Is the distribution of these quantities calculated by the computer as it could be expected considering the data in the database?

The first question can be answered by a simple "yes". The second question can also be answered by a "yes", but here more detailed statistical analysis becomes necessary. It has been done by a Kolmogoroff-Smirnoff test.

4 CONCLUSION

The aim of our contribution is twofold.
Firstly, it should make clear, why the apparent complexity of general dynamic nets cannot be relaxed in favour of the simpler cellular automata, which are often used for similar purposes but can be handled in a mathematical simpler way and used in simulations in a simpler way.
Secondly, the reader should become acquainted with a quite unusual application of a well-known formalism, the model of dynamic nets. This is not only fruitful for "classical"

applications, like computer science or physics, but also for linguistics and other domains indeed. A powerful, mathematically sound and appropriate description of change mechanisms is useful and has been looked for for decades now.

The benefits of net-based approaches are numerous: Approaches in the direction of sociology with similar ideas as the present paper, but different means and different aims are described for example in [6] and [7]. In physics they are used for the description of phase changes (magnetisation, formation of ice, etc.). A major field of application is medicine, in particular theoretical epidemiology, as described in [8]. Also the spread of bush fires and the quantity of sexual contacts on American university campuses have been successfully investigated by means of dynamical nets, to name some fields further apart from our application.

Also in linguistics there exist net-based methods. However, the techniques are very different from our approach. As yet, nodes have not been speakers or groups, but words with various interpretations of what "interaction of words" means. This is the idea in [14]. Language universals, properties common to the majority of language, are investigated in [12,13]. Properties of nets are used to explain characteristica of certain language groups in [4], to name just a few of many successes of such methods.

Our different technique is a new attempt to answer the question, how and why changes appear. Our work gives researchers tools to tackle questions like

- How successful can a central language policy be?
- If there are certain dialects of a language, what can be said about the distribution of the quantities of their speakers?
- Which impact has the homogeneity of a society on the stability of the language?
- Which is the distribution of second languages? Which ones will be preferred?
- Are there sudden shifts in the language use that can be explained by network analysis?
- Can mathematical constants (like the *b* in the logistic differential equation) be explained by network analysis rather than being introduced for mathematical reasons without a clear linguistic meaning?
- Does the structure of the net reflect structures of the language, as hints are given in [4] and elsewhere?
- Do net structures reflect brain structures, as hints are given in [14]?

in a strictly mathematical way. There are hints to solutions in the general theory of nets. These could be exploited in future.

References

[1] A.-L. Barabási, R. Albert: Emergence of scaling in random networks. Science 286, 509-512 (1999).

[2] K.-H. Best: Spracherwerb, Sprachwandel und Wortschatzwachstum in Texten. Zur Reichweite des Piotrowski-Gesetzes. Glottometrics 6, 9-34.

[3] S.N. Dorogovtsev, J.F.F.Mendes: Evolution of Networks. Oxford UP, Oxford 2003.

[4] R. Ferrer y Cancho, R.V. Solé, R. Köhler: Patterns in syntactic dependency networks. Phys. Rev. E 69, 051915, 2004.

[5] G. Hentschel: Zur „Seuche" des deutschen Lehnwortes im Polnischen und zu den „Selbstheilungskräften" dagegen. In: Bochnakowa, A. & . St. Widłak (eds.), Munus Amicitiae. Studia linguistica in honorem Witoldi Mańczak Septuagenarii. Kraków, 69-78.

[6] J. Klüver: Soziologie als Computerexperiment. Vieweg – Serie Wissenschaftstheorie, Wissenschaft und Philosophie, Bd.41, Braunschweig, Wiesbaden 1995.

[7] Q. Michard, J.-P. Bouchaud: Theory of collective opinion shifts: from smooth trends to abrupt swings. New Scientist, 7.5.2005, 15-25.

[8] M.E.J. Newman: The spread of epidemic disease on networks. Phys. Rev. E 66, 016128 (2002)

[9] A.A. Piotrovskaja, R.G. Piotrovskij: Matematiceskie Modeli diachronii i tekstoobrazovanija, in: Statistika reci i avtomaticeskij analiz teksta. Leningrad, Nauka 1974, 361 – 400.

[10] M. Schenke: Sprachliche Innovation – lokale Ursachen und globale Wirkungen. Südwestdeutscher Verlag, Saarbrücken (2008).

[11] M. Schenke, A.Jhosan: Spread of Innovations in Language Nets. Proc. Intern. Conf. InterSymp 2011 (Baden-Baden, August 1-5, 2011) (to appear).

[12] R.V .Solé: Syntax for free? Nature 434, 289 (2005).

[13] R.V. Solé, B.C.Murtra, S.Valverde, L.Steels: Language Networks: Their structure, function and evolution. http://www.santafe.edu/research/publications/workingpapers/05-12-042.pdf

[14] M. Steyvers, J.P. Tenenbaum: The large-scale structure of semantic networks: statistical analyses and a model of semantic growth. Cognitive Science 29, 41-78 (2005).

[15] S. Wolfram: A new kind of science. Wolfram Media 2002.

[16] A. de Vincenz, G. Hentschel: Wörterbuch der deutschen Lehnwörter im Polnischen (to appear).

Zbornik 14. mednarodne multikonference

INFORMACIJSKA DRUŽBA – IS 2011

Proceedings of the 14th International Multiconference

INFORMATION SOCIETY – IS 2011

Internet in Slovenija: 1985 - 1995

Internet and Slovenia: 1985 - 1995

Uredili / Edited by

Andrej Brodnik, Matjaž Gams, Rok Sosič

http://is.ijs.si

13. oktober 2011 / October 13th, 2011
Ljubljana, Slovenia

PREDGOVOR

Letos mineva dvajset let ne samo od osamosvojitve Slovenije, ampak tudi nekako 20 let od prve nekomercialne povezave med Slovenijo in svetom. To je bil zadosten razlog, da smo se odločili pripraviti konferenco na tematiko interneta v Sloveniji v letih 1985 do 1995. V omenjenih desetih letih se je od skupka poskusov povezave med različnimi računalniškimi centri v Sloveniji razvilo trdno akademsko omrežje, za razvoj katerega je skrbel ARNES. Ob tem ne smemo pozabiti na komercialna omrežja, ki so sledila.

V pričujočem zborniku konference Internet in Slovenija 1985-1995 je zbranih osem prispevkov, ki opisujejo začetke omrežnega povezovanja v Sloveniji:

- *Tomaž Kalin*, **Razvoj akademskega računalniškega omrežja v Evropi in pri nas (Osebni pogled)**.
- *Borka Jerman-Blažič*, **Prihod prve internetne linije v Slovenijo I del**.
- *Borka Jerman-Blažič*, **Prihod prve internetne linije v Slovenijo II del**.
- *Denis Trček*, **Dan D - zgodovina neposrednega operativnega povezovanja Slovenije v internet**.
- *Rok Vidmar*, **Internet 1985-1995**.
- *Marko Bonač*, **Izzivi ob ustanavljanju ARNESa**.
- *Avgust Jauk*, **Medmrežje v Sloveniji - od začetkov do eksplozije interneta**.
- *Davor Šoštarič*, **Pogled Univerze v Mariboru in Instituta informacijskih znanosti 20 let pozneje**.

Prispevki uravnoteženo podajajo poglede na nastajanje interneta pri nas in le-tega umeščajo tehnološko in politično v širši kontekst. Verjamemo, da bo pričujoč zbornik pomenil vsaj majhen doprinos v pisanju zgodovine računalniških komunikacij v Sloveniji.

Andrej Brodnik, Matjaž Gams, Rok Sosič

RAZVOJ AKADEMSKEGA RAČUNALNIŠKEGA OMREŽJA V EVROPI IN PRI NAS.

(Osebni pogled)

Tomaž Kalin
Tel: +386 41 662798
e-mail: tomaz@kalin.si

Povzetek

Ta dokument poskuša prikazati razvoj akademskega računalniškega omrežja v Evropi in pri nas. Pri tem začenja s "prazgodovino" teh prizadevanj, da je moč razumeti okolje, v katerem se ja razvijalo znanje in potrebe za povezovanje v svetovne informacijske tokove.

Pri pisanju sta bili posebaj koristni dve deli: H. Davies and B. Bressan "A History of International Networking" in raziskava T. Oblak–Črnič "O začetkih Interneta na Slovenskem"

1 UVOD (prazgodovina)

Slovenski raziskovalci so že v zgodnjih sedemdesetih letih začeli sodelovati v prvem tovrstnem mednarodnem projektu COST11 European Informatics Network, eksperimentalni računalniški mreži, ki je povezovala vozlišča v petih evropskih univerzah, oziroma raziskovalnih centrih: National Physical Laboratory (Velika Britanija), INRIA (Francija), ETH (Švica), Politenico di Milano (Italija) in Joint Research Centre (Ispra, Italija). Norveška, Portugalska, Švedska, Nizozemska in Jugoslavija pa so sodelovale pri razvojnem delu, kajti za EIN je bila definirana in razvita celotna paleta lastnih protokolov: od nepovezavno orientiranega paketnega, do posebej izdelanega transportnega protokola. Pri tem delu so bili prisotni tudi sodelavci iz našega okolja. Žal je delo zelo ovirala odločitev evropskih PTT, da vpeljejo serijo X25 protokolov in z njimi zgradijo javno podatkovno omrežje. Še predno so bili zmožni nuditi zanesljivo storitev, so prekinili najete povezave med vozlišči EIN, ki smo jih morali nadomestiti s tuneli skozi X25 povezave. Toda kljub temu je vseevropska demonstracija rezultatov, leta 1978, pri kateri je sodelovala tudi Slovenija, preko klicne povezave do enega od vozlišč, zelo lepo uspela. Glede na stališča, takrat še monopolnih PTT-jev, in hitrega razvoja Interneta v ZDA, ni bilo smiselno še naprej razvijati svojega nabora protokolov in lastne infrastrukture. Zato se je nadalnjevalni projekt COST11bis (1981 do 1985), ki ga je vodil avtor tega sestavka koncentriral pretežno na aplikacije in njihove protokole. Pri tem pa smo seveda bili čedalje pogosto seznanjeni s storitvami, ki so potekale preko omrežij, kot elektronska pošta, centralni depozitorij dokumentv itn. Ob tem se je čedalje bolj krepilo zavedanje prednosti teh storitev pred klasičnimi, kot recimo telefaks, in nujnosti, da se vključimo v evropska prizadevanja za njihov vzpostavitev na evropskem nivoju. Jasno je bilo tudi, da je treba graditi lastno omrežje, saj javne storitve, ki so jih nudil PTTji niso zadovoljevale po zmogljivosti, fleksibilnosti in ekonomičnosti.

2 COSINE

COSINE (Cooperation for Open Systems Innterconnection Networking in Europe). Pri projektu v okviru EUREKA smo začeli sodelovati že leta 1987, zahvaljujoč takratnemu glavnemu tajniku RSS dr.E. Pirkmajerju, takrat še v okviru Jugoslavije, vendar v celoti z našimi ljudmi v Upravnem odboru, kot pri izvedbi priključka na evropsko infrastrukturo.

Enoletno obdobje namenjeno specifikaciji je bilo končano leta 1987, nato je bila podpisana nova triletna pogodba za izvedbeno fazo. Proračun je bil 30 milijonov EUR, od teh je EC prispevala dve tretjini sredstev. Za operativno vodenje projekta je bil izbran RARE (združenje evropskih raziskovalnih mrež), izvedba pa je bila na mednarodnem razpisu poverjena Nizozemskemo telekomu KPN. Ta je zgradil omrežje z imenom IXI (International X25 Infrastructure), ki je, kot že ime pove temeljila na X25 protokolu. Treba se je zavedati, da je Evropska komisija zahtevala, da se v Evropi dosledno uporabljajo protokoli serije " X " in je s tem pogojevala svoj finančni prispevek.

IXI je začela poskusno obratovanje konec leta 1989. Takrat so naši kolegi na Inštitutu Jožefa Štefana pripravili YU vozel in se priključili na IXI, začasno preko X25 povezave, ki jo je nudila slovenska PTT. Z ostalimi interesenti v Jugoslaviji smo dosegli dogovor, da je bila dostopna točka locirana na IJS, kar je bilo izredno pomembno pri prehodu na samostojno državo, ki je prevzela obveznosti do COSINE projekta. Sodelavci IJS, pod vodstvom M. Bonača ki so se nato leta 1992 pridružili novo ustanovljenemu ARNESU, so seveda v tem obdobju pridobili znanje in izkušnje, ki so bile ključne pri nadaljnem razvoju storitev za raziskovalne inštitucije, univerze, šole in druge uporabnike v Sloveniji. Od tega vozlišča so šle povezave na Hrvaško, Srbijo in BiH. Povezavo preko X25 mreže PTT Slovenije ni bila kvalitetna, zato smo zaprosili za najet vod 64 kb?sek, kar pa je kar dolgo zaviral Zvezni sekretariat za promet in

zveze (pravzaprav je zavlačeval KOS). Končno je bila ta zveza vzpostavljena in je bila pomembna med drugim tudi za slovenske napore za obveščanje mednarodne javnosti v času samosvajanja.

3 EMPB

Zaradi neverjetnega uspeha Internet protokola v ZDA, se je tudi v Evropska Komisija začela zavedati, da ni smiselno vztrajati na X25 kot izključni tehnologiji za raziskovalno mrežno infrastrukturo je pripeljalo do spremembe v IXI mreži. Dobavitelj vozlišč je jih je dopolnil, tako da so podpirala tudi IP protokol. Tako se je IXI transformiral v EMPB (European Multi-Protocol Backbone) pri katerem so se kapacitete povezav v hrbtenici povišale do 8 Mb/sek. EMPB je deloval do oktobra 1995. V tem času je nastopil nov element z imenom EBONE, čista IP infrastruktura, ki je povezovala v paralelen sistem manjše število evropskih nacionalnih akademskih mrež. To dejstvo je seveda vzpodbudilo relativno hitro transformacijo iz IXI mreže v EMPB, ki je nudil tudi IP vmesnike za priključek.

4 DANTE

Kot že rečeno je operativno vlogo v začetku prevzel RARE, kar pa se je izkazalo za problematično, saj je bil ta interesno združenje akademskih mrež, ki je težko prevzemalo vodenje nekaj 10 milijonskih projektov. Zato smo se odločili, (T. Kalin je bil takrat generalni sekretar RARE) da organiziramo neodvisno enoto "Operational unit ", ki bo skrbela za upravljanje infrastrukture. Po dokaj kompliciranem postopku je končno uspelo marca leta 1994 ustanoviti neprofitno organizacijo s sedežem v Cambridgeu (VB), ki je bila last evropskih deležnikov pri razvoju infrastrukture, vključno z ARNESOM. Ta enota je dobila ime DANTE (Delivery of Advanced Network Technology to Europe). Imenovana sta bila dva enakopravna direktorja Dai Davies in Howard Davies, od katerih je prvi skoraj 20 let vodil razvoj evropske raziskovalne in akademske računalniške infrastrukture (H. Daviesa po njegovi upokojitvi za krajše obdobje 2002 do 2004 zamenja T.Kalin). Vztrajnosti, diplomatskim občutkom za delovanje v izredno kompeksni in pogosto kontroverzni evropski sceni in trdemu delu obeh " Daviesov " se imamo zahvaliti za nesluten razvoj v zadnih 18 letih. DANTE ves čas operativno vodi mrežno infrastrukturo in je pomemben element pri vseh tehničnih in poslovnih odločitvah pri zagotavljanju storitev.

5 EuropaNET

Ko se je približeval konec pogodbe za EMPB, je bilo treba poiskati rešitev za naslednje obdobje. Mednarodni razpis je najbolje zadovoljil British Telecom, ki je ponudil rešitev kjer je IP osnovni protocol, X25 pa teče preko IP povezav. V hrbtenici mreža so bile povezave s kapacitetami od 2 Mb/sek do 8 Mb/sek. Julija 1994 je bilo priključenih že 18 nacionalnih akademskih mrež. Pristopne hitrost pa so bile od 64 kb/s pa do 2 Mb/sek. ARNES je imel dva priključka, 128 kb/sek IP in 64 kb/sek preko Telekomove X25 mreže.

6 TEN34 (Trans European Network 34)

Za ta naslednji korak je značilno zelo težavno pogajanje z EC, saj je v tistem času potekalo nekaj projektov, ki jih je financirala EC, ki so bili prepričani, da so sposobni nuditi usluge mrežnih povezav za raziskovalce. Eden od takih je bil NICE (National Hosts Interconnection Experiments), kjer se niso zavedali, da ni mogoče zamenjati delujočega servisa, pa čeprav s pomankljivo kapaciteto, z eksperimentalno akcijo. Po dolgih tehničnih pogajanjih je bilo odločeno, da akademske mreže in DANTE, sami nadaljujejo z gradnjo infrastrukture. Poseben problem so bili še vedno monopolni PTTji, vendar je eden od mednarodnih operaterjev pristal, da dobavi najete vode med večjimi državami, sledili pa so tudi drugi. Tako se je gradnja TEN34 lahko začela, vendar je žal Evropska Komisija delila mnenje z Evropskimi PTT, da je ATM "tehnologija prihodnosti " in tako je v TEN34 Internetni pritokol tekel direktno na najetih vodih med 6 članicami zadruge, pri ostalih pa na ATM povezavah. TEN34 je obratoval od marca 1997 do novembra naslednjega leta, ko se mesečni promet povečal od 7000 na 40000 GBytov na mesec. Treba je bilo začeti načrtovati novo povečanje kapacitet. Ob tem dosedanjem pregledu dogajanj bi se lahko vprašali, zakaj je bilo treba napredovati v tako majhnih korakih? Ali ne bi bilo efektivnejše in ceneje, če bi delali večje korake, ki bi zadoščali za dlje časa. Razlog za tako počasno napredovanje so bili tudi v razvitem delu Evrope še vedno zelo močni monopoli telekomunikacijskih operaterjev, saj je splošna liberalizacija trga, ki jo je zahtevala EC stopila v veljavo šele leta 1998. Vendar so se visoki okopi začeli že rušiti in tako je bl možen naslednji korak:

7 TEN155

Ime pove, da so se kapacitete omrežja zvišale na 155 Mb/s, na nekaterih povezavah v osrednjem obroču (Paris, Briselj, London, Amsterdam in Frankfurt) pa so že novembra 2000 vzpostavljene zvez s hitrostmi 622 Mb/sek.

TEIN155 je treba smatrati kot nujno povišanje kapacitet, ni pa prinesel nekih povsem novih rešitev , pač pa je pripravil teren in izkušnje med sodelavci DANTEJA in evropskih akademskih mrež, ki so omogočale končni preboj, ki se uresničil z gradnjo GEANTA.

8 RAG

Koncem 1999 je EC imenovala ekspertno skupino, v kateri je sodeloval tudi T.Kalin, ki so jo imenovali Requirements Action Group. Skupina je pripravila dokument za dolgim naslovom "High-Level Requirements for Broadband Interconnection of National Research, Education and

Training Networks and Testbeds “. V tem dokumentu je bilo nekaj zelo jasnih stališč, med katerimi je najvažnejše naslednje:

“ Potrebno je zgraditi in upravljati najbolj napredne trans - evropske povezave med nacionalnimi raziskovalnimi in izobraževalnimi mrežami, ki bodo zadostovale za pokrivanje zahtev evropskih akademskih in industrijskih raziskovalcev. To vključuje povečanje obstoječih kapacitet tako, da se bodo približale gigabitom/sek, vpeljavo različnih nivojev QoS in izboljšanje povezav s tretjimi državami. “

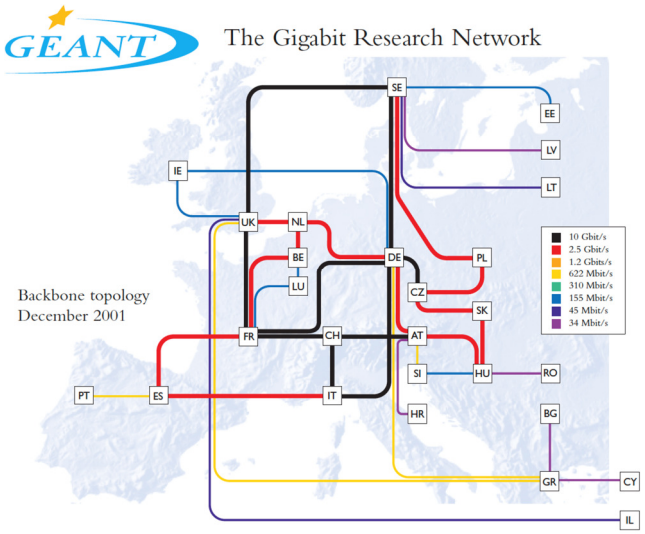
Dokument je predvideval, da se morajo pristopne kapacitete povišati na najmanj 2.5 Gb/sek, s tem da mora infrastruktura podpirati tudi nekaj 10 Gb/sek, pa tudi uporabo najetih neosvetljenih optičnih vaken.

Zadnja verzija dokimenta ima datum 15 januarja 2000, DANTE in nacionalne mreže pa so morali pripraviti predlog projekta do 31. marca 2000, da se je uvrstil v financiranje v 5 okvirnem programu.

9 GEANT

Konzorcij nacionalnih mrež in DANTE so se odzvali na ta izziv in pripravili predlog projakta, za evropsko infrastrukturo ki so jo imenovali GEANT in je prva, od dosedaj treh inkarnacij, v globalnem smislu najpomembnejše tovrstne organizacije. Hrbtenico so sestavljali vodi s hitrosmi do 10 Gb/sek, ki so povezovali Juniper usmerjevalnike različnih zmogljivosti. Na to hrbtenico je bilo povezanih 27 evropskih nacionalnih mrež, polega tega pa partnerji v ZDA, Aziji in Južni Ameriki

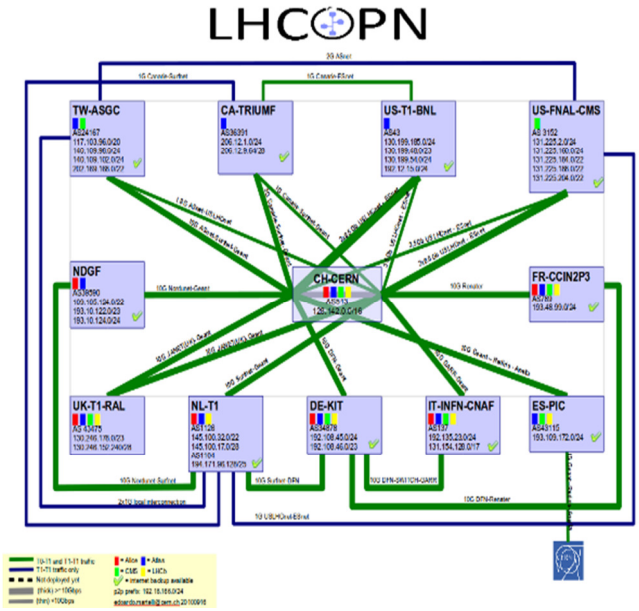
V dobrem letu je zrastla hrbtenica, ki je zadoščala potrebam akademske in raziskovalne skupnost od leta 2001 za 4 leta:



Slika 1. Toplogija GEANTA koncem leta 2001

10 GEANT2

GEANT2 se od svojega predhodnika razlikuje v pomembnem elementu. Ta je bil čista IP mreža, ki je tekla na najetih do 10 Gb/sek in je vsak od partnerjev imel po en IP priključek. GEANT2 temelji v svojem večjem delu na uporabi najetih optičnih vlaken, na katerih je relativno poceni mogoče vpeljati dodatne “lambde”, to je komunikacijske kanale s hitrostmi najmanj 10 Gb/sek, pa do 40 Gb/sek. Ti dodatni kanali so lahko namenjeni IP prometu, ki ga seveda posredujejo usmerjevalniki, ali pa neposredno direktnim povezavam med dvema ustanovama. Zato tej verziji infrastrukture rečemo hibridna mreža, ker nudi poleg Interneta svojim uporabnikom tudi hitre direktne povezave. Najbolj znana uporaba takih zvez je povezana z obdelavo ogromne količine podatkov, ki jih generira LHC (Large Hadron Collider) v CERNU in ki jih v samem centru ne bi mogli obdelati. Zato se podatki preko hitrih 10Gb/sek povezav prenašajo v 11 raziskovalnih centrov po vsem svetu, ki imajo zelo zmogljive računalnike, ti podatke sami obdelujejo, poleg tega pa jih še naprej posredujejo na 160 centrov na naslednjem nivju. Tako je s to svetovno mrežo računalnikov mogoče obvladovati masovne podatke.



Slika 2 LHCOPN omrežje za prenos podatkov

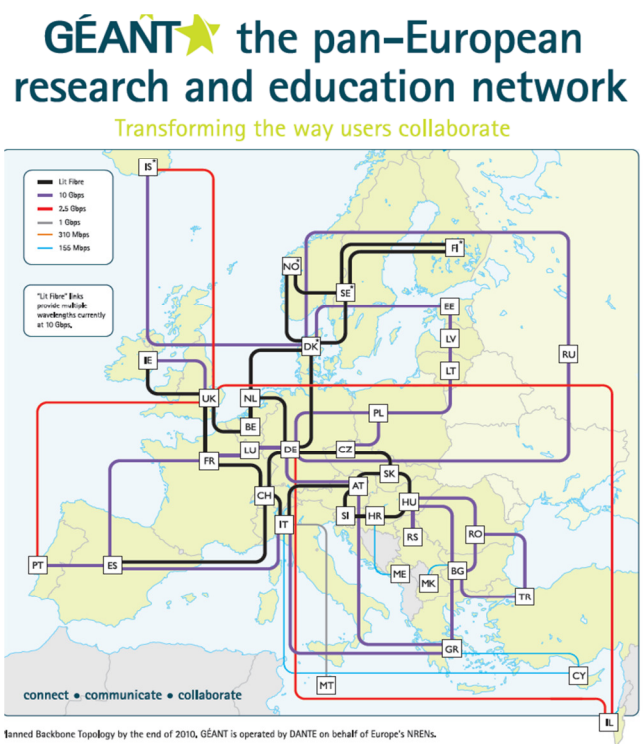
Nekaj manjše potrebe imajo tudi raziskovalci, ki z radijskimi teleskopi peoučujejo vesolje in morajo sproti izmenjevati podatke preko hitrih povezav, in še mnogi drugi.

11 GEANT3

GEANT3 ki je začel delovati aprila 2009 nadaljuje delo ki se je začelo z njegovim predhodnikom uporabljajoč tehnologijo in servisne elemente razvite v GEANT2. Na nekaterih relacijah se je po najetih optičnih vlaknih dvignila hitrost prenosa na 40Gb/sek.

Ena od pomembnih tem, ki jih obravnava upravni odbor konzorcija je vprašanje, ali je treba spremeniti model organiziranja za zagotavljanje čedalje večjih pa tudi čedalje bolj specifičnih potreb uporabnikov.

Stanje mreže se je v zadnih 10 letih spremenila le v toliko, da odraža povečanje kapacitet, vendar ob nižjih stroških, saj se je začel kazati vpliv liberalizacije telekomunikacijskega trga po vsem svetu.



Slika 3. Topologija GEANT3

12 GLOBALNI VPLIVI

Vpliv, ki je posledica odlično razvite raziskovalne mrežne infrastrukture, ki ji ni para na svetu, se ne ustavi na pragu Evrope. Partnerji GEANTA, DANTE in Evropska komisija so vedno vlagali sredstva v globalno povezanost s tretjimi deželami. Poleg hitrih povezav z raziskovalnimi infrastrukturami v ZDA so izpeljani razvojni projekti, ki širijo povezanost raziskovalcev vsega sveta:

- EUMEDCONNECT2 - mediteransko področje
- ALICE2 - povezava z RedCLARA združenjem Latinskoameriških mrež
- Dežele Južnega Kavkaza
- UbuntuNet - Podsaharske dežele
- CAREN - centralna Azija
- TEIN3 - dežele južne Azije in Pacifika
- ORIENT - Kitajska

Pri teh projektih DANTE opravlja vlogo vodje projekta, finančna sredstva pa prispeva EC, pri večini projektov pa tudi evropske nacionalne mreže, članice GEANTA.

APLIKACIJE

Pri ocenjevanju razvoja na opisanem področju ne smemo tudi pozabiti na razvoj kopice aplikacij in orodij, ki so nastale s sodelovanjem vrste nacionalnih mrež med seboj pa tudi v širšem okolju. Tu je prav gotovo potrebno omeniti:

- eduroam - enostavno gostovanja v tujih omrežjih
- perfSONAR - sistem za preverjanje karakteristik in delovanja povezav od konca do konca
- in mnoge druge

ZAHVALA

Ob pregledu neverjetnega napredka, ki ga je doživelo povezovanje vseh raziskovalcev s pomočjo elektronskih komunikacij ne smemo spregledati prispevka kolegov iz nešega okolja, ki imajo ves čas pomembne funkcije v delovnih telesih in pri upravljanju projektov.

Mnogo ljudi je zaslužnih za uspehe na tem področju, in jih je preveč, da bi se jim lahko vsakemu posebej zahvalili, vendar ne moremo mimo pred kratkim umrlega Klausa Ullmanna, direktorja nemške akademske mreže, ki je bil vedno z vsem srcem na prvi fronti, ter je pomembno vplival na razvoj in obeh, že navedenih direktorjev DANTE Howarda in Daia Davies.

Doma pa smo lahko neskončno hvaležni Marku Bonaču, direktorju ARNESA, ki je z neverjetno vztrajnostjo, ne glede na mnogokrat neugodno okolje gnal napredek pri nas, seveda pa to velja tudi za njegovo ekipo, ki že skoraj 20 let vztraja pri tem mnogokrat nehvaležnem poslu. Njihova zaluga je, da imamo pri nas infrastrukturo, ki je vsaj enakovredna če ne boljša kot v enako razvitih državah.

PRIHOD PRVE INTERNETNE LINIJE V SLOVENIJO I DEL

Borka Jerman-Blažič
ISOC Slovenija
Slovenia
Tel: +386 1 4773 408; fax: +386 1 477 3395
e-mail: borka@e5.ijs.si

POVZETEK

Prispevek podaja na kratko potek razvoja Interneta in vojno med arhitekturnimi koncepti.

1 UVOD

Razumevanje današnjega Interneta in načrtovanje njegove prihodnosti sloni predvsem na zgodovinskemu razvoju omrežnih konceptov in dolge tridesetletne vojne med njimi, ki se je končala z zmago internetnega koncepta. Bralci uvoda v to zgodovino lahko mislijo, da vse to danes nima nobenega pomena, ko vendar imamo Google in Facebook! Vendar lahko zatrdimo, da je vse to zelo pomembno. Najprej je pomembno, da vemo, kako je potekal razvoj današnjega Interneta, kakšno je njegovo dejansko stanje in kakšna bo njegova evolucija v prihodnosti. Trditev, da je vse, kar je bilo narejeno, le rezultat tehnike in inženirstva in da je bila to edina pot, ne drži popolnoma, ker se pri tem pozabljajo nastali miti in napake, ki lahko še danes vplivajo na razvoj prihodnjega Interneta. Dejstvo je, da kljub velikim naporom v zvezi z deregulacijo in liberalizacijo telekomunikacijskega trga storitev do velikih sprememb na tem trgu ni prišlo. Tehnologija in nomenklatura se je spremenila, vendar so napetosti še tukaj, ker so ista vprašanja še zmeraj v "zraku": Kdo bo kaj prodajal? in Kdo bo nadziral dohodek?. Znani politični in gospodarski akterji vodijo razvoj tehnologije in vsak vleče na svoj konec, ne glede na to, kako ga kdo imenuje: ali novi multimedijski internetni sistem IMS (Internet Muldimedia System) ali pa novi model za dostop do storitev Interneta ISP (Internet Service Provider model). Ponujene nove storitve v omrežju z novimi inteligentnimi usmerjevalniki, ki zagotavljajo nove, čudovite omrežne aplikacije, so samo poizkus vzpostavitve mej glede tega, kdo lahko prodaja storitve in kdo tega ne sme delati.

2 ZAČETEK VOJNE DVEH ARHITEKTURNIH KONCEPTOV

Razvoj podatkovnih komunikacij v letih 1985–1995 in relativno pozen prihod Interneta v Slovenijo je rezultat dogajanj v razvitih državah sveta: Evrope in ZDA, ter v poznih osemdesetih letih pri nas doma. Razvoj vede, danes znane kot znanost o omrežjih (network science) se je oblikoval med dolgoletno vojno za prevlado enega od dveh razvijajočih se arhitekturnih modelov podatkovnih komunikacij. Ta dva modela sta znana kot model Mednarodne organizacije za standardizacijo ISO/OSI (International Standard Organization, Open Systems Interconnection) in kot arhitekturni model omrežja ARPANet, ki je dobilo ime agencije ARPA Ministrstva za obrambo ZDA (Advance Research Project Agency), ki je projekt financirala in nadzorovala. Projekt ARPANet in omrežje, ki je ob tem nastalo, je pozneje v akademskih krogih dobilo ime Internet. Vojna teh konceptov je trajala približno 30 let, od poznih šestdesetih do zgodnjih devetdesetih, ki jih imamo za obdobje komercialnega interneta in kot obdobje zmagoslavnega pohoda Interneta po vseh celinah sveta. Koncept modela ISO/OSI so podpirali predvsem tedanji monopolisti na področju telekomunikacij, znani kot podjetja za pošto in telekomunikacije (PTT). Model ARPANet pa je postal klub dobaviteljev in proizvajalcev računalniške opreme, predvsem iz ZDA. Vojna se je končala v prvih letih zadnjega desetletja 20. stoletja z zmago modela raziskovalnega omrežja Internet, ki se je v l. 1992 preoblikoval v komercialno omrežje in se razširil povsod po svetu. Ves čas trajanja te vojne niti ena od obeh arhitektur, ki sta bili predmet spora, ni bila ne dorečena in ne popolnoma razvita in tako je podobno kot v drugih zgodovinskih primerih znanosti in tehnike tudi ta vojna imela pozitiven učinek na razvoj temeljev in splošno sprejetih konceptov nove vede, vede o omrežjih. Oba arhitekturna modela sta izhajala iz dveh nasprotujočih si konceptih podatkovnih komunikacij in sta oba prispevala k nastanku tehniške vede o omrežjih. Temelji te

vede so pozneje postali izhodišče za razvoj implementacij, ki jih danes poznamo kot storitve elektronskih komunikacij, med katere uvrščamo prenos govora, videovsebin, elektronske pošte, spleta ter pridruženih družabnih spremljevalcev. Med njimi pa so vsekakor najbolj znana družabna omrežja na trgu 21. stoletja s kapitalsko vrednostjo, ki gre v nebo.

3 OMREŽNE ARHITEKTURE

Temelj obeh konceptov pa so t. i. »sloji« v podatkovnih komunikacijah, ki so se kot koncept pojavili v zgodnjem obdobju računalništva konec šestdesetih let in v začetku sedemdesetih letih prejšnjega stoletja. Obdobje sedemdesetih let je tisto, ki je ponujalo uporabnikom računalnike z najmanj 30 različnimi računalniškimi operacijskimi sistemi (med njimi so Multics/UNIX, VMS, IBM OS, DOS) in ki ga je najbolj zaznamoval utemeljitelj računalništva Norbert Weiner, ki je v novo znanstveno disciplino – računalništvo (computer science) vpeljal pojem »črne škatle« in jo opredelil kot sistem storitev na njenem obrobju, ki jih zagotavlja drugim komponentam v računalniku. Delovanje črne škatle v notranjosti oziroma način, kako je bila narejena, ni imel nobenega vpliva na njeno vlogo v računalniškem sistemu. Razvoj slojevite omrežne arhitekture izhaja iz koncepta črne škatle in vpeljanega koncepta abstrakcije na področju tehnike, tako tipičnega za do tedaj klasične znanstvene vede (med prvimi, ki so vpeljali abstrakcijo, je bil Galilej z enačbami o poteh topovskih krogel ter veliko pozneje tudi Isaac Newton), vendar malo manj uporabljana na področju tehniških ved. Vojna dveh omrežnih konceptov se je vrtela okrog abstrakcije pri določanju načina delovanja mrežnih sistemov, vendar le-ti po naravi niso bili čisto znanstveno-tehnične rešitve, ampak predvsem politične, zgodovinske in, še kar je najslabše pri vsem tem, predvsem ekonomske. Ekonomska plat izbire tehnične rešitve ali koncepta je najslabša mogoča rešitev pri razvoju znanstvenih konceptov, ker se te rešitve rojevajo z namenom spremeniti poslovni model in tako posredno odločajo o tem, kdo bo na trgu služil denar.

V 30-letni vojni so se bitke vodile okrog »arhitekture nanizanih zrn na žici« (imenovanje, ki ga je dal John Day v svoji knjigi Vrnitev k temeljem, vzorci v omrežnih arhitekturah, 2011) in »arhitekturnega modela rahlo povezanih omrežnih sistemov« ter o najbolj primernem protokolu za prenos podatkov: povezavnega in nepovezavnega. Povezavnega so zagovarjali pristaši modela »nanizanih zrn na žici«, drugega pa pristaši Interneta.

4 RAZISKOVALNA OMREŽJA

Prva omrežja za prenos podatkov, ameriški ARPANet in francoski CYCLADES, so zgradili računalniški strokovnjaki za operacijske sisteme in ne eksperti za telekomunikacije. Med nastajanjem slojevite arhitekture operacijskega sistema, ki ga je zasnoval v l. 1968 znanstvenik E. W. Dijkstra, je bila disciplina, znana kot programsko inženirstvo, v povojih in operacijski sistemi so tedaj bili najbolj kompleksna programska oprema na svetu. V tem času je bil tudi projekt ARPANet zamišljen kot omrežje, ki naj bi zagotovilo porazdeljeno uporabo računalniških virov, da bi se izognili koncentraciji računalniških kapacitet na eni lokaciji in dali možnost izvrševanja zadanih nalog na daljavo. Sistem je bil zaradi enostavnosti načrtovanja zasnovan kot slojevit arhitekturni model, ki je imel za osnovo razumevanje omrežnega sloja kot Wienerovo črno škatlo. Bistvena lastnost sloja v omrežni arhitekturi je bila v tem, da se funkcionalnosti, ki jih zagotovi en sloj omrežja, ne da ponoviti v kakšnem drugem sloju omrežja. V l. 1974 je s prehodom ARPANet-a v akademsko sfero (v Stanford Reserach Institute in druge) omrežje dobilo ime Internet in njegova arhitektura je bila določena kot sistem, sestavljen iz 4 slojev: fizični sloj, sloj podatkovne linije, relejni sloj, ki je posredoval podatkovni promet do naslova sprejemnika podatkov (Inter NET), in aplikacijski sloj, ki je moral opraviti konkretno delo za uporabnika (V l. 1974 sta Bob Khan in »oče« Interneta Vint Cerf objavila znamenito publikacijo v reviji ACM, ki je opredelila temeljni internetni protokol IP. Oba sta zato 20 let pozneje dobila ACM-nagrado, ki jo v svetu računalništva imajo za enakovredno Nobelovi nagradi za znanost.)

Ob tem je treba omeniti, da so dejansko zgradili prva komunikacijska omrežja telefonska podjetja in organizacije. Med razvojem Interneta so te organizacije imele že stoletne izkušnje pri gradnji omrežij velikih razsežnosti z lastno arhitekturo, prilagojenih storitvam prenosa govora. Arhitektura, znana kot »sistem za preklapljanje linij«, je bila predvsem primerna za gospodarsko in politično okolje, v katerem so delovale PTT-organizacije. Do obdobja intenzivnega razvoja podatkovnih komunikacij v poznih sedemdesetih letih prejšnjega stoletja so ta omrežja dejansko imela le en sloj: fizičnega, ker je vso preklapljanje povezav med napravami potekalo s preklapljanjem fizičnih linij. Telefonsko omrežje pa je imelo dve komponenti: eno za prenos govornih podatkov in drugo za nadzor in izvajanje preklapljanja. Dejanska potreba po večslojni arhitekturi ni obstajala. Takšen tip omrežja je oblikoval koncept sveta, ki je temeljil na povezavno zasnovanih protokolih, namreč s preklapljanjem linij se je vzpostavila direktna povezava med

uporabniki, ki je trajala do konca storitve in so bili protokoli, ki so povezavo vzpostavili, znani kot povezavni. Poleg monopola na trgu govornih komunikacij so imele PTT-organizacije še en monopol na področju nabave omrežne opreme, ki so jo ali same proizvajale, kot je bil primer tedanjega ameriškega giganta AT&T, ali pa je ta bila narejena na podlagi njihovih specifikacij. Standardi za komunikacijo so bili narejeni le zaradi potreb po povezovanju različnih škatel in so zato predvsem definirale vmesnike, ki so omogočali povezovanje z uporabnikom (odjemalcem storitve) ali s podobnim dobaviteljem te iste storitve. Ta situacija se je temeljito spremenila z deregulacijo, ki je naprej nastala v poznih osemdesetih v ZDA in z reorganizacijo velike ameriške korporacije AT&T ter veliko pozneje tudi v Evropi, ko so se državni PTT-monopoli preoblikovali v Telekome in je EU sprejela zakonodajo o liberalizaciji telekomunikacijskega trga storitev. Iz teh zgodovinskih dejstev izhaja izvir spora med omrežnimi modeli in koncepti. Namreč, glavni motiv nastanka obeh modelov je bil določiti, kdo bo lastnik česa in kdo bo obvladoval trg.

Zato ni presenetljivo, da se je model »nanizanih zrn na žici« razvil v omrežni model za prenos podatkov državnih telekomov, ko je bila digitalizacija informacijske tehnologije zlasti pri prenosu podatkov na pohodu. In tukaj je nastal prvi nesporazum. V večslojnem modelu računalniškega omrežja je vmesnik definiran tako, da poveže in deluje med dvema slojema iste naprave, v modelu »nanizanih zrn na žici« pa vmesnik deluje oziroma povezuje povezane škatle z ustrezno opremo. Pojav večslojnega modela za omrežja v ZDA sredi sedemdesetih letih je povzročil veliko zmede in odpora v klubu, ki je zagovarjal koncept povezavnih protokolov, saj so računalniški proizvajalci stopili na “tuje ozemlje”. Strokovna in znanstvena sfera je pričakovala, da obe tedaj zgrajeni raziskovalni računalniški mreži (ARPANet v ZDA in CYCLADES v Evropi sta bili operativni že v l. 1974) naj ne bi uspešno delovali. Vendar sta obe omrežji delovali zelo uspešno in s tem pokazali, da novi koncept deluje z najetimi vodi in brez poštnih organizacij. V Evropi so bile te organizacije del vladnih organizacij in so zato telekomi postavljali pravila delovanja na trgu. Med temi pravili so bila tudi tista, ki so dovoljevala priklop na omrežje le računalnikov v lasti PTT in nič drugega (primer naprave PAD (Packet Assembly and Dissasembly) v javnih omrežjih za prenos podatkov po standardu X.25). Vsi vpleteni v 30-letni vojni so se sicer zavedali, da bo prihajajoči trg podatkovnih komunikacij velik. Kako velik pa bo, nihče ni vedel. Še v poznih osemdesetih letih so evropske PTT-organizacije trdile, da obseg govornega prometa po telefonskem omrežju ne bo nikoli presežen z obsegom podatkovnega prometa. Kako so se motile! V tem času so mednarodni odbori za standardizacijo telekomunikacij razvijali t. i. telekomunikacijske storitve z dodano vrednostjo (teletext in vidoetext), vendar idejo, da lahko druge organizacije zgradijo svoja lastna omrežja, je bila zavržena v kali. V ZDA je bila situacija drugačna. AT&T je bila monopolna, vendar zasebna družba in je novo nastalo situacijo drugače ocenila. Pojav računalniških omrežij, zasnovanih na večslojnem modelu, je dojela kot priložnost za vstop na obetavni trg novih storitev in je v l. 1992 prevzela v upravljanje razvito infrastrukturo (najete in povezane vode z omrežnimi usmerjevalniki) raziskovalnega omrežja Internet, ki ga je do takrat financirala in vzdrževala vlada ZDA oziroma njena znanstvena agencija NSF. To letnico imamo tudi za začetek komercializacije Interneta.

ZDA in njene institucije so verjele v prihodnost Interneta, ki je bil desetletja pozneje razglašen tudi za največjo inovacijo 20. stoletja. Prevzeli so takoj evropski izum hipertekstovnega protokola in označevalskega jezika. Ta evropski izum, ki je nastal l. 1991 v CERN-u, je bil takoj prenesen v ZDA (brskalnik NCSA Mosaic) in je po enem letu začel komercialni pohod po svetu (Netscape). To je dokončno zagotovilo zmago omrežnemu konceptu, po katerem se je razvijal Internet. V tem času so imele evropske PTT-organizacije težave glede uveljavitve svojih storitev z dodano vrednostjo (ki niso bile dovolj prijazne do uporabnika), predvsem zaradi zaprtosti v svoji arhitekturi. Druga težava, za katero so iskale rešitev, je bila, kako še naprej uveljavljati model podatkovnih komunikacij s povezavo z zasnovanim protokolom, ki je imel zasnovo v sistemu s preklapljanjem linij.

Arhitektura, ki jo je zagovarjal klub PTT, je ob razvoju podatkovnih komunikacij in slojevite arhitekture omrežij vodila v slepo ulico. Vendar so PTT-organizacije in nova populacija računalniških inženirjev, ki so uporabljali internet, postali nekoliko bolj dojemljivi glede slojevite arhitekture in so ta princip začeli implementirati v svojih omrežjih. Tako je na primer protokol X.25, ki se je v Evropi in pri nas v osemdesetih letih (omrežje JUPAK) uporabljal kot temeljni protokol za javna omrežja za prenos podatkov, že vseboval tri ločene sloje. Med njimi tudi sloj, ki je skrbel za komunikacijo med vmesnikom podatkovne enote (DTE, ki je pripadala uporabniku) in komunikacijsko enoto omrežja (DCE, ki je pripadala lastniku omrežja PTT). Pri JUPAK-u in pozneje, v zgodnjih devetdesetih pri SIPAX-u, je bil problem v tem, da omrežje ni imelo pričakovanega prometa zaradi manjkajočih aplikacijskih storitev, ki naj bi potekale v gostiteljskih računalnikih uporabnikov, priključenih na JUPAK. Vendar leti teh standardnih aplikacij niso imeli, ker so vsebovali

lastniške operacijske sisteme (VAX od Digital-a, IBM-a ipd.) in njihove rešitve za prenos podatkov. Pristop vmesnika v protokol X.25 najdemo tudi v znanih tehnologijah, kot so ISDN, ATM, MPLS, WAP (ki je izjemno podoben Videotexovem standardu), ki so se pojavile veliko kasneje in ki se še danes intenzivno uporabljajo v omrežjih, na primer MPLS.

5 ZA ZAKLJUČEK – POMEN VOJNE ZA PRIHODNOST INTERNETA

Lahko se seveda vprašamo, zakaj je to danes pomembno in zakaj to temo odpiramo ob 20-letnici prihoda Interneta v Slovenijo? Odgovor je preprost: Zato ker vojna traja še danes v raznih ameriških in evropskih znanstveno-raziskovalnih iniciativah o prihodnjem Internetu in o tem, kako naj bo ta spremenjen. Vsi novi predlogi za nadaljnji razvoj svetovnega medmrežja, znani kot sestavni deli novega »prihodnjega Interneta« se lahko uvrstijo med koncepte enega ali drugega zagovornika znanih nasprotnikov iz tridesetletne vojne. Kar pa je zaskrbljujoče, je predvsem dejstvo, da kljub finančni in drugi podpori različnih lobijev rešitev ni na vidiku in je najverjetneje v kratkem ne bo, vsaj do trenutka, ko bo prišlo do streznitve in do določene sinergije, ki bo vodila do skupnega novega modela. Novi model bo nastal na podlagi sinteze najbolj primernih rešitev in enostavne implementacije, primerne za dobavitelje internetnih storitev in njihovih uporabnikov.

Reference

1. Day, J., Patterns in Network Architecture, A Return to fundamentals, Pearson Education Inc, December 2008
2. A Brief History of the Internet, Article by the pioneers for Internet Society,(URL: www.isoc.org/internet/history/brief.shtml)

PRIHOD PRVE INTERNETNE LINIJE V SLOVENIJO II DEL

Borka Jerman-Blažič
ISOC Slovenija
Slovenia
Tel: +386 1 4773 408; fax: +386 1 477 3395
e-mail: borka@e5.ijs.si

POVZETEK

Prispevek podaja potek dogodkov in akcij, ki so potekali pri prihodu prve internetne linije v Slovenijo

1 UVOD

V času vojne med koncepti je bila Jugoslavija in njena znanstvena sfera navezana na ameriško znanost na splošno. Znanstveniki so odhajali v ZDA zaradi izpopolnjevanja in pridobivanja izkušenj. ZDA je bila vodilna znanstvena sila na svetu. Tudi vsa računalniška oprema, ki je prihajala v Jugoslavijo, je bila ameriškega izvora. Vendar je bilo računalništvo v povoju in kakšnega resnega sodelovanja med ZDA in Jugoslavijo, zlasti na strateško pomembnih področij, ki jih je financirala ARPA, ni bilo. V začetku osemdesetih letih se je pojavil prvi osebni računalnik, in to je bil tudi začetek poznejših slovenskih in jugoslovanskih naporov za vstop na ta obetajoča področja tehnologije in trga (Iskra in pozneje Delta). Vendar se je domača politika v osemdesetih letih nekoliko spremenila. V Sloveniji ne le med intelektualci družboslovci, ampak predvsem v inženirskih in gospodarskih panogah so se začela spogledovanja z Evropo in zlasti z EFTA (Skupnost za prosto trgovino 6 evropskih držav). ISO je razvijal predstavitveni sloj (šesti) omrežnega modela ISO/OSI in je potreboval znanje jezikov in znakov, ki niso pripadali latinskemu pisanju. Zaradi odprtosti in uporabe cirilice je Jugoslavija postala idealna članica ISO za sodelovanje v Tehničnem odboru za nabor kodiranih znakov, ki je skrbel za 6. sloj ISO/OSI predvsem pri razvoju 8-bitnih kodnih tabel za vzhodno Evropo in pozneje za cirilico. Sodelovanje ekspertov jugoslovanskega odbora za kodirane znake se je nadaljevalo tudi pri zametkih nastanka večoktetnega zapisa za znake, znanega kot standard ISO 10 646 ali Unicode, ter pozneje pri internacionalizaciji računalniške tehnologije. Glede uvoza računalniške tehnologije je bila Jugoslavija v tem času v nekakšnem vmesnem položaju v primerjavi z Zahodom in z njenimi sosedami na vzhodu, zlasti glede dostopa strateške računalniške tehnologije, med katero se je uvrščala omrežna tehnologija. Sosednje države na vzhodu so bile glede nabave in uporabe naprednih tehnologij izpostavljene oceni ameriškega odbora COCOM, ki je določal, kaj lahko te države kupijo od dobaviteljev naprednih tehnologij v ZDA. Tako na primer Madžarski v devetdesetih letih ni bilo dovoljeno kupovati omrežne opreme s protokolom X.25. Za Jugoslavijo so bile omejitve precej milejše, in tako je Jugoslavija po sprejetju Zvezne resolucije o družbenem sistemu informiranja v l. 1981, ki je bila v času nastanka izjemno napredna strateška odločitev, potrebovala javno podatkovno omrežje za realizacijo njenih ciljev. Vodilne republike pri nastajanju javnega omrežja po protokolu X.25 sta bili Slovenija in Hrvaška. Omrežje je dobilo ime JUPAK in je postalo strateška infrastruktura države, ki pa zaradi pomanjkanja uporabniških rešitev iz modela ISO/OSI (elektronska pošta, prenos datotek ipd.) ni imela dovolj prometa. X.25 je omrežni protokol, ki nosi le podatke med povezanimi končnimi sistemi. Za generiranje prometa in funkcionalno uporabo potrebujejo ti končne sisteme v omrežju uporabniških aplikacij, ki generirajo promet. Teh pa ni bilo! V tem času je v Sloveniji in delno po Jugoslaviji delovalo le računalniško omrežje DECNET računalniškega dobavitelja Digital, ki je imel lastniško tehnologijo, ki ni omogočala povezavo z omrežji, ki niso imela opremo istega dobavitelja. Naslovni prostor je bil nestrukturiran, raven in je dovoljeval izmenjavo elektronske pošte le med uporabniki istega omrežja, kar pomeni istega lastnika in istega upravitelja. V primeru DECNET-a je bila to slovenska raziskovalna sfera, ki je lahko komunicirala le med seboj in s kolegi iz drugih republik Jugoslavije. Edina prednost tehnologije Digital je bila v tem, da je delovala v javnem prenosnem omrežju JUPAK, ker je imela programsko in strojno opremo z implementiranim protokolom X.25.

Nekaj elektronske pošte je konec osemdesetih let v Sloveniji pritekalo v okviru skupnosti EUnet, ki ni imela prenosnega omrežja, vendar je pošto dostavljala z občasnim prenosom datotek, ki je potekal med računalniki z operacijskim sistemom UNIX in protokolom UUCP (Unix to Unix Copy Protocol). Povezave med računalniki so se vzpostavljale le občasno s klici po telefonskem omrežju v dogovorjenem času. Del EUneta Jugoslavije je deloval na Fakulteti za elektrotehniko v Ljubljani (Leon Mlakar, Ivan Pepelnjak in Marjan Bradeško).

2 RAZISKOVALNI PROJEKTI

Evropa, ki se je odločila podpreti model ISO/OSI, je ugotovila, da na tem področju nekoliko zaostaja za ZDA in da zato potrebuje evropsko raziskovalno omrežje, ki bo delovalo s tehnologijo, razvito po konceptu OSI. Tako se je komisija Evropske skupnosti v l. 1986 dogovorila s članicami skupnosti EFTA, da sproži raziskovalni projekt v sklopu programa EUREKA z zaporedno številko 8. Projekt EUREKA 8 je imel naslov Cooperation of Open System Interconnection ali na kratko COSINE in je imel za cilj vzpostavitev akademskega in raziskovalnega omrežja Evrope z vsemi razpoložljivimi tehnologijami in rešitvami modela ISO/OSI. Treba je bilo pokazati, da Evropa ne zaostaja, da zmore več in da lahko s konceptom, za katerega se je opredelila, naredi enakovredno ameriškemu svoje raziskovalno omrežje. V času nastanka projekta COSINE je v Evropi obstajala dobro delujoča oprema za X.25, v okviru omrežij evropskih PTT uporabniških aplikacij pa je manjkala. Raziskovalno omrežje COSINE je bilo zamišljeno kot idealen okvir za razvoj uporabniških storitev in za potrditev učinkovitosti koncepta ISO/OSI. Sodelovanje slovenskih raziskovalcev v evropskih raziskovalno-znanstvenih akcijah COST je bilo zaradi spogledovanja z EFTA od slovenske politike podpirano in želeno. Povabilo, da Jugoslavija stopi v EUREKO 8 glede na cilje Evrope in družbena prizadevanja v Sloveniji, ni bila presenetljiva akcija Evropske skupnosti. Vabilo je prišlo na Institut "Jožef Stefan", ker so njegovi raziskovalci aktivno sodelovali v raziskovalnih akcijah evropskega programa COST. EUREKA 8 je bil projekt, ki so ga same financirale sodelujoče države Evropske skupnosti (ta je imela 12 članic) in članice za prosto trgovino EFTA (takrat je imela EFTA 6 članic). Jugoslavija je postala 19. članica projekta COSINE in je morala za svojo udeležbo (kotizacijo) sama priskrbeti sredstva. Na Institutu "Jožef Stefan" smo bili tega vabila veseli, ker smo že pred tem v skupini za računalniške mreže pod vodstvom dr. B. Jerman - Blažič delali pri projektu za slovensko PTT in imeli pripravljene osnove za nadaljnji razvoj JUPAK-a z uporabniškimi storitvami, ki naj bi omogočile, da promet v omrežju zaživi. Logično nadaljevanje teh prizadevanj je bilo delo pri implementaciji uporabniških aplikacij ISO/OSI v okviru projekta COSINE.

Zvezna vlada Jugoslavije je sredi osemdesetih let ustanovila nov razvojni sklad za spodbujanje tehnološkega razvoja, s katerim je upravljal minister za razvoj dr. Matić, profesor Fakultete za elektrotehniko iz sarajevske univerze. Pridobivanje sredstev iz tega sklada je potekalo z javnim razpisom in prijavitelji so morali pripraviti projekt v skladu z določenimi pogoji, ki so bili zelo podobni sedanjim zahtevam v razpisih EU iz okvirnih programov. Na kratko: sodelovanje entitet iz najmanj treh republik, sodelovanje industrije in akademskih ustanov, izdelava natančnega delovnega programa za 4 leta za vsako entiteto ter skupni delovni program. Tudi organizacija prijave je potekala podobno, kot to počne danes EU v svojih okvirnih programih. Organizirani so bili informacijski dnevi v prestolnicah tedanjih republik, na katera je prišel osebno tudi minister Matić.

Celotno vsebino prijave jugoslovanskega projekta COSINE je pripravila dr. B. Jerman - Blažič in njena sodelavka mag. Irena Fabič na podlagi predhodnih dogovorov o sodelovanju s Kemijskim inštitutom Boris Kidrič, z Računskim centrom Univerze v Mariboru (RCUM) in računskim centrom ljubljanske univerze ter s fakulteto za elektrotehniko iz Sarajeva, s podjetjem Energoinvest iz Sarajeva, Računskim centrom Univerze v Zagrebu – SRCE in fakulteto za elektrotehniko iz Beograda. Projekt je bil odlično ocenjen in so bila ustrezna sredstva za delovanje in razvoj projekta zagotovljena. Kotizacijo oziroma članarino za prvo leto sodelovanja v COSINE je plačala zvezna vlada.

V ekipi projekta COSINE na Institutu "Jožef Stefan" so sodelovali: Avgust Jauk, ki je bil mladi raziskovalec (MR) pod mentorstvom dr. B. Jerman - Blažič in je bil odgovoren, da kot sistemski administrator skrbi za računalnik VAX II, mag. Marko Bonač, ki je pod vodstvom dr. Kalina pripravljal doktorat kot MR in je bil odgovoren za sodelovanje v delovni skupini COSINE za X.400, in dr. B. Lavrenčič, ki je bil zveza med raziskovalci fizike in skupino za računalniške mreže v Odseku za digitalne komunikacije. Iz RCUM-a je sodeloval D. Šoštarič, iz Računskega centra Univerze v Ljubljani pa Rok Vidmar. SRCE je zastopal dr. Aurer, Energoninvest dr. G. Babić in elektrofakulteto iz Beograda mag. Zvonko Oršolić. V tem času je bila vojna med zagovorniki obeh omrežnih konceptov skoraj na višku. V istem času je načrtovalcem omrežja COSINE postalo jasno, da akademsko omrežje po ISO/OSI ne morejo graditi po modelu, ki ga je določal ISO/OSI, in so zato naredili prvo izjemo – omrežje, ki se je gradilo med 19 državami Evrope, je uporabljalo le protokol X.25. Protokol za medsebojno povezavo lastniško različnih X.25-omrežij, znan kot protokol X.75, je bil zavržen in tako je infrastruktura IXI delovala med državami Evrope po istem protokolu – protokolu X.25.

3 POTEK VZPOSTAVITVE MEDNARODNEGA VODA

Povezava z IXI za YUNAC- akademsko mrežo Jugoslavije je bila planirana, da poteka med Dunajem in Ljubljano. Moja ideja, da ima YUNAC le zasebno linijo do IXI ni bila podprta in tako so se zadeve za pridobivanje dovoljenj v Beogradu za najem voda za povezavo z JUPAK-om zapletle. . Ozadje tega stališča je bilo dvojno: držati v lastnih rokah ves nadzor in razvoj v zvezi z omrežji (ter podpirati uradna vladna stališča) in hkrati se izogniti stroškom za gradnjo lastne akademske komunikacijske infrastrukture (akademsko omrežje bi uporabljalo za prenos omrežje JUPAK-a in bi vsa zadeva bila nadzorovana). Tokratno omrežje DECNET, ki je povezovalo slovensko akademsko srenjo, je uporabljalo za povezavo med računalniki v omrežju storitve JUPAK-a. Rešitev je bila enostavna: namesto gradnje lastnega akademskega omrežja bi z uporabniškimi sistemi ISO/OSI in JUPAK-om lahko nadomestili obstoječe storitve omrežja

DECNET. Vendar je bil s tem razvoj akademskega omrežja in prihod Interneta zamujen najmanj za dve leti ali več. Vloga za dovoljenje za najetje mednarodne linije je bila poslana v Beograd že na začetku projekta COSINE v l. 1988. Zvezni komite za promet v Beogradu, ki naj bi izdal dovoljenje, je to pogojeval z izdajo soglasja Zajednice PTT Jugoslavije in Saveznog sekretariata za odbranu. JUPAK je bil strateška oziroma kritična državna infrastruktura in so za priklop v tuje omrežje zahtevali odobritev za to pooblaščenih organov. Po pričakovanju soglasja in dovoljenja v dveh letih ni bilo.

4 PRIPRAVLJALNA DELA ZA PRIHOD INTERNETA

Med čakanjem na dovoljenje iz Beograda so bila vsa pripravljalna dela za prihod Interneta. Že v l. 1989 sem osebno registrirala v tedanjem edinem centru za domene NIC (Internet Network Information Center) vrhnjo domeno za .YU, ki nam je omogočila tudi vzpostavitev mednarodnega vstopnega vozlišča za elektronsko pošto po standardu X.400 in tako tudi zagotovila posredovanje naše elektronske pošte poleg v omrežje X.400 COSINE tudi v omrežje Internet. V letu 1990 je domena .YU imela registrirani dve računalniški domeni, obe sta bili na Institutu "Jožef Stefan". Vsaka država članica COSINE je imela svoje nacionalno vstopno vozlišče za X.400, znano kot "Well Known Entry Point" ali WEP. To vozlišče je zbiralo vso pošto akademske srenje tiste države in jo pošiljalo naslovnikom v tujino, enako je delalo z naslovniki v lastni državi. Vstopno vozlišče za X.400 je bilo instalirano na računalniku VAX II, in za njega je skrbel kot sistemski administrator Avgust Jauk.

Vendar je bila najbolj pomembna lastnost tega vozlišča, tj. možnost, da se je naša pošta po preoblikovanju v drugo standardno obliko dostavljala tudi v omrežje Interneta in BITNET-a. Ta storitev je bila Edina mednarodna storitev. Druge, ki so takrat bile na voljo, na primer virtualni terminal ali telnet, prenos datotek ali FTP, storitev za iskanje dokumentov na omrežju ali Gopher in Veronika, so bile za nas nedosegljive. Na univerzi v Berkley-ju v Kaliforniji je bila vzpostavljena datoteka za imenski prostor Interneta (DNS) z registriranimi pod-domene pod vrhnjo domeno .YU. Naše vozlišče WEP (Well Known Entry Point) za elektronsko pošto, ki je uporabljalo za vstop v tujino prenosni protokol X.75 (zaradi neobstoja naše direktne linije X.25 do IXI smo uporabljali JUPAK-ov protokol X-75), je bilo prek infrastrukture IXI povezano z vozliščem omrežja EARN (evropski del omrežja BITNET), ki se je nahajalo v Parizu. To vozlišče omrežja EARN je imelo instaliran pretvornik protokolov, ki je pošto, pripravljeno s protokolom X.400, preoblikovalo v elektronsko pošto po internetnem standardu RFC 822 in je z internetnim protokolom za elektronsko pošto SMTP (Simple Mail Transport Protocol) poslalo našo pošto na Univerzo Columbia (CUNY) v New York in od tam v ameriški Internet. Nasprotno pot je opravila pošta iz Interneta v naš WEP in prišla do naših poštnih predalov z naslovi X.400. V tem času je to že bila nekakšna revolucija, Jugoslavija je lahko prek vozlišča na Institutu "Jožef Stefan" komunicirala z elektronsko pošto z razvitim svetom ZDA in Evrope v omrežju Internet.

Vzporedno z razvojem omrežne tehnologije so potekale v evropskih državah tudi institucionalne spremembe. Nastajale so institucije, znane kot raziskovalne in akademske mreže, ki so imele različno pravno obliko. Če so bile ustrezno organizirane, so postale tudi uradne partnerice projekta COSINE. Najbolj pogosto so to bili konzorciji univerz ali pa samostojne pravne osebe, ki so jih le-te ustanovile. Vendar so skoraj vse imele vladno podporo in so bile članice RARE – združenja akademskih mrež Evrope, ki se je potem preimenovalo v TERENA (Trans European Research and Academic Network Association). V ZDA je fundacija NFS (National Science Fundation) ustanovila in financirala organizacijo MERIT ter akademsko omrežje NFSNet (National Science Fondation NET), ki je skrbela za operativno delovanje raziskovalnega omrežja Internet in koordinirala raziskave, ki so potekale na ameriških univerzah in na Raziskovalnem institutu v Stanfordu (SRI. Tako je tudi v l.1989 stekel postopek ustanavljanja Jugoslovanske akademske mreže s partnerji iz projekta COSINE. YUNAC je bil organiziran kot delniška neprofitna družba, registrirana na Okrožnem sodišču v Ljubljani.

Med tem smo mi v Ljubljani še naprej čakali na dovoljenje iz Beograda za priklop JUPAK-a na IXI; od naše prve vloge je minilo več kot dve leti. Niso pomagale ne urgence Slovenskega komiteja za promet in zveze in ne predsednika Zveznega komiteja za promet in zveze, ki je bil naše gore list. Vojna konceptov in politike, ki se je odvijala na svetovnem prizorišču je imela "mlade" tudi pri nas. Nenadoma, zgodaj spomladi l. 1991 sem s pomočjo svojega dobrega sodelavca in prijatelja iz Zajednice PTT Jugoslavije, s katerim sem delala pri standardizaciji jugoslovanskih abeced v 7-mo, 8-mo in 16-bitnih kodnih tabelah za potrebe predstavitvenega sloja ISO/OSI v aplikacijah Videotex in Telektekst, dobila njihovo soglasje za priklop JUPAK-a na IXI. Potrebovali smo le še soglasje Saveznog sekretariata za narodnu odbranu. Aprila 1991 je na Institut "Jožef Stefan" prišlo sporočilo, da naj osebno pridem v Beograd na razgovor s tov. Ražnjatovićem v zvezi z našo zahtevo za dovoljenje (urgenca prek Zveznega instituta bezbednosti, s katerim je sodelovala Iskra Delta in so moje osebne povezave očitno še delovale in uredile dostop do osebe, ki je o tem odločala). In tako sem nekega sončnega jutra aprila 1991 odpotovala z jutranjim letalom Adrie v Beograd in se takoj napotila v Zvezni secretariat za narodnu odbranu, kjer so me pričakovali. Odpeljali so me do Ražnjatovića. Bila sem lepo sprejeta, vendar sem morala najprej z njim na prazen želodec zgodaj zjutraj piti slivovko in kavo in se šele nato z njim pogovoriti o naših potrebah. Čutila sem v zraku, da se ve, da Slovenija počasi odhaja. Vendar je bil rezultat pogovora pomemben. Dobila sem dovoljenje za priklop na IXI v imenu jugoslovanskega akademskega omrežja YUNAC. Po povratku v Ljubljano sem takoj stopila do slovenske PTT do g. Pavleta Mešeta, ki nam je veliko

nesebično pomagal pri vseh prizadevanjih za razvoj akademskega omrežja. Vozlišče X.25 je bilo na Institutu "Jožef Stefan" in ni bilo več nobenih ovir za testiranje povezave z dunajskim vozliščem IXI.
Napori in prizadevanja so končno dala rezultat. Povezava je začela delovati tri tedne pred osamosvojitvijo Slovenije, junija 1991. Linija je bila pod našim nadzorom in nihče nam je ni mogel vzeti. V prvih dneh osamosvojitve je nadzorni center omrežja COSINE IXI, ki je bil v Amsterdamu, prvi ugotovil, da se v Sloveniji nekaj dogaja in so me kontrolorji omrežja poklicali prve dne osmosvojitve po telefonu in spraševali, kaj se pri nas dogaja. Nadzorni zasloni njihovega centra za nadzor omrežja so kazali, da naša linija do Dunaja ni ves čas operativna. Razlog za to je bil v napadu letal JLA na oddajnike na Krvavcu, ki so komunikacijsko povezovali Jugoslavijo z zahodno Evropo.

5 PRIHOD PRVE INTERNETNE POVEZAVE

Vojna v Sloveniji se je sicer hitro končala, vendar Slovenija v l. 1991 in v prvi polovici l. 1992 še ni imela priznanja ZN, kar je imelo za posledico, da nismo imeli niti kodo/domeno za Internet (.SI). Za pridobitev te kode v ISO so zahtevali sprejetje v članstvo ZN. Ves čas vodenja projekta COSINE sem bila v tesnih stikih s svojimi kolegi iz CERN-a (Mednarodni center za jedrsko fizikov Ženevi), NIKHEF-a (Institut za matematiko in fiziko v Amsterdamu) in z drugimi znanstvenimi ustanovami, ki so vzporedno z razvojem COSINE na podlagi modela in koncepta OSI razvijali in uporabljali tudi raziskovalno omrežje Internet. To delovanje pri projektu COSINE je bilo neuradno, vendar široko sprejeto med nekaterimi raziskovalci COSINE (nizozemskimi, skandinavskim in angleškimi), kljub uradnim načrtom gradnje omrežja ISO/OSI. Glavna zamera mojih evropskih kolegov, zagovornikov koncepta Interneta, je bila, da ima YUNAC svoj sistem za imenski prostor za Internet (.YU) v ZDA in da je to za evropsko državo nedopustno. Vendar, kako naj DNS prenesem v domovino, če je bila uradna doktrina vladajočih uporaba le koncepta ISO/OSI? Ves čas moledovanja pri zveznih organih v Beogradu za dovoljenje za najem mednarodne povezave na IXI je naša elektronska pošta X.400 v tujino potovala s protokolom X.75, ki pa ne omogoča uporabe nepovezavnega protokola IP. Tako je bila vzpostavitev linije za X.25 v IXI za YUNAC dejansko odrešitev. Imeli smo direktno linijo, ob nabavi postaj SUN SPark 1+ z operacijskim sistemom UNIX pa sem zagotovila tudi programsko opremo za X.25, ki je lahko opravila enkapsulacijo protokola IP znotraj X.25 in se začela dogovarjati s kolegi iz NIKHEF-a in CERN-a, da mi dovolijo uporabo zasebnih vodov omrežja EASInet, ki bi omogočilo, da naš promet, ki je nosil internetne naslove, prišel do lokacije v Amsterdamu, in sicer do Instituta za fiziko in matematiko NIKHEF-a. Formular za dovoljenje za priklop je oktobra 1991 poslal moj kolega dr. Olivier Martin iz CERN-a, ki je potem z ustreznimi podatki za YUNAC bil poslan v Berlin-administrativni sedež EASIneta. V NIKHEF-u so imeli že moderni usmerjevalnik CISCO, mi na Institutu "Jožef Stefan" v Laboratoriju za odprte sisteme in mreže pa le UNIX-ov operacijski sistem na delovni postaji SUN. Če bi imeli tudi vozlišče CISCO, bi bila vsa zadeva veliko enostavnejša. Testiranje konfiguracij in nastavitev z minimalno opremo v Ljubljani je potekalo na daljavo, zato se je testiranje vleklo kar več dni brezuspešno in povezava ni in ni hotela steči. Vendar je naslednji teden povezava bila vspostavljena . Tako je naša dolgo pričakovana internetna povezava stekla. Bil je temen in mrzel dan v novembru, vendar je nam posijalo sonce. Hitro nato smo prenesli datoteko in zapise za DNS za .YU in jo dopolnili z dodanimi zapisi. In smo imeli v še nepriznani Sloveniji lastno internetno povezavo v okviru Laboratorija za odprte sisteme in mreže na Institutu "Jožef Stefan", ki smo jo postavili tiho in po gverilsko, brez vednosti kakih avtoritet in samooklicanih voditeljev, brez fotografiranja, slik, okroglih miz, objav in brez proslav. Elektronska pošta s protokolom SMTP, ki je bil vgrajen v OS UNIX, je nato takoj stekla, ker je bil koncept Interneta pragmatičen, enostaven in učinkovit. In to mu je prineslo zmago v 30-letni vojni konceptov. Nato so spomladi l. 1992 vsa slovenska lokalna omrežja, ki so imela internetne številčne naslove, ne glede na vrsto ustanove (akademske ali komercialne), prosila za priklop na to našo Internetno linijo.

6 ZAKLJUČEK

Naslednje leto se je dejansko končala Internetna zgodovina kot raziskovalnega projekta. Uprava akademskega omrežja ZDA NFSnet je predala prenosno raziskovalno omrežje Internet (njegovo infrastrukturo) v vzdrževanje korporaciji AT&T, ki je telekomunikacijske stroške začela zaračunavati univerzam, tako kot je zaračunavala telefonske komunikacije. Dve leti nato so NFSnet tudi ukinili. Univerze v ZDA pa so začele graditi svoja lokalna omrežja in razvijati uporabniške storitve Interneta v skladu z zmagovitim omrežnim konceptom, ki ga je zastavil že prof. Pouzin, da se storitve dogajajo na obrobju mreže, v uporabniških sistemih. Stanford Research Institute je istega leta (1992) objavil prvo knjigo o Internetu »Internet: Geting Started«. Timothy Berners-Lee je tudi v tem pomembnem letu, spomladi 1992, na konferenci Združenja akademskih mrež Evrope - RARE (pozneje TERENA) v Innsbrucku pokazal prvič raziskovalni skupnosti za omrežja svoj izdelek za izmenjavo hipertekstovnih datotek, ki so ga nato imenovali World Wide Web. S tem je začelo novo obdobje omrežja Internet. Internetni koncept računalniških komunikacij je zmagal in v l. 2011 dosegel 1,5 milijarde uporabnikov.
V Sloveniji je razvoj uporabe internetnih storitev in ustrezne infrastrukture potekal nekoliko drugače, počasneje. Prvo spletno stran so izdelali relativno hitro na gradbeni fakulteti Univerze v Ljubljani pod vodstvom dr. Žiga Turka in v računalniškem centru Instituta "Jožef Stefan" kot znano spletno mesto Mat Kurja v l. 1992/1993, vendar je širjenje internetnega omrežja potekalo počasi. Zlasti to velja za uporabo storitev v poslovne in druge produkcijske namene.

Reference

1. A Brief History of the Internet, Article by the pioneers for Internet Society,(URL: www.isoc.org/internet/history/brief.shtml)

2. Krtaka zgodovina Interneta, (URL:www.isoc-drustvo.si/?p=304

3. Prihod Interneta v Slovenijo, osebni zapiski, (URL:www.isoc-drustvo.si/prihod-interneta/

DAN D – ZGODOVINA NEPOSREDNEGA OPERATIVNEGA POVEZAVOVANJA SLOVENIJE V INTERNET

Denis Trček
Laboratorij za e-medije
Fakulteta za računalništvo in informatiko
Tržaška cesta 25, 1000 Ljubljana, Slovenija
Tel: +386 1 4768918
e-pošta: denis.trcek@fri.uni-lj.si

POVZETEK

Pričujoč prispevek podaja avtorjev pogled na začetke interneta v Sloveniji, kjer je na tehnično-operativnem nivoju na Institutu Jožef Stefan (IJS) vzpostavil prvo povezavo IP za takrat še celotno bivšo Jugoslavijo. Avtor je poleg same povezave prvih nekaj mesecev upravljal tudi s sistemom DNS (angl. Domain Name System) in vrhovno domeno za takrat še celotno Jugoslavijo. V sestavku je tako navedenih kar nekaj zanimivih dejstev iz neposredne zgodovine pred "dnem D", ko je stekla prva povezava IP, pa do zaključka pionirskih časov, ko so vse tovrstne storitve prešle v profesionalno opravljanje z ustanovitvijo javnega zavoda ARNES. Prispevek (poleg ostalih) podaja tudi dva nova dokumentirana vira, ki do sedaj nista bila (širše) poznana in osvetljuje vlogo nekaterih posameznikov, ki do sedaj prav tako ni bila dovolj poznana.

1 UVOD

Danes se sploh ne zavedamo več, kako močno je internet zasidran v vse pore našega življenja. V globini je zaznamoval naš vsakdan in množico vzorcev našega udejstvovanja, npr. načine, kako vzpostavljamo stike, se družimo, nakupujemo, se izobražujemo, itd. In tudi na širšem kolektivnem, državnem nivoju so spremembe velike. Si bi pred dobrimi desetimi leti sploh lahko predstavljali, da je varnost neke države lahko tako zelo odvisna od med sabo komunicirajočih računalnikov, da se pogovarjamo o realnosti kibernetskega vojskovanja?

V Sloveniji ima opisana situacija neposredne korenine predvsem na Institutu "Jožef Stefan", širše gledano pa je k vzpostavitvi in prodoru internet v Sloveniji prispevala velika množica posameznikov, ki so pri nas uvajali računalništvo in kasneje računalniške komunikacije. Vsem slednjim se opravičujem, ker jih glede na omejeno dolžino prispevka ne bom uspel navesti. Tako se bom osredotočil le na neposredne korenine vzpostavitve povezave IP.

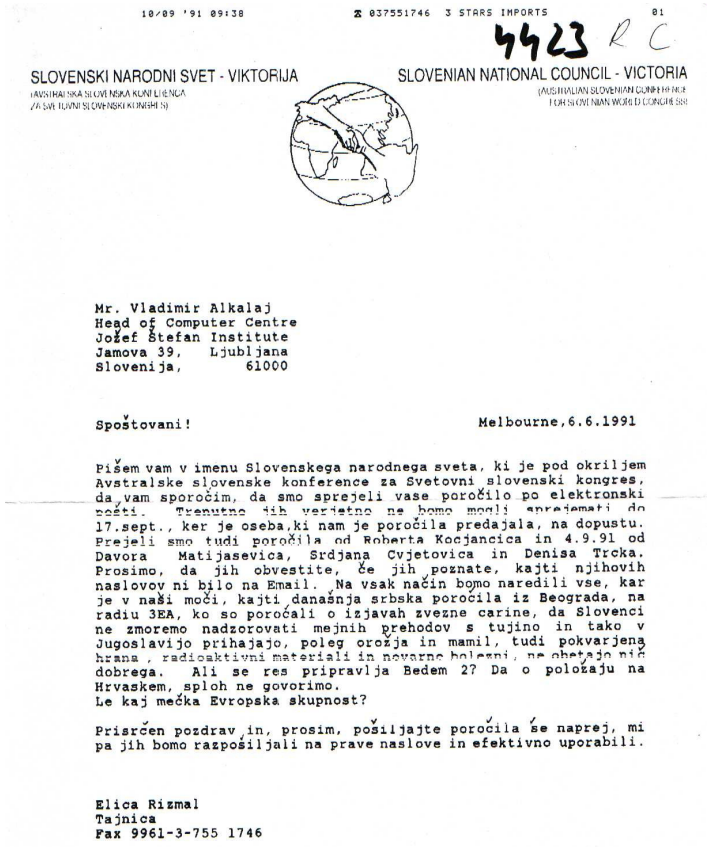
2 POMEMBNEJŠI NEPOSREDNI AKTERJI

Slovenski del zgodbe o internetu je v izdatni meri vezan na Institut „Jožef Stefan" (IJS). Zgodba ima svoje začetke v osemdesetih letih prejšnjega stoletja, ko je direktor instituta **dr. Kalin Tomaž** zaslutil potencial računalniških komunikacij in začel vzpostavljati stike z drugimi akterji na tem področju v evropskem znanstveno-raziskovalnem prostoru. Kmalu zatem, to je koncem osemdesetih let, so prišli prvi tovrstni mednarodni projekti. Prevzel jih je Odsek za digitalne komunikacije in mreže, ki ga je takrat vodil **dr. Korenini Janez**. S tem se je na omenjenem odseku IJS začela akumulacija znanja na področju računalniških komunikacij, kjer so delovali predvsem sledeči posamezniki: **mag. Bonač Marko, dr. Džonova Jerman Blažič Borka, mag. Jauk Avgust, dr. Kapus-Kolar Monika, dr. Rugelj Jože, dr. Pučko Marjeta, dr. Trček Denis, mag. Tvrdy Iztok in mag. Tvrdy Helena.** Od drugih sodelavcev IJS pa velja omeniti še fizika, pokojnega **dr. Boruta Lavrenčiča.**

3 NEKAJ TEHNOLOŠKE ZGODOVINE

Zgoraj omenjena skupina je spočetka ustvarjala akumulacijo izključno na področju tehnologije OSI (Open Systems Interconnection), ki je bila osnovana na naboru standardov za računalniške komunikacije, sprejetem s strani mednarodnih standardizacijskih organizacij CCITT in ISO. Čeprav sta to mednarodni organizaciji, pa standardi OSI niso dobili pravega odziva v ZDA. Postali so na nek način točka homogenizacije Evrope napram ZDA na področju računalniških komunikacij.

Nekaj let je bilo v svetu moč opažati skoraj ideološke delitve med pripadniki tehnologije OSI (prisotni so bili večinoma v Evropi), in tehnologijo TCP/IP (ti pa so bili prisotni večinoma v ZDA). Kaj se je zgodilo, vemo - ZDA so tudi tokrat zmagale. Tehnologija TCP/IP, ki je osnova interneta, je bila preprosto boljša. Povedati je tudi treba, da so široko-področna omrežja takrat v Evropi sicer že obstajala, vendar so bila relativno zaprta in omejena na specifične tehnologije. Največje tovrstno omrežje je bilo BITNET, za katerim je stal IBM, na nivoju lokalnih in mestnih omrežij pa je bil precej prisoten Decnet podjetja Digital.



10/09 '91 09:38 ☎ 037551746 3 STARS IMPORTS 01

4423 R C

SLOVENSKI NARODNI SVET - VIKTORIJA
(AVSTRALSKA SLOVENSKA KONFLIENCA ZA SVETOVNI SLOVENSKI KONGRES)

SLOVENIAN NATIONAL COUNCIL - VICTORIA
(AUSTRALIAN SLOVENIAN CONFERENCE FOR SLOVENIAN WORLD CONGRESS)

Mr. Vladimir Alkalaj
Head of Computer Centre
Jožef Štefan Institute
Jamova 39, Ljubljana
Slovenija, 61000

Spoštovani! Melbourne,6.6.1991

Pišem vam v imenu Slovenskega narodnega sveta, ki je pod okriljem Avstralske slovenske konference za Svetovni slovenski kongres, da vam sporočim, da smo sprejeli vase poročilo po elektronski pošti. Trenutno jih verjetno ne bomo mogli sprejemati do 17.sept., ker je oseba,ki nam je poročila predajala, na dopustu. Prejeli smo tudi poročila od Roberta Kocjancica in 4.9.91 od Davora Matijasevica, Srdjana Cvjetovica in Denisa Trcka. Prosimo, da jih obvestite, če jih poznate, kajti njihovih naslovov ni bilo na Email. Na vsak način bomo naredili vse, kar je v naši moči, kajti današnja srbska poročila iz Beograda, na radiu 3EA, ko so poročali o izjavah zvezne carine, da Slovenci ne zmoremo nadzorovati mejnih prehodov s tujino in tako v Jugoslavijo prihajajo, poleg orožja in mamil, tudi pokvarjena hrana , radioaktivni materiali in nevarne bolezni, ne obetajo nič dobrega. Ali se res pripravlja Bedem 2? Da o položaju na Hrvaskem, sploh ne govorimo.
Le kaj mečka Evropska skupnost?

Prisrčen pozdrav in, prosim, pošiljajte poročila še naprej, mi pa jih bomo razpošiljali na prave naslove in efektivno uporabili.

Elica Rizmal
Tajnica
Fax 9961-3-755 1746

Slika 1: Dokument o aktivnostih pripadnikov akademsko-raziskovalne sfere prek e-pošte ob osvobajanju Slovenije

Zgoraj omenjena skupina je tako sprva delovala skladno z evropsko usmeritvijo in proučevala tehnologije OSI. Kmalu ji je v sodelovanju s takratnim PTT-jem (Pošta, telegraf, telefon) uspelo vzpostaviti prve mednarodne računalniške povezave, kjer je za osnovo služilo jugoslovansko paketno omrežje Jupak. Ena prvih storitev, ki je delovala globalno, je bila elektronska pošta po standardu X.400. Programska oprema je tekla v okolju računalnikov VAX podjetja Digital (pri pridobitvi programske rešitve gredo zasluge večinoma dr. Borki Džonovi Jerman Blažič). Logično pa takratne rešitve niti približno niso bile podobne današnjim. Njihova operativna usposobitev je zahtevala kar nekaj sistemskih in programerskih znanj, kjer sta dolga leta imela ključno vlogo mag. Marko Bonač in mag. Avgust Jauk.

In verjetno bodo bralci presenečeni, da je omenjena programska rešitev odigrala nezanemarljivo vlogo med osamosvajanjem Slovenije. Raziskovalci na IJS so takrat intenzivno izkoristili ta medij (to je elektronsko pošto) za obveščanje svetovne javnosti v času agresije JLA, predvsem ZDA, kjer je bilo največ omrežnih priključkov. In ko se je začela vojna za Slovenijo, je zadeva dosegla kulminacijo. Učinek je bil več kot opazen. Radijski poročevalci iz ZDA so navajali (mislim, da konkretno g. Uroš Lipušček), da so bile Združene države na zahodu preplavljene z elektronsko pošto iz Slovenije. Vendar je tovrstna elektronska pošta preplavila dodobra celi svet. Odzivi so prihajali od vsepovsod – celo iz oddaljene avstralske Viktorije in tamkajšnjega Slovenskega narodnega sveta, kjer so naši rojaki skrbeli za nadaljnje informiranje avstralskih oblasti (glej sliko 1). Zato je tudi tisk Miloševićevega režima kmalu reagiral in tam se je navajalo predvsem ime mag. Marka Bonača [1]. Veliko pa so prispevali tudi naši takratni doktorski študentje na tujih univerzah, predvsem. **dr. Andrej Brodnik** in pa **dr. Rok Sosič** [1].

4 DAN D

Če skočimo še malce nazaj v zgodovino - leta 1990 je na IJS nastal Laboratorij za odprte sisteme. Laboratorij se je začel intenzivno povezovati s tistimi deli evropske akademsko-raziskovalne sfere, ki so prisegali na Internet in ključno tovrstno evropsko združenje je bilo RIPE (Réseaux IP Européens).

Večletni napori na tem področju so pripeljali do tega, da je bila s pomočjo povezav v RIPE 27. 11. 1991 vzpostavljena prva naša nekomercialna internetna povezava med Ljubljano in Amsterdamom (in od tam dalje s celotnim internetom). Približno sočasno s tem je bil vzpostavljen še sistem imenskih strežnikov (angl. Domain Name System, DNS) za celotno takratno Jugoslavijo. Povezava je temeljila na nižjih nivojih omrežja OSI, to je omrežju X.25, preko tega pa smo tunelirali promet IP. Organizacijsko upravljavski del je slonel na dr. Borki Džonovi Jerman-Blažič, vodji laboratorija, tehnično pa je sistem vzpostavil in upravljal dr. Denis Trček. Pri tem gre ponovno omeniti pomoč partnerjev v Amsterdamu, z vodstvenega vidika gredo zasluge predvsem dr. Rob Blokzijl-u, s tehničnega pa Maartenu Terspstri in Pietu Beertemi. Še to – takratna vlada v Beogradu, predvsem vojaški vrh, se je bala izgube nadzora, ki bi jih ta nov komunikacijski kanal, sloneč na takrat visoki tehnologiji, prinesel. In pri pridobivanju dovoljenj za vzpostavitev »substrata« za povezavo IP, to je mednarodne podmreža X.25 (ki je del standardov OSI), je imel pomembno vlogo dr. Tomaž Kalin. Dr. Kalin pa je imel pomembno vlogo tudi na drugi, evropski strani. Pripravil je namreč teren za pogajanja o priklopu na OSI hrbtenico v Evropi, ki se je razvijala v okviru evropskega mednarodnega projekta COSINE (Cooperation for Open Systems Interconnection Networking in Europe) [2].

Sčasoma se je tako nabrala obilica koordinacijsko-strateških povezav in tehničnega znanja, ki je bila predvsem v domeni sledečih posameznikov: dr. Tomaža Kalina, mag. Bonač Marka, dr. Borke Džonove Jerman Blažič, mag. Jauk Avgusta, dr. Tomaža Kalina in dr. Trček Denisa. Čas je bil zrel za vzpostavitev slovenske nacionalne akademske mreže Arnes. Ta je bila ustanovljena kot javni zavod z odlokom vlade z dne 15. 5. 1992. Z njegovo ustanovitvijo se je začel intenziven prodor interneta v šolsko, akademsko in raziskovalno okolje v Sloveniji, posredno pa tudi v komercialno okolje (akademsko raziskovalno okolje je vključevalo - in še vključuje - tudi raziskovalne enote v industriji). Omenimo na tem mestu tudi predhodnico Arnesa, organizacijo YUNAC. Ta je bila zastavljena že za časa nekdanje Jugoslavije in je imela prav tako zametke in sedež na IJS, vendar so bile aktivnosti YUNAC-a kratkega daha, saj je Jugoslavija takrat že začela razpadati.

4 VZPOREDNE ZASEBNE POBUDE

Internet se je hitro širil tudi v komercialni domeni. Tu najprej ne moremo mimo dejstva, da je Slovenija tudi v komercialni sferi po zaslugi podjetja Nil d.o.o. iz Ljubljane v približno istem času prišla do priklopa na internet, kjer velja poudariti predvsem delo **mag. Pepelnjaka Ivana**. Po podatkih, dostopnih avtorju, je ta povezava bila vzpostavljena prek EUneta, uporabljala pa je protokol UUCP (Unix-Unix CoPy).
Nedolgo zatem pa so se začeli v komercialni sferi pojavljati še drugi pomembni igralci, ki so ponujali dostop do interneta. Mednje sodi podjetje ABM d.o.o., ki ga je vodil g. B. Horvat, pa K2.net, ki ga je vodil g. D. Bulat in pa Quantum d.o.o., ki ga je vodil g. L. Zadravec. Večinoma so bila ta podjetja odvisna od najetih vodov, ki jih je nadzoroval Telekom. Izjema je bil v tem pogledu Quantum d.o.o., ki je imel še paralelno povezavo prek satelita z angleškim ponudnikom tovrstnih storitev, podjetjem PIPEX.

Komercialno gledano pa Telekom Slovenije, zanimivo, na ta trg razmeroma dolgo ni vstopil (čeprav ga je budno opazoval in se pripravljal nanj), tako da je preteklo kar nekaj časa, preden je začel konkretno slediti in je pognal podjetje SiOL.

5 ZATON SVETA OSI

Vzporedno z razvojem tehnologije IP smo na IJS še vedno vzpostavljali tehnologijo po standardih OSI. V ta namen smo pridobili programski paket ISODE (ISO Developement Environment), kjer je bila implementirana večina visokonivojskih storitev OSI: navidezni terminal (angl. Virtual Terminal, VT), manipulacija datotek (angl. File Transfer Access and Manipulation, FTAM), manipulacija procesov (angl. Job Transfer and Manipulation, JTAM), itd.

Softver ISODE je bil pisan za operacijski sistem SunOS, dobili pa smo ga na magnetnem traku, kolutu, v formatu TAR, za katerega bralne enote nismo imeli. Traku ni bilo moč prebrati niti na IJS.

Tako smo našli računalnik z bralno enoto za naš softver na Fakulteti za elektrotehniko Univerze v Ljubljani in sicer v laboratoriju dr. Janeza Trontlja. Tam so trak prebrali na računalnik z operacijskim sistemom UNIX, nakar smo ga prek DECNET-a prenesli na računalnik VAX z imenom Cathy na IJS, od tam pa prek lokalnega omrežja na končni računalnik Sun Sparcstation 1+, kjer je tekel SunOS operacijski sistem.

Potem smo lahko začeli z inštalacijo v upanju, da bo programski paket ISODE v nekaj urah preveden in nameščen. Žal temu ni bilo niti približno tako. Naše

programsko okolje (operacijski system) je imelo nekaj specifik (drugačna različica), tako da je prevajanje rezultiralo v več sto opozorilih in napakah... Paketa ni bilo moč prevesti. Kot posledica tega je bilo potrebno na roko iti skozi več MB izvorne kode in jo prilagajati, kar je na koncu zahtevalo kar nekaj mesecev, preden se je (kljub množici opozoril pri prevajanju) paket končno zadovoljivo prevedel in omogočil zagon storitev VT, FTAM [2].

Kmalu zatem smo pognali tudi direktorij po standard X.500 ter ga odprli svetu, ob ustanovitvi ARNES-a pa je vsa ta dejavnost prešla tja [3]. ARNES je direktorij vzdrževal še nekaj let, nakar ga je (tako kot IJS) približno sredi 90-ih let prenehal ponujati.

Kar se tiče elektronske pošte velja kot zanimivost povedati še to, da jo je (po standard X.400) vzpostavil in ponujal tudi Telekom Slovenije, vendar se ni nikoli prav prijela.

6 POPULARIZACIJA INTERNETA

Vzporedno z omenjenimi dejavnostmi je potekala tudi popularizacija in izobraževanje s področja tehnologije interneta. Običajni uporabniki so takrat imeli na svojih osebnih računalnikih operacijski sistem Microsoft Windows 3.X, ki ni podpiral protokolov TCP/IP, je pa nek računalniški navdušenec iz Avstralije napisal manjkajoč del softvera, imenovan Trumpet in ta je tem uporabnikom omogočal priklop v Internet [4, 5]. Dodatno tehnološko možnost za tovrstne uporabnike je omogočal tudi softver KA9Q [6].

7 ZAKLJUČEK

Da bi ta kratka zgodovina dobila še nekaj dodatne popestritve, navedimo enega prvih, mednarodno odmevnih in večkrat nagrajenih projektov v okviru pravkar spočetega slovenskega interneta. To je bil projekt Mat Kurja, ki je nastal v Računalniškem centru IJS. Center je vodil mag. **Alkalaj Mišo**, gonilna sila s tehničnega - in tudi širšega projektnega - vidika pa je bil **Martinec Mark** (podpora tega centra, predvsem Marka Martinca, je bila tudi sicer pomemben element v celotni zgodbi).

Poleg omenjenega centra velja na koncu omeniti še g. **Leona Mlakarja** s takratne Fakultete za elektrotehniko in računalništvo Univerze v Ljubljani. G. Mlakar je upravljal hrbtenico omrežja EUnet za takratno celotno Jugoslavijo [7] (omenjena knjiga založbe Prentice Hall iz leta 1993 poleg g. Mlakarja navaja kot kontaktne pionirske osebe na našem ozemlju še dr. Borko Jerman-Blažič, mag. Marka Bonača, mag. Avgusta Jauka in dr. Denisa Trčka).

Vse to pa se je odvijalo že proti sredini devetdesetih, ko se je internet z velikimi koraki začel prebijati v komercialno okolje. In nekaj let zatem so se v svetu in pri nas pojavili evangelisti, ki so nas zaradi fascinantnih učinkov Interneta začeli poučevati „novo“ ekonomijo. Ampak to je že druga zgodba…

Viri in literatura

[1] D. Trček, Lastna dokumentacija (zapiski, e-pošta, pogovori), 1990-2006.

[2] T. Oblak-Črnič, O začetkih Interneta na Slovenskem, Javnost – The Public, Št. 15, str. 151 - 174, Ljubljana, 2008.

[3] D. Trček, Se bomo pustili prehiteti vsem kar po desnem pasu? - varnost na informacijski avtocesti, Delo, letnik 37, št. 282, str. 11, 6. Dec. 1995.

[4] D. Trček, Ves svet je mreža / I, *Moj mikro*, letnik 8, str. 16-19, Delo, 1992.

[5] D. Trček, Ves svet je mreža / II, *Moj mikro*, letnik 8, str. 24-25, Delo, 1992.

[6] D. Trček, S KA9Q v Interent, *Moj mikro*, letnik 9, št. 3, str. 20-21, Delo, 1993.

[7] April M. et al., "Internet: Getting Started," Prentice-Hall, New York, 1993.

INTERNET 1985-1995

Rok Vidmar

Rok.Vidmar@GMail.Com

Povzetek: *V prispevku opišem dogodke, za katere menim, da so ključno prispevali k uvedbi Interneta v Slovenijo, kot sem jih doživel in se jih spomnim iz časa, ko sem delal na Računalniškem centru univerze v Ljubljani.*

V Sloveniji smo začeli graditi računalniška omrežja takoj, ko se je dalo. Ne zaradi potrebe, pač pa iz radovednosti in da se naučimo novega. Takrat seveda še slutili nismo, v kako veliko in pomembno stvar se bo to razvilo.

Prvi poskus, za katerega vem, se je odvijal na Ekonomski fakulteti leta 1980. Tam je imel svoje prostore Računalniški center univerze (RCU) z novo pridobitvijo, sistemom DECsystem 10, podjetje Delta pa je imela v kleti nedaleč proč postavljen računalnik PDP 11/30. Seveda smo poskusili med njima vzpostaviti povezavo z DECnetom, takrat še faze III, a je bilo nekaj narobe s programsko opremo na strani DECsystem 10. Poglobili smo se v iskanje napake - to je bil še čas, ko je bila programska oprema na voljo tudi v izvorni obliki - a iskanje smo morali zaradi višje sile (počila je cev in voda je zalila PDP) prekiniti.

Leto ali dve kasneje so z DECnetom faze IV bili v resnici povezani sistemi IJS, NBS, RCUM in še nekaj drugih podjetij. Omrežje, ki je tako nastalo, se je sčasoma širilo in kasneje smo ga poimenovali SLON (Slovenija ON line). Uporabljali smo ga predvsem sistemci in si pri tem brusili kremplje za pravo stvar: za skrb za nemoteno delovanje omrežja in servisov, ki jih je omogočalo, in za varnost na omrežju.

SLON je z osnovnim protokolom DECnet omogočal tudi prenos drugih protokolov, med drugimi X.25 in TCP/IP. To je bilo pomembno že leta 1986, ker je Jupak, ki ga je spočela slovenska PTT, s protokolom X.25 omogočal povezave po vsej Jugoslaviji, pa tudi nalaganje drugih protokolov na X.25 in zaradi tega se je SLON v tem času močno razmahnil.

Leta 1987 je postavljeno omrežje s protokolom TCP/IP na Fakulteti za elektrotehniko in računalništvo (FER) in RCU ga je preko DECneta priključil na Jupak.

To leto začne RCUM graditi knjižnični sistem COBISS. Knjižnice poveže z DECnetom in tako SLON nenadoma dobi pravo vlogo in delo. COBISS uporablja ta protokol skoraj desetletje, preden v celoti preide na TCP/IP.

Leta 1990 se RCU s programskim paketom Multinet, ki na operacijskem sistemu VMS implementira protokol TCP/IP, preko DECneta (in Jupaka) občasno povezuje v omrežje Inštituta “Rudjer Bošković” v Zagrebu, ki se občasno na zahtevo povezuje s protokolom TCP/IP v omrežje EUNET, ki je takrat že del Interneta.

Na pobudo RCU je bil tega leta programski paket MX, ki na operacijskem sistemu VMS implementira prenos elektronske pošte s protokolom SMTP, dopolnjen tako, da je omogočal SMTP tudi preko DECneta. S tem je postala elektronska poštna povezava s tem protokolom dostopna na vseh vozlih v omrežju SLON, ki pa seveda ni bilo povezano nikamor.

Leta 1989 na pobudo raziskovalcev z IJS ustanovimo YUNAC, jugoslovansko raziskovalno in akademsko omrežje, kar omogoči registracijo vrhnje domene “.yu” za Jugoslavijo. Na IJS steče elektronska pošta po protokolu X.400, a še vedno nismo imeli neposredne povezave z Internetom. Takrat Dr. Jerman-Blažič v COSINE izposluje, da se na vozlišče IXI (X.25) na Dunaju lahko priključi Jugoslavija a traja vse do začetka leta 1991, da Dr. Kalin na Skupnosti PTT Jugoslavije in Zveznem sekretarijatu za obrambo izposluje, da se povezava z IXI konča v Ljubljani na IJS - in ne v Beogradu ali Sarajevu.

Junija 1991 steče promet po povezavi IXI med Ljubljano in Dunajem, še vedno po protokolu X.25.

Med osamosvojitvijo je programska oprema MX na RCU omogočila usmerjanje elektronske pošte z mreže Univerze v Ljubljani tako v omrežje X.400 kot tudi v Bitnet preko vozlišča v Beogradu, kar je omogočilo nemoteno povezavo, ko je bila na Krvavcu zaradi zračnih napadov JLA prekinjena povezava z IXI.

V začetku leta 1992 IXI poleg protokola X.25 na svoji opremi omogoči tudi protokol TCP/IP. Uporabi tega protokola se je ostro in odločno uprla generalna

sekretarka YUNACa Dr. Jerman-Blažič, zato jo je skupščina YUNACa v aprilu 1992 zamenjala z Dr. Bonačem.

Maja leta 1992 je bil ustanovljen javni zavod Akademska in raziskovalna mreža Slovenije (ARNES), ki kot raziskovalna in izobraževalna infrastruktura za razvoj, organiziranje in vodenje akademske in raziskovalne računalniške mreže v RS med drugim registrira domeno ".si" in poskrbi za vklop v Internet s protokolom TCP/IP. Šele takrat postane računalniško omrežje zanimivo za vse uporabnike.

Kasneje leta 1992 je bila v Internet preko RCU in ARNESa stalno povezana tudi Univerza v Ljubljani, s protokolom TCP/IP, seveda. Kmalu za tem se pojavijo tudi prvi ponudniki internetnih storitev v samostojni Sloveniji, uporaba protokolov DECnet in X.25 počasi zamira.

Na prizorišče pridejo nove generacije, stare se različno hitro umikajo, Internet je v Sloveniji trdno zasidran in se razvija, raste in bogati podobno, kot po vsem svetu.

IZZIVI OB USTANAVLJANJU ARNESA

Marko Bonač
Arnes
Tehnološki park 18, 1001 Ljubljana, Slovenia
Tel: +386 1 4798877; fax: +386 1 4798878
e-mail: marko.bonac@arnes.si

POVZETEK

V prispevku je opisana ustanovitev in začetek delovanja javnega zavoda »Akademska in raziskovalna mreža Slovenije«. Pomembno vlogo je imela »Komisija za akademske mreže«, ki je po navodilu ministra za znanost in tehnologijo določila osnovne organizacijske zasnove bodočega javnega zavoda. V nadaljevanju je opisan postopek ustanovitve Arnesa in prve težave. Pomembni so bili dogovori z glavnimi akterji na področju računalniških mrež. Opisana so prizadevanja za čimprejšnjo vključitev v mednarodne organizacije in projekte ter opisane dileme pri zagotavljanju mednarodnih povezav.

1 UVOD

Leto 1991 je bilo izredno dinamično. Na področju računalniških mrež sta se poleg protokolov posameznih proizvajalcev računalniške opreme vedno bolj uveljavljala dva sistema standardov: ameriški (TCP/IP), ki so ga razvijale predvsem ameriške univerze in evropski, ki je sledil standardom mednarodnih organizacij CCITT in ISO in ki ga je tedaj želela uveljaviti Evropska komisija. Kasneje je prevladal preprostejši ameriški model, kar takrat še ni bilo razvidno. Zaradi množice protokolov je bilo že zagotavljanje delovanja elektronske pošte med različnimi sistemi izredno zahtevna naloga.

Pred osamosvojitvijo Slovenije se je mednarodno računalniško povezovanje odvijalo formalno v jugoslovanskem okviru in tudi stroški so bili delno pokriti iz zveznih sredstev. Ob samostojnosti se je bilo potrebno čim hitreje znova vključiti v mednarodne organizacije in projekte, sedaj formalno pod imenom Slovenija. Pri tem je poleg tehnoloških delitev (TCP/IP proti ISO/CCITT), obstajala še vrsta političnih delitev: tako avstrijska kot italijanska vlada sta želeli, da bi Avstrija oziroma Italija postali središči računalniškega povezovanja v srednji Evropi in sta bili pripravljeni v ta namen prispevati določena finančna sredstva. Potem so bili tu še projekti Evropske komisije, ki pa je takrat še postavljala kot pogoj za sofinanciranje uporabo ISO/CCITT tehnologije. Zaradi finančnih težav slovenske akademske sfere je bilo zelo pomembno izbirati partnerje tako, da bi bili stroški kar najnižji.

V Sloveniji je bilo največ znanja in opreme za računalniško povezovanje na IJS, IZUM-u in obeh univerzah. IZUM je že gradil povezave po Sloveniji za knjižnični projekt KIS/SZTI. Poleg tega je vsaka od teh institucij želela vzpostavljati svojo mednarodno povezavo. Po mnogih pogovorih se je izoblikovalo stališče, da je smiselno imeti po Sloveniji eno hrbtenico, ki bi prenašala promet za vse storitve v akademskem in knjižničarskem okolju ter centralizirano, čim močnejšo mednarodno povezavo. Ti dve nalogi je prevzel novoustanovljeni javni zavod Arnes.

2 STANJE PRED USTANOVITVIJO ARNESA

Jugoslavija se je vključila v svetovno računalniško omrežje jeseni leta 1989, ko je bil v Ljubljani vzpostavljen mednarodni IXI priključek s kapaciteto 48 kb/s in na Inštitutu Jožef Stefan (IJS) centralno vozlišče (WEP) za X.400 elektronsko pošto. To je bil rezultat sodelovanja IJS v projektu COSINE/EUREKA, ki je po zaslugi direktorja IJS Tomaža Kalina odprl vrata Jugoslaviji že leta 1987. Po ustanovitvi jugoslovanske akademske mreže YUNAC d.o.o. je oktobra 1990 predstavništvo v projektu prešlo nanj. Potrebne aktivnosti na nacionalnem nivoju (pri vzdrževanju X.400 WEP za elektronsko pošto in pri registraciji domen v naslovnem prostoru) so opravljali raziskovalci in mladi raziskovalci na IJS (Borka Jerman Blažič, Denis Trček, Avgust Jauk in Marko Bonač). Borka Jerman Blažič je bila najzaslužnejša za ustanovitev YUNAC-a in njegova generalna sekretarka. Plačilo za mednarodno linijo in potrebno koordinacijo na mednarodnem nivoju je Jugoslavija plačevala s kotizacijo v projektu COSINE in s članarino mednarodnemu združenju akademskih mrež RARE, za WEP se je uporabljal računalnik µVAX II (last odseka E6 na IJS), večino dela so opravili mladi raziskovalci, dodatna sredstva pa je za razvoj akademskih mrež prispeval zvezni »Matičev sklad«. Storitev elektronske pošte, ki so ga omogočala ta sredstva, se je uporabljal v največji meri v Sloveniji, v precej manjši meri na Hrvaškem in samo simbolično v Bosni in Hercegovini in Srbiji. Ob vojni v Sloveniji poleti 1991 je spoznanje o koristnosti obveščanja svetovne javnosti s komuniciranjem preko svetovnih računalniških mrež spodbudilo Ministrstvo za znanost in tehnologijo (MZT) k nakupu računalnika (VAX 4000/300) za potrebe akademske elektronske pošte. Od

poletja 1991 je služil ta računalnik kot slovenski X.400 WEP.
Leta 1991 so prenehala plačila iz Beograda tako v tujino (za COSINE in RARE) kot za delo in opremo doma, financiranje MZT pa se je začelo šele z ustanovitvijo Arnesa maja 1992. Vkljub temu je servis elektronske pošte tudi v tem vmesnem obdobju deloval nepretrgano. MZT je kasneje poravnal stroške, ki jih je imel IJS, dolgovi tujini pa so ostali neporavnani.

3 KOMISIJA ZA AKADEMSKE MREŽE PRI MZT

Minister za znanost in tehnologijo Peter Tancig je v začetku leta 1992 ustanovil Komisijo za akademske mreže. Člani so bili: Tomaž Kalin (IJS/RARE), predsednik, Pavel Meše (Telekom Slovenije), Branko Zebec (IZUM), Davor Šoštarič (Računski center Univerze v Mariboru), Tone Vidmar (Univerza v Ljubljani). Sodelovali so še Ciril Baškovič (MZT), Marko Bonač (IJS/Arnes), Avgust Jauk (IJS/Arnes), Iztok Tvrdy (MZT), Helena Tvrdy (MZT) in Rok Vidmar (Računski center Univerze v Ljubljani). Komisija se je v letu 1992 sestala petkrat in na osnovi zgledov drugje po Evropi postavila osnovne organizacijske koncepte slovenskega akademskega omrežja. Po njenih priporočilih se je formiralo področje dejavnosti Arnesa in konceptualno razmejila odgovornost med Arnesom, ki naj skrbi za slovensko hrbtenico in mednarodne povezave ter univerzami, IZUM-om in raziskovalnimi inštituti, ki naj na svojih področjih skrbijo za povezovanje fakultet, knjižnic ter gradijo informacijske sisteme. Komisija je tudi predlagala, naj Arnes čim prej pomaga razširiti internetne storitve v slovensko akademsko sfero.

4 USTANOVITEV ARNESA

Dne 8.5.1992 je Vlada RS z odlokom ustanovila javni zavod Akademska in raziskovalna mreža Slovenije in imenovala Marka Bonača za v.d. direktorja. Ministrstvo za šolstvo in šport je 29.7.1992 imenovalo svoja dva člana Upravnega odbora: Tomaža Pisanskega (Univerza v Ljubljani) in Bogomirja Horvata (Univerza v Mariboru). Ministrstvo za znanost in tehnologijo pa je dne 9.9.1992 imenovalo svoje tri člane Upravnega odbora: Ceneta Bavca (MZT), Vladimirja Alkalaja (IJS) in Dušanko Janežič (Kemijski institut). Upravni odbor se je sestal na prvi seji septembra 1992.
V Arnesu sta se leta 1992 zaposlila Marko Bonač in Benjamin Zwittnig, v začetku leta 1993 pa sta se jima pridružila še Avgust Jauk (kot tehnični direktor) in Mihael Dimec. Upravni odbor je leta 1994 izvolil, Vlada RS pa potem potrdila Marka Bonača za direktorja. Do leta 1995 je število zaposlenih naraslo na devet.
Julija 1993 se je formiral Strokovni svet Arnesa. Na predlog direktorja je Upravni odbor potrdil naslednjo sestavo: Tomaž Kalin, predsednik (RARE), Borka Jerman Blažič (IJS), Emil Hudomalj (Univerza v Ljubljani), Tone Vidmar (Univerza v Ljubljani), Davor Šoštarič (Univerza v Mariboru), Branko Zebec (IZUM), Matej Penca (MZT), Avgust Jauk (Arnes) in Pavel Meše (Telekom Slovenije).
Glede na želje uporabnikov je Arnes takoj ob ustanovitvi predložil MZT predlog za nakup potrebne opreme za internetne storitve. Zaradi proračunskih težav je Arnes prva investicijska sredstva dobil šele konec januarja 1993. Da je lahko servis nemoteno tekel in da je Arnes lahko takoj ponudil tudi nove internetne storitve (telnet, ftp), si je moral ustrezno opremo izposoditi, kar je oteževalo pogoje dela.
V prvih letih je imel Arnes stalno probleme s pomanjkanjem finančnih sredstev. Število uporabnikov Arnesa je namreč skokovito naraščalo, dodajale so se nove storitve, cena zakupljenih vodov v Sloveniji in v tujino pa je bila zaradi tedaj monopolnega položaja Telekoma na tem področju izredno visoka. Postavka Arnes je bila v proračunu MZT sicer zelo majhna, je pa vsako leto naraščala, kar je predstavljalo vsakoletni problem pri potrditvi letnega finančnega načrta.

5 DOGOVORI MED GLAVNIMI AKTERJI

Decembra 1992 je Arnes sklenil dogovor z Univerzo v Ljubljani, januarja 1993 pa z Univerzo v Mariboru o sodelovanju pri povezavi njihovih lokalnih računalniških mrež v svetovno računalniško omrežje. Ti dogovori so predvideli, da morajo imeti računalniki na univerzah dostop do drugih raziskovalnih institucij v Sloveniji preko DECnet, IP in X.25 protokolov in do institucij v tujini preko IP in X.25 protokolov. Povezava mora biti vzpostavljena do vseh pomembnih svetovnih akademskih in raziskovalnih omrežij (npr. do BITNET/EARN). Za elektronsko pošto se uporablja MAIL-11, X.400 in SMTP/RFC822 tehnologija. Vsak uporabnik bo lahko uporabljal FTP (za prenos datotek) in TELNET (za oddaljeni dostop) aplikacije. Raziskovalcem mora biti omogočen dostop do superračunalnika na Dunajski univerzi (ki ga je ta tedaj nekaj časa nudila slovenskim raziskovalcem brezplačno), dostop do NEWS storitve in dostop do COBISS-a. Specificirane so bile naloge partnerjev, da bi se to čim prej realiziralo.
Januarja 1993 so Arnes, IZUM, Univerza v Ljubljani in Univerza v Mariboru sprejele »Program razvoja, organizacije in vzdrževanja komunikacijske infrastrukture akademske in raziskovalne sfere Slovenije v letu 1993«. Dogovorjena je bila topologija omrežja in tip opreme. Sprejeto je bilo dvo-nivojsko načelo, to pomeni, da je naloga Arnesa izgradnja in vzdrževanje multiprotokolarne hrbtence z enajstimi vozlišči v Sloveniji, naloga univerz, da povežejo fakultete in naloga IZUM-a, da zagotovi povezave večjih knjižnic na vozlišča hrbtenice in pomaga pri vzdrževanju opreme. Dogovorjeno je bilo tudi, da se medkrajevne povezave, ki jih je doslej zgradil IZUM za potrebe projekta KIS/SZTI, uporabijo kot del hrbteničnega omrežja za vse storitve akademskega in knjižničnega sektorja. Za mednarodne povezave je bilo sklenjeno, da Arnes zagotavlja glavno mednarodno povezavo, IZUM pa še dodatno satelitsko povezavo do Rusije.

Z IJS je Arnes sklenil pogodbo o najemu prostorov ter pogodbo o povrnitvi stroškov, ki jih je imel IJS s posojanjem opreme in drugimi storitvami v obdobju, ko Arnes še ni bil ustanovljen, storitve pa so se izvajale. Po drugi strani pa je bila preknjižena tista oprema iz IJS na Arnes, ki jo je za ta namen kupil MZT še pred ustanovitvijo Arnesa in se je tedaj začasno knjižila na IJS.
Ti sporazumi so nastali na podlagi vrste sestankov, kjer so sodelovali predvsem Tomaž Seljak (IZUM), Branko Zebec (IZUM), Tone Vidmar (Univerza v Ljubljani), Rok Vidmar (Univerza v Ljubljani), Davor Šoštarič (Univerza v Mariboru), Marko Bonač (Arnes) in Avgust Jauk (IJS/Arnes).

6 VKLJUČEVANJE ARNESA V MEDNARODNE ORGANIZACIJE IN PROJEKTE

Pogodbe in dogovori s tujino za delovanje mednarodnih povezav in servisov so se v začetku leta 1992 še vedno nanašali na YUNAC d.o.o. in Jugoslavijo, prav tako je bil YUNAC d.o.o. tudi član ustreznih mednarodnih asociacij. Ob ustanovitvi Arnesa je MZT zahteval, da se vsaka aktivnost v Sloveniji na tem področju opravlja le skozi Arnes.
Arnes je takoj začel s pripravami za sklenitev pogodb in dogovorov s tujino in za včlanitev v ustrezne mednarodne asociacije. YUNAC/Jugoslavija je imel privilegiran status (glede na druge centralno in vzhodno evropske države) v evropskih projektih/asociacijah na področju akademskih in raziskovalnih mrež. Arnes/Slovenija žal ni mogel uveljavljati pravic kot naslednik YUNAC-a/Jugoslavije, ampak si je moral svoj položaj priboriti samostojno. Nebolečemu prehodu iz YUNAC-a/Jugoslavija v Arnes/Slovenija so pripomogla naslednja dejstva:

- Upravni odbor YUNAC-a in Skupščina YUNAC-a sta marca 1992 sklenili, da se lahko nacionalni akademski in raziskovalni mreži Arnes in CARnet (hrvaška akademska in raziskovalna mreža) včlanita samostojno v mednarodne asociacije in organizacije, da naj bo bodoča vloga Arnesa in CARnet-a operativna, vloga YUNAC-a pa le še usklajevalna na področju bivše Jugoslavije. Nato je upravni odbor YUNAC-a junija 1992 izbral za v.d. generalnega sekretarja M. Bonača, ki je bil mesec dni pred tem imenovan za v.d. direktorja Arnesa.
- Embargo za raziskovalno in akademsko povezovanje s Srbijo in Črno goro in pretrganje PTT povezav z Bosno in Hercegovino so skrčili operativno delovanje YUNAC-a na Slovenijo in Hrvaško.
- T. Kalin je bil tedaj generalni sekretar v RARE.

Vsa ta dejstva so pripomogla, da se je Arnes lahko v kratkem času včlanil v mednarodne organizacije kot predstavnik Slovenije in sklenil z organizacijami v tujini ustrezne pogodbe in dogovore, obenem pa so storitve ves čas nemoteno tekle.
Tako je še v letu 1992 Arnes postal član združenja akademskih in raziskovalnih mrež Evrope RARE, sodeloval je pri ustanovitvi neprofitne organizacije za operativno delovanje evropskega akademskega omrežja »Operational Unit« (kasneje preimenovano v DANTE), sodeloval je v organizaciji RIPE NCC za pridobivanje IP naslovnega prostora in bil je aktiven v COSINE projektu pri razvoju X.400 elektronske pošte.
Ker sta tedaj tako Avstrija kot Italija kazali željo po koordinaciji aktivnosti na področju akademskih mrež v vzhodni Evropi in sta imeli tudi sredstva za pomoč, je Arnes sodeloval tako pri avstrijski iniciativi »Cooperation of Networking in Central and Eastern Europe« kot pri italijanski iniciativi »Hexagonal Initiative«.

7 UPRAVLJANJE SI NASLOVNEGA PROSTORA

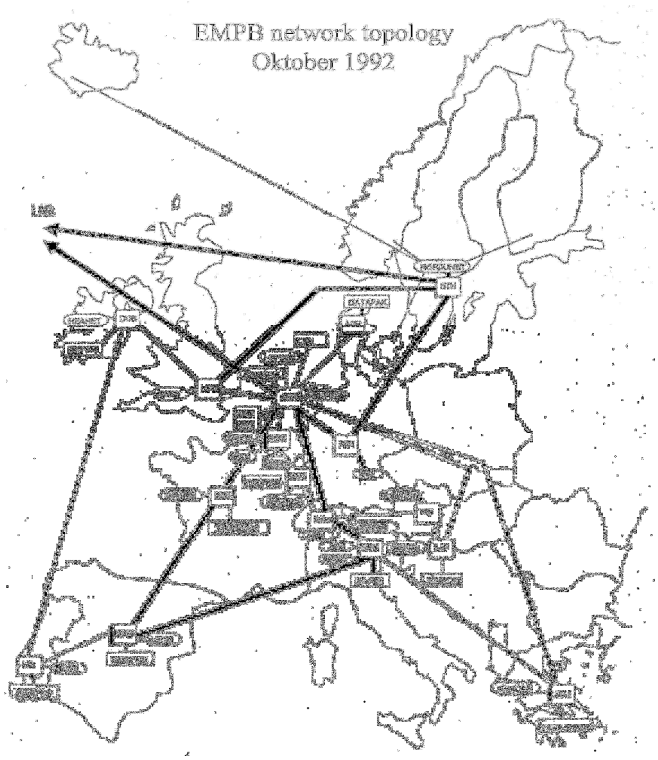
ISO 3166 Maintenance Agency je spomladi leta 1992 odobrila Republiki Sloveniji dvočrkovno oznako SI. Željena koda SL je bila namreč že rezervirana za državo Sierra Leone. MZT je v dogovoru z Ministrstvom za promet in zveze, Sestavljenim PTT podjetjem in Uradom za standardizacijo in meroslovje zapisalo v ustanovitveni akt Arnesa, da upravlja slovenski naslovni prostor. Dejanska registracija domen se je lahko začela, ko je IANA pooblastila Arnes za vodenje tega dela naslovnega prostora. Arnes je pridobil prve bloke naslovov v RIPE NCC, začel z registracijo domen in postavil strežnike DNS.
Obenem je bilo potrebno spremeniti naslove v mrežah po ISO OSI standardu iz country=yu na oznako country=si. Ker sta nekaj čas vzporedno delovali tako elektronska pošta po standardu X.400 kot elektronska pošta po standardu RFC822, je bilo potrebno skrbeti za preslikavo med x.400 in RFC822 naslovnima prostoroma. Tabele s preslikavami je potem Arnes redno pošiljal vsem gateway-em na svetu, ki so tedaj skrbeli za povezavo med x.400 in RFC822 sistemi.

8 MEDNARODNE POVEZAVE

Mednarodna povezanost Slovenije je bila v letu 1992 do 31.9.1992 zagotovljena preko IXI priključka kapacitete 48 kb/s. Pri iskanju možnosti za mednarodno povezavo od 1.10.1992 dalje je imel Arnes na razpolago dve možnosti: najeta linija do Dunaja in tam dostop do EBONE omrežja ali najem EMPB pristopne točke v Ljubljani.
Avstrijska vlada je preko avstrijske nacionalne akademske mreže ACOnet ponudila akademski in raziskovalni sferi v Sloveniji pomoč pri vzpostavitvi povezave v EBONE. Zakupila bi se 64 kb/s linija Ljubljana-Dunaj. Avstrija je bila pripravljena plačati stroške za zakup linije na avstrijski strani in plačilo prometa v EBONE. Strošek Arnesa bi bil le slovenski del zakupljene povezave do Dunaja.
Druga možnost je bila vključitev v EMPB. Evropska komisija je izvedla javni razpis za izgradnjo multiprotokolarne hrbtenice med akademskimi mrežami članic Evropske unije. Na natečaju je dobil PTT Telecom Nizozemska, ki je ponudil storitev EMPB (European Multiprotocol Backbone). Ta storitev je omogočala preko priključnih točk kapacitet od 9.6 kb/s do 2 Mb/s s protokoli X.25, IP in CLNP dostop do vseh drugih EMPB priključnih

točk v Evropi. Vsaka priključna točka je bila zaradi potreb po brezhibnem delovanju povezana z dvema linijama do drugih priključnih točk. Slabost te možnosti je bila v tem, da ni vključevala povezave v EBONE in povezave do ZDA. Zanimivo je bilo, da je bila druga možnost stroškovno dosti ugodnejša (približno petkrat). Razlog je bil v tem, da so bile cene zakupljenih vodov (posebej mednarodnih) zaradi monopola Telekoma Slovenije zelo visoke. Delitev stroškov znotraj EMPB pa je upoštevala kot kriterij le zmogljivost priključka in ne stroškov zanj. To pomeni, da so visoko ceno slovenskega dela povezav do Dunaja in Amsterdama pomagali pokriti vsi partnerji.

Dokončna odločitev za drugo možnost je bila sprejeta, ko je Arnes uspel doseči dogovor z akademsko mrežo Velike Britanije (JANET), da mu bo JANET brezplačno nudil IP dostop do EBONE omrežja v Evropi in IP dostop preko njihove transatlantske linije do ZDA. V ta namen je Arnes uporabil prenos IP preko X.25. Po krajšem obdobju testiranja je Arnes konec leta 1992 vzpostavil stabilni X.25 in IP povezavi s tujino. Arnes je bila takrat edina akademska mreža izven zahodne Evrope, ki je bila vključena v EMPB (Slika 1).



Slika 1: *Topologija omrežja EMPB*

Kapaciteta priključka v EMPB je bila na začetku 64 kb/s. EMPB se je naslednje leto preimenoval v EuropaNET in upravljanje omrežja je prevzela, s strani evropskih akademskih mrež ustanovljena, neprofitna organizacija DANTE. V letu 1994 se je zmogljivost priključka dvignila na 128 kb/s, v letu 1995 na 1 Mb/s. Hitro povečanje zmogljivosti je omogočila tudi finančna pomoč, ki jo je Arnes dobil skozi evropski projekt PHARE.

9 UPORABNIKI ARNESOVIH STORITEV IN HRBTENICA PO SLOVENIJI

Glavni uporabniki v letu 1992 so bili Univerza v Ljubljani, Univerza v Mariboru, Institut Jožef Stefan, Kemijski institut, IZUM in nekaj večjih knjižnic. Ocenjevali smo, da je bilo rednih, končnih uporabnikov približno 2000.

V prihodnjih treh letih je število uporabnikov izredno naraslo. V decembru 1995 je imel Arnes vozlišča v Ljubljani, Kranju, Kamniku, Mariboru, Celju, Ravnah na Koroškem, Velenju, Novem mestu, Kopru, Postojni, Novi Gorici in Portorožu. Preko zakupljenega voda je bilo povezano 140 raziskovalnih in izobraževalnih organizacij ter knjižnic. Poleg tega je imel Arnes še 144 modemov v vseh omrežnih skupinah (zaradi tedaj veljavnega tarifnega sistema), ki so omogočali 2500 individualnim uporabnikom komutirani dostop do omrežja.

Razlogov za hitro rast je bilo več. MZT je s prihodom Tomaža Seljaka na mesto vodje oddelka za informatizacijo leta 1994 pospešil več slovenskih IKT projektov. Arnes je začel sodelovati z Zavodom za šolstvo, predvsem Janezom Čačem, koordinatorjem dela na področju računalniških omrežij v programu računalniškega opismenjevanja, kar je okrepilo delo Arnesa v šolstvu. Fundacija SOROS pa je tedaj nekaj let finančno pomagala šolam pri nakupu opreme za povezavo v omrežje Arnes.

10 ZAKLJUČEK

Leta 1991 in 1992 so zaznamovale burne spremembe na področju računalniških omrežij – tako na področju tehnologije kot na področju organiziranja na evropskem nivoju. Poleg tega smo dobili svojo državo. Kljub temu smo takrat uspeli zagotavljati akademski sferi enake storitve kot drugje v zahodni Evropi in to brez prekinitev. Prav tako nam je uspelo postaviti temelje za izobraževalno in raziskovalno omrežje, ki so omogočili hitri nadaljnji razvoj. Uspešni smo bili predvsem zaradi sodelovanja vseh strokovnjakov na tem področju.

Viri

[1] Zapisniki komisije za akademske mreže pri MZT.
[2] Odlok o ustanovitvi javnega zavoda Arnes.
[3] Zapisniki sej Upravnega odbora Arnes.
[4] Poročila javnega zavoda Arnes.
[5] Dogovor o sodelovanju med Univerzo v Ljubljani in Arnesom pri povezavi lokalnih računalniških mrež na Univerzi v nacionalno in svetovno računalniško omrežje, 1992.
[6] Dogovor o sodelovanju med Univerzo v Mariboru in Arnesom pri povezavi lokalnih računalniških mrež na Univerzi v nacionalno in svetovno računalniško omrežje, 1993.
[7] Arnes, IZUM, Univerza v Ljubljani, Univerza v Mariboru: Program razvoja, organizacije in vzdrževanja komunikacijske infrastrukture akademske in raziskovalne sfere Slovenije, 1993.

Medmrežje v Sloveniji – od začetkov do eksplozije interneta

Avgust Jauk
ARNES
Arnes p.p. 7, 1001 Ljubljana, Slovenia
Tel: +386 1 4798877; fax: +386 1 4798878
e-mail: jauk@arnes.si

POVZETEK

V prispevku je podan pregled razvoja računalniških omrežij in storitev v Sloveniji v osemdesetih in začetku devetdesetih let prejšnjega stoletja, ki so pripravila teren za uvedbo interneta. Sledi opis uvedbe in predvsem širitve interneta po Sloveniji, s poudarkom na tehničnem in operativnem nivoju. Pojasnjeni so določeni predpogoji, ki so omogočili njegovo relativno hitro širitev, ter ovire, ki jih je bilo treba pri tem premagovati. Prispevek je pisan na podlagi osebnih izkušenj avtorja, ki je v teh procesih intenzivno sodeloval od samega začetka, najprej kot mladi raziskovalec na Institutu Jožef Stefan, nato pa kot tehnični direktor Arnesa.

1. UVOD

Ko je konec leta 1991 Slovenija, oz. takrat še Jugoslavija, dočakala prvo povezavo v internet, to ni bilo nekaj revolucionarnega. Računalniška omrežja so namreč bila pri nas prisotna že precej časa, tudi mednarodno povezavo smo imeli že od leta 1989, le na drugih standardih je bila zasnovana. Uporabniki, takrat predvsem v akademskem okolju, smo že imeli na voljo vrsto storitev, kot so elektronska pošta, prenos datotek, diskusijske oz. novičarske skupine, oddaljen dostop do strežnikov in pošiljanje kratkih sporočil.

Konec osemdesetih ter v začetku devetdesetih let prejšnjega stoletja internet (medmrežje, zasnovano na protokolih TCP/IP) ni ponujal bistveno večjega oz. bolj zanimivega nabora storitev kot konkurenčna omrežja. Njegova glavna prednost je bila v tem, da ni bil omejen zgolj na eno državo, niti na opremo določenega proizvajalca. Počasi se je širil iz ZDA na ostale kontinente in tako postal globalno omrežje. Z rastjo omrežja se je hitro začel večati tudi nabor vsebin, ki so bile na njem dostopne, še posebej z uvedbo spletne tehnologije (web). Ko se je kasneje razširil izven raziskovalno-izobraževalnega okolja, pa je njegova prevlada nad ostalimi omrežji postala popolna.

V primerjavi z internetom in njegovimi storitvami so sicer bila nekatera omrežja, ki so že bila prisotna pri nas, omejena na geografsko območje Slovenije oz. Jugoslavije (npr. DECnet) oz. na Evropo (akademsko omrežje). Toda ta omejenost ni nujno veljala tudi za vse storitve teh omrežij. Preko posebnih prehodov do drugih omrežij smo namreč imeli na voljo globalno elektronsko pošto in z njeno pomočjo tudi prenos datotek ter dostop do nekaterih globalnih diskusijskih skupin, zasnovanih na elektronski pošti.

V nadaljevanju je najprej podrobneje opisano stanje in razvoj omrežij v Sloveniji ter delno v Jugoslaviji pred uvedbo interneta, nato pa sledi podrobnejši opis širitve interneta po Sloveniji, **s poudarkom na tehniki in strokovnjakih, ki so zanjo skrbeli**. Zaradi časovne oddaljenosti in obširnosti tematike so določeni opisi nepopolni, prav tako pa so verjetno izpuščeni posamezni strokovnjaki, ki so aktivno sodelovali. Vsem prizadetim se opravičujem.

2. DOGAJANJE PRED UVEDBO INTERNETA

Mreženje se v Sloveniji ni začelo z internetom. Pred tem smo namreč že imeli za tiste čase relativno dobro razvito omrežno sceno. Na voljo je bilo nekaj različnih omrežnih tehnologij oz. protokolov ter množica storitev. Okolje je bilo mnogo bolj heterogeno kot danes, ko je na voljo zgolj internet. To je po eni strani dalo uporabnikom na voljo več možnosti, po drugi strani pa je bila uporaba storitev v tako heterogenem okolju bolj kompliciranа, ker so uporabniki morali biti poučeni o topologiji omrežij ter njihovih medsebojnih povezavah oz. prehodih med njimi.

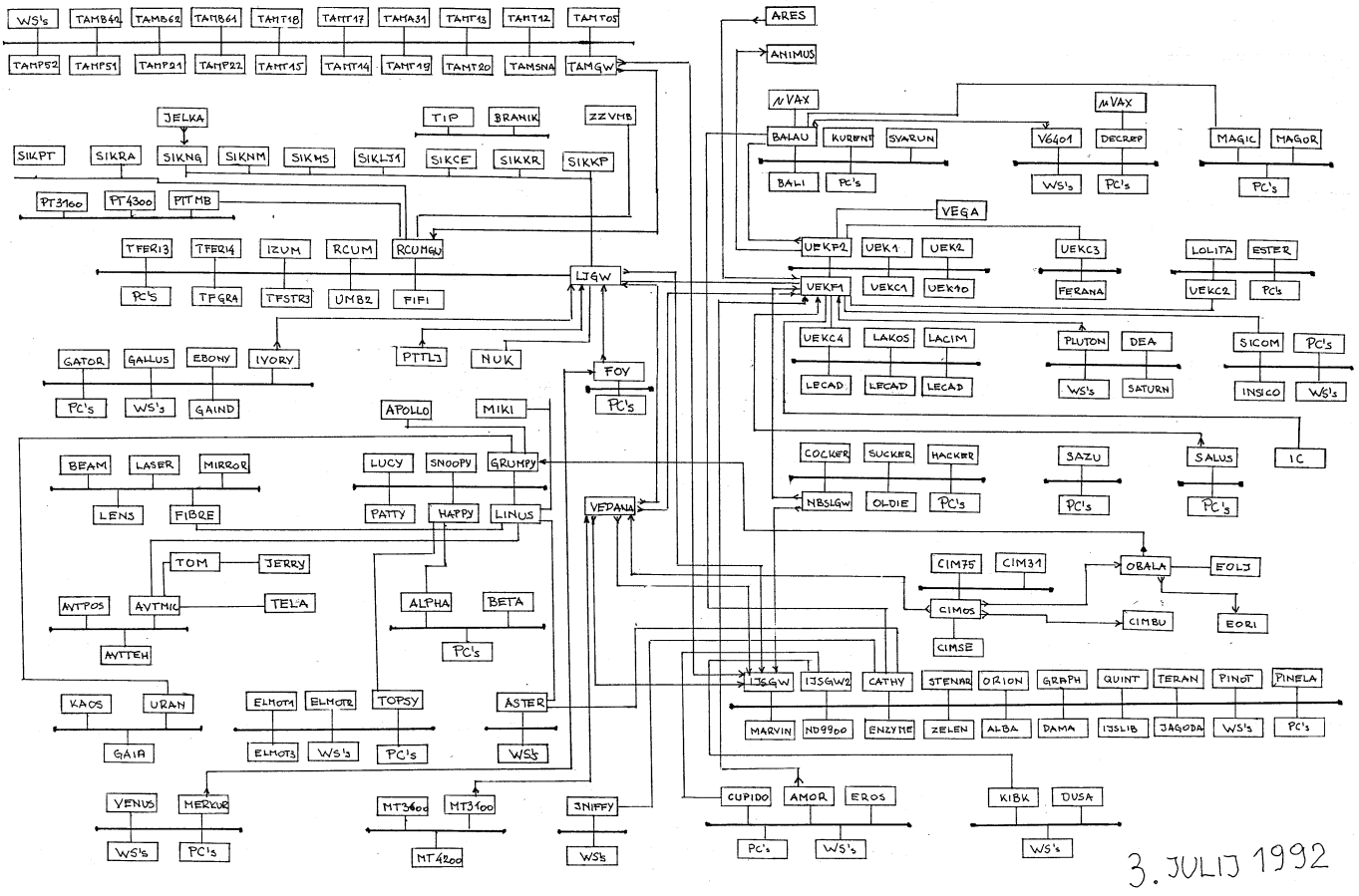
Na grobo lahko takrat prisotne tehnologije/protokole oz. omrežja delimo na v tri skupine:

- Omrežja, zasnovana na lastniških protokolih posameznih proizvajalcev IT opreme: omrežji DECnet (DEC - Digital Equipment Corporation) in EARN /BITNET/NETNORTH (IBM).
- Omrežja, zasnovana na uradno priznanih mednarodnih standardih organizacij CCITT in ISO: omrežja SIPAX, JUPAX, IXI. Najbolj poznani standardi: X.25, X.400, X.500, FTAM, X.29.
- Omrežja in storitve, zasnovane na "de facto" standardih ter nestandardnih rešitvah: EUnet (protokoli UUCP), BBS.

Velja poudariti, da se nismo ozirali na "primernost" posameznih protokolov, ampak smo uporabili vse, ki so bili razpoložljivi, funkcionalno primerni ter stroškovno učinkoviti. Tako smo npr. elektronsko pošto X.400

prenašali preko X.25, DECneta in IP-ja. Na razpoložljivost posameznih protokolov pa je občasno vplivala politika (podpora ISO/CCITT protokolov s strani Evrope in njenih mehanizmov sofinanciranja).

V Sloveniji oz. v Jugoslaviji so bili na institutih, univerzah in tudi v industriji najbolj razširjeni računalniki proizvajalca DEC (t.i. mini računalniki PDP in VAX). Njihovo povezovanje se je začelo na pobudo sistemskih inženirjev institutov in univerz, kasneje pa tudi drugih javnih institucij ter dela industrije, v omrežje **DECnet**. Pri tem so v začetni fazi igrali vodilno vlogo predvsem strokovnjaki Instituta Jožef Stefan – IJS (Mark Martinec, Matej Wedam, Iztok Tvrdy), Univerze v Ljubljani (Rok Vidmar, Matjaž Rihtar), Univerze v Mariboru (Davor Šoštarič, Nikolaj Sajko, Saš Bibič, Slavko Breznik, Polde Podlogar, Boris Boštar), ASTER (Henrik Krnec, Robert Reinhardt), NBS (Bob Marčan), MZT (Marjan/Piko Brezovar), ter še mnogi drugi.



Slika 1: Topologija DECneta, julij 1992

Prvi povezavi sta bili vzpostavljeni okrog leta 1984, in sicer med Institutom Jožef Stefan in NBS ter med Univerzo v Ljubljani in Univerzo v Mariboru. Prepustnost povezav je bila sprva zgolj 1200 oz. 2400 bit/s. Razvoj tehnologije modemov in vpeljava omrežja SIPAX/JUPAK sta omogočili hitro rast omrežja [1]. Slika 1 prikazuje topologijo DECneta sredi leta 1992. Poudariti velja, da slika prikazuje predvsem gostiteljske računalnike, ne pa tudi posameznih osebnih računalnikov, ki so tudi bili del omrežja – skupaj s PC-ji je bilo v omrežju več kot 640 računalnikov, glavnina iz Slovenije, precej tudi iz Hrvaške, Univerze v Beogradu, v ostalih republikah pa predvsem vozlišča knjižničnega omrežja KIS/SZTI, ki ga je vzpostavljal IZUM. Tipična zmogljivost povezav je bila 9,6kbit/s. Z uvedbo omrežja ARNES v letih 1992/1993 se je topologija bistveno spremenila, s prehodom na internet pa je DECnet postopoma odmrl. Glavne storitve omrežja DECnet so bile elektronska pošta, prenos datotek, oddaljen dostop, prenos kratkih sporočil in diskusijske skupine oz. forumi, poznani kot VAX Notes. Slednji so bili še posebej popularni, od strokovnih pa vse do kuharskih in političnih.

EARN je bilo evropsko akademsko omrežje, sofinancirano s strani IBM in zgrajeno na osnovi njegovih protokolov (RSCS/NJE). Leta 1990 je povezovalo okrog 750 računalnikov [2]. Povezano je bilo s sorodnim omrežjem v ZDA (BITNET) in Kanadi (NETNORTH). Centralno vozlišče tega omrežja za Jugoslavijo s povezavo 9,6kbit/s do Linza je leta 1989 na Univerzi v Beogradu postavila Mirjana Tasić. Iz Slovenije smo lahko do tega vozlišča

dostopali preko DECneta. Ker je bilo v nasprotju z DECnetom povsem akademsko, je imelo zelo striktna pravila uporabe – neposredna ali posredna komercialna uporaba ni bila dovoljena. Glavne storitve so bile elektronska pošta, prenos datotek, prenos kratkih sporočil in oddaljen dostop. Popularna je bila uporaba strežnikov LISTSERV, NETSERV, KERMIT in TRICKE. Slovenija se v omrežje EARN ni priključila, čeprav je bilo to spomladi 1991 planirano kot ena od aktivnosti "Servisnega centra akademske in raziskovalne mreže Slovenije" [3]. Z uvedbo interneta je namreč povsod postopoma odmrlo.

SIPAX je bilo slovensko javno omrežje za prenos podatkov, zasnovano na protokolu **X.25**. Upravljala ga je PTT Slovenije (pred osamosvojitvijo Slovenije smo uporabljali komplementarno jugoslovansko omrežje **JUPAK**). Kljub dragi uporabi in relativno počasnim linijam (tipična prepustnost priključkov je bila 9,6kbit/s ali 19,2 kbit/s) smo ga občasno uporabljali tudi v raziskovalno-izobraževalnem okolju, npr. za povezave omrežja DECnet, izmenjavo elektronske pošte X.400 ter oddaljen dostop.

Z vključitvijo v podprojekt **COSINE** (Cooperation for Open Systems Interconection Networking in Europe) projekta EUREKA, kjer je po zaslugi direktorja Tomaža Kalina Institut Jožef Stefan sodeloval že od leta 1987, smo se začeli intenzivno spoznavati s standardi ISO in CCITT [12]. (T.Kalin je kasneje opravljal tudi funkcijo generalnega sekretarja evropskega združenja akademskih mrež RARE ter bistveno pripomogel k razvoju komunikacij v Sloveniji) Na osnovi pridobljenih znanj smo se vključili v omrežje evropskih nacionalnih raziskovalnih omrežij in ponudili prve storitve: elektronsko pošto po standardu **X.400**, imenik po standardu X.500, FTAM in X.29. Mednarodna povezava je bila zasnovana po standardu X.25 in je potekala preko javnega omrežja SIPAX/JUPAK ter namenske povezave v evropsko raziskovalno omrežje **IXI** (International X.25 Infrastructure) zmogljivosti 48 kbit/s in kasneje 64 kbit/s. Osebno sem na Odseku za digitalne komunikacije in mreže na mikrovaxu z imenom Commie instaliral priključek X.25 ter nato še **EAN**, sistem za elektronsko pošto po standardu X.400. Pri tem sem se lahko naslonil na teoretično znanje sodelavcev Marka Bonača, Iztoka Tvrdyja, ter izkušnje sistemcev Marka Martinca in Mateja Wedama. V vlogi upravitelja centralnega akademskega vozlišča X.400 za Jugoslavijo (WEP – Well known Entry Point) sem skrbel za mednarodno povezavo, delovanje centralnega vozlišča, ter širitev in delovanje X.400 po Jugoslaviji. EAN smo instalirali na obe slovenski univerzi, Univerzo v Zagrebu, Univerzo v Beogradu, Iris Energoinvest v Sarajevu ter Inštitut Ruđer Bošković v Zagrebu. Ker je bila programska oprema instalirana na centralnih, gostiteljskih računalnikih, je bila e-pošta na voljo vsem, ki so na teh računalnikih imeli uporabniške račune – tako rekoč vsi raziskovalci, profesorji in študentje. S pomočjo prehodov do omrežij EARN/BITNET, internet in UUCP, ki so nam jih velikodušno nudila druga akademska omrežja (za BITNET in internet DFN, za UUCP pa Switch), smo dobili globalno storitev. V EAN vgrajen prehod do elektronske pošte VMS Mail je omogočil komunikacijo z vsemi uporabniki omrežja DECnet. Elektronska pošta se ni uporabljala zgolj za medosebno komunikacijo, temveč preko posebnih prehodov tudi za brskanje po internetu ter za prenos datotek. Za povezavo z raziskovalci in študenti v tujini sta pomembno vlogo odigrali distribucijski listi RokPress (skrbnika Rok Sosič in Igor Benko) ter Pisma-Bralcev (skrbnik Andrej Brodnik).

Med vojno za Slovenijo je količina e-pošte tako narasla, da smo morali računalnik Commie na hitro nadomestiti z bistveno močnejšim računalnikom. 3.7.1991 so si izposodili microVAX 4000-200 od Reaktorskega računskega centra IJS, 10.7.1991 pa smo dobili nov microVAX 4000-300, poznan kot Stenar. Zaradi pomena elektronske pošte smo vozlišče med alarmom za zračni napad preselili v zaklonišče IJS. Pri kolobocijah z menjavami računalnikov je veliko pomagal Robert Reinhardt s podjetja Aster d.o.o., pri nakupu Stenarja pa Iztok Tvrdy v vlogi svetovalca republiškega sekretarja za raziskovalno dejavnost.

Računalnik Stenar je igral pomembno vlogo tudi zaradi klicnih modemov, ki so uporabnikom omogočali dostop od doma s pomočjo programov za emulacijo terminalov VT100/VT220 (npr. WinQVT). Na ta način povezani uporabniki so lahko uporabljali na Stenarju instalirane programe: na začetku predvsem VMS Mail, VAX Notes, DECNET Copy, EAN, kasneje pa še različne internetne storitve, kot so Gopher, WWW, USENET News (ANU-News), FTP, irc, ter programe za prenos datotek preko klicne povezave (Zmodem, Kermit) [4]. Na voljo je bil tudi repozitorij prosto dostopne programske opreme PCSoft. Tik pred njegovo zamenjavo leta 1997 je imelo na njem uporabniška imena 11.000 uporabnikov iz raziskovalne in izobraževalne sfere[5]. Za njegovo sistemsko administracijo je zaslužen predvsem Primož Gabriječič.

Podobno vlogo kot Stenar za uporabnike iz raziskovalno-izobraževalne sfere, so za splošno javnost igrali sistemi **BBS** (Bulletin Board System). Ponujali so terminalski dostop preko klicnih povezav. Glavne storitve so bile elektronska pošta, izmenjava datotek ter klepetalnice oz. forumi. V Sloveniji so bili najbolj znani IUS-info BBS (Anton Tomažič, Rastislav Vintar, Darko Bulat), ABM-BBS (Boris Horvat, Dalibor Cerar), MojsteR (Matjaž Potrč) in MicroArt (Gregor Rudolf) in Krpan. Da bi povečali uporabnost, so se BBS-i povezovali v omrežja (AdriaNet na področju Jugoslavije in svetovno omrežje Fidonet) [6].

Leon Mlakar s Fakultete za elektrotehniko in računalništvo Univerze v Ljubljani je upravljal centralno vozlišče omrežja **EUnet** za Jugoslavijo. Uporabljali so nabor protokolov **UUCP** (Unix-to-Unix Copy). Ponujalo je predvsem elektronsko pošto in izmenjavo datotek. Kasneje je slovenski del omrežja EUnet prevzelo podjetje NIL d.o.o.,

ki je po zaslugi Ivana Pepelnjaka postalo prvi komercialni ponudnik interneta v Sloveniji.

3. ŠIRITEV INTERNETA V AKADEMSKEM OKOLJU

Pomembnosti uvedbe interneta v Slovenijo smo se zavedeli že relativno zgodaj ter o tem začeli opozarjati pristojne. Tako je bila npr. v predlogu "Servisnega centra akademske in raziskovalne mreže Slovenije" [3] z 25.maja.1991 vzpostavitev povezave v internet identificirana kot prioritetna (začetek delovanja čimprej).

Večina aktivnosti za priklop na internet je potekala (takrat še v imenu YUNAC-a, jugoslovanskega akademskega omrežja s sedežem na Institutu Jožef Stefan) v Laboratoriju za odprte sisteme na IJS pod vodstvom Borke Jerman Blažič, ki je bila zelo aktivna tudi v projektu COSINE ter opravljala funkcijo generalnega sekretarja YUNAC-a. Do realizacije prve povezave je prišlo 27.11.1991, ko je Denis Trček ob pomoči Maartena Terspstre z Nikhefa na Nizozemskem v internet preko omrežja IXI povezal prvi računalnik (SUN SparcStation).

S tem priklopom je bil storjen prvi korak. Sledila je povezava celotnega omrežja Instituta Jožef Stefan ter kasneje še ostalih raziskovalnih in izobraževalnih organizacij. Zavedanje pomena interneta in s tem tudi podpora je bila široko prisotna tudi na političnem nivoju. Tako se je sklep 4. seje Komisije za akademske mreže pri MZT, katere člani so bili Tomaž Kalin (IJS/RARE), Pavel Meše (PTT), Branko Zebec (IZUM), Davor Šoštarič (Računski center Univerze v Mariboru) in Tone Vidmar (Fakulteta za elektrotehniko in računalništvo, Univerza v Ljubljani), glasil: "Servise Interneta je treba čimprej pripeljati v slovenski prostor. ARNES naj poišče čim boljšo začasno rešitev" [7].

Pri širitvi interneta po Sloveniji so na operativnem nivoju sodelovali sistemski administratorji posameznih organizacij, osebno pa sem bil zadolžen najprej za razširitev interneta na celoten Odsek za digitalne komunikacije in mreže na IJS, vključno s Stenarjem, ter kasneje za konfiguriranje interneta na hrbtenici omrežja ARNES, na dostopovnih usmerjevalnikih, ki so posamezne organizacije povezovali na hrbtenico, ter za upravljanje mednarodne povezave.

Posebej je potrebno izpostaviti vlogo Izuma, ki je začel graditi omrežje za potrebe knjižničnega informacijskega sistema KIS/SZTI. Pri tem so bili zaslužni predvsem Tomaž Seljak, Branko Zebec in Igor Baš. Del tega omrežja je kasneje prerasel v hrbtenico omrežja Arnes, del pa v omrežji obeh univerz.

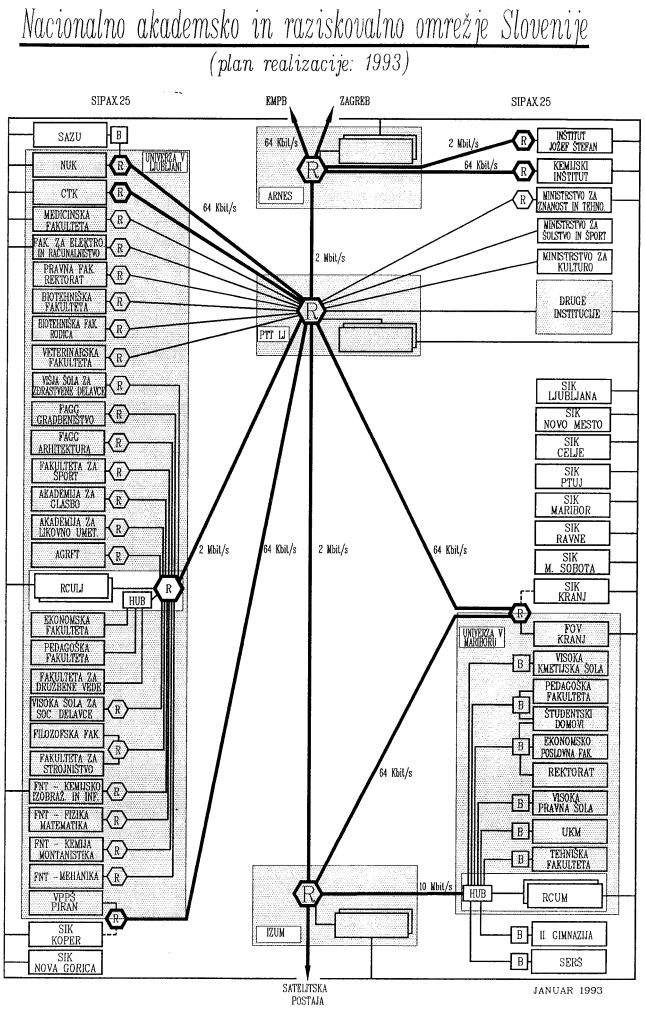
Seveda pa širitev inteneta ni bila tako enostavna kot danes. Računalniki oz. njihova osnovna programska oprema zvečine še ni podpirala protokola IP, obstoječe implementacije pa so bile precej nerodne. Tako je npr. že prva naloga – povezava celotnega oddelka na IJS, predstavljala izziv. Kot usmerjevalnik sem uporabil PC z operacijskim sistemov DOS ter programom v javni lasti za usmerjanje interneta PCroute. Zapletlo se je, ker je PCroute podpiral zgolj zelo omejen nabor ethernet kartic, obenem pa ni bilo možno hkrati instalirati dveh enakih gonilnikov. Celoten projekt se je prevedel na iskanje primernih ethernet kartic, ki so takrat bile precejšnja redkost - čeprav smo na IJS že uporabljali osebne računalnike, zvečine še niso bili povezani v lokalno omrežje, ampak zgolj preko asinhrone povezave kot terminal na računalnike VAX. Zato smo morali, da bi čim več uporabnikom na enostaven način ponudili dostop do interneta, na gostiteljske računalnike instalirati podporo za internet, kar je v večini primerov pomenilo nakup dodatnih licenc, npr. za Stenar licenco za produkt Multinet.

Z ustanovitvijo Arnesa 15.maja.1992 [8] bil podan organizacijski okvir za vzpostavitev enotnega raziskovalno-izobraževalnega omrežja. Tehnične zahteve za hrbtenično omrežje so bile usklajene med Arnesom, Izumom, IJS in obema univerzama junija 1992. Določeno je bilo, da mora biti hrbtenica enotna, pokrivati mora tako potrebe knjižnično informacijskega sistema KIS/SZTI, ki ga je gradil Izum, kot tudi vse ostale storitve, potrebne v akademskem okolju. Podana je bila zahteva po hkratni podpori več protokolov: DECnet (faza IV in faza V oz. CLNS), X.25 in IP [9]. V skladu s programom dela Arnes za 1992 [10] smo začeli z vpeljavo storitev.

V juniju 1992 smo (kot drugi v Evropi!) od RIPE n.c.c pridobili blok naslovov IP (193.2.0.0/16) in ga začeli dodeljevati raziskovalno-izobraževalnim organizacijam.
Sledila je uvedba domene .SI. V juliju 1992 je avtor ob organizacijski pomoči Marka Bonača postavil strežnike DNS za domeno .SI. S tem je Arnes začel z registriranjem domen pod .SI ter vzdrževanjem vrhnjega DNS za .SI. Začel se je proces prehoda iz domene .YU na domeno .SI za e-pošto. DNS in sisteme za e-pošto posameznih organizacij smo skonfigurirali tako, da so sprejemali pošto na naslove pod .SI in .YU, oddajali pa z naslovov pod .SI. Arnes je upravljal strežnike DNS za .YU vse do junija 1994, ko je upravljanje prevzela Mirjana Tasić iz YUNET-a (nove organizacije Srbije in Črne gore). Po dogovoru z Mirjano so zapisi MX, potrebni za dostavo pošte na obstoječe naslove pod ac.mail.yu, delovali še vse do poletja 1995.

Za potrebe mednarodne povezave in razpeljave interneta po Sloveniji smo že v jeseni 1992 realizirali nakup usmerjevalnika prometa (Cisco AGS+/4). 1.10.1992 smo se s povezavo prepustnosti 64kbit/s priključili na novo generacijo mednarodnega akademskega omrežja EMPB, ki je podpiralo X.25, IP in CLNS. Postopoma smo gradili hrbtenično omrežje po Sloveniji in priključevali organizacije. Slika 2 prikazuje plan omrežja za leto 1993, usklajen s potrebami Izuma in obeh univerz. Razvidni sta tudi interni omrežji obeh univerz. Povezave s prepustnostjo

64 Kbit/s in 2Mbit/s so pomenile kvalitativen preskok v primerjavi s povezavami omrežja DECnet. S tem smo postavili odlično osnovo za nadaljnji razvoj. Seveda smo morali kasneje povezave še mnogokrat nadgraditi, hrbtenično omrežje pa razširiti tudi v druge kraje. K temu je veliko prispeval Mihael Dimec. Ustrezno smo povečevali tudi prepustnost mednarodnih povezav. Med drugim smo novembra 1992 med prvimi v Evropi in svetu ob pomoči Tonya Batesa in Martena Terpstre iz RIPE n.c.c. vpeljali protokol BGP-4, ki je omogočal optimizacijo velikosti usmerjevalnih tabel v internetu.



Slika 2: plan akademskega omrežja za leto 1993

Hkrati z gradnjo hrbteničnega omrežja smo gradili tudi storitev ponujanja individualnega dostopa od doma preko klicnih povezav. Po vseh omrežnih skupinah v Sloveniji smo instalirali klicne modeme in tako omogočili klicanje po lokalni tarifi. Sprva smo omogočali zgolj terminalski dostop do Stenarja, kasneje pa dodali tudi klicni dostop do interneta (protokol PPP). Ker operacijski sistem Windows sprva ni imel vgrajene podpore za IP, smo morali napisati priročnike in instalacijske programe, npr. za preiskusni program Trumpet Winsock [11]. Ko je Telekom Slovenije ponudil ISDN, smo ponudili rešitev, ki je omogočala dostop tako preko ISDN kot tudi preko analognih telefonskih priključkov. Za klicni dostop je skrbel Janko Kersnik, kasneje pa je področje prevzela Ksenija Furman, za priročnike in navodila pa je skrbel Tomi Dolenc.

Vzporedno z graditvijo omrežja smo nadgrajevali in vpeljevali nove storitve. Oktobra 1992 smo se pridružili konzorciju akademskih mrež, ki je distribuiral diskusijske skupine Usenet News. V ta namen smo preko omrežja EMPB vzpostavili neposredno IP povezavo z švicarsko akademsko mrežo Switch. Uporabili smo programski paket C-news. Kasneje smo program za news instalirali tudi na Stenarja (ANU-News). Za storitev je najprej skrbel avtor, nato pa je administracijo prevzel Tomaž Borštnar.

Z vzpostavitvijo povezave v internet se je odprla tudi možnost za nadaljni razvoj e-pošte. V letu 1992 smo postavili lasten prehod med X.400 in SMTP (internetna e-pošta) s programom PP. Upravljanje storitve e-pošta je prevzel sodelavec Benjamin Zwittnig. Na PP smo preselili tudi centralno vozlišče za X.400 (WEP). S pomočjo programa Multinet smo na Stenarju ponudili prehod med sistemom VMS Mail in SMTP. Ko so uporabniki postopoma prešli na uporabo vmesnikov za internetno e-pošto, sta X.400 in VMS Mail odmrla. Toda s tem se razvoj ni končal. Prišle so nove storitve, kot je WebMail ter označevanje in razvrščanje e-slame (spam) ter filtriranje virusov.

Internet, kot ga mnogi razumejo danes, se je v Sloveniji začel s postavitvijo prvega spletnega strežnika**,** ki ga je na računalniškem centru Instituta Jožef Stefan v okviru projekta Mat'kurja pod vodstvom Miše Alkalaja postavil Mark Martinec [12]**.**

Zaradi rastočega zanimanja za spletno tehnologijo smo uporabnikom Stenarja omogočili postavitev lastnih spletnih strani, kasneje pa dodali še mnoge druge storitve. Toda to že presega namen tega prispevka.

4. INTERNET V KOMERCIALNI SFERI

Kot začetek komercializacije interneta v ZDA se omenja letnica 1994. Slovenija tudi takrat ni dosti zaostajala, vsaj ne kar se tiče aktivnosti prvih komercialnih ponudnikov. Razlika v primerjavi z ZDA je bila še najbolj očitna pri hitrosti razvoja in prepustnosti linij, kjer smo zaostajali zaradi toge zakonodaje na področju telekomunikacij.

Prvi komercialni ponudnik je bil NIL d.o.o. z omrežjem SInet/EUNET. Za to ima glavne zasluge Ivan Pepelnjak. Sledili so Quantum d.o.o. (Lojze Zadravec) in ABM d.o.o. (Boris Horvat, Darko Bulat), K2.net (Darko Bulat) in Perftech (Damijan Hafner in Uroš Rezar). Telekom Slovenije je vstopil relativno pozno s podjetjem SIOL d.o.o., hitro pa si je pridobil prednost pred konkurenti zaradi svoje moči na trgu telekomunikacij ter z nakupom

omrežja SInet/EUNET. Z leti se je število Internet ponudnikov v Sloveniji postopoma večalo, nekaj ponudnikov so kupili večji, nekaj pa jih je iz različnih razlogov prenehalo z delovanjem.

Pomembno vlogo za razvoj interneta v Sloveniji je igral (in jo še igra) tudi SIX - Slovenski Internet eXchange. Mednarodne linije so bile zelo drage, zato je bilo nujno vzpostaviti vozlišče, ki bo ponudnikom interneta omogočilo izmenjavo prometa v Sloveniji. Ker je bil takrat PTT Slovenije (predhodnik Telekoma Slovenije) monopolni ponudnik zakupljenih vodov v Sloveniji, smo februarja 1994 SIX vzpostavili kar v okviru vozlišča Arnes, ki se je nahajalo v prostorih PTT na Cigaletovi 10 v Ljubljani. Preko te lokacije so potekali vsi glavni telekomunikacijski vodi in je bila zato tako tehnično kot tudi cenovno najbolj ugodna. Na začetku je služil za izmenjavo prometa med Arnesom in edinim komercialnim ponudnikom Interneta v Sloveniji SInet/EUNET, postopoma pa so se pojavili in pridružili še drugi. SIX je bil leta 1994 eno prvih tovrstnih vozlišč v Evropi.

S postopno liberalizacijo trga telekomunikacij so se pojavili alternativni ponudniki telekomunikacijskih vodov. Podobno kot v tujini so alternativni ponudniki tudi pri nas ponudili v zakup optična vlakna, preko katerih je bilo moč relativno poceni doseči nekaj razredov večje hitrosti prenosa podatkov kot preko približno enako dragih bakrenih zakupljenih vodov Telekoma Slovenije. Ker so potrebe po kapacitetah z razvojem Interneta v Sloveniji zelo hitro naraščale, so bili ponudniki interneta prisiljeni poseči po optičnih vlaknih. Edino na ta način so lahko zagotovili kolikor toliko ugodne cene prenosnih poti in posredno svojih storitev. Žal Telekom Slovenije alternativnim ponudnikom telekomunikacijskih vodov ni dovolil izgradnje optičnih vodov do SIX-a na Cigaletovi. To je postavilo ponudnike Interneta (razen SIOL-a) v težaven položaj. Telekom Slovenije ni hotel ponuditi optičnih vlaken, dvomegabitni zakupljeni vodi so bili prepočasni, ATM tehnologija pa bistveno predraga. Zato smo bili SIX prisiljeni seliti iz prostorov Telekoma Slovenije.

Trenutno je SIX porazdeljeno vozlišče, zasnovano na tehnologiji Ethernet. Vozlišče SIX v prostorih Arnesa na lokaciji Tehnološkega parka Ljubljana je z redundantnima povezavama povezano na vozlišče SIX na lokaciji Instituta Jožef Stefan. Ima 17 članov, petminutna povprečja vsote prometa preko SIX pa presegajo 30 Gbit/s [13].

5. ZAKLJUČEK

Od prvih korakov pri uvajanju DECneta do danes tako rekoč vseprisotnega interneta smo prehodili dolgo pot. Pri tem je sodelovalo veliko število organizacij in posameznikov. Uspešni smo bili predvsem zaradi sodelovanja vseh akterjev. Upam, da bomo tudi v prihodnosti znali ohraniti enako kombinacijo entuziazma, profesionalizma in sodelovanja. Glede na aktivnosti in uspehe pri uvajanju IPv6 v Slovenijo lahko trdim, da smo še vedno na pravi poti, da se duh sodelovanja ni izgubil, pa čeprav morajo danes pri tem med sodelovati tudi organizacije, ki med seboj na trgu tekmujejo za preživetje.

Pomembno je tudi, da internet ostane odprt, kajti zgolj na ta način bomo ohranili okolje, v katerem je možen hiter razvoj in uvedba novih storitev. Vsaj za raziskovalno-izobraževalno sfero je to bistvenega pomena.

Viri

[1] M.Wedam, D. Šoštarič, A.Jauk. Lastna dokumentacija, zapiski, e-pošta, pogovori.

[2] Pravila korišćenja EARN mreže, EARN/86/07, Ženeva, 1986.

[3] M. Bonač, A. Jauk. Pregled dejavnosti in stroškov servisnega centra akademske in raziskovalne računalniške mreže Slovenije. Ljubljana, 25.maj.1991. Dopis direktorja instituta Jožef Stefan prof.dr. Tomaža Kalina republiškemu sekretarju za raziskovalno dejavnost in tehnologijo, prof. dr. Petru Tancingu, 25.5.1991.

[4] Arnes. Uporaba storitev omrežja internet na računalniku Stenar, priročnik, april 1995.

[5] A.Jauk. Predaja računalnika STENAR Muzeju računalništva, gradivo ob 10 letnici Arnesa, Ljubljana, December 2002.

[6] V. Djurdjič. Pozabljeni predhodnik interneta, Monitor, Januar 2011

[7] Zapisnik 4. Seje Komisije za akademske mreže pri Ministrstvu za znanost in tehnologijo, 3.7.1992

[8] Odlok o ustanovitvi javnega zavoda Akademska in raziskovalna mreža Slovenije. Uradni list RS, št. 23, str. 1799-1800, 15.5.1992

[9] A.Jauk, I.Baš, S.Bibič,D.Šoštarič,R.Vidmar,M.Wedam. Zahteve za bodočo slovensko akademsko mrežo, 24.6.1992

[10] Arnes. Program dela za leto 1992. Gradivo za 1. sejo Upravnega odbora Arnes,. 21.9.1992

[11] Arnes. Priročnik: Programska oprema za polni internet dostop preko telefonskega omrežja za MS Windows, oktober 1995.

[12] T. Oblak-Črnič. O začetkih Interneta na Slovenskem, Javnost – The Public, Št. 15, str. 151 - 174, Ljubljana, 2008.

[13] SIX. http://www.arnes.si/infrastruktura/six-sticisce-omrezij.html

POGLED UNIVERZE V MARIBORU IN INSTITUTA INFORMACIJSKIH ZNANOSTI 20 LET POZNEJE

Davor Šoštarič
Institut informacijskih znanosti
Prešernova 17, 2000 Maribor
e-mail: davor.sostaric@izum.si

POVZETEK

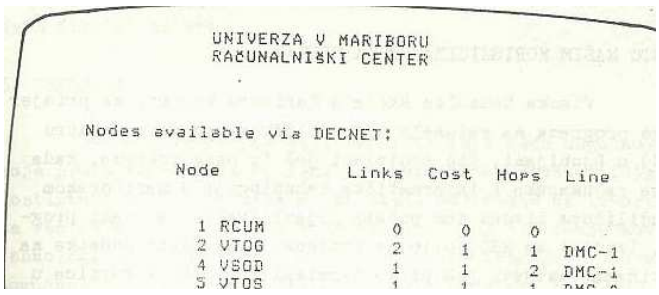
V prispevku se iz vidika Univerze v Mariboru in IZUM-a spominjamo začetkov vstopa v internetni svet.

ABSTRACT

In the article, we reminisce the beginning of entering the Internet world from the point of view of the University of Maribor and IZUM.

1 UVOD

Univerza v Mariboru se je v zgodnjih osemdesetih letih prejšnjega stoletja po vzoru na svoje sestre odločila za skupni računalniški center, ki je postal znan kot RCUM. Dve bistveni želji sta botrovali temu spočetju – enotno urejanje računalniške podpore pedagoškemu in raziskovalnemu procesu ter drzna ideja o internih informacijskih sistemih na področju finančnega poslovanja in študentske evidence. Na srečo so se začetni nabavni postopki nekoliko zavlekli in prvotno načrtovani mastodonstki »mainframe« je padel v vodo. Najboljša možna alternativa je bila nabava treh samostojnih računalnikov (»mini« v takratni terminologiji), povezanih v mrežo (da se pravilno reče »omrežje«, smo se naučili šele kasneje). Kakor hitro je bila odločitev o nabavi znana, so se začele aktivnosti o zamenjavah posameznih fakultetnih sistemov, da bi bili kar se da hitro med seboj kompatibilni in bi lahko univerzitetno omrežje spodobno naraslo. Ne pozabimo – govorimo o času, ko je pojem »mreža« pomenil centralni računalnik s kopico terminalov, tiskalnikov in morda še kakšno eksotično periferno enoto!


UNIVERZA V MARIBORU
RAČUNALNIŠKI CENTER

Nodes available via DECNET:

Node	Links	Cost	Hops	Line
1 RCUM	0	0	0	
2 VTOG	2	1	1	DMC-1
4 VSOD	1	1	2	DMC-1
5 VTOS	1	1	1	DMC-1

Slika 1: Pogled na omrežje leta 1985

2 ZAČETKI

Od samega začetka je bilo jasno, da je omrežje tisto, ki bo prineslo preboj v miselnosti. Storitve, ki so bile na voljo na za današnje pojme skromnem univerzitetnem omrežju so bile praktično enake današnjim: poznali smo elektronsko pošto, klepete, forume, prenos datotek, oddaljen dostop, deljenje virov in celo koncept »strežnik – odjemalec«. Študentje, učitelji in raziskovalci so te storitve takoj osvojili v svojih vsakodnevnih opravilih. Bistveno pa je bilo, da so razvojniki in programerji te zadeve začeli uporabljati tudi v aplikacijah univerzitetnih informacijskih sistemov – pred pojavom interneta pri nas in po protokolih, ki so definirani čisto drugače kot današnji, ki temeljijo na IP. Kmalu se je na Univerzi v Mariboru pojavil nov informacijski sistem, ki je bil usmerjen v knjižničarski svet in ki je prestopil meje mariborske univerze. Iz tega je kasneje zrasel COBISS, iz RCUM-a je izšel Institut informacijskih znanost – IZUM, ki se je kasneje izdvojil in postal samostojni infrastrukturni raziskovalni zavod z vlogo nacionalnega knjižničnega servisa.

Univerza v Mariboru ni bila osamljena. Podobne zamisli so imeli tudi drugi, pretežno v akademskem in raziskovalnem svetu. Celo računalniške platforme so si bile podobne in ni bilo treba dolgo čakati, da so se posamezni omrežni otočki in otoki povezali med seboj. Omrežje je raslo kot ameba; vsak novi član je pripeljal vse svoje bogastvo in ga zlil s celoto. Imeli smo torej »omrežje omrežij«. Najprej je vsako vozlišče moralo poznati vsa druga. Potem se je to pravilo nekoliko omililo, ampak kljub temu se je bilo potrebno dogovarjati za pasove posameznih naslovnih intervalov. Še vedno pa je vse temeljilo na ideji protokolno monolitnega računalniškega omrežja. Kot večina drugih omrežij, ki so temeljila na praviloma zgolj enem proizvajalcu in bolj ali manj internih standardih, je bila usoda v bodočnosti neizprosno jasna – ali bo zmagal najmočnejši ali pa se bodo pojavile črne škatle, ki bodo na stičiščih raznorodnih omrežij spravljale zadeve v navidezni red.

Kot vemo, se je zgodilo nekaj tretjega – prišel je internet in pometel z vsemi dotakratnimi principi. Saj ne, da ne bi povezovanja po protokolu IP poznali že prej, tudi s pridom smo ga uporabljali, a navzven smo se še vedno šli preverjeno zanesljivega starega načina. Dogovori o intervalih naslovov

niso več bili pomembni; pojavili so se povsem drugi gospodarji, od katerih je bilo potrebno pridobiti novo vrednost – kopico IP naslovov. Naenkrat je bilo potrebno spoštovati neka nova pravila igre.

3 VSTOP V INTERNET

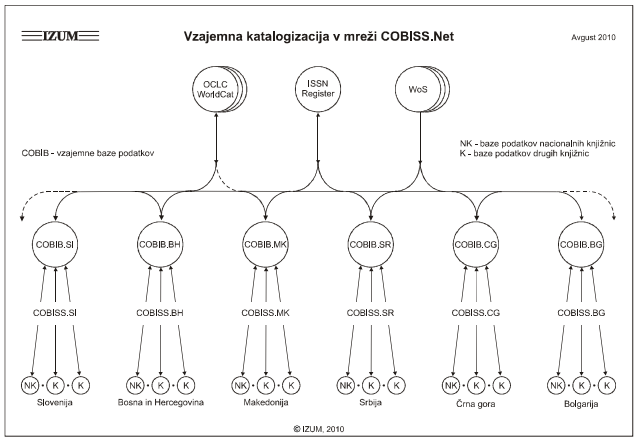
Univerza v Mariboru je ostala zvesta svoji poti; njen RCUM je bil zadolžen za urejanje vseh omrežnih relacij navznoter in navzven. To je bilo uzakonjeno z odločitvami vseh glavnih organov univerze, kakor koli so se že takrat imenovali. RCUM se je resno lotil načrtovanja omrežja v novih – internetnih – okvirih. Kot so se nekoč risale sheme terminalov in tiskalnikov, so nastajale posamezne konklave z natanko določenimi karakteristikami in poenotenem principu poimenovanja, številčenja, povezovanja in ločevanja. Sledila je drzna odločitev – zahtevati enoten velik blok IP naslovov, ki bodo zadostovali Univerzi v Mariboru za nekaj desetletij. To je bilo potrebno podpreti z realnimi analizami, trendi in tudi vizijami. Uspelo je – Univerza v Mariboru je dobila blok naslovov »razreda B« (danes bi rekli /16). Na posameznih fakultetah oziroma oddelkih je bilo kar nekaj nejevolje, češ zakaj omrežje graditi, načrtovati in upravljati iz enega mesta, vendar je zamisel ostala. Za načrtovanje interne univerzitetne topologije je RCUM ostal zadolžen in odgovoren vse do danes.

Univerza v Mariboru si je z vstopom v internetni svet hitro našla svoj novi prostor pod soncem. Vsem uporabnikom (vključno s študenti) so bile omogočene vse internetne storitve. Z nediskriminatornim odnosom do študentov se je na mlade rodove prenašalo novo znanje in odprl širok svet novih spoznanj. Ko se je pojavil še splet, je bila Univerza v Mariboru med prvimi v Sloveniji, ki se je z vsem svojim potencialom predstavila svetu, pod mentorskim delov mladih zagnancev RCUM-a pa so se veščine novih omrežnih tehnologij hitro širile tudi drugod. Fokus ni ostal usmerjen zgolj na univerzo ampak tudi na širše družbeno okolje.

Seveda ni šlo vse gladko. Objektivne težave smo znali predvideti in se z njimi tudi spopasti: nova oprema, izobraževanje, načrtovanje procesov, profesionalno upravljanje omrežja in podobno. Kot običajno pa je bil velik problem tudi v nekaterih glavah - študent naj lepo kupuje skripte svojih učiteljev, pridno dela seminarske naloge in se na podlagi uveljavljenih metod uči za izpit, ne pa da si išče znanje po nekih čudnih poteh. Niso bili vsi imuni na »šaltersko hrčkarijo« v smislu pridobitev nekakšnih dovoljenj za dostope na internet in njegovih storitev ter posledično potreb po obrazcih, obrazcih in obrazcih. Ideja, da bi vsakem študentu ob vpisu in vsakemu zaposlenemu od začetku dela avtomatično pripadlo ustrezno internetno okolje s konkretnim sistemom pravic, je ostala v oblakih. Tudi velik del konzervativne strokovne javnosti ni bil naklonjen novim paradigmam, v javnosti pa so krožili številni povsem iz trte zviti podatki. V takrat vrhunski prilogi o znanosti pomembnega slovenskega dnevnika je še nekaj let po povsem uradni in operativni prisotnosti interneta v Sloveniji pisalo o deset tisoč internetnih priključkih v ZDA, treh na Kitajskem in še nobenem pri nas. Ampak v splošnem smo z vztrajnim prepričevanjem in dokazovanjem ugodnosti počasi ustvarili okolje, kjer se posamezniki sploh več ne zavedajo, kakšna ogromna zgradba se skriva za možnostmi, ki jih imajo. Pravzaprav smo še danes hvaležni vsem tistim, ki so nam metali polena pod noge, saj je s tem naša trma dobila nov zalet.

4 IZUM, COBISS, SICRIS …

Knjižničnega informacijskega sistema COBISS in informacijskega sistema o raziskovalni dejavnosti SICRIS (nad obema bdi danes IZUM) si ne moremo zamisliti brez internetne podpore. COBISS je od samega začetka bil eksistenčno odvisen od internetnih temeljev. Ne gre zgolj za prenos podatkov in oddaljeni dostop. Nevidne povezave med posameznimi zbirkami podatkov, klasične in spletne aplikacije, prepletene navigacijske poti, takšni in drugačni relacijski modeli med podatki in sistemi so nekaj, kar brez sodobne omrežne filozofije enostavno ne bi delovalo. Kdo neki sploh ve, kje se nahajajo posamezni strežniki (no ja, zdaj je modno govoriti o oblakih …)!



Slika 2: Pogled na omrežje danes

Sistem COBISS.Net je zanimiv izvoz slovenskega znanja v tujino, saj kot sodobno zasnovana informacijska mreža, ki v državah Jugovzhodne Evrope deluje kot katalizator inoviranja teh družb in kot infrastruktura za izmenjavo informacij o intelektualni produkciji, s tem pa kot spodbujevalec medkulturnega dialoga v prostoru, kjer so še nedavno prevladovali konflikti. Te države sprejemajo COBISS.Net kot informacijsko infrastrukturo, ki jih povezuje v skupni evropski raziskovalni, izobraževalni in kulturni prostor. Za uveljavitev COBISS-a v ciljnih državah pa je odločilna slovenska razvojna pomoč sodelujočim institucijam. Gre za izrazito dolgoročen projekt s trajnimi pozitivnimi posledicami za ugled in vpliv Slovenije.

Danes je v to veliko družino vključenih preko 400 slovenskih knjižnic, skupaj s tujimi pa dosežemo številko 650. V slovenskem delu sistema COBISS se na normalni delovni dan izvede preko štiri in pol milijona transakcij in pretoči 20 milijard bitov. Skozi SICRIS se dnevno sproži po 2500 izpisov osebnih bibliografij, javno dostopni katalog OPAC pa v konici zabeleži tudi do 1300 sočasnih zahtev. Brez interneta bi bile te številke verjetno precej nižje, komunikacijski stroški pa višji.

Z drugimi besedami, brez interneta tega sistema ne bi bilo.

5 KAJ VIDITA UMB IN IZUM V INTERNETU

Tako Univerza v Mariboru kot Institut informacijskih znanosti vidita v internetu predvsem množico razpršenih informacijskih virov. Če odmislimo tehnikalije in računalniški žargon, potem nam internet predstavlja medij za prenos informacij. Milijonom ljudi po vsem svetu omogoča pogovor, razumevanje in sodelovanje in ima zato še posebno sociološko noto. Zato je internet več kot samo omrežje računalniških omrežij in je več kot še tako sofisticiran informacijski servis. Tisto, kar je najpomembnejše, so Ljudje (z veliko začetnico). Internet je prvi globalni forum in prva globalna knjižnica v zgodovini človeštva. Vsakdo lahko sodeluje in to kadarkoli; internet nikoli ne zapre svojih vrat (in nima kolektivnega dopusta). Kdorkoli ste, vedno ste dobrodošli. Nikoli ne boste odrinjeni ali celo izključeni zaradi neprimernih oblačil, napačne barve kože, vere ali denarnega položaja. Internet nima zakonov (čeprav obstajajo nekatera pravila igre), nima policije in nima vojske, in nihče - ne organizacije, fundacije, vlade ali OZN - si ne more lastiti monopola nad njim, kaj šele kontrole. Zahvaljujoč internetu smo prvič v zgodovini človeštva dobili možnost neomejenega medsebojnega komuniciranja in pokazalo se je, da je v naši naravi, da smo komunikativni, pomagajoči, radovedni, obzirni in pozorni. Na tem smo na Univerzi v Mariboru in Izumu gradili našo bodočnost.

Nismo se ustavili samo na lastni produkciji oziroma na uporabi internetnih komunikacijskih zmožnosti zgolj za lastne storitve. Vseh dvajset let smo aktivno sodelovali na ozaveščanju in vzpodbujanju ter uveljavljanju novih tehnologij v vse pore življenja. Bili smo med prvimi s profesionalno narejenimi spletnimi stranmi, ki so bile namenjene vsemu svetu. Hitro smo se zavedali neslutenega potenciala in se prilagodili novim paradigmam. Med prvimi v Sloveniji smo uveljavili koncepte varnih in zanesljivih brezžičnih dostopov Eduroam in Libroam. Z vso pozornostjo spremljamo nove trende in iščemo poti za čim uspešnejše prilagajanje.

6 ZAKLJUČEK

Pravijo, da je v naših genih zakodirano, da ne moremo živeti sami in izolirano. Vpetost človeka v njegovo okolico, njegova soodvisnost z naravnim, družbenim in socialnim okoljem, njegova ne-utilistična podvrženost estetiki, kulturi in tradiciji, to potrjuje. In ko je točno to že pred slabega pol tisočletja opisal angleški pesnik John Donne (*Noben Človek ni Otok, povsem sam zase; vsak človek je kos Celine, del kopne zemlje; če morje odplavi grudo prsti, je Evrope manj, prav tako, kakor da je bil Rtič, prav tako, kakor da je bilo Posestvo tvojih prijateljev ali tvoje lastno; ob smrti vsakega človeka je mene manj, zakaj včlenjen sem v Človeštvo ...*), kdo ve, ali ni v svoji preroški viziji slutil ta izjemni razmah nove civilizacijske dobrine na poti človeške evolucije.

In še zanimivost za konec: vsa specifična omrežja so prepoznavna po svojih karakterističnih oblikah naslovov posameznih elementov. Pred začetkom interneta je bilo za naše omrežje značilno dvojno dvopičje; danes je z uvedbo protokola IPv6 dvojno dvopičje spet tu.

Zbornik 14. mednarodne multikonference

INFORMACIJSKA DRUŽBA – IS 2011

Proceedings of the 14th International Multiconference

INFORMATION SOCIETY – IS 2011

Vzgoja in izobraževanje v informacijski družbi

Education in Information Society

Uredili / Edited by

Vladislav Rajkovič, Tanja Urbančič, Mojca Bernik

http://is.ijs.si
http://lopes1.fov.uni-mb.si

14. oktober 2011 / October 14th, 2011
Ljubljana, Slovenia

PREDGOVOR

Svet vre. Države, razen redkih, so prezadolžene. Ljudje so na ulicah. Nezaposlenost raste. Tudi mladi in šolani so brez dela. Zmanjšujejo se socialne pravice. Kaj je narobe? Kje so resnični vzroki in kakšne so rešitve?

Herbet Simon, dobitnik Nobeleove nagrade za ekonomijo, je že leta 1976 dejal: »Obstaja le model realnosti.« Pa vendar, se učimo iz modelov, ki smo jih oblikovali v preteklosti. V šolstvu govorimo o vzgojno izobraževalnih modelih. Iz izkušenj gradimo nove, ki naj bi reševali probleme, s katerimi se soočamo. Starih modelov ne gre le popravljati, ampak potrebujemo predvsem nove inovativne modele. Kako do njih?

Veliko je novega, tehnološke rešitve na čelu z računalnikom, ki ni več le pripomoček, ampak nam narekuje, kako se stvari delajo od znanosti in umetnosti preko poslovnega sveta do šole. Tehnologija omogoča nove povezave med ljudmi. Gre za bliskovito izmenjavo informacij po svetu, ki vpliva na vrednote ljudi in omogoča nove vergige vrednosti. To so nekako danosti od zunaj. Znotraj pa se tako posameznik kot šolski sistemi in države soočamo z obstoječimi modeli, ki ne ustrezajo več sodobnemu človeku in organizaciji družbe, v kateri živimo.

Je model, ki predvideva, da gredo prvošolci v šolo s tabličnim računalnikom prava rešitev? Mislimo, da je. Vendar ta model zahteva spremembo drugih obstoječih modelov, npr. poslovnih modelov založnikov učnih gradiv. Seveda ostajajo temeljni cilji vzgoje in izobraževanja. Že podcilji pa so prej naravnani na izobraževanje za industrijsko družbo, kot za družbo znanja. Govoriti, da se s tabličnim računalnikom v šoli nič ne spreminja, je nesmisel. More in mora se veliko spremeniti.

Naše dojemanje sveta dohiteva tehnologijo z zamikom. Z drugimi besedami lahko rečemo, da človek sam pri sebi zaostaja za tehnološkimi inovacijami. Tehnološke inovacije se dogajajo zunaj nas. Inovacija 'poslovnega' modela pa je notranja stvar posameznika, ustanove, podjetja, korporacije, države ipd.

Spremembe moramo ponotranjiti in jih udejaniti v svojem neposrednem okolju. Ker pa so spremembe prelomne, ne zadostuje le 'lepotno' spreminjanje modelov, ampak ustvarjanje prelomno novih inovativni modelov tudi na področju vzgoje in izobraževanja.

Inovativnost je domena človeka. Naša glava pa ne more biti inovativna, če je prazna. Torej ne gre, da bi vse imeli na računalniku oz. omrežju, učenje pa bi zanemarili. Seveda se postavlja vprašanje, kaj in kdaj se učiti ob novih tehnoloških možnostih, da bomo najbolje izkoristili 'računalnik', ki ga imamo med obema ušesoma. Na ta in podobna vprašanja bomo iskali odgovore tudi na letošnji že 14. konferenci Vzgoja in izobraževanje v informacijski družbi.

Vladislav Rajkovič, Mojca Bernik, Tanja Urbančič

PREFACE

The world and people are at their feet. A large number of countries is heavily indebted. Even the young and the educated are unemployed. The rate of unemployment is rising. Social rights are shrinking. What is going on? What is wrong? What are the real causes and solutions?

Herbert Simon, Nobel Prize winner in Economics in 1976, said that only a model of reality exists. Still, we learn from the models we created in the past. In education we refer to education models. From experience we build new ones that are supposed to solve the problems we are facing. The old models need not only to be fixed but also above all replaced by new innovative models. How do we arrive at these models?

There are many novelties, technological solutions based on computers are no longer mere aids but provide us with new directions in science, art, business, and schools. Technology also enables new relations among people. Rapid exchange of information worldwide has an impact on people's values and makes new value chains possible. These are external factors. On the inside, individuals as well as school systems and countries are faced with existing models that no longer fit the modern human and the society we live in.

Is a model, which predicts that first graders take tablet computers to school, the right solution? We believe it is. However this model requires other existing models to change, for instance publishing houses' business models for textbooks. Of course there are basic educational goals. Then again, already their subordinate goals are better suited for the education in the industrial society than the knowledge society. To claim that tablet computers in schools do not change anything does not make sense. A lot has to change and it can change.

Our perceptions of the world are catching up with the technology with a delay. In other words, humans are intrinsically lagging behind technological innovations that are occurring on the outside. Innovation of a 'business' model occurs internally within an individual, institution, company, corporation, country etc.

We need to internalize the changes and directly implement them in our environment. Since the changes are disruptive, a mere 'face-lift' does not suffice. Needed are new disruptive and innovative models, including in the field of education.

Capability to innovate is a human domain. Our head cannot be innovative if empty. It is not an option to store everything on the computer or Internet and forget about learning. Naturally, the question that arises is, what and when to learn with all the new technological advancements in order to best put to use the ‘computer’ between our ears. This question along with similar questions will be addressed also at this year's already 14th conference Education in information society.

Vladislav Rajkovič, Mojca Bernik, Tanja Urbančič

PROGRAMSKI ODBOR / PROGRAMME COMMITEE

Predsednik / Chair
Vladislav Rajkovič
Univerza v Mariboru, Fakulteta za organizacijske vede
in Institut Jožef Stefan

Alenka Adamič
Zavod MIRK

Vladimir Batagelj
Univerza v Ljubljani,
Fakulteta za matematiko in fiziko

Igor Bernik
Univerza v Mariboru,
Fakulteta za varnostne vede

Mojca Bernik
Univerza v Mariboru,
Fakulteta za organizacijske vede

Janez Bešter
Univerza v Ljubljani,
Fakulteta za elektrotehniko

Lea Bregar
Univerza v Ljubljani,
Ekonomska fakulteta

Andrej Brodnik
Univerza v Ljubljani
Fakulteta za računalništvo in informatiko

Borut Čampelj
Ministrstvo za šolstvo in šport RS

Dejan Dinevski
Univerza v Mariboru, Medicinska
fakulteta

Saša Divjak
Univerza v Ljubljani,
Fakulteta za računalništvo in informatiko

Tomi Dolenc
ARNES

Mitja Jermol
Institut Jožef Stefan

Peter Kokol
Univerza v Mariboru,
Fakulteta za zdravstvene vede

Alenka Krapež
Gimnazija Vič

Nives Kreuh
Zavod RS za šolstvo

Tanja Logar
Center za poklicno izobraževanje

Andreja Lenc
Center za mobilnost in evropske programe
izobraževanja in usposabljanja

Gregor Mohorčič
Zavod RS za šolstvo

Mojca Orel
Gimnazija Moste

Milan Pagon
University of Iowa, CIMBA

Jože Rugelj
Univerza v Ljubljani, Pedagoška fakulteta
in Institut Jožef Stefan

Niko Schlamberger
Slovensko društvo Informatika

Tomaž Skulj
Zveza za tehnično kulturo Slovenije

Poslovna komunikacija in poslovni tuji jezik v Moodlu

Business Communication and German as a Foreign and Business Language in Moodle

Jelka Bajželj
ESIC Kranj, Višja strokovna šola
jelka.bajzelj@guest.arnes.si

Povzetek

Preko komunikacije v spletni učilnici se lahko učimo bontona in pravil uspešnega (poslovnega) komuniciranja, delo v spletnem učnem okolju Moodle pa prinaša še številne druge prednosti. Avtorica v prispevku predstavi uporabo spletnih učilnic na primeru predmetov Poslovno komuniciranje in Poslovni tuji jezik (nemščina). Skrb za nenehno izboljševanje na področju izobraževanja pomeni še en korak k višji kakovosti in vseživljenjskemu učenju.

Ključne besede: (poslovna) komunikacija in poslovni tuji jezik (nemščina), spletna učilnica, kakovost

Abstract

By using an e-classroom as means of communication, we can learn the successful (business) communication etiquette and rules. Besides that, using Moodle is accompanied by other advantages. The authoress represents the use of e-classrooms for the subjects Business Communication and German as a Foreign and Business Language. The meticulous improvement in the field of education represents another step towards higher quality and lifelong learning.

Keywords: (business) communication, german as a foreign and business language, e-classroom, quality

Otrok in računalnik v vrtcu

Child and Computer in Kindergarten

Bojana Benkovič
VVE Bogojina, Vrtci Občine Moravske Toplice
Levstikova 11, 9226 Moravske Toplice
bojana.benkovic@gmail.com

Povzetek

V tem članku želim prikazati delo z računalnikom v skupini otrok, starih od štiri do šest let, in pomen dela z računalnikom v zgodnjem otroštvu. Otroci so skozi igro pridobili nova znanja o računalniku in njegovi uporabi. Delo z računalnikom je bilo načrtovano pri vseh vzgojnih področjih postopno in primerno predznanju otrok. Z upoštevanjem otrokove želje, volje in interesa sledimo njegovemu razvoju. Pričakujemo, da bo otrok sprejel ter usvojil nova spoznanja in znanja, tako kot je načrtovano v našem celoletnem projektu Otrok in računalnik v vrtcu.

Ključne besede: predšolski otrok, računalnik, vrtec

Abstract

In this article I want to present the education using a computer in a group of children aged from four to six years and the importance of working with computers in early childhood. Through playing the children have gained new knowledge about a computer and its use. Work with the computer has been planned in all fields of education gradually and suitably to prior knowledge of the children. Taking into account the child's desires, willingness and interest we follow the child’s development. It is expected the child will adopt new knowledge and skills as planned in our whole year project Child and computer in kindergarten.

Keyword: Pre-school child, computer, kindergarten

Informacijska podpora dokumentaciji sistema vodenja kakovosti v šolstvu

Information Technology Support of a Documentation Quality Management System in the Education Sector

Nastja Beznik
Tehniški šolski center Kranj
Višja strokovna šola za mehatroniko, informatiko in elektroenergetiko
nastja.beznik@guest.arnes.si

Povzetek

V dobi informatizacije in uveljavljanja komunikacijskih tehnologij se organizacije srečujejo s številno dokumentacijo in informacijskimi tokovi. Sistem vodenja kakovosti je eden izmed elementov, ki iz gospodarstva vedno bolj prehaja tudi na negospodarske inštitucije. V prispevku so opredeljene vloga, pomen in ključne značilnosti sistema vodenja kakovosti s poudarkom na praktičnem primeru računalniškega spremljanja dokumentacije. Z informacijsko tehnologijo podprtim sistemom vodenja kakovosti obvladujemo in pospešujemo razvoj poslovanja ter preko sistema uvajanja stalnih izboljšav ohranjamo konkurenčno prednost na tržišču. V drugem delu prispevka je prikazan, v poslovnem okolju že uveljavljen, primer informacijske rešitve, s katerim zadostimo tako zakonskim regulativam, zahtevam sistema vodenja kakovosti po standardih družine ISO 9000 kakor tudi potrebam managementa po točnih, popolnih, logično usklajenih, pravočasnih, edinstvenih in veljavnih informacijah, ki so osnova za poslovno odločanje.

Ključne besede: Sistem vodenja kakovosti, procesni pristop, dokumentacijski sistem, bistveni proces, podporni proces.

Abstract

In the era of computerization and communication technologies, organizations are faced with numerous documents and information flows. Quality management systems are one of the elements of the real sector which are increasingly being applied also in non-economic institutions. This paper defines the role, importance and key features of a quality management system with emphasis on the practical case of computerized documentation monitoring. A quality management system supported by information technology allows us to manage and strengthen operations through the introduction of continuous improvements, which allow us to maintain a competitive advantage in the marketplace. The second part of the paper illustrates the IT solution already established in a business environment which enables us to meet statutory regulations, quality management requirements observing the ISO 9000 family of standards, as well as the needs of management for accurate, complete, logically coherent, timely, unique and valid information, which form the basis for business decisions.

Keywords: Quality Management System, Process Approach, Documentation System, Primary Process, Secondary Process.

Prihodnost uporabe IKT v izobraževanju skozi prizmo bodočih učiteljev

Future Teacher's Attitude toward ICT in Education

Tomaž Bratina
tomaz.bratina@uni-mb.si

Povzetek

Sodobnega in učinkovitega pouka si danes ni več mogoče zamisliti brez uporabe IKT. Pri tem gre tako za tehnološke pripomočke kakor za sodobne oblike učnih vsebin, njihove priprave in distribucije. Da bi temu zadostili, potrebujemo kompetentnega učitelja, ki ga moramo na uporabo IKT pripraviti skozi proces izobraževanja bodočih učiteljev. V ta namen smo v pripravili učni predmet IKT, ki obsega učne vsebine vezane na uporabo IKT v izobraževanju. Sodobnega pouka si brez želje ter pripravljenosti učitelja za uporabo IKT pri poučevanju ni mogoče predstavljati. Zato mora bodoči učitelj že med izobraževanjem izražati pozitiven odnos do tovrstnih učnih vsebin in tako spoznati možnosti, ki mu jih uporaba IKT ponuja. Tak odnos bo posledično zbudil tudi željo in pripravljenost uporabljati IKT in s tem narediti pouk sodoben in učinkovit. Priljubljenost učnih vsebin s področja IKT v izobraževanju in odnos ter pripravljenost za uporabo IKT v poučevanju smo preverjali pri dveh generacijah študentov, bodočih učiteljev. Izidi so pokazali dokaj pozitiven odnos do vsebin in visoko mero pripravljenosti za uporabo IKT pri delu učitelja.

Ključne besede: IKT, izobraževanje, učitelj, e-gradiva, digitalna kompetenca

Abstract

The contemporary learning process inevitably requires the application of ICT means irrespective of their form. ICT in education comprehends both the technological resources as e-learning materials with their specific manner of distribution. To meet the expectations of contemporary learning process the ICT competent teacher is required. The necessary competencies are to be achieved during the educational process of future teachers within learning contents of the studying subject ICT in education. Despite the learning contest certain level of acceptance and positive attitude of future teachers toward ICT in education is necessary. The adequate acceptance and attitude will later stimulate the application of modern learning principles and methods in performing the learning process of present future teachers. During the learning process the levels of acceptance and attitude of two generations of future teachers were investigated. The results were encouraging showing acceptance, positive attitude and readiness to apply the ICT in the learning process.

Keywords: ICT, education, teacher, e-learning material, digital competence

Sodelovalno delo v spletni učilnici Moodle na primeru obravnave domačega branja

Cooperative Learning in Moodle Learning Environment Based on the Example of Required Reading Discussion

Dragica Brinovec
Poženelova ulica 7, 3270 Laško
dragica.brinovec@guest.arnes.si

Povzetek

Prispevek prikazuje primer praktične uporabe učnega okolja Moodle v 7. razredu osnovne šole pri predmetu slovenščina, in sicer na primeru obravnave domačega branja. Z učnim okoljem Moodle združujemo klasično in spletno učilnico, spodbujamo motivacijo za branje pri osnovnošolcih ter hkrati razvijamo digitalno pismenost – temeljno učno kompetenco v novem učnem načrtu za slovenščino. Kombinirano učenje in poučevanje (s klasično in spletno učilnico) omogoča brezmejno ustvarjalnost, tako učencem kot učiteljem, spodbuja sodelovalno delo, razvija pozitivne odnose med učenci kakor tudi med učenci in učiteljem ter v vzgojno-izobraževalni proces uvaja partnerski odnos kot novo dimenzijo učenja in poučevanja.

Ključne besede: učno okolje Moodle, 7. razred OŠ, slovenščina, domače branje, motivacija za branje, digitalna pismenost, pozitivni odnosi, sodelovalno delo, partnerski odnos

Abstract

This paper introduces the practical use of Moodle Learning Environment in 7th grade of primary school in Slovene language teaching on the example of required reading discussion. Moodle learning environment combines "classic" and digital classroom increases reading motivation in primary school children and helps develop digital literacy – the fundamental learning competence in a new curriculum for Slovene language. Combining classic and digital learning environment enables the shift in student – teacher roles. Students are no longer passive recipients of information transmitted by a teacher or textbook. On the other hand, the teacher is not the centre of attention but takes the role of facilitator – sets the goals, provides support and resources. Consequently, the combination of classic and digital learning environment enables creativity in the learning process (for both parties), cooperative learning, development of positive relationships (between students and the teacher as well as among students) and introduces cooperative relationship as a new dimension of learning –teaching process.

Keywords: Moodle Learning Environment, 7th grade of primary school, required reading, Slovene language, digital literacy, positive relationships, cooperative learning, cooperative relationship

Multimedijska tehnologija v pomoč izobraževanju

Multimedia Technology in Support of Education

Polona Brišar
lona.brisar@gmail.com

Dominik Snedec
dominik@snedec.si

Povzetek

Izobraževanje je eno izmed področji, ki v 21. stoletju na podlagi razvoja tehnologije spreminja naravo in dobiva novo obliko. Poleg ostalih področij, v katerih je ravno tako zaznati velik napredek v razvoju, tudi informacijsko komunikacijske tehnologije vedno bolj prodirajo v panogo šolstva in izobraževanja, kjer svojo dodano vrednost nudijo z novimi možnostmi pri procesih učenja in izobraževanja. Vsesplošna dosegljivost in prilagodljivost tehnologije je samo še dodatni argument za lažje in hitrejše vključevanje v izobraževalne institucije. V nadaljevanju avtorja navajava nekaj primerov multimedijskih tehnologij, katere lahko znatno pripomorejo pri izobraževanju in motivaciji učencev. Vsebine, ki se podajajo v obliki gibljive slike in zvoka, učenci (slušatelji) veliko lažje in hitreje razumejo, hkrati pa si jih tudi kvalitetnejše zapomnijo. Internet nam ponuja neomejene možnosti komuniciranja in medsebojnega sodelovanja. Je osnova za delovanje tehnologije, ki služi za prenos zvoka, slike in multimedijskih vsebin. Premoščanje prostorskih razlik je ena izmed možnosti, ki jih ponujajo te tehnologije, v prenekaterih primerih pa celo ključna, saj pomeni prihranek pri času in denarju. Nove tehnologije nam namreč omogočajo, da je lahko isti predavatelj prisoten v dveh ali več predavalnicah hkrati.

Ključne besede: multimedija, telekonferenca, videokonferenca, izobraževanje, multimedijske tehnologije, izobraževanje v oblaku

Abstract

Education is one of the areas which has changed significantly in the 21st century due to technology development. This induced changes in its nature and gained new meaning to education. Among other areas, where we can also detect great progress in development, information technology is increasingly penetrating into schools and education sector. Its added value is mostly in offering new methods in learning and in the education process. Another valid two arguments for easier and faster involvement of technology into educational facilities are its widespread availability and adoptability. In this paper we wanted to point out several examples of multimedia technologies, which can help significantly in the process of education and in motivation for students.

Internet offers limitless possibilities of communication and interwork and is also the foundation for the technology which enables transmitting of sound, images and multimedia content. Bridging the gap in distance is one of the advantages offered by this technology, which is crucial in many cases, because it constitutes in saving time and money.

Keywords: multimedia, tele-conference, video-conference, education, multimedia technology, cloud education.

Medpredmentno povezovanje štirih slovenskih gimnazij pri tematizaciji alkohola z uporabo IKT

Application of Information Technologies in an Interdisciplinary Approach of Four Slovenian High Schools to the Theme of Alcohol

Nika Cebin
Rozmanova 19, 1330 Kočevje
nika.cebin@gmail.com

Maja Lebar Bajec
Gimnazijska cesta 9, 1420 Trbovlje
maja.lebar@guest.arnes.si

Branka Klemenčič
Ulica padlih borcev 7, 8340 Črnomelj
branka klemenčič@gimnm.org

Peter Prhavc
Bezenškova 16, 1000 Ljubljana
peter.prhavc@gmail.com

Tomaž Šmid
Valjavčeva 4, 4000 Kranj
tomaz.smid@guest.arnes.si

Povzetek

V prispevku je opisan skupni projekt Gimnazije Ledina in še treh slovenskih gimnazij, ki je bil osredotočen na danes precej aktualne teme, alkohol, njegovo uporabo in zlorabo, posledice pitja za posameznika in njegovo zdravje ter vpliv pitja alkoholnih pijač na širšo družbo. Kot izhodišče projekta so bila uporabljena književna dela. Izbrani avtorji so iz različnih časovnih obdobij in prostorov, saj je bil eden od namenov projekta osvetliti tematiko iz različnih zornih kotov in v različnih kontekstih ter pokazati, da tema pitja alkohola pri nas in po svetu obstaja že stoletja. Dijaki so morali po prebrani knjižni predlogi opraviti različne naloge, za kar so bili tudi ocenjeni. Literarna dela so služila le kot izhodišče, skupaj z angleščino so bila povezovalni element z drugimi predmeti: s psihologijo, kemijo in nemščino, česar brez podpore informacijske tehnologije ne bi bilo moč storiti. Za to je služila povezava na spletu preko e-učilnice, foruma, interaktivnega sestavljanja slovarčka, uporaba gradiv z medmrežja, interaktivna tabla in videokonference za povezovanje med šolami.

Ključne besede: Alkohol, književnost, medpredmetno povezovanje, timsko poučevanje, videokonferenca, digitalne tehnologije.

Abstract

In this article we are describing a joint project of Ledina high school and three other Slovenian high schools, focusing on current issues of alcohol, its use and abuse, the consequences of both on an individual, their health and also general public. The basis of the project are numerous literary texts, dating from different periods and describing diverse societies with a view to highlighting the problem from many aspects, thus showing that the problem in question has been with us for centuries already. After having read the texts, the students completed numerous tasks, which were also assessed. Literary works and the English language acted as a common thread between Psychology, Chemistry and German. All that was enabled by means of information technologies in the form of e-classrooms, forums, interactive dictionaries, interactive board and videoconferences.

Keywords: Alcohol, literature, interdisciplinary learning, team teaching, videoconference, digital technologies.

Razširjenost spletnih učilnic in dostopnost e-gradiv na spletnih straneh slovenskih osnovnih šol

The Prevalence of Online Classrooms and Accessibility of E-Materials on the Websites of Slovenian Primary Schools

Stanka Damjan
OŠ Kamnica
stanka.damjan4@gmail.com

Povzetek

V okviru razvoja e-izobraževanja na OŠ Kamnica smo želeli ugotoviti, kakšno je stanje na drugih osnovnih šolah v Sloveniji. Pregledali smo spletne strani slovenskih osnovnih šol (N=450) z namenom, da ugotovimo, koliko spletnih učilnic na šolah že deluje in ali so e-gradiva na njihovih spletnih straneh prosto dostopna. Raziskavo smo opravili v aprilu 2011. Ugotovili smo, da ima že skoraj polovica slovenskih šol spletne učilnice (195), od tega ima 72 šol e- gradiva prosto pristopna. Vsebine ostalih so dostopne z gesli. Razširjenost kaže na stalno povečevanje rasti in uporabe e– gradiv v vzgojno izobraževalne namene.

Ključne besede: IKT pri pouku, e gradiva, spletne učilnice.

Abstract

Within the development of e-learning at primary school Kamnica we wanted to determine the situation in other elementary schools in Slovenia. We have reviewed the websites of the Slovenian elementary schools (N = 450) in order to figure out how many online classrooms at schools have already been operating and whether e-learning materials on their websites have been freely accessible. The research was conducted in April 2011. We found that almost half of Slovenian schools already have online classroom (195), of which 72 schools have e-learning materials freely accessible. Other contents are accessible by password. The prevalence points to the continued growth and increasing use of e-learning materials in the educational purposes.

Keywords: ICT in the classroom, e material, online classroom.

Ali so hospitacije učiteljem nujno zlo?

Are the Teacher's Observations Necessarily a Bad Option?

Nevenka Dolenc
Srednja šola Jesenice
nevenka.osredkar@gmail.com

Povzetek

V prispevku predstavljam odnos učiteljev Srednje šole Jesenice (SŠJ) do hospitacij. Pri večini anketiranih (19 od 61 učiteljev) v kolektivu ima hospitacija negativen prizvok. Za nekatere je celo kot nujno zlo, ki povzroča velik stres. Večina se ne zaveda osnovnega pomena hospitacij, ki ni kontrola in iskanje napak pri posameznem profesorju, ampak imajo hospitacije bistveno globlji pomen.

Ključne besede: učitelj, ravnatelj, hospitiranje učiteljev

Abstract

This paper illustrates the teacher′s relationship to classroom observations at the Secondary School Jesenice (SSJ). The majority of the teachers who responded to the questionnaire (19 out of 61 teachers) had a negative connotation. For some, even as a necessarily bad option, this causes great stress. Most teachers don′t understand the basic importance of observations, which is not controlling and looking for mistakes of the individual teacher, but they have much deeper meaning.

Keywords: teacher, principal, classroom observations

Didatkični vidiki informatizacije slovenskega šolstva

Educational Aspects of Informatization Slovenian Education System

Ivan Gerlič
Univerza v Mariboru, Fakulteta za naravoslovje in matematiko
ivan.gerlic@uni-mb.si

Povzetek

Računalnik oz. informacijsko-komunikacijska tehnologija (IKT) že imata svoje pomembno mesto v slovenskem izobraževalnem sistemu, v nekaterih šolah in izobraževalnih stopnjah in usmeritvah bolj, v drugih nekoliko manj uspešno. Mnogo je strokovnih in znanstvenih diskusij o možnostih sodobnejšega in kvalitetnejšega pouka in poučevanja, ki ga ta tehnologija omogoča, o možnostih uspešnejše individualizacije in diferenciacije, možnostih prehoda od pouka, ki temelji na pomnjenju obilice podatkov, k reševanju problemov, ki zahtevajo kreativno mišljenje in kot rezultat tudi takšno znanje. Za naš izobraževalni sistem je to izrednega pomena, vendar primanjkuje konkretnih lastnih raziskav, ki bi dejansko pokazale, kako se da ta pričakovanja doseči, kaj se ne da doseči, katere didaktične oblike in metode dela ter katera dodatna znanja pri učencih in učiteljih terja uporaba računalnikov oz. informacijsko-komunikacijske tehnologije pri pouku, če naj bi z njimi zares miselno in motivacijsko razgibali učence in se predvsem izognili morebitnim negativnim spremljevalnim učinkom. V prispevku je predstavljen krajši pregled zgodovine informatizacije slovenskega šolstva in s tem v zvezi primerjalni prikaz najnovejših rezultatov (šolsko leto 2009 in delno tudi 2011) raziskave »Stanje in trendi uporabe IKT v slovenskih osnovnih in srednjih šolah«, ki jo avtor izvaja že vrsto let za potrebe Ministrstva za šolstvo in šport. Predstavljen je povzetek rezultatov s področja didaktične problematike pouka Informatike in Računalništva ter širše uporabe računalniške in sodobne informacijske tehnologije v slovenskih osnovnih in srednjih šolah.

Ključne besede: IKT, izobraževanje, didaktična problematika e-izobraževanja, informatizacija šolstva.

Abstract

Information and communication technology (ICT) already have their important place in the Slovenian education system in some schools and educational levels and more choices, the other somewhat less successful. There are many professional and scientific discussions about the possibilities of a more modern and better quality teaching and learning, which enable this technology, the possibilities of individuation and differentiation, and the possibilities of transition from school-based memory wealth of information to solve problems that require creative thinking and as the result of such knowledge. For our educational system is this very important, but lack of concrete research that could actually be shown how to achieve these expectations, what cannot be attained, which didactic forms and methods of work and additional knowledge for students and teachers requires the use of computers or. ICT in the classroom if it is to them and really thought move student's motivation and especially avoid

any possible negative accompanying effects. In the paper is presented a short overview of the history of computerization of the Slovenian education and comparative picture of the latest results (school year 2009 and partly 2011) study "Status and trends of ICT in Slovenian primary and secondary schools", which the author carried out a number of years for the Ministry of Education and Sport. Presented is a summary of results from the field of didactic instruction, issues of information technology and computer technology and wider use of computer and modern information technology in Slovenian primary and secondary schools.

Keywords: ICT, education, didactics problems of e-education, computerization of educational system.

Oblikovanje odzivnih strategij za kakovostno izvedbo praktičnega izobraževevanja študentov višjih in strokovnih šol

Response Strategies to Enhance the Quality of Practical Education of Students at Vocational Colleges.

Branka Jarc Kovačič
Tehniški šolski center Kranj, Višja strokovna šola
branka.jarc-kovacic@guest.arnes.si

Branka Balantič
Tehniški šolski center Kranj, Višja strokovna šola
branka.balantic@guest.arnes.si

Povzetek

Pomemben del višješolskih študijskih programov predstavlja praktično izobraževanje študentov v delovnem procesu. V podjetjih študenti nadgradijo teoretične vsebine, ki jih pridobijo v šoli, dobijo vpogled v realne delovne razmere, pogoje dela ter različne delovne naloge. Uspešno opravljeno praktično izobraževanje študentom predstavlja priložnost za kasnejšo zaposlitev. Prispevek uvodoma na kratko povzema praktično izobraževanje študentov na Višji strokovni šol, TŠC Kranj ter kompetence, ki naj bi jih študent usvojil na PRI. Poudarek je na spremljanju in merjenju zadovoljstva udeležencev z organizacijo in izvedbo praktičnega izobraževanja v delovnem procesu. Ugotovitve spremljanja so povzete v SWOT analizi in nadgrajene s TOWS matriko. Prepoznane medsebojne relacije šoli predstavljajo izhodišče za oblikovanje odzivne strategije za še kakovostnejšo izvedbo praktičnega izobraževanja študentov.

Ključne besede: Višje strokovno izobraževanje, praktično izobraževanje, kakovost, SWOT, TOWS

Abstract

Practical education in working process presents an important part of vocational programmes of higher education. Working in enterprises enables students to upgrade theoretical content gained in school, get inside of a real working situations, conditions and different working assignments. Successfully finished practical training enhances students' possibilities for later employment. The article in the beginning shortly describes practical training on Vocational college, Kranj school centre for technical sciences and skills which are gained during trainings on practical education. Emphasis is made on measuring participants' satisfaction with organization and implementation of practical education in a working process. Measurement conclusions are summed in SWOT analysis and upgraded with TOWS matrix. Relations which were identified presented starting point for school to manage response strategy to further enhance quality of students practical education.

Keywords: Vocational education, practical education, quality, SWOT, TOWS

Internet in medpredmetno povezovanje pouka angleščine in francoščine

The Internet and a Liaison of English and French Language Teaching

Helena Jošt
Jezerska cesta 106, 4000 Kranj
helena.jost@guest.arnes.si

Povzetek

Informacijska tehnologija s številnimi možnostmi učiteljem omogoča, da pouk bolj kot kdajkoli prej posnema resnično življenje in je s tem v večji meri usmerjen v usposabljanje za spopadanje z dejanskimi življenjskimi situacijami. Internet na primer omogoča bolj dinamičen, samostojen pouk, pa tudi neposredno medpredmetno povezavo pouka dveh tujih jezikov, angleščine in francoščine, s tem da eden od njiju postane orodje za učenje drugega jezika. Kljub določenim oviram in pomanjkljivostim ustvarjajo internet in nove metode poučevanja pogoje, zaradi katerih se lahko učitelj približa novi vlogi, ki mu jo narekujejo potrebe sodobne družbe.

Ključne besede: internet, medpredmetna povezava angleščine in francoščine,vloga učitelja

Abstract

Multiple possibilities of information technology enable teachers to make class more than ever before become an imitation of real life and thus to a greater extent oriented to the training to tackle actual life situations. For instance, the internet ensures more dynamic and independent teaching, as well as a direct liaison of two foreign languages teaching, English and French, during which one of the languages becomes a tool for learning the other language. Despite certain obstacles and drawbacks, the internet and the new teaching techniques create conditions directing the teacher closer to his new role being imposed by the demands of the modern society.

Keywords: internet, liaison of English and French language, role of the teacher

Wikispaces – uporabna spletna mesta

Wikispaces – Useful Web Sites

Alenka Jurančič
Osnovna šola Matije Valjavca Preddvor
Šolska ulica 9, 4205 Preddvor
alenka.jurancic@guest.arnes.si

Povzetek

Pri pouku vedno pogosteje uporabljamo informacijsko-komunikacijsko tehnologijo. Medmrežja ne uporabljamo le za iskanje informacij, ampak tudi za objavo vsebin, za kar potrebujemo primerno orodje. V prispevku bo prikazano, kako v nekaj korakih ustvarimo spletno mesto wikispaces in kaj vse nam omogoča. Uporabnost takega spletnega mesta smo že preizkusili v mednarodnih projektih in v okviru tehniških dni.

Ključne besede: wiki, wikispaces, projekti

Abstract

ICT technology is becoming more and more useful in education. Internet is not used for searching information only but for publishing as well. In order to do this we need certain web tools. This article deals with the creation of wikispaces web site in a few steps and the possibilities of it. The useful of a web site was tested in our international projects and on technology days.

Keywords: wiki, wikispaces, projects

Uporabnost IKT pri pouku Slovenščine v 1. razredu Osnovne šole: nadaljujmo zgodbo

The Usefulness of ICT in Teaching the Mother Tongue (Slovene) in the First Class of Primary School: Let's Continue the Story'

Bernarda Juvan
OŠ Franceta Prešerna Kranj
bernarda.juvan@guest.arnes.si

Povzetek

Predmetu slovenščina je v prvem razredu namenjenih 210 ur na leto (tj. 6 ur tedensko). Pri tem dajemo velik poudarek učenčevemu govoru, razvijanju sporazumevalnih zmožnosti, poslušanju, branju in pisanju. Za aktivno in ustvarjalno poučevanje slovenščine si pomagamo tudi z informacijsko komunikacijsko tehnologijo. Z učenci smo pripravili govorni nastop, kjer so učenci po poslušanju sami nadaljevali zgodbo in jo ilustrirali. Učence sem želela še dodatno motivirati, zato smo pri tem delu uporabili informacijsko komunikacijsko tehnologijo. V prispevku vam bom predstavila primer dobre prakse pri poučevanju slovenščine v prvem razredu, in sicer nadaljevanje zgodbe, ki so jo narisali učenci v nekaj slikah ter njen prenos na računalnik, tako da je nastala digitalna zgodba.

Ključne besede: informacijsko komunikacijska tehnologija, slovenščina, govor, poslušanje, branje, pisanje, zgodba, govorni nastop.

Abstract

In this article I would like to present an example of a good praxis in teaching the mother tongue. The pupils have to continue the story by drawing pictures. Then we download the pictures to a computer and make a digital story. In the first class of primary school pupils attend 210 lessons (6 lessons per week) of Slovene language. In the lessons the main focus is on pupils' speech and on developing the communicative skills, such as listening, reading and writing. For active and creative teaching we use ICT. The pupils have to speak in front of the others, they have to listen to them and continue the story by illustrating it. We additionally motivate the pupils by using ICT. In this article I try to present a case of a good praxis in teaching the Slovene language.

Keywords: information-communication technology (ICT), Slovene language, speech, listening, reading, writing a story, speaking in front of the others.

E-gradiva za okolje in trajnostni razvoj za interaktivno poučevanje okoljskih vsebin in vsebin iz področja trajnosti

E-Materials for the Environment and Sustainable Development for Environmental Interacitve Teaching Content and Content of Sustainability

Mojca Kokot Krajnc,
Ana Vovk Korže, Nina Globovnik,
Jerneja Križan, Nina Hriberšek
Mednarnodni center za ekoremediacije
Filozofska fakulteta, Univerza v Mariboru
mojca.kokot@uni-mb.si

Povzetek

V slovenskem šolstvu se v zadnjem času dogajajo številne novosti, ki so povezane z učnimi metodami in pristopi pri pouku kot tudi izbirnostjo predmetov. V šolstvu se velik poudarek namenja izobraževanju za okolje in trajnostni razvoj, s pomočjo katerega želimo mlade ljudi izobraziti in nenazadnje vzgojiti v odgovrne ljudi, ki bodo znali ceniti pomen in vse funkcije naravnih virov, ki nas obdajajo. Pri takšni vzgoji se daje pomen predsvem interkativnemu učenju in izkustvenemu učenju, s pomočjo katerega učenci sami na bolj predstavljiv in inovativen način prihajajo sami do novih spoznanj, na podlagi katerih sami razvijajo svoj odgovoren odnos do okolja, vendar za to potrebujejo tudi privlačna učna gradiva, ki jih v slovenskem šolstvu na področju okolja primanjkuje. V ta namen smo na Filozofski fakulteti Univerze v Mariboru v okviru Mednarodnega centra za ekoremediacije razvili e-gradiva za Okolje in trajnostni razvoj. S pomočjo e-gradiv bodo udeleženci izobraževalnega procesa tako imeli možnost poglabljati in razširjati svoje znanje o okoljski problematiki na inovativen način, ki zagotavlja preplet klasičnih oblik pouka z delom v skupinah, individualnim delom, delom v paru, hkrati pa e-gradiva zagotavljajo tudi uporabo inovativnih metod dela, kot so delo na terenu, demonstracija, aktualizacija, razgovor, delo s tekstom.

Ključne besede: e-gradiva, okolje, trajnostni razvoj, inovativno učenje, izkustveno učenje

Abstract

In the Slovenian education system there is recently undergoing a number of innovations, which are related to teaching methods and approaches in the classroom as well as with elective subjects. In education system there is a large emphasis on education for environment and sustainable development, through which we educate young people and ultimately raise the responsibility of people, who will appreciate the importance and features of natural resources that surround us. In such education the importance is given to the interactive and experiential learning through which students in a more innovative and imaginable way come to new knowledge on which, they develop their own environmental responsibility. For this they need an attractive learning materials about environment, but those materials are lacking

in Slovenian education system. For this purpose, have we at the University of Maribor in part of the International Centre for ecoremediation, developed e-learning materials for environment and sustainable development. With the help of e-materials training process, participants will have the ability to deepen and extend their knowledge of environmental issues in an innovative way, providing mix classical forms of instruction by working in groups, individual work, work in pairs. Also providing e-learning materials, help use the innovative methods of work, such as field work, demonstrations, updating, interview and working with text.

Keywords: E-materials, environment, sustainable development, innovative learning, experimental learning

Razvijanje digitalne spretnosti učencev v tretjem vzgojno – izobraževalnem obdobju osnovne šole

Training for Digital Skills of Pupils in the third Triad of Elementary School

Olga Koplan
OŠ Ivana Groharja
Podlubnik 1, 4220 Škofja Loka
olga.koplan@guest.arnes.si

Povzetek

Informacijska pismenost je pomembna ne le za posameznike, da znajo informacije najti, pravilno izbirati, vrednotiti in pravilno uporabiti, ampak tudi za gospodarsko konkurenčnost držav. Najdaljšo tradicijo na področju razvijanja in praktičnega izvajanja informacijskega opismenjevanja imajo ZDA. Njihovi standardi so zgled in dajejo smernice državam po svetu. O pomenu in razvoju informacijske pismenosti govorijo številni slovenski uradni dokumenti. Program informacijskega opismenjevanja na osnovnih šolah izvajajo šolski knjižničarji in učitelji drugih predmetnih področij. Gre za učno obliko, kjer je učitelj pobudnik, usmerjevalec in animator. Učenec je v vzgojno-izobraževalnem procesu bolj aktiven, saj informacije zbira, analizira, vrednoti in teoretična znanja preoblikuje v praktična. V članku predstavljam vlogo in izvedbo medpredmetnega povezovanja v tretjem vzgojno-izobraževalnem obdobju osnovne šole, primer učenja pisanja pisnih izdelkov po mednarodnih standardih ISO in poenotenje v navodilih učencem na ravni šole.

Ključne besede: informacijska pismenost, e-izobraževanje, osnovna šola, medpredmetno povezovanje, informacijska tehnologija, mednarodni standardi ISO

Abstract

Digital literacy helps individuals in proper and efficient information management, e.g. it enhances finding, selection, evaluation and use of information. On the other hand the digital literacy also supports competitiveness of national economies. The United States have the longest tradition in systematic digital literacy development and implementation. Their standards have were been accepted at least as the guidelines in many countries all over the world. The importance and development of the digital literacy has been emphasized also in several Slovenian official documents. The digital literacy programmes in the elementary schools are promoted by librarians and also by other subjects teachers. The teacher plays the roles of initiator, guide and facilitator. The pupil takes a more active role in the learning process by collection, analysis and evaluation of information and by transformation of theoretical knowledge into the practical work.

This contribution presents the role and realization of the cross-curricular learning in the third triad of elementary school, a sample lesson to prepare a written article according to international ISO - standards and the unification of the instructions on the level of the school as a whole.

Keywords: digital literacy, e-education, elementary school, cross - curricular learning, information technology, ISO - standards

Z IKT do kakovostnega strokovnega sodelovanja med šolami - Projekt Dobre vesti iz naše šole in našega mesta

With ICT to Quality Professional Cooperation between Schools - Project Good News from Our School and Our Town

Ljubica Kosmač
Osnovna šola Ledina
Komenskega ulica 19, 1000 Ljubljana
ljubica.kosmac@guest.arnes.si

Povzetek

Projekt Dobre vesti iz naše šole in našega mesta, ki ga vodi Osnovna šola Ledina iz Ljubljane, med sodelujočimi osnovnimi šolami iz Slovenije in tujine že vrsto let s pomočjo informacijske komunikacijske tehnologije (IKT) vzpostavlja kakovostno strokovno sodelovanje, ki je dodana vrednost šolskemu delu. V pričujočem članku predstavljamo tako cilje projekta kot tudi načine dela, ki nam jih za uspešno sodelovalno delo med šolami omogoča IKT. Projekt se je v desetih letih svojega obstoja nadgrajeval, skladno z razvojem na tem področju, njegova rast in njegova uspešnost pa je bila sad vzajemnega dela mentorjev z vseh sodelujočih šol. Že samo desetletno delovanje projekta je dovolj velik pokazatelj, da so učitelji v njem s kakovostnim delom ustvarili nadstandardno obliko dela, ki jih izpolnjuje in bogati njihovo osnovno delo pri pouku. V prispevku predstavljamo sodelovalno delo med šolami kot primer dobre prakse medšolskega povezovanja in sodelovanja na strokovnem področju slovenskega jezika.

Ključne besede: informacijska komunikacijska tehnologija, projekt, sodelovalno delo

Abstract

Project Good news from our school and our town, that has been led from Primary school Ledina, Ljubljana, between cooperating schools from Slovenia and abroad with information and communication technology (ICT). It established quality professional cooperation and became a special added value to school work. In the article we present the goals and how we use ICT for successful cooperation between schools. The project upgraded during ten years of its existence, consistent with development in this area. Its growth and its success was the result of interactive work of mentors from all cooperated schools. Ten years long operation of the project is a big enough indicator, that the cooperating teachers have created a superior way of work that fulfils them and enrichens their basic work in Education. In the article we present cooperating work between the schools as an example of good practice of integration and cooperation between schools on the professional field of slovenian language.

Keywords: information and communication technology, project, cooperated work

Informacijska pismenost v visokem šolstvu v Sloveniji

Information Literacy in Higher Education in Slovenia

Mirjam Kotar
mirjam.kotar@fdv.uni-lj.si

Silva Novljan
silva.novljan@telemach.net

Karmen Stopar
karmen.stopar@bf.uni-lj.si

Povzetek

Prispevek obravnava pomen učinkovite in etične uporabe informacij, ki predstavlja temelj za vseživljenjsko učenje. Izobraževanje za uporabo informacij – razvijanje informacijske pismenosti – je danes sestavni del sposobnosti in veščin, ki jih razvijamo vse življenje, pri čemer sta temeljni opori njenega razvijanja izobraževalni sistem in knjižnice. Izpostavljena vloga visokošolskih ustanov ustreza njihovi odgovornosti za razvijanje sposobnosti in znanja pri skupini, ki to znanje v največjem obsegu prenaša v okolje. Informacijsko pismen študent je že nekaj časa strateški cilj na univerzah po svetu, ki se zavedajo, da informacijska pismenost izboljšuje kakovost izobraževalne in raziskovalne dejavnosti na univerzah in prispeva k uspešnosti diplomantov. Rezultati raziskave kažejo, da se v Sloveniji koncept informacijske pismenosti v visokem šolstvu razvija prepočasi, nesistematično in fakultativno. Ob analizi dosedanjih aktivnosti za uveljavitev informacijskega opismenjevanja prispevek predstavi dokument Merila in kazalci informacijske pismenosti v visokem šolstvu. V nadaljevanju pa predstavi možnosti za nadaljnji strateški razvoj in sistemsko uveljavljanje programov informacijske pismenosti v izobraževalnem procesu.

Ključne besede: informacijska pismenost, izobraževanje, standardi, informacije, študenti, knjižnice, knjižničarji, visoko šolstvo, univerze, študijski proces, Slovenija

Abstract

The article aims to discuss the importance of efficient and ethical use of information as the basis of lifelong learning. The education for the use of information – advancing of information literacy – is now the integral part of competences and skills developing throughout life with the basic support of development stemming from the system of education and libraries. The exposed role of higher education institutions complies with their responsibility to develop competences and knowledge of the population group which tends to disseminate such knowledge to the fullest extent into the environment. For quite some time, educating an information literate student has been one of the strategic goals of universities around the world, especially those being aware that the information literacy improves the quality of educational and research activities in universities and contributes to graduates' success. The survey results report that in Slovenia, the concept of information literacy in

higher education is developing rather too slowly, unsystematically and optionally. Along with the analysis of current activities in terms of implementing the information literacy, the article also aims to present the document entitled Criteria and indicators of information literacy in higher education. It concludes with presenting the options for further strategic development and systematic implementation of information literacy programmes in the system of education.

Keywords: information literacy, education, standards, information, students, libraries, librarians, higher education, universities, Slovenia

Uporaba e-pošte in Power Point predstavitev pri projektnem učnem delu v 5. razredu

The Use of E-Mail and Power Point Presentation on Project Based Learning in the5th Class

Slavka Kozel
OŠ Trzin, Mengeška 7b, 1236 Trzin
slavka.kozel@gmail.com

Povzetek

Ta članek prikazuje vključitev projektnega učnega dela pri poučevanju v 5. razredu osnovne šole s pomočjo e-pošte in računalniškega programa Microsoft Power Point. V članku so prikazane faze projektnega učnega dela od uvodne motivacije do zaključne predstavitve z modno revijo. Takšno učenje se je izkazalo za učinkovito, zanimivo, kreativno in vzpodbudno za nove izzive. Učenci so pridobili spretnosti za samostojno uporabo elektronskih medijev, pripravo podatkov in komunikacijo.

Ključne besede: Microsoft Power Point, e-mail, projektno učno delo, plastične vrečke, modne kreacije, modna revija.

Abstract

This article presents the integration of Project based learning in teaching in the 5th class of elementary school through e-mail and computer program Microsoft Power Point. The article shows phases of Project based learning from the initial motivation for the final presentation with a fashion show. Such learning has proven to be effective, interesting, creative and stimulating for new challenges. Pupils have acquired skills for independent use of electronic media, preparation of information and communication.

Keywords: Microsoft Power Point, e-mail, Project based learning, plastic bags, fashion creation, fashion show.

Učenje programiranja s programom Scratch

Learning Programming with Scratch

Sonja Lajovic
Oš Kolezija
sonja.lajovic@gmail.com

Povzetek

Učitelji računalništva se velikokrat sprašujemo, kateri programski jezik je najprimernejši za učenje programiranja v osnovni šoli. Od leta 2007, ko je bil narejen programski jezik Scratch, je odgovor na to vprašanje precej lažji. Scratch je nastal namensko, prav za učenje programiranja. Otroci lahko z njim programirajo lastne računalniške igre, izdelujejo animacije ali interaktivne zgodbe. Program je novejši, zato ni na voljo dovolj literature o tem, na kakšen način poučevati, kako združiti prijetno s koristnim, da bi se otroci naučili osnovnih pravil računalniškega programiranja, hkrati pa delali računalniške igre, kar večino pri delu z računalnikom najbolj veseli. V prispevku bodo predstavljeni možni načini poučevanja programiranja s programom Scratch. Eden je poučevanje s programiranjem iger, drugi način pa je bolj podoben klasičnemu, sistematičnemu učenju posameznih programskih konceptov, podobno kot pri programu Logo. Prikazanih bo nekaj primerjav z drugimi programskimi jeziki ter izkušnje s poučevanjem Scratcha. Na kratko bodo predstavljene tudi osnovne značilnosti programa Scratch.

Ključne besede: Scratch, učenje, programiranje, igre, naloge

Abstract

IT teachers often ask themselves which programming language to use when teaching programming in primary school. The programming language Scratch, which started to be used in 2007, has helped to answer the question. Scratch was created to be used to teach programming. Children can use it to create their own computer games; they can make animations or interactive stories. Because the programming language is relatively new, not much has been written about how to use it effectively, so that the children would, on the one hand, learn the basics of programming and, on the other hand, learn to create their own games, which they usually enjoy the most. This article presents different ways of how the program Scratch can be used. One can use it for programming games but also in a more traditional way – to systematically learn concepts pertaining to programming like in the programming language Logo. In this article, Scratch will be compared to some other programming languages; furthermore, examples of its usage in teaching programming will be presented. Main characteristics of Scratch will also be presented.

Keywords: Scratch, teaching, programming, games, exercise

Astrologija za hec – prevzete besede pa za res! Obravnava prevzetih besed z IKT-tehnologijo

Astrology for Fun – Acquired Words Seriously! Application of Acquired Words with ICT Technology

Sanja Leben Jazbec
OŠ Solkan
Šolska ulica 25, 5250 Solkan
sanja.leben-jazbec@guest.arnes.si

Povzetek

Današnji učenci so t.i. spletna generacija: so zelo spretni uporabniki sodobnih digitalnih orodij, kar velja izkoristiti tudi pri pouku za doseganje ciljev. Nujno je, da učitelji modro posegamo po uporabi IKT, izkoristimo prednosti, ki jih tehnologija ponuja, s premišljenim načrtovanjem pa poskušamo minimalizirati eventuelne pomanjkljivosti. V prispevku predstavljamo, kako učenci 8. razreda osnovne šole usvajajo in utrjujejo znanje prevzetih besed s pomočjo IKT-tehnologije: uporabljajo spletne slovarje, rešujejo in sestavljajo interaktivne vaje s programom Hot Potatoes, izdelujejo slovar in komunicirajo na forumu (vse v spletni učilnici v okolju moodle), uporabljajo splet.

Ključne besede: prevzete besede, elektronski priročniki, Hot Potatoes, spletna učilnica v okolju Moodle, svetovni splet

Abstract

Today, at the beginning of the second decade of the twenty-first century, educators should no longer question themselves, ICT - yes or no?, but our basic dilemma should be: ICT - When? Where? How? To what extent? Today's students, the so-called web generation (net-generation), are very adept users of modern digital tools, which can also be exploited in class to achieve objectives. It is imperative that teachers use ICT wisely, exploit the advantages that technology offers and try to minimize the eventual shortcomings with well considered planning. The paper presents how 8th grade students learn and consolidate their knowledge of acquired words during their Slovenian lessons with the help of ICT-technologies:

- By using online dictionaries (e-SSKJ, e-SP, Nova beseda).
- By solving and composing interactive exercises using the program Hot Potatoes.
- They learn by making a dictionary as well (in the online classroom in Moodle), searching for the right words for personality traits.
- They communicate with each other and the teacher on the forum (the online classroom in Moodle).

They use the World Wide Web to search for information and learn to be critical of the flood of information on the web. Pupils are brought up with the idea that the Internet in itself is not a bad thing, that the huge mass of information can be put to good use; but that it is very wrong to uncritically accept anything and everything found on the web. Students are reminded to use standard language and to abide netiquette.

Keywords: acquired words, e-books, Hot Potatoes, online classroom in Moodle, the World Wide Web

6 brezplačnikov za pouk slovenščine

6 Free Programs for Teaching Slovene

Tatjana Lotrič Komac
OŠ Naklo
Glavna cesta 47, 4202 Naklo
tatjana.lotric@guest.arnes.si

Tina Žagar Pernar
OŠ Naklo
Glavna cesta 47, 4202 Naklo
tina.zagar1@guest.arnes.si

Povzetek

Informacijsko-komunikacijska tehnologija v razredu sama po sebi še ne spodbuja učenja, zato je ključno, kako jo učitelji znamo in zmoremo vključiti v učni proces. Prispevek se zato osredotoča na to, da bi vse dejavnosti, povezane z uporabo IKT, ponujale možnost interakcije, delovale motivacijsko in bi bile hkrati dobro didaktično premišljene in strokovne. V prispevku je tako nanizanih nekaj brezplačnih, prosto dostopnih, enostavnih programov, katerih prostočasna funkcija je nadgrajena z učno in didaktično vlogo. Predstavljena je njihova uporaba v različnih fazah učnega procesa, prav tako so nakazane ideje, kako te programe vključiti v obravnavo umetnostnih in neumetnostnih besedil ter jih uporabiti tudi npr. pri urah oddelčne skupnosti, novinarskega krožka, multimedijske delavnice, sodelovalnega dela med učitelji.

Ključne besede: brezplačni programi, slovenščina, interaktivna učna gradiva, ideje za poučevanje

Abstract

The use of ICT in the classroom is not guarantee on its own. It is very important that the teacher implements ICT in learning process reasonably and efficiently. Paper is focused on ICT activities that give possibility of interaction, are motivating and therefore didactically well considered and professional. Authors present some simple free online programs, that are basically used for fun, but in this case upgraded with learning and didactical function. Their role is introduced in different phases of learning process, what is more, the article shows ideas how to integrate them in reading and analysis of fiction or nonfiction texts and also in other segments of learning and teaching process, such as class teachers' lessons, journalistic club, multimedia club, teachers' cooperative work.

Keywords: free programs, Slovene language, interactive learning resources, ideas for teaching

Modernizem napove vojno klasicizmu angleškega prevajanja s stališča učencev

Modernism Declares War against Classicism of English Translation from Pupils' Point of View

Mirjana Meško
Osnovna šola Ormož
mirjana.mesko@gmail.com

Povzetek

Tisti, ki naše učence ženemo naprej, smo mi, učitelji. Vendar, pri tem se nikoli ne smemo riniti v ospredje. Ko naši učenci nastopajo na odru smo učitelji tisti, ki stojimo tik za zaveso in njihovim dosežkom ploskamo z največjo frekvenco. V tem kontekstu želim pomagati svojim učencem preplezati zid, pred katerega so postavljeni vedno, ko želijo zapisati besedilo v angleščini z enako vnemo in bogatim besediščem, kot to zmorejo v maternem jeziku. Orodje, ki jim pri tem pomaga, je slovar. V želji, da bi moji učenci premagali sami sebe, sem jim zastavila nalogo. Učence sem glede na njihovo predznanje razdelila na različne ravni, ki so prevajale odlomek iz knjige M. Twaina Prigode Toma Sawerja. Polovica učencev je odlomek v angleščino prevajala s pomočjo tiskanega slovarja, druga polovica pa je pri tem uporabila IKT sredstvo – spletni prevajalnik. Rezultati kažejo, da modernizem zmaguje v vojni, napovedani klasicizmu. Članek navaja prednosti uporabe spletnega slovarja v primerjavi s tiskanim slovarjem.

Ključne besede: angleški jezik, prevajanje, osnovna šola, spletni slovar, tiskani slovar

Abstract

The thing that drives our pupils forward is us – the teachers. But we mustn't put ourselves in the forefront. When our pupils act on stage, teachers are the ones that stand behind the curtain applauding with the highest frequency. In this context I wish to aid my pupils climb the wall which they face every time they want to write in the English language with the same amount of enthusiasm and rich vocabulary as they are capable in their native language. The tool that gives them this ability is a dictionary. In desire, for my pupils to overcome themselves, I gave them a task. I divided the pupils into levels of knowledge and then a half of the pupils in each level had to translate a passage from Mark Twain's book The Adventures of Tom Sawyer into English language using a paper dictionary, the other half used an ICT tool – an online dictionary. The results show that modernism defeated classicism. The article states the advantages of using an online dictionary in comparison to a paper dictionary.

Key words: English language, translating, primary school, on-line dictionary, paper dictionary

Konflitki v šoli

Conflicts at School

Damjana Oblak
Ekonomska gimnazija in srednja šola Radovljica
Gorenjska cesta 13, 4240 Radovljica
damjana.oblak@gmail.com

Povzetek

Vedno več otrok in mladostnikov preživlja večino prostega časa v socialni izolaciji, pred televizijo in za računalnikom. Prikrajšani so za pomembne izkušnje in spretnosti v medsebojnih odnosih, ki naj bi jih pridobili v interakcijah z vrstniki in v družini. Strokovnjaki opisano povezujejo z razvojem neprimernih vedenjskih vzorcev mladih, ki v šoli vse pogosteje vodijo v konflikte, kar potrjujejo tudi številne raziskave. Na pojav konfliktov v šoli pa vplivajo tudi številni drugi dejavniki, kot so kvalitetna organizacija dela, šolska klima, jasen sistem vrednot v šoli in natančna ter učencem smiselna šolska pravila. Ena temeljnih veščin, ki naj bi jo učenci pridobili na tem področju, je sposobost za konflikt oz. sposobnost razreševanja konfliktov. Za pozitiven odnos v šoli sta odgovorna oba, učitelj in učenec, a učiteljeva odgovornost je večja. Učitelj naj ne želi spreminjati učencev, ampak naj se usmeri vase in naj bo pozoren na odnose, prepričanja in vedenjske vzorce, ki so na splošno avtomatizirani in trdno zakoreninjeni v naših izobraževalnih strukturah. To od učitelja zahteva, da v svoje delo vloži večjo stopnjo načrtnega zavedanja, da na novo premisli cilje in včasih opusti dolgo cenjene vrednote, ki nam danes ne služijo več. Ta proces vsebuje preokvirjanje koncepta discipline in preusmeritev iz kaznovalnih vedenj, ki poudarjajo napake učencev, v preventivna vedenja, ki gradijo odnose, ki se že vnaprej izogibajo konfliktom. Za skupno, ustvarjalno reševanje konfliktov je potrebno znati se pogovarjati, aktivno poslušati, se vživljati v položaj drugih, razumevati njihova čustva in biti pripravljeni na spremembe.

Ključne besede: šola, vzgoja, konflikti, komunikacija, strategije reševanja konfliktov

Abstract

More and more children and adolescents spend most of their free time in social isolation in front of television and computer. They are deprivated of important experiences and skills in their (inter)personal relationships, which should be obtainted in interactions with their peers and family members. The experts link the described situation with with development of unsuitable behaviour models of young people, which leads increasingly to conflicts at school, as also numerous studies have evidenced. On the occurrence of conflicts at school, however, affected also numerous other factors like quality work organization, school climate, clear system of values at school and accurate and meaningful school rules for pupils. One of the basic skills to be acquired by pupils in this area is capacity for conflict or ability to resolve conflicts. For a positive attitude at school both teacher and pupil are responsible, but a teacher's responsibility is greater. The teacher should not want to change the pupil, but only focus on him-/herself and be attentive to relations, beliefs and behavioral patterns that are generally automated and firmly rooted in our educational structures. This requires the teacher

to incorporate into his/her work a higher degree of intentional awareness that reflects the new objectives and sometimes abandons long appreciated values that are nowadays no longer in use. This process includes the concept of discipline and shift from punitive behaviors that emphasize pupils' errors in preventive behaviors that build relationships that already in advance avoid conflicts. For common and creative conflict resolution should someone be able to talk, to listen actively, to practice empathy in the situation of others, to understand other people's feelings and to be willing for changes.

Keywords: school, education, conflicts, communication, conflict-solving strategies

Odnos med učiteljem, učenci in starši

The Relationship between Teachers, Pupils and Parents

Damjana Oblak
Ekonomska gimnazija in srednja šola Radovljica
Gorenjska cesta 13, 4240 Radovljica
damjana.oblak@gmail.com

Povzetek

Vzajemno spoštovanje in zaupanje sta temelja pozitivnega razrednega vzdušja. Konstruktiven odnos med učiteljem in učenci je odnos medsebojnega dopolnjevanja in vključevanja ter ustvarjalnega dialoga med njimi, ki iščejo skupno in hkrati ohranjajo svojo individualno ter različno osebnost. Učiteljev odnos do učenca je usmerjen v spodbujanje odgovornosti, učitelj je pozitivno naravnan do učenca in mu skuša dajati vedenjske zglede. Zaupa učenčevi sposobnosti odločanja in mu prepušča odgovornost za lastno vedenje in posledice. Učitelj, ki načrtno podpira razvoj dobrih odnosov med otroki v razredu, organizira veliko medsebojne pomoči, sodelovalnega učenja in tudi neformalnih srečanj, odločilno prispeva k uspešnemu socialnemu razvoju otrok. Strokovnjaki pripisujejo velik pomen odkrivanju in prepoznavanju močnih področij učencev. Šolsko okolje (Magajna idr., 2008) predstavlja varovalni dejavnik za učence, če je stabilno in varno. Pomembni so ustrezni načini komunikacije in uspešno reševanje sporov. Učenci so deležni visoke ravni čustvene podpore, sprejetosti in pomoči. Šoli je potrebno dati 'družinski' predznak. To pomeni, da učencem ponudimo čustveno varno okolje, da upoštevamo njihova čustva in jih tudi sami pokažemo. Starši lahko učitelju nudijo pomembno podporo in podkrepitev in hkrati začutijo pristno povezanost z njim, mu zaupajo in verjamejo ter se v njegovi družbi dobro počutijo.

Ključne besede: šola, vzgoja, medosebni odnosi, komunikacija,učitelj kot zgled, močna področja učencev, partnerstvo med učiteljem in starši

Abstract

Mutual respect and trust are the cornerstones of a positive classroom atmosphere. A constructive relationship between teacher and pupils is a relationship of mutual complementarity and integration and constructive dialogue between them, looking for common, while maintaining their individual and different personality. Teacher attitudes towards students are aimed at promoting accountability, teacher is positively aimed to pupils and he/she tries to give behavioral examples. Teacher trusts pupil’s decision-making ability and he/she leaves the responsibility for their own behavior and consequences. The teacher who systematically supports the development of good relationships among children in the classroom, organizing a lot of mutual aid, cooperative learning, as well as informal meetings, a decisive contribution to the success of social development of children. Experts attach great importance to detect and identify pupils' strengths. School environment (Magajna et al., 2008) represents a protective factor for pupils, if it is a stable and secure. Important methods of communication are appropriate and effective dispute resolution. Pupils receive a high level of emotional support, acceptance and support. School should be given 'family' sign.

This means that pupils have a safe emotional environment to consider their feelings and show themselves. Parents can provide important support for teachers and also feel a genuine connection with the teacher they trust the teacher feel comfortable.

Keywords: school, education, interpersonal relations, communication, teacher as an example, student strengths, partnership between parents and the teacher

Socialna vključenost učencev z učnimi težavami z uporabo programa Microsoft Mouse Mischief

Social Integration of Pupils with Learning Disabilities using the Program Microsoft Mouse Mischief

Tina Pajnik
OŠ Vide Pregarc, Ljubljana
tina.pajnik@gmail.com

Sonja Strgar
OŠ Vide Pregarc, Ljubljana
sonja.strgar@gmail.com

Povzetek

Sodobna družba na eni strani načrtno razvija individualne pristope in diferenciacijo v pedagoških teorijah, v praksi pa kljub temu skozi prikriti kurikulum iz socialne mreže izključuje posameznike, ki so učno neuspešni ali pa imajo nek učni primanjkljaj. Socialni stiki so silno pomembni za mlajše učence, ki jim šola predstavlja prvo izkušnjo sekundarne socializacije. V želji, da bi se tudi socialno šibkejši vključili v mrežo pozitivnih socialnih stikov ter razvijali pozitivno samopodobo, imajo učitelji danes veliko orodij in pripomočkov, s katerimi lahko popestrijo pouk in skozi izobraževalne vsebine pletejo socialno mrežo. S tem namenom so učitelji na osnovni šoli Vide Pregarc v prvem triletju uporabili izobraževalno tehnologijo na različne načine. V tretjem razredu je bilo izvedenih deset ur z uporabo programa Microsoft Mouse Mischief, kjer so učenci v PowerPoint programu z opcijo Multiple Mouse reševali različne kvize v heterogenih skupinah, ki so omogočale vključevanje učencev z učnimi težavami.

Ključne besede: socialna vključenost z uporabo IKT, Microsoft Mouse Mischief, opcija Multiple Mouse

Abstract

Modern society planns development and differentiation of individual approaches in educational theory, on the other hand through the hidden curriculum excludes individuals, which have a learning disabilities. Social skills and social contacts are very important for younger pupils. For them school represents the first form of socialization. All children should have opportunity to successfully integrate into society and to develop a positive self-image. Teachers today have a lot of gadgets and tools that can enrich the teaching and facilitate the involvement of the group. Therefore, teachers in primary school Vide Pregarc in the first three classes used educational technology in order to improve the social class structure. The third class had 10 sessions, where the program Microsoft Mouse Mischief was used. Pupils solve quizzes with option Multiple Mouse in heterogeneous groups in order to include pupils with learning disabilities.

Keywords: social integration with educational technology, Microsoft Mouse Mischief, Option Multiple Mouse

Intuitivno spletno orodje za podporo procesom odločanja

Supporting the Decision Making Process using an Intuitive Web Based Solution

Taj Pelc

Frenk Ten Sedmak Nahtigal

Mirjana Kljajić Borštnar
mirjana.kljajic@fov.uni-mb.si
University of Maribor, Faculty of Organizational Sciences

Abstract

There are many decision methodologies and tools for supporting decision-making processes, for example AHP, DEX, system dynamics models and others. Although they are well researched in theory, they are not commonly used in everyday practice. There are few reasons for that: lack of methodological knowledge and misunderstanding of complex decision situations (conflicting criteria, misperception of feedback information, group decision-making etc.). In this paper we discuss the development of an intuitive web application, based on the multiple-criteria methodology in a collaborative social media environment, called ODESYS. A preliminary experiment was conducted to explicate the usefulness of the application's guidance of decision-making process. There were 26 subjects participating in the preliminary experiment to test the usefulness of the application. Their task was to solve the presented decision problem using the tool. After the experiment their opinions were gathered by a questionnaire. Results suggest that the perceived usefulness as well as the ease of use were rated high, which gives optimistic outlook for further work.

Keywords: decision-support systems, multi-criteria decision making, web system, collaborative environment

Odvisnost kakovosti vzgojno – izobraževalnega procesa v osnovni šoli od kompetenc operativnega managementa šole

Correlation between the Quality of Educational Process in Elementary School and Competences of Operative Management of a School

Polonca Petrica Ponikvar
polonapponikvar@gmail.com

Jože Florjančič
joze.florjancic@fov.uni-mb.si

Mojca Bernik
Univerza v Mariboru, Fakulteta za organizacijske vede
mojca.bernik@fov.uni-mb.si

Povzetek

Članek prikazuje problematiko današnjega izobraževalnega sistema na nivoju operativnega managementa na šolah. Z raziskavo med top managementom, učitelji – operativnimi managerji in uporabniki storitev – starši in učenci smo skušali priti do odgovora, katere kompetence so pomembne za uspešni operativni management na šolah in posledično za kvalitetno in konkurenčno slovensko šolo v evropskem merilu. Rezultat raziskave: najbolj pomembna so informacijska znanja, managerska znanja - inovativnost in ustvarjalnost ter komunikacijska znanja.

Ključne besede: operativni management, izobraževalni system, managerske kompetence, informacijske kompetence, komunikacijske kompetence

Abstract

This article deals with the problems of today's educational system concerning the operational management of a school. For the research we asked top managers, teachers – operational managers, parents and pupils what are the most important competences of a successful teacher – operational manager. The result of a survey research: the most important are informational competences, management competences - innovation and creativity and communication competences.

Keywords: operational management, educational system, management competences, informational competences, communication competences

Urejanje šolske prehrane z informacijsko komunikacijsko tehnologijo

Organization of School Meals with the Help of Information Communication Technology

Tatjana Podbevšek Sajevic
ESIC Kranj, Gradbeno-storitvena šola
Cesta Staneta Žagarja 33, 4000 KRANJ
tatjana.sajevic@siol.net

Povzetek

Današnji ritem in način življenja sta neposredno povezana z informacijsko-komunikacijsko tehnologijo (IKT). V svetu globalizacije in hitrih sprememb je IKT v šolah pomemben dejavnik družbenega razvoja. Prispevek obravnava vpletenost IKT v organizacijo šolske prehrane po spremembah, ki jih je prinesel novi Zakon o šolski prehrani. Predstavljena sta dva koncepta evidentiranja dijakov pri prijavljanju in odjavljanju od šolske malice: s čipom in preko spleta. Primerjava neodjavljenih in neprevzetih obrokov malice na obeh šolah pokaže na velike razlike med sistemoma evidentiranja v prid spletnega načina prijave in odjave malice. Kratka anketa med dijaki na ESIC Kranj, kjer za prijavo dijaki uporabljajo elektronski čip, za odjavljanje malice pa elektronsko pošto, pa je pokazala, da bi bilo odjavljanje doslednejše, če bi dijaki lahko uporabljali spletni način prijave in odjave.

Ključne besede: evidentiranje malice, neodjavljena malica, elektronski čip, svetovni splet, e-pošta

Abstract

The pace and way of life today is directly connected with the use of ICT. Because of globalization and fast world changes the use of ICT has become an important part of school social development. The article talks about the organization of school meals with the help of ICT, after the changes, introduced by the latest passed Law dealing with school meals. There are presented two ways of how to keep records of registrations and cancellations of students' meal orders: with the help of the micro chip or the internet. The comparison of the number of uncancelled or not picked up meals at both schools shows a big difference between the two systems of keeping the meal number records; in favour of the internet registration and cancellation. A brief survey among students of ESIC (School Centre of Economics and Trade, Kranj) Kranj, where the elektronic chip is used for registration and e-mail for cancellation, shows, that cancellation would be more consistent if students could use the internet for both - registration and cancellation.

Keywords: recording of school meal orders, uncancelled meals, the elektronic chip, the world wide web, e-mail

Izobraževanje v virtualnem svetu Second Life

Learning in Second Life Virtual World

Irena Podveržen
Šolski center Velenje
irena.podverzen@gmail.com

Dejan Dinevski
Univerza v Mariboru, Pedagoška fakulteta
dejan.dinevski@uni-mb.si

Povzetek

Virtualni svet, kot rečemo navideznim svetovom, je računalniško ustvarjeno okolje, ki živi na internetu. Takšna okolja so prikazana v tri ali dvodimenzionalni grafiki in so zelo podobna resničnemu svetu. V njih veljajo podobna pravila (gravitacija, gibanje, čas, komunikacija …), dogajajo se podobne stvari. Uporabniki v njem ustvarijo želeno podobo, se z njo predstavljajo drugim in z njimi komunicirajo. Čas se v virtualnem nikoli ne ustavi, kar pomeni, da je navidezni svet aktiven in dostopen 24 ur na dan, vsak dan. Virtualni svet, kot je Second Life, odpira nove možnosti in priložnosti na različnih področjih in eno od teh področij je tudi učenje in razvoj posameznika. Aplikacija predstavlja eno največjih socialnih mrež na svetu v 3D-prostoru, ki omogoča opravljanje raznovrstnih, predvsem pa koristnih nalog pri procesu izobraževanja in usposabljanja. Je nov in bolj privlačen način izobraževanja in usposabljanja ter ponuja veliko možnosti oziroma alternativo sedanjemu učenju na daljavo, predvsem pa predstavlja nov sistem, ki ga lahko univerze ponudijo kot dodatno izobraževanje. Second Life ne uporabljajo le univerze in druge izobraževalne ustanove, ampak jo za različne namene uporabljajo tudi podjetja. Nekatera se bolj posvečajo zgolj oglaševanju in promociji, nekatera pa izkoristijo vse možnosti, ki jih ponuja. Ena največjih prednosti uporabe je odstranitve geografskih in časovnih ovir, ki jih ima neko globalno podjetje.

Ključne besede: virtualni svet, Second Life, izobraževanje na daljo, priložnosti za izobraževanje, virtualna knjižnica

Abstract

Virtual world is a computer-generated environment, which resides on the Internet. Such environments are displayed in three or two-dimensional graphics and are very similar to the real world. They are subject to similar rules (gravity, motion, time, communication ...), similar things are happening. Users must create a desired avatar, representing it to others and communicate with them. Time never stops in a virtual world, which means that the virtual world is active and accessible 24 hours a day, every day. The virtual world like Second Life opens up new possibilities and opportunities in various fields and one of these areas is learning and development. The application is one of the largest social networks in the world in 3D-space, giving access to diverse, especially useful tasks in the process of education and training. It is a new and more attractive way of education and training, and offers many

options and alternatives to the present distance learning, in particular, a new system that can be offered as an additional education at a university. Second Life is not used only by universities and other educational institutions, but also for different purposes by companies. Some are more concerned with advertising and promotion, and some take advantage of the opportunities it offers. One of the biggest advantages is the removal of geographical and time barriers, which a global company has.

Keywords: Virtual world, Second Life, distance learning, learning opportunities, virtual library

Določanje težnega pospeška

Assessing Gravitional Acceleration

Neža Poljanc
OŠ Križe, Cesta Kokrškega odreda 16, Križe
neza.poljanc@gmail.com

Povzetek

Težni pospešek lahko določimo na več načinov. V učbeniku za 9. razred fizike je predlagano, naj ocenimo pospešek padanja pri poskusu z brnačem. Poskus smo izvedli z bolj sodobno metodo z uporabo fotoaparata in z računalniško obdelavo slike s programom Photoshop. Izračunani pospešek s tako dobljenimi podatki je precej bolj natančen kot z uporabo brnača, metoda dela pa je otrokom bolj zanimiva in razumljiva.

Ključne besede: težni pospešek, prosti pad, fotoaparat

Abstract

We can assess gravitational acceleration in various ways. The textbook suggests assessing the acceleration of the fall by performing the simple buzzer experiment. We have done an experiment by using an advanced method which includes using a camera and computer editing of the photo with the Photoshop programme. As a result, the derived acceleration is by far more exact than the one using the buzzer. Furthermore, children seem to find the method interesting as well as intelligible.

Keywords: gravitational/weight acceleration, free fall, camera

Uporaba računalnika pri mlajših učencih

Computer Use among Younger Pupils

Betka Potočnik
Osnovna šola Križe
betipot67@gmail.com

Povzetek

Razvoj računalniške tehnologije z uporabo različnih računalniških orodij in medijev omogoča vrsto novih možnosti za popestritev in dopolnitev pouka. Računalnik in splet nam nudita različne možnosti širjenja, utrjevanja, ponavljanja znanj v šoli in doma. Učitelji ugotavljamo, da so učenci izredno motivirani za delo z računalniškimi orodji. Učenci že v prvi triadi obvladajo spretnosti pri delu z različnimi informacijskimi tehnologijami. Menim, da se učitelji te situacije ne smemo prestrašiti, ampak moramo širiti in poglabljati naša znanja na tem področju in učence spodbujati pri delu z informacijsko komunikacijsko tehnologijo. Tudi v pri mlajših učencih je uporaba računalnika postala del pouka. V OPB učitelji učencem pomagamo in jih spodbujamo pri uporabi spletnih učilnic, ki so jih pripravili njihovi učitelji. Seveda pa pri vsem tem ne smemo pozabiti, da učenci prihajajo iz različnih okolij, z različnimi predznanji in z različno možnostjo dostopa do IKT.

Ključne besede: znanje, IKT, spletna učilnica, učenec, učitelj.

Abstract

The development of computer technology accompanied by the use of different computer tools and media enables various possibilities of making lessons interactive and interesting. Personal computer and internet offer numerous possibilities of upgrading, repeating and revising knowledge at school and at home. Teachers share the opinion that the computer plays an important motivational and educational role when children are in question. Pupils in the first three years of schooling already possess skills required for dealing with informational technology. My firm conviction is that we should not react as scared individuals, but rather brighten and deepen our knowledge and encourage children when working with ICT. Computer use is also included in lessons for younger pupils. Teachers of afternoon care help and encourage pupils when working with online classrooms set especially for them. Nevertheless, the fact that pupils have different backgrounds, knowledge and access to ICT should not remain ignored.

Keywords: knowledge, ICT, online classroom, pupil, teacher.

Utrinki zgodovine od začetka 20. Stoletja do druge svetovne vojne – ponavaljanje in utrjevanje z uporabo izobraževalnih zgoščenk

Highlights of History from the Beginnings of 20th Century to World War II with Educational CD-ROM

Savina Radišek Kuhar
OŠ Primoža Trubarja Laško
Trubarjeva ulica 20, 3270 Laško
savina.radisek@guest.arnes.si

Povzetek

Uporabnost ene prvih izobraževalnih zgoščenk, Zgodovinski časovni trak, je pri pouku zgodovine še vedno velika. Učencem ponuja celovit pregled zgodovine. Novejši zgodovinski izobraževalni zgoščenki in spletni aplikaciji pokrivata obdobje prve svetovne vojne in Slovence v času prve svetovne vojne. Obogateni sta z interaktivnimi vsebinami in nalogami, kar poveča njuno pestrost in dinamičnost v primerjavi z Zgodovinskim časovnim trakom. Vsebina zgoščenk ustreza zahtevam učnega načrta tako za osnovne šole, kakor tudi učnim načrtom za pouk zgodovine v gimnazijah, srednjih strokovnih šolah, kakor tudi za poučevanje družboslovja v poklicnih šolah. Učitelju ob uporabi Zgodovinskega časovnega traku, novejše izobraževalne zgoščenke, ki obravnavajo posamezna zgodovinska dogajanja, odlično služijo pri metodi dela z računalnikom ali interaktivno tablo oziroma i-tablo, kakor tudi večjemu razumevanju in poglabljanju znanja.

Ključne besede: Izobraževalna zgoščenka, aplikacija, interaktivna vsebina, učni načrt, učni pripomoček

Abstract

Applicability of one of the first educational CD – roms "Zgodovinski časovni trak" ("History timeline") at the history lessons is still great. It enables a complete overview of the history. The latest educational CD – roms and e-application refer to the period of World War I and the history of Slovene people during that time. Both CD – roms contain e-learning content and tasks, which make them more varied and dynamic than the former "Zgodovinski časovni trak" ("History timeline"). The content of CD – roms meets the demands of history curriculum in primary schools, general secondary school and secondary technical schools. In addition to this it is also appropriate for the teaching of social studies in secondary vocational schools. The latest educational CD – roms, which treat individual historical events, together with the first educational CD – roms "Zgodovinski časovni trak" ("History timeline"), are ideal for the computer assisted teaching and interactive board. What is more they facilitate understanding of the learning content and enable learners to deepen their understanding and knowledge of history.

Keywords: Educational CD – rom, application, e-learning content, curriculum, learning aid

Virtualni laboratorij pri pouku naravoslovja – eksperimentalna študija razumevanja, reprodukcije in uporabe osvojenega znanja

Virtual Laboratory in Science Class – Experimental Study of Understanding, Reproduction and Application of Acquired Knowledge

Nataša Rizman Herga
Osnovna šola Ormož
Dobravska 2, 2270 Ormož
natasa_herga@yahoo.com

Dejan Dinevski
Univerza v Mariboru, Pedagoška fakulteta
dejan.dinevski@uni-mb.si

Povzetek

Tradicionalni pouk pogosto ne omogoča aktivnejše vloge učenca pri pouku. V kemiji in naravoslovnih vedah nasploh je eksperimentalno in laboratorijsko delo eno izmed najučinkovitejših metod za pridobivanje znanja. Eksperimentalno delo lahko izvajamo tudi v virtualnem svetu. Virtualni laboratorij prinaša nekatere pomembne prednosti. Razumevanje kemije vključuje sposobnost razmišljanja na treh nivojih: makroskopskem nivoju, simbolnem nivoju in nivoju delcev, kjer kot učinkovito orodje lahko uporabimo virtualni laboratorij. Na tej osnovi smo izvedli didaktični eksperiment, da bi preverili učinkovitost uporabe virtualnega laboratorija, in sicer z vidika znanja učencev. V eksperiment so bili vključeni učenci 7. razreda (N = 38). Ob tem smo si zastavili vprašanje, ali so učni rezultati učencev po eksperimentalni zasnovi pouka z virtualnim laboratorijem boljši kot pouk naravoslovja brez vizualizacijskih elementov. Rezultati didaktičnega eksperimenta so pokazali, da je z vidika osvajanja znanja uporaba virtualnega laboratorija učinkovitejša kot pouk, kjer niso vključeni vizualizacijski elementi, ki so ključni pri učenju in razumevanju kemije.

Ključne besede: virtualni laboratorij, naravoslovje, kemijska vizualna pismenost, znanje

Abstract

Traditional teaching often does not allow very active involvement of pupils in class. In chemistry and natural sciences in general, experimental and laboratory work is one of the most effective methods for acquiring knowledge. Experimental work can also be exercised using virtual world. Virtual laboratory offers some important advantages. Understanding chemistry involves the ability of cognitive comprehension on three levels: the macroscopic level, the symbolic level and the level of particles, where the virtual laboratory can be an effective tool. On this basis, a didactic experiment was performed in order to verify the effectiveness of virtual laboratory from pupils' knowledge point of view. The experiment

involved seventh grade pupils (N = 38). Furthermore, we tried to answer the question whether the learning results of pupils, according to the experimental design of classes using a virtual laboratory, are better than results gained through classical science classes without visualization tools. The results of didactic experiments have shown that the knowledge acquiring is more effective when using the virtual laboratory instead of classical teaching (in the case when classical approach does not include visualization elements crucial for learning and understanding chemistry).

Keywords: virtual laboratory, science, chemical and visual literacy, knowledge

Uporaba WII-ja kot možnost uresničitve diferenciacije in medpredmetnega povezovanja v 1. razredu

The Use of the WII as a Possibility for Differentiation and Cross-Curricular Integration in the First Grade of Elementary School

Anita Smole
Ločica ob Savinji 84, 3313 Polzela
anita.smole@gmail.com

Sonja Strgar
Brigadirska cesta 3, 1360 Vrhnika
sonja.strgar@gmail.com

Povzetek

V 1. razredu smo preizkusili možnost uporabe WII-remota za doseganje interaktivnosti brez I-table in prevelikih materialnih stroškov. Priprava pred uporabo v razredu je bila sestavljena iz tehnične in didaktične priprave. Sledila je izvedba. Učenci so s pomočjo IR-pisala pri pouku slovenščine izvajali grafomotorične vaje, pri likovni vzgoji mešali barve, pri spoznavanju okolja opazovali in po navodilih izvajali dejavnosti opazovanja slike in sodelovali so pri medpredmetno zasnovanih dejavnostih. Dejavnosti s pomočjo WII-ja so bile pestre, pogosto so se menjavale. Glavni nosilci aktivnosti so bili učenci, ki so dosegli številne učne cilje. Učitelj je bil v vlogi usmerjevalca. Pri izvedbi dejavnosti pa smo naleteli na nekaj težav – poiskati je bilo potrebno ustrezno mesto za postavitev WII-remota (drugače kalibracija ni bila uspešna, zamikalo je sliko), paziti je bilo potrebno na sence (ki so ovirale povezavo med WII-remote in IR pisalom), učenci so se morali navaditi na IR-pisalo (ki deluje na dotik s podlago). V članku so navedene rešitve omenjenih težav. WII se je vseeno izkazal kot dobra naprava za ustvarjanje interaktivne učilnice z zelo malo denarja (žal pa manj kakovostna kot interaktivna tabla).

Ključne besede: WII-remote, IR-pisalo, 1. razred, medpredmetnost, diferenciacija, interaktivna učilnica

Abstract

We tried using the WII-remote for achieving interactivity without the I-table and large material expenses in the first grade of elementary school. Technical and didactic preparation was needed before the use in the classroom. Pupils executed graphomotorical exercises with the IR pen in Slovene class. They mixed colours in Art class, observed pictures and performed activities according to teacher's instructions in Nature class. In addition, they participated in cross-curricular activities. Activities with the WII were rich and diversified. The main performs of activities were pupils. They achieved numerous learning aims. The teacher was only in the role of an organizer. We also faced some problems during the implementation

phase – we had to find a suitable standing spot for the WII-remote. Otherwise, the calibration was not successful. The main problem was image lag. We had to mind the shadows, because they obstructed the connectivity between the IR pen and the WII-remote. The students had to get used to the IR pen, because it only works by touching the surface. The solutions of various problems are listed in the article. The WII has proved itself a great gadget for making an interactive classroom with little money. However, it does not quite reach the quality of an interactive table.

Keywords: WII-remote, IR pen, first grade of elementary school, differentiation, cross-curricular integration, interactive classroom

Določanje učne poti v Moodlu 2.0

Defining of the Learning Way in Moodle 2.0

Sašo Stanojev
Ekonomsko-storitveni izobraževalni center Kranj (ESIC Kranj)
Ekonomsko-trgovska šola, Kidričeva cesta 65, 4000 Kranj
sašo.stanojev@guest.arnes.si

Povzetek

Pri postavitvi predmeta v spletni učilnici mora učitelj upoštevati različnosti učencev. V učilnico mora umestiti raznovrstne vire in naloge za spremljanje poteka usvajanja znanja. Ti so povezani s sprotnim preverjanjem znanja in vodenjem učenca skozi aktivnosti na podlagi njihovih rezultatov. Moodle 2.0 učitelju omogoča določanje pogojev za prehode med posameznimi viri in dejavnostmi, medtem, ko učencu prepušča njihovo izbiro prek lastne učne poti. Poleg pogojev za dokončanje posameznega predmeta je učencu na voljo tudi spremljanje poteka njegovega izobraževanja in možnost samostojne presoje pri potrjevanju pričakovanih učnih ciljev.

Ključne besede: Moodle 2.0, pogojne aktivnosti, učna pot

Abstract

When build an online classroom teacher must take into account the diversity of his students. He must prepare multiple sources in the classroom and accompany the acquisition of knowledge. These are linked to assessment of knowledge and student activities through the basis of his results. Moodle 2.0 allows teacher determining the conditions for transitions between individual resources and activities, while leaves students the choice of their own learning paths. In addition to the conditions for completion of each course the student can also monitor the progress of its education and have opportunities for self-assessment with validating the expected learning goals.

Keywords: Moodle 2.0, conditional activities, learning way

Izdelava in reševanje domačih nalog pri računalniškem programiranju

Creating and Doing Homework at Computer Programming

Gašper Strniša
Tehniški šolski center Kranj
Kidričeva 55, 4000 Kranj,
gasper.strnisa@gmail.com

Povzetek

Prvi koraki pri učenju računalniškega programiranja, so za marsikoga zelo naporni, saj je ta proces sestavljen tako iz logičnega razmišljanja, kot tudi faktografskega učenja. Za uspešno osvojitev novih znanj, je potrebno precej motivacije in ustreznih nalog (vaj), ki pomagajo pri razumevanju celotnega koncepta. V prispevku je predstavljeno problemsko stanje učenja in poučevanja programiranja v srednjih šolah, ter teoretična izhodišča, predlog rešitve in možnost za nadgradnjo omenjenih rešitev, ki dvigujejo nivo znanja programiranja naših dijakov.

Ključne besede: računalniško programiranje, domače naloge, motivacija, dijaki, življenjske spretnosti.

Abstract

The first steps in learning computer programming can be very difficult for many students since the process is a combination of logical thinking and factographic knowledge. To gain new knowledge successfully a lot of motivation and appropriate tasks (exercises) are needed to help students fully understand the concept as a whole. In the paper the problems of learning and teaching computer programming in high schools are discussed together with the theoretical frameworks. A solution to the problem and an upgrade to the already existing solutions raising the level of computer programming knowledge of our students are also suggested.

Keywords: computer programming, homework, motivation, students, life skills

Projekt Spodbujanje aktivnega državljanstva s študijami primerov iz zgodovine in sodobnega sveta

Project Ecouraging Active Citizenship with Study Cases from History and Modern World

Marjeta Šifrer
Osnovna šola Naklo
Glavna cesta 47, 4202 Naklo
marjeta.sifrer@guest.arnes.si

Povzetek

V prispevku je predstavljen primer interdisciplinarne povezave, ki se je v OŠ Naklo izvajala s projektno organizacijo pouka. Integriran skupni cilj ob konceptu migracij so učenci uresničili tudi s pomočjo pouka zgodovine in državljanske in domovinske vzgoje ter etike. Na Osnovni šoli Naklo smo v šolskih letih 2008/2009 in 2009/2010 sodelovali v dvoletnem projektu Spodbujanje aktivnega državljanstva s študijami primerov iz zgodovine in sodobnega sveta, ki sta ga podprla Ministrstvo za šolstvo in šport ter Evropski socialni sklad. Glavni cilj projekta je bil okrepiti socialne in državljanske kompetence z aktivnim poučevanjem vsebin na temo aktivnega državljanstva ter spodbujanje narodne zavesti, zasnovan pa je bil interdisciplinarno.

Ključne besede: aktivno državljanstvo, demokracija, migracije, narodna zavest

Abstract

This article introduces an example of interdisciplinary linking which was carried out in Primary School Naklo within project oriented lessons. Integrated common goal was achieved also with the help of History and Citizenship lessons. In school years 2008/09 and 2009/10 Primary School Naklo took part in a two year project called Encouraging active citizenship with case studies from history and modern world. The project was supported by the Ministry of Education and Sport and European Social Fund. The main goal of the project was to strengthen social and civil competence by actively studying various contents about active citizenship and encouraging national consciousness. The project was planned interdisciplinary.

Keywords: active citizenship, democracy, migration, national consciousness.

Kombinirano učenje z učbenikom Video košarka v sedmem razredu osnovne šole ob podpori e-okolja Moodle

Blended Learning with the Textbook Video Basketball in the 7th Grade of Primary School by Using a Support of an E-Environment Moodle

Viljem Škornik
Osnovna šola Hruševec Šentjur
Gajstova 2a, 3230 Šentjur, Slovenija
viljem.skornik@gmail.com

Povzetek

V prispevku je predstavljen model kombiniranega poučevanja košarke v sedmem razredu osnovne šole ob podpori Moodle okolja. Uporabljen je model kombiniranja na ravni aktivnosti. Kombiniranje na tej ravni je prisotno takrat, ko pouk vsebuje tako tradicionalne elemente učenja kot tudi elemente e-izobraževanja, zasnovane na uporabi informacijsko-komunikacijske tehnologije (v nadaljevanju IKT).

Ključne besede: kombinirano poučevanje, košarka, sedmi razred, osnovna šola.

Abstract

In the article a model of blended learning for teaching basketball in the 7 th grade of Primary school at Moodle invironment is represented. The model of blended learning on this kind of level is present when the lesson includes the traditional elements of learning as well as the elements of education based on use of information and communication technology.

Keywords: blended learning, basketball, 7th grade, Primary school.

S podatki povezana učna gradiva za družboslovje

The Data Driven Learning Resources for Social Sciences

Janez Štebe
Arhiv druzboslovnih podatkov
Fakulteta za družbene vede, Univerza v Ljubljani

Povzetek

V prispevku je predstavljeno bogastvo ponudbe s podatki povezanih učnih gradiv, ki se navezuje na uporabo podatkovnih arhivov pri poučevanju ali samoučenju. Izhodišče je struktura ponudbe v slovenskem Arhivu družboslovnih podatkov, sledi pa prikaz mednarodno prosto dostopnih gradiv. V drugem delu je prikazanih nekaj primerov uporabe v lokalnem okolju. Poudarek je na pomenu učenja skozi primere in lastno preizkušanje. Zaključek poudarja pozitivne izkušnje uporabe in spodbuja nadaljnje širjenje uvajanja podatkov v različne stopnje in smeri izobraževanja.

Ključne besede: raziskovalni podatki, Arhiv družboslovnih podatkov, statistična pismenost, podatkovna pismenost, s podatki povezana učna gradiva, samoučenje

Abstract

The abundance of data-driven learning resources that facilitate the use of Social Science Data Archives in classroom teaching or self-education is presented in the paper. A review of what is available through Slovene Social Science Data Archive is a starting point. This is then extended with a review of related open learning resources available internationally. Second part shows selected cases of use of resources in the local environment. The potential of learning through examples, extended by one's own testing of new concepts on data available, is emphasised. The conclusion stresses on positive experiences of use of secondary data and related learning resources, and encourages its further integration into various curriculum activities in wide spectrum of educational programmes.

Keywords: research data, Slovene Social Science Data Archives, statistical literacy, data literacy, data-driven learning resources, self-education

E-kultura – digital(izira)ne kulturne vsebine na medmrežju

E-Culture – Digital Culture Content on Internet

Janja Šuster, Dejan Dinevski
INVEL Inkubator za razvoj podjetništva in
Pedagoška fakulteta Univerza v Mariboru
janja.suster@gmail.com, dejan.dinevski@uni-mb.si

Povzetek

V prispevku opredeljujemo e-kulturo kot pomembno paradigmo za širjenje kulturnih vsebin in informacij o kulturi. Digitalne kulturne vsebine in javna dostopnost niso bistvenega pomena le za področje kulturne industrije in turističnega gospodarstva, temveč predstavljajo osnovo za izobraževanje, usposabljanje, vseživljenjsko učenje in raziskovanje. V članku so poleg Digitalne knjižnice Slovenije in Europeane predstavljeni še drugi portali kulturnih vsebin in stanje na področju digitalne kulture v Sloveniji. Digitalne zbirke dokumentov predstavljajo velik potencial znanja in dosežkov kulture posameznega naroda in njegove znanosti, zato je pomembno, da stremimo k čim večjemu obsegu in dobri organizaciji digitaliziranih kulturnih vsebin. Na spletu je trenutno dostopno relativno veliko vsebin e-kulture, vendar bi težko rekli, da delujejo celovito, premišljeno, urejeno ali pregledno.

Ključne besede: E-kultura, digitalne kulturne vsebine, digitalizacija, digitalna knjižnica, dLib, Europeana

Abstract

In this feature we are defining e-culture as an important paradigm to expand cultural content and information about culture itself. Digital culture content and public accessibility are not important only for culture and tourism industry, but are representing the basis of education, training, lifelong learning and research. Furthermore, this feature, beside the Digital Slovenian Library and Europeana, also represents other portals of cultural contents and the state of digital culture in Slovenia. Digital collections of documents represent a great potential of knowledge and achievements of one nation's culture and its science, therefore it is important to strive for a great extent and good organization of digitalised culture contents. There are relatively many contents of e-culture on-line; however it is hard to say that they are comprehensive, thought-out or transparent.

Keywords: E-culture, digital cultural contents, digitalisation, digital library, dLib, Europeana

Uporaba računalniške virtualizacije v srednji šoli

The Use of Computer Virtualization in High School

Zoran Tkavc
Cesta prvih borcev 21, 8280 Brestanica
zoran.tkavc@gmail.com

Svetlana Novak
Kvedrova cesta 31, 8290 Sevnica
svetlanan@gmail.com, svetlana.novak1@guest.arnes.si

Povzetek

Računalniška virtualizacija je pojem, ki v velikem obsegu vstopa tudi na področje informatizacije v šolah. O virtualnih strežnikih, virtualnih delovnih postajah in aplikacijah slišimo v povezavi z varčevanjem energije ter kakovostnim vzdrževanjem strojne in programske opreme. Na kakšen način pa se lahko srednja šola spopada z vsakdanjimi težavami, kako do ustrezne računalniške opreme na čim cenejši način in kako kvalitetno vzdrževati storitve povezane s poukom, vam bomo odgovorili v naslednjem članku, ki temelji na rešitvah, ki smo jih realizirali na naši šoli, ŠC Krško-Sevnica. V članku je opisana uporaba in možnosti, ki jih ponuja računalniška virtualizacija v srednji šoli. V prvem delu članka je opisana uporaba virtualnih računalnikov pri poučevanju strokovnih predmetov za poklic računalniškega tehnika, v drugem delu pa predstavljena ekološko naravnana računalniška učilnica, ki uporablja virtualizacijo z namenom zmanjšanja potreb po strojni opremi in porabe električne energije. Članek torej predstavlja rešitev uporabe računalnika za lokalno delo, ki s pomočjo virtualizacije omogoča delo več uporabnikov hkrati.

Ključne besede: virtualizacija, programska oprema, strojna oprema, ekološko naravnana računalniška učilnica, VirtualBox.

Abstract

Computer virtualization is a concept, which is rapidly entering into the school informatization area. We talk about virtual servers, desktop computers, applications in conjunction with saving energy, quality computer hardware and software maintenance. The articles tries to answer the foundametnal questions schools are dealing with today such as how to buy cheaper computer equipment and how to maintain computer services. The article describes the solutions applied in our school, ŠC Krško-Sevnica. The use and possibilities which virtualization offers in a secondary school are thus described here. In the first part the use of virtual computers in teaching computer technics in the secondary school is described and in the second part the ecological computer classroom is presented.

Virtualization helps to reduce hardware costs and energy consumption resulting in saving money as well as energy. The article therefore presents a conceptual solution using only one local computer, which, with the support of the virtualization, transforms into a multiseat workstation.

Keywords: virtualization, software, hardware, ecological classroom, VirtualBox.

Vloga uredniškega odbora pri kreiranju šolskih spletnih strani

The Role of the Editorial Board at Creating School Webpages

Polona Tomac Stanojev
Ekonomsko-storitveni izobraževalni center Kranj (ESIC Kranj)
Ekonomsko-trgovska šola, Kidričeva cesta 65, 4000 Kranj
polona.tomac@guest.arnes.si

Povzetek

Ko smo v letošnjem šolskem letu postavili nove spletne stran Ekonomsko-trgovske šole ESIC Kranj, smo se odločili za spremembe, ki zadevajo tako njihovo prenovoljeno celostno podobo kot tudi drugačen sistem za urejanje spletnih strani. Izbrali smo spletno aplikacijo ExpressionEngine, ker podpira novosti, ki smo jih uvedli predvsem s formiranjem uredniškega odbora, v katerega so vključeni posamezni uredniki z različnimi zadolžitvami in kompetencami glede objavljanja in urejanja spletnih vsebin. Ker aplikacija podpira možnost, da se posameznim urednikom dodelijo različni nivoji pravic glede urejanja, so vsebine na spletnih straneh vidne šele, ko jih jezikovno pregleda lektor in vsebinsko potrdi tudi glavni oz. odgovorni urednik. Takšen nivojski način urejanja strani se je pokazal kot smiseln in učinkovit zato, ker omogoča večjo preglednost in nadzor nad objavami ter hkrati podpira dinamično timsko uredniško politiko.

Ključne besede: dinamične (šolske) spletne strani, uredniški odbor

Abstract

As we worked on setting up of the new webpage of The Economics-Trade School ESIC Kranj, we decided to introduce changes that refer to both: its overall updated look and its different system of editing webpages. We have chosen the web application ExpressionEngine because it supports the novelties regarding our new approach to the editorial board which consists of individual editors who are competent and are in charge of publishing and editing various web contents. The application enables the possibility that each editor can be assigned a different level of editing rights, so that the contents appear on the webpage only after they have been proof-read and checked by the proof-reader and by the editor in chief. Such a differentiated way of webpage editing has proven to be meaningful and effective because it enables a much higher level of surveyability along with the control over the published items; moreover, it also encourages dynamic editorial team policies.

Keywords: dynamic (school) webpages, editorial board

E-učilnica – bližnjica pri pripravi na splošno maturo iz slovenščine

E-Classroom – a Shortcut to the Preparation for General Matura

Polona Tomac Stanojev
ESIC Kranj, Kidričeva cesta 65, 4000 Kranj
polona.tomac@gmail.com
Naslov e-učilnice: http://esola.esic.si/course/view.php?id=11

Povzetek

E-učilnica se je zaradi možnosti, ki jih s svojimi viri in dejavnostmi ponuja Moodle, pokazala kot zelo primerna in koristna aplikacija tudi pri pripravi na splošno maturo iz slovenščine. Objava in uporaba številnih virov je pripomogla k nazornejšemu in zanimivejšemu pouku; Moodlove dejavnosti pa so nam omogočile tudi sodelovalno delo na daljavo, kar je posledično prispevalo tudi k enostavnejši in učinkovitejši (pred)pripravi na zrelostni izpit iz materinščine.

Ključne besede: e-učilnica, matura, slovenščina, Moodle

Abstract

Due to its possibilities offered by Moodle with its sources and activities, e-classroom has proved to be a very convenient and useful application for the preparation of students for the general matura, as well. The publication and usage of numerous sources have contributed to a more explanatory and interesting class. Moodle activities have also enabled a remote collaborative work , which has consequently contributed to a simpler and more effective (preliminary) preparation for the native language maturity exam.

Keywords: e-classroom, Matura, Slovenian, Moodle

Analiza uspešnosti šolske spletne strani na platformi Joomla

Analyzing Success of Joomla as School's Web Page

Marina Trkman,Peter Trkman,
Maja Ramšak, Branko Kaučič
marina.trkman@gmail.com

Povzetek

V članku smo se osredotočili na Joomlo, ki predstavlja orodje Web 2.0. Na podlagi študije primera šolske spletne strani smo analizirali tako implementacijo kot uporabo. Uporabnike spletne strani smo anketirali in intervjuirali ter na podlagi rezultatov prišli do pomembnega sklepa: izobraževanje uporabnikov je zelo pomembno pri doseganju visokega nivoja uporabe. Naše priporočilo je, da vrhnji menedžment pripravi dva tipa izobraževanj za uporabnike in sicer pred in po implementaciji; oboje za namen višje verjetnosti intenzivne uporabe.

Ključne besede: Joomla, Web 2.0, implementacija, uporaba, izobraževanje

Abstract

The paper focuses on a tool of Web 2.0 called Joomla and highlights challenges of implementation and use. We have conducted a case study of a primary school which implemented Joomla as a school's web page. With the answers from a survey and interviews we show the outmost importance of training of the users for the success of the use. Our recommendations are that higher managers should provide two types of training: before and after the implementation; both done for the purpose of greater likelihood of intensive use.

Keywords: Joomla, Web 2.0, implementation, use, training

Učinkovito kombiniranje besedilnega, grafičnega, zvočnega in video učnega medija ob upoštevanju učnih stilov

Effective Combination of Textual, Graphic, Audio and Video Learning Media Considering the Learning Types

Živa Trstenjak
Višja prometna šola Maribor
ziva.trstenjak@gmail.com

Povzetek

Uporaba različnih medijev v poučevanju in učenju omogoča nazornejše prikazovanje procesov in bolj učinkovito usklajevanje načinov poučevanja z učenčevim stilom spoznavanja in učenja. Kadar se odločamo za uporabo različnih oblik multimedije pri e-izobraževanju je potreben temeljit razmislek. Za učinkovito uporabo besedilnega, grafičnega, zvočnega ali video posredovanja znanja je potrebno upoštevati osebne značilnosti in sposobnosti učencev ter njihove učne stile. V tem prispevku bomo poskusili predstaviti učinkovito uporabo različnih oblik posredovanja znanja glede na različne teorije o učnih stilih in primer semantičnega spleta, ki način podajanja znanja prilagodi posameznemu učencu.

Ključne besede: multimedija, učni stili, semantični splet

Abstract

Use of various media in teaching and learning process enables a more effective adaptation of teaching methods to the needs and predominant learning style of the student. Teaching strategies and the choice of e-materials requires careful consideration. For efficient use of textual, graphic, audio and/or video media for the knowledge transfer, the personal characteristics and abilities of students and their learning styles have to be taken into account. In this paper we tried to present the effective use of various forms of transmission of knowledge regarding various theories of learning styles and an example of the semantic web, which provides students with personal education packages.

Keywords: multimedia, learning styles, semantic web

Uporaba metod in tehnik mehanike iger v izobraževanju

Methods and Techniques of Game Mechanics Used in Education

Marko Urh
Univerza v Mariboru, Fakulteta za organizacijske vede
Kidričeva 55a, 4000 Kranj, Slovenija
marko.urh@fov.uni-mb.si

Povzetek

Prispevek prikazuje možnosti uporabe mehanike iger v izobraževanju. V uvodnem delu so predstavljene značilnosti izobraževanja, ki so zaradi vse hitrejših tehnoloških sprememb, drugačne kot so bile v preteklosti. Sledi področje e-izobraževanja kot posledice razširjenosti informacijsko-komunikacijske tehnologije, ki predstavlja osnovo za implementacijo mehanike iger v izobraževanje. V nadaljevanju so navedeni osnovni elementi mehanike iger s svojimi značilnostmi in lastnostmi. Prikazane so možnosti, za uvedbo mehanike iger v izobraževanje z namenom doseganja večje motiviranosti udeležencev v izobraževalnem procesu. Prispevek prikaže tako pozitivne kot negativne učinke, ki jih moramo upoštevati pri uvajanju mehanike iger v izobraževanju. V zadnjem delu so podane ugotovitve, priporočila in izkušnje na področju uporabe mehanike iger v izobraževanju.

Ključne besede: izobraževanje, učenje, e-učenje, motivacija, informacijski sistemi, igre, mehanika iger

Abstract

This paper deals with the possibility of using the game mechanics in education. The first part of the paper presents the characteristics of education that are changed because of rapid technological development. E-learning is becoming very important because of expansion brought by information and communication technology and this represents foundation for implementing game mechanics in education. The paper deals with the basic elements of game mechanics with its own characteristics and properties. Higher motivation among students can be achieve with implementation of the game mechanics in education process. Both negative and positive effects are presented in this paper that must be taken into account when establishing the game mechanics in education. The final section of the paper presents findings, recommendations and experiences of game mechanics in education.

Keywords: education, learning, e-learning, motivation, information systems, games, game mechanics

Uvedba sistema ugotavljanja in zagotavljanja kakovosti vzgojno – izobraževalnih organizacij – samoevalvacija: izbira ciljev z uporabo lupine eskpertnega sistema DEX-i

Introducing System to Find and to Assure Quality in Educational Institutions – Self Evaluation: Option of Aims by Using Shell of the Expert System DEX-i

Barbka Vidmar
Srednja šola Zagorje
Cesta zmage 5, Zagorje ob Savi, Slovenija
barbi_vidmar@yahoo.com

Povzetek

V prispevku je predstavljena izbira kurikularnega cilja na šoli v okviru usposabljanja za samoevalvacijo. Izbor ciljev smo določili s pomočjo metode večparametrskega odločanja in lupine ekspertnega sistema DEX-i na podlagi analize ciljev, ki so jih v skupinah predlagali učitelji. S tem smo olajšali proces izbire ciljev in hkrati omogočili transparentno razlago izbire.

Ključne besede: samoevalvacija, kakovost v šoli, kurikularni cilji, lupine ekspertnega sistema DEXi.

Absrtact

In our contribution we introduced choice of curicular aim as part of qualification for self-evaluation. We defined the selection of aims using the shall of the expert system DEXi on the base of analysis aims, which were proposed in the groups of teachers. We alleviated the process of choosing aims and to make possible transparent explanation of choose.

Keywords: self-evaluation, quality in school, curicular aims, shell of the expert system DEXi.

Uporaba spletnega foruma za delo na daljavo na OŠ Predoslje

Uporaba spletnega foruma za delo na daljavo na OŠ Predoslje

Branka Vodopivec
Osnovna šola Predoslje Kranj
branka.vodopivec@guest.arnes.si

Eva Traven
Osnovna šola Predoslje Kranj
eva.traven@guest.arnes.si

Povzetek

Spletni forumi predstavljajo eno od oblik sodobne komunikacije. V prispevku je predstavljena ena od možnosti uporabe foruma za potrebe učnega procesa, med drugim tudi za pripravo na pisanje razlagalnega spisa na tekmovanju za Cankarjevo priznanje. Glavni namen asinhrone učne metode (foruma) je bil povezava med sodelovalnimi oblikami dela v razredu s projektnim delom na daljavo, kjer bi s pomočjo izkustvenega učenja učenci lahko izmenjali izkušnje, mnenja in ideje.

Ključne besede: Spletni forum, učenje na daljavo, priprava na Cankarjevo tekmovanje, izmenjava idej

Abstract

Web-based forum represents a form of modern communication. The paper presents a possibility of using the forum for learning process, among others a method of practising how to write an explanatory essay for the Cankar Prize competition. The main goal of the asynchronous learning method (forum) was bringing together cooperative working methods in the classroom and distance project work where pupils could exchange experiences, opinions, and ideas.

Keywords: web-based forum, distance learning, practising for Cankar Prize competition, exchange of ideas

Odziv slovenskih učiteljev informatike na on-line tečaj »EUCIP«

Response of Slovene Informatics Teachers to the EUCIP On-Line Course

Margareta Vrtačnik
Univerza v Ljubljani, Naravoslovnotehniška fakulteta
Vegova 4, 1000 Ljubljana
metka.vrtacnik@guest.arnes.si

Danica Dolničar
Univerza v Ljubljani, Naravoslovnotehniška fakulteta
Vegova 4, 1000 Ljubljana
danica.dolnicar@ntf.uni-lj.si

Niko Schlamberger
Slovensko društvo INFORMATIKA
Vožarski pot 12, 1000 Ljubljana
niko.schlamberger@gmail.com

Špela Švoljšak
IPMIT d.o.o.
Kotnikova 30, 1000 Ljubljana
spela.svoljsak@ipmit.eu

Povzetek

V prispevku poročamo o izkušnjah in rezultatih on-line tečaja »Zasnova in priprava spletnih učnih gradiv, osnovanih na sistemu certifikatov EUCIP«, ki smo ga izvedli v sklopu mednarodnega projekta Sloop2desc sklada Leonardo da Vinci. Tečaj je bil namenjen predvsem izobraževalcem poklicev v informatiki. Nanj se je prijavilo 64 udeležencev, od tega največ učiteljev informatike in računalništva v osnovnih in srednjih šolah, od katerih je slaba polovica uspela priti do zadnjega modula tečaja. Pokazalo se je, da je za uspeh on-line tečajev ključna nenehna komunikacija med udeleženci tečaja in tutorji ter podajanje natančnih in konkretnih navodil za izdelavo nalog. Želimo si, da bi tečajniki, ki so uspešno končali tečaj, postali ambasadorji sistema EUCIP, za katerega upamo, da ga bomo uspeli uvesti tudi v Sloveniji.

Ključne besede: EUCIP, on-line tečaj, rezultati

Abstract

In the paper we present our experience and results related to the design and management of the on-line course »Development of e-learning resources based on the EUCIP certification

scheme", which was organized under the auspices of the international Leonardo da Vinci project Sloop2desc. The course was primarily aimed at teachers of informatics. Out of 64 participants that applied for the course, slightly less than 50 % managed to reach the last course module. In line with other research, our experience showed that in order to minimize participants' drop-out rate, a sustained communication flow between tutors and participants was necessary and specific assignments had to be provided by the tutors, along with guidance. We expect the participants who successfully finished the on-line course to become a some kind of ambassadors for the introduction of the EUCIP scheme in Slovenia.

Keywords: EUCIP, on-line course, results

Znanje v družbi znanja

Knowledge in the Knowledge Society

Rado Wechtersbach
Zavod RS za šolstvo, Poljanska 28, 1000 Ljubljana
rado.wechtersbach@zrss.si

Povzetek

Članek razširja perspektive pedagoških razmišljanj prav do epistemoloških obzorij sodobne znanosti. Dejstvo namreč je, da v današnji družbi, kot tudi v družbi bližnje prihodnosti, bistvena vprašanja izhajajo iz problemov, ki jih moramo rešiti. Zato je v mladih potrebno razviti znanje, s katerim bodo premostili razliko med teorijo in prakso, med vedeti in delovati in ki ga lahko opredelimo z atributoma celostno in kompleksno. Značilnost tako opredeljenega modela je interdisciplinarni pristop, ki pogojuje uporabo računalniške tehnologije in zahteva drugačno ocenjevanje izhajajoč iz nove paradigme kompleksnega znanja.

Ključne besede: šola, znanje, kompleksnost, družba znanja, digitalna kompetenca

Abstract

The paper extends the perspective of educational thinking as far as to epistemological horizons of the contemporary science. In the society of the present days as well as in the near future essentially issues are conceived from the aspects of the problems which have to be solved. Therefore, new thinking must be developed which will bridge youth the difference between theory and practice, between knowing and acting, and which can be described with attributes such as systemic and complex. Characteristic of this model is interdisciplinary approach which binding computer technology and requires a new assessment practices that are geared to the complexity knowledge paradigm.

Keywords: school; knowledge; complexity; knowledge society; digital competences

Šola v oblaku?

A School in the Cloud?

Alenka Zabukovec
Srednja ekonomska šola Ljubljana, Roška 2, 1000 Ljubljana
alenka.zabukovec@guest.arnes.si

Tomaž Ferbežar
Šolski center Novo mesto, Srednja elektro šola in tehniška gimnazija
Šegova ulica 112, 8000 Novo mesto
tomaz.ferbezar@guest.arnes.si

Povzetek

Računalništvo v oblaku se v zadnjih letih razvija zelo hitro. Storitve, ki jih ponujajo različne vrste računalniških oblakov, uspešno uporablja vse več uporabnikov in organizacij. Ker oblak gotovo ponuja več koristi kot pomanjkljivosti, je lahko ena od možnih rešitev za organizacije, ki imajo premalo tako človeških kot tudi finančnih virov za prenovo svoje informacijske infrastrukture. V prispevku so predstavljene storitve različnih vrst računalniških oblakov, prednosti, nevarnosti in možnosti uporabe za slovenske izobraževalne zavode.

Ključne besede: Računalništvo v oblaku, računalniški oblak, storitve v oblaku, Google Apps

Abstract

Cloud computing has been developing very fast during last few years. The services, offered by different types of computing clouds, are successfully used by many users and organizations. As the cloud certainly offers more benefits than disadvantages, it may be one possible solution for organizations that are understaffed and lack financial resources for the renewal of its IT infrastructure. The paper presents various types of cloud comoputing services, benefits, risks and opportunities for the Slovenian educational institutions.

Keywords: Cloud computing, cloud computing services, Google Apps

S spletno učilnico do kakovostnejšega znanja?

With Virtual Classroom to a Higher Quality Knowledge?

Alenka Zabukovec
Srednja ekonomska šola Ljubljana, Roška 2, 1000 Ljubljana
alenka.zabukovec@guest.arnes.si

Dobrila Lazović
OŠ Riharda Jakopiča, Derčeva 1, 1000 Ljubljana
dobrila.lazovic@guest.arnes.si

Povzetek

Spletna učilnica je ena od oblik vključevanja IKT v sodobno izobraževanje. Omogoča vso podporo procesom poučevanja in pridobivanja znanja. Izkušnje učiteljev, ki jo uporabljajo v slovenskem prostoru, so pozitivne, saj naj bi znanje učečih, pridobljeno na tak način bilo kakovostnejše in trajnejše. V prispevku so opisane možnosti uporabe spletnih učilnic v slovenskem prostoru (pridobitev za zavod in seminarji) za učitelje in možnost uporabe spletne učilnice za posameznega učitelja. Predstavljeno je razmišljanje o vplivu uporabe spletnih učilnic pri pouku na kakovost znanja.

Ključne besede: Spletna učilnica, pouk, izobraževanje na daljavo, pridobivanje znanja, e-šolstvo.

Abstract

Virtual classroom is a way of integrating IT in modern education. It enables full support for the process of teaching and gaining knowledge. Teachers using virtual classrooms in Slovenia, report of positive experience - the knowledge gained this way is of higher quality and longer duration. The paper discusses the use of online classrooms in Slovenia (improvement for the institution and seminars) for teachers and the possibility of using the virtual classroom for each teacher. Presented is a reflection on the impact of virtual classrooms for lessons and the quality of knowledge.

Keywords: Virtual classroom, education, online education, gaining knowledge, e-learning.

(Multi)medijski projekti v izobraževanju

Multimedia Projects in Education

Srečo Zakrajšek
Inštitut in akademija za multimedije
Leskoškova 12, Ljubljana
sreco.zakrajsek@iam.si

Miha Zakrajšek
Inštitut in akademija za multimedije
Leskoškova 12, Ljubljana
miha.zakrajsek@iam.si

Povzetek

Priprava multimedijskih projektov sodi med najbolj motivacijska in želena dela v izobraževalnih institucijah, na tem principu pa temeljijo tudi številne delavnice, s katerimi vključujejo mlade v izobraževanje in usposabljanje tudi v neformalnem izobraževanju. Elemente multimedijske produkcije je mogoče zelo uspešno uporabiti v izobraževalnih projektih in to praktično v vseh fazah nastajanja, realizacije in promocije. Izdelava in uporaba multimedijskih elementov, produktov in storitev je tesno povezana z uporabo IKT (Informacijsko komunikacijske tehnologije) kar skupaj s timskim in interdisciplinarnem delom in uporabi zanimivih tem, pomeni velik izziv in motivacijo za delo mladih. Posebej če so rezultati ustrezno promovirani in celo nagrajeni. V prispevku je predstavljeno sodelovanje študentov IAM (Inštituta in akademije za multimedije) na multimedijskih projektih, ki so nastali v sodelovanju s podjetjem Petrol in so obravnavali sodobne teme – racionalna raba energije, sočna energija, varovanje okolja idr. V celoten sistem priprave, promocije, ocenjevanja in nagrajevanja projektov pa so bila vključena tudi družabna omrežja, izobraževalna platforma, e- konference, raznovrstne oblike predstavitve ipd. Celoten potek projekta smo spremljali z aplikativno raziskavo in nekateri rezultati so tudi podani v prispevku.

Ključne besede: multimediji, projektno delo, (multi)medijska produkcija,

Abstract

Preparation of multimedia projects is one of the most motivational and desired positions in educational institutions and many workshops are based on the same principle where young people are included in education and in training also in non-formal education. Elements of multimedia production can also be very successfully used in educational projects in almost all phases of its making, realization and promotion. Making and using multimedia elements, products and services is tightly connected with the use of the ICT (Information and Communication Technologies) which, combined with team work and interdisciplinary work, while using interesting topics, means a great challenge and motivation for the work of young people; especially when the results are suitably promoted and even rewarded.

This article presents the cooperation of IAM students (Institute and Academy of Multimedia) on multimedia projects that were created in cooperation with the Petrol Company using current topics – rational energy use, solar energy, environmental protection, etc. In the process of preparation, promotion, evaluation, and rewarding of the project were also included social networks, educational platform, e-conferences, various forms of presentation etc. We have followed the project with an applicative research. Some results are also available in the article.

Keywords: multimedia, project work, (multi)media production

Likovno izražanje ob računalniku

Using Computer at Art

Metka Žnidar
Srednja šola Jesenice
Ulica bratov Rupar 2, 4270 Jesenice
metka.znidar@gmail.com

Povzetek

Računalnik je postal nenadomestljiv tudi v šoli. Pri pouku umetnostne zgodovine in likovnega izražanja dijakom omogoči dostop do slikovnega gradiva. Boljša je predstavitev kontrastov in barv različnih umetniških del. Glede na primere, ki sem ji obdelala, lahko zaključim, da je uporaba računalnika pri razlagi umetnostno zgodovinskih obdobij priporočljiva in tudi rezultati glede razumevanja posameznih obdobij so boljši pri dijakih. Po predavanju so si dijaki in dijakinje različno predstavljali obdobje kubizma. Nekateri ga enačijo z ekspresionizmom. Po ponovnem ogledu slik pred risanjem ob uporabi računalnika so slike podobne slikam tega obdobja. Pri risbah, ki so jih delali dijaki brez računalnika, še bolj pride do izraza njihova domišljija, spomin in osebna izraznost. Dijaki srednje tehnične šole so po razlagi narisali grški tempelj vseh oblik. Ob dostopu do računalnika pa so narisali grški tempelj z tehnično natančnostjo. Ko so dijaki risali po opisu kip Avgusta in Palade Atene, so risbe v podrobnostih različne, z dostopom do računalnika, pa sta kipa resnična. Ko so dijaki po spominu risali Ostržka, so izkazali svojo izraznost, domišljijo in spomin. Ob uporabi interneta nas Ostržek spominja na junake Disneyjevih risank. Nekateri dijaki in dijakinje so preveč dobesedno prerisali dostopno sliko na računalniku, čeprav naj bi jim primeri služili le kot primer za njihovo lastno risbo ali sliko. So dijaki, ki so nadarjeni, vendar so z izdelkom zadovoljni le, če natančno prerišejo iz videnega na računalniku. Treba jih je spodbuditi, da vključijo svoje sposobnosti in domišljijo. Tukaj vidim slabost. Opažam, da nekateri dijaki v primeru, ko jim dam določeno temo ali motiv, ki ga morajo narisati, nekateri na skrivaj vzamejo v roke mobilne telefone in preko interneta iščejo motive in teme. Lažje jim je to, kot da bi se sami potrudili in premislili, kaj naj narišejo.

Ključne besede: likovno izražanje, umetnostna zgodovina, slikovno gradivo, domišljija, spomin, osebna izraznost, risba

Abstract

Computer has become a must at school. It helps students to get access to the materials they need. The presentation of contrasts and colours of different paintings is much better with the use of computer. From my experience I can recommend the use of computer because it students perform better. After lectures the students had different impressions of cubism, some mixed it with expressionism. After taking another look at the paintings with the help of computer, their paintings looked much more than those of cubism.When the students did thier drawing without using the computer first, their drawings included much more imagination,memory,and personal expression. After hearing the lecture the students of secondary technical school drew a Greek temple in many diefferent ways, but after using the

computer their drawings were technically acczrate ... When the statue of Augustus and Palada Atena were described to the studends, their drawins differed in details a lot. After using the computer, they were very real. After hearing the description of Pinocchio, students again used their imagination, . After using the computer, their drawings were similar to Disney's characters. Some students copied too much the picture they saw on the computer although they should use it only as example for their own paintings. Some students are gifted, but they seem to be satisfied only if thei work is an accurate copy of the computer picture.. They need encouragement to include their imagination and skills.. This is one of the disadvantages of this method. I have noticed that some students after being given the theme secretly use their mobiles to find motives and themes. They find it more easily than trying to think and do the task by themselves.

Indeks avtorjev / Author index

Konferenca / Conference
Uredili / Edited by

Inteligentni sistemi /
Intelligent Systems
Marko Bohanec, Matjaž Gams

Izkopavanje znanja in podatkovna skladišča /
Data Mining and Data Warehouses (SiKDD 2011)
Dunja Mladenić, Marko Grobelnik

Sodelovanje, programska oprema in storitve v informacijski družbi /
Collaboration, Software and Services in Information Society
Marjan Heričko

Kognitivne znanosti /
Cognitive Sciences
Urban Kordeš, Matjaž Gams, Olga Markič

Kognitonika /
Cognitonics
Vladimir A. Fomichov, Olga S. Fomichova

Robotika /
Robotics
Jadran Lenarčič, Leon Žlajpah, Andrej Gams

Internet in Slovenija: 1985 - 1995 /
Internet and Slovenia: 1985 - 1995
Andrej Brodnik, Matjaž Gams, Rok Sosič

Vzgoja in izobraževanje v informacijski družbi /
Education in Information Society
Vladislav Rajkovič, Tanja Urbančič, Mojca Bernik