Virtual Assistant Aggregator for the Project Electronic and Mobile Health

Primož Kocuvan
Fakulteta za elektrotehniko
Tržaška cesta 25
1000, Ljubljana, Slovenia
primoz.kocuvan@outlook.com

Aleš Tavčar
Jožef Stefan Institute
Jamova cesta 39
1000, Ljubljana, Slovenia
ales.tavcar@ijs.si

Gregor Grasselli
Jožef Stefan Institute
Jamova cesta 39
1000, Ljubljana, Slovenia
gregor.grasselli@ijs.si

Matjaž Gams
Jožef Stefan Institute
Jamova cesta 39
1000, Ljubljana, Slovenia
matjaz.gams@ijs.si

ABSTRACT

Electronic and mobile health (English: EMH, Slovene: EMZ) sub-project was part of the EcoSMART Smart Specialization program. During the EMZ sub-project, we developed several prototypes, products, and services that were integrated into a platform with the use of the Rocket.Chat application, which was developed as part of the AS-IT-IC project. Currently, ten entities and web services are integrated into the platform. Some of them like the municipality assistant consist of 200 single assistants. Others, like HEP-Y, an application for hepatitis, were designed before and represent a single entity. The EMZ platform simplifies the interaction with each of the services either by clicking or through natural language. When asked about healthcare, the platform searches through hundreds of possible replies and proposes the most promising.

Keywords

AS-IT-IC platform, Jaccard index, lemmatization, working agents, virtual assistants, Electronic and mobile health

1. INTRODUCTION

This paper describes a unifying platform addressing the specific needs of the healthcare domain and was developed as part of the EcoSMART S4 program [3]. It is based on the AS-IT-IC [6] platform which was developed in the cooperation of Austrian and Slovenian developers. AS-IT-IC, short for Austrian-Slovenian Tourist Information Center is a web application and a platform for assisting tourists who are traveling across Austria and Slovenia to plan a trip. We have reorganized the internal structure of the source code and added new functionalities to meet the needs of the health care system. The EMZ platform implements several sub-applications, for example, ASPO [1] and HEP-Y [2], which are services for diagnosing and recognizing sexually transmitted diseases and hepatitis, respectively. The most important part of our platform is the ability to answer questions formulated in natural language by the user. The whole ecosystem is built and integrated into a popular and open-source application for messaging called Rocket.Chat.

During past projects, the Jožef Stefan Institute has already developed over 200 so-called chatbots or virtual assistants to be used on the web pages of the Institute and Slovenian municipalities. We integrated all of the existing chatbots into a single platform. Our platform is called in Slovenian language ”Platforma za elektronsko in mobilno zdravje” or English ”Platform for electronic and mobile health”. The platform is for now in the Slovenian language, so it is intended for end-users who speak and understand Slovenian. The platform has three functionalities:

- It can answer questions posed in natural language.
- It can provide a link to the information we need.
- It can search through the database of Institute employees by applying lemmatization and stemming services to the search string and the data in the database.

In the following sections and subsections, we describe how the assistants, which are the building blocks of the platform, work, and which technologies we used to get the expected functionalities. Finally, we present some figures of the graphical interface and examples of conversations with the platform agent.

2. SYSTEM DESCRIPTION

Our top-level working agent in the EMZ platform works as follows:

- First, when a new question is received, the system processes the input. It asks all the working agents that are available about healthcare or the Eco-SMART project. These are the Summoner - Izbirčnež, the healthcare waiting queue assistant, the Eco-SMART assistant and the municipality assistants. Moreover, it also queries the database of prototypes and domains.
- If the question concerns assistants and not the database of prototypes and domains it sends the whole string by a POST or GET request to the online assistants. The assistants then perform lemmatization and stemming of the text, search for the best answer, assign weights

to all possible answers, and return the one with the highest weight.

- If the question concerns prototypes and domains, the system queries the database directly. However, before that, another set of lemmatization operations is applied. This process differs from the first since we integrated the search locally on the server where we read it from a file that is regularly updated.
- All the received answers are collected into a table. We filter the answers by selecting only those that the agent thinks are relevant to the initial query. Finally, we show them in the Rocket.Chat dialogue window.
- The user can select the most appropriate answer and start the conversation with the agent that provided the most appropriate response by clicking on one of the displayed responses. This functionality will be further described in the following text.

2.1 Assistans

Here we describe the assistants that provide possible answers to the requests in the EMZ platform. All are web applications and RESTful applications. REST is short for representational state transfer. This term means that we can communicate between different programs or processes by using HTTP requests. All of the agents are using JSON for the response format.

We will now provide a short description of the assistants.

2.1.1 Assistant for Queues in Healthcare [1]

The assistant identifies all medical institutions in Slovenia, where someone can apply for a medical procedure. It then displays the waiting time for the procedure in question at the institutions in a specific region, ordered by the waiting time from the shortest to the longest, to make it easier to identify where is it possible to get the procedure the fastest. The information about a medical facility includes contact information and the address. The system actively inquires the user for additional search constraints, such as the region in which we would like to receive the treatment and the urgency. It provides buttons that enable the user to see the available choices and to simplify communication.

2.1.2 Eco-SMART Assistant [2]

The assistant can answer general questions about the Eco-SMART project. It responds to questions about the institutions and companies which are involved in the project. Next, it can provide organizational data from the project or information or services related to the execution and results of the project.

2.1.3 JSI Assistant [3]

The JSI assistant called Robi is the first and most important assistant which was developed at the Institute. It's the fundamental building block for all other assistants. At first, it was implemented in Python 2.7, but we updated the code to Python 3.6. The main task was to delete all redundant code which was repeating. Now the assistant is slim and small, but we kept all functionalities from the older version. The assistant can answer questions about the people who are working on the Institute; it can open relevant web pages in the background for further searching. It has a multitude of available applications such as the Slovenian dictionary, a computer terminology dictionary and information about the menu at the canteen at the Institute. It also provides a quick way of informing the maintenance personnel of any problems requiring their attention.

2.1.4 Assistants for municipalities in Slovenia [4]

The Republic of Slovenia has 212 municipalities, and 200 municipalities have their assistant. It can answer general questions about the mayor, municipal council, environment and transport in the municipality and several other from the municipality related domain. The assistants are based on the original assistant Robi which works on the official web page of the institute. This system was integrated into the EMZ platform since it was agreed with the project partners.

2.2 Rocket.Chat

One of the reasons why we selected Rocket.Chat, as the underlying communication platform and for the integration of text messaging and communication with the assistant, is that its a free and open-source project. The application provides hooks that make it easy to integrate a custom assistant. It makes it possible to send HTTP requests with the payload containing message inputs from the user, to a predefined address, and it displays the response to this request in the same chat room as the original inquiry. This mechanism was used to connect Rocket.Chat, which serves in the role of the front-end, with the developed back-end application written in Python using Flask.

Rocket.Chat also supports direct messaging, multiple rooms, and public channels. Since we are hosting Rocket.Chat ourselves, we have no limitations as we would if we were to use for example Slack, which limits the number of messages retained in history when using the free plan.

2.3 Lemmatization

Each text query to the platform is processed. First, it is split into individual words, producing a set of words. Second, we perform simple POS tagging on the text to identify nouns. This process is performed because nouns carry the most information when identifying the relevant entries in the prototypes and domain databases. Also, all punctuation is removed. Next, we apply lemmatization on the nouns and get a set of lemmas or dictionary from words. Then we calculate the Jaccard index on the lemmas, comparing words in the database to the words in the question string. The Jaccard index is described in more detail in the next subsection. Finally, we return the search result.

We have chosen lemmatization over stemming because it produced better results in our tests. Stemming only cuts off the extensions of the words. We need words in the dictionary form. The downside of lemmatization is that it is slower and

[1] https://df-chatbox.herokuapp.com/

[2] https://ekosmart.docker-e9.ijs.si/

[3] http://www.projekt-asistent.si/ijs

[4] https://asistenti-website.docker-e9.ijs.si/post/seznam-asistentov/

requires additional data in the form of a dictionary, which in our case was a model for Slovenian language, to apply morphological analysis. For this purpose we used a library that was also developed at JSI.

2.4 Jaccard index

We can measure the similarity between two sets A and B with the Jaccard statistical formula (Equation 1). First, we calculate the intersection between two sets and count the number of members. That number is our numerator. Then we calculate the union on the same two sets, count the number of members and put that number in the denominator. A result is always a number between 0 and 1 (as stated in Equation 2). If we multiply it with 100, we get a percentage. This method produces the Jaccard index, the higher the number we get, the more similar the two sets are. Complementary of the Jaccard index is the Jaccard distance. It measures how different two sets are.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \tag{1}$$

$$0 \leq J(A,B) \leq 1 \tag{2}$$

If both sets are empty, then we consider the Jaccard index to be equal to one. We have applied the Jaccard index to the set of words obtained from lemmatization of the question string and the set of concepts stored in the database file. The higher the Jaccard index, the more similar wee deem the question with the prototype or domain name. The system only returns the item with the highest score.

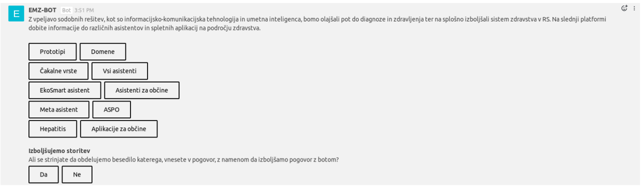
The Jaccard index is commonly used as an error measure when training convolutional neural networks to perform bounding box segmentation of images.

2.5 Blocking the other assistants when talking to just one

To prevent confusion when talking to multiple bots at the same time, we have implemented a function to focus the conversation to a selected assistant. This is achieved by tracing the value of the "talking" field associated with a particular assistant. Its value may be either True or False. When True then this assistant is also receiving questions, otherwise the working agent is "blocked" and does not participate in the conversation. The user sets the right value by clicking on the button to limit the conversation to a particular assistant. By clicking "Yes, I want to talk to the assistant" we set the "talking" field to True and for other assistants to False.

Besides allowing the user to have a more consistent conversation this also enables the users to select the conversation partner they believe will provide the right information even when the system is incapable of determining which answer to present correctly.

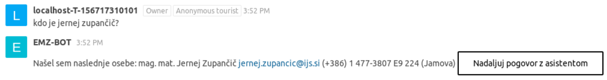
3. DEMONSTRATION OF THE PLATFORM



EMZ-BOT Bot 3:51 PM
Z vpeljavo sodobnih rešitev, kot so informacijsko-komunikacijska tehnologija in umetna inteligenca, bomo olajšali pot do diagnoze in zdravljenja ter na splošno izboljšali sistem zdravstva v RS. Na slednji platformi dobite informacije do različnih asistentov in spletnih aplikacij na področju zdravstva.

Izboljšujemo storitev
Ali se strinjate da obdelujemo besedilo katerega, vnesete v pogovor, z namenom da izboljšamo pogovor z botom?

Figure 1: Example of an EMZ-BOT introductory greeting.



localhost-T-156717310101 Owner Anonymous tourist 3:52 PM
kdo je jernej zupančič?

EMZ-BOT 3:52 PM
Našel sem naslednje osebe: mag. mat. Jernej Zupančič jernej.zupancic@ijs.si (+386) 1 477-3807 E9 224 (Jamova)

Figure 2: Example of an EMZ-BOT answer for looking up a person from JSI.

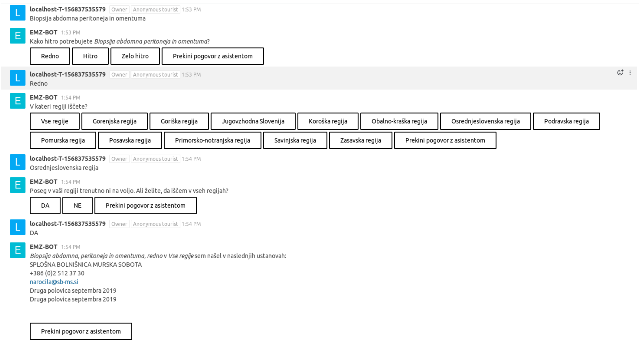
The platform works similarly as AS-IT-IC. It has two components, first the web page which provides information about projects, prototypes, and links to the relevant working agents. Second, the Rocket.Chat widget where we can search through the database or ask a question. Here we are talking to a bot, named EMZ-BOT. When we load the Rocket.Chat widget, we get an initial greeting from the bot (Figure 1). Here we can choose which working agent we want to talk to. If we don't choose an agent, the EMZ-BOT provides answers from all working agents or assistants that responded.

In the remainder of this section, we present 4 examples of interactions with the platform and communication with the different assistants.

We can ask the bot about the staff or people who work at the Institute. We receive an answer as shown in Figure 2.

Next, if we have any medical problems, we can type it into the system. And the bot replies accordingly (see Figure 3).

Also, we can search through the database of prototypes and domains. In Figure 4 we performed a search for heart failure. We get two answers, one from prototypes and one from domains or data. By comparing the answers, we can choose the assistant we want to talk to.



localhost-T-156837535579 Owner Anonymous tourist 1:53 PM
Biopsija abdomna peritoneja in omentuma

EMZ-BOT 1:53 PM
Kako hitro potrebujete *Biopsija abdomna peritoneja in omentuma*?

localhost-T-156837535579 Owner Anonymous tourist 1:53 PM
Redno

EMZ-BOT 1:54 PM
V kateri regiji iščete?

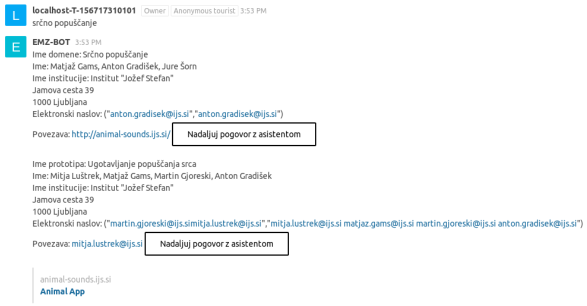
localhost-T-156837535579 Owner Anonymous tourist 1:54 PM
Osrednjeslovenska regija

EMZ-BOT 1:54 PM
Poseg v vaši regiji trenutno ni na voljo. Ali želite, da iščem v vseh regijah?

localhost-T-156837535579 Owner Anonymous tourist 1:54 PM
DA

EMZ-BOT 1:54 PM
Biopsija abdomna, peritoneja in omentuma, *redno* v *Vse regije* sem našel v naslednjih ustanovah:
SPLOŠNA BOLNIŠNICA MURSKA SOBOTA
+386 (0)2 512 37 30
narocila@sb-ms.si
Druga polovica septembra 2019
Druga polovica septembra 2019

Figure 3: Example of a chat with EMZ-BOT when we need abdominal biopsy.



localhost-T-156717310101 Owner Anonymous tourist 3:53 PM
srčno popuščanje

EMZ-BOT 3:53 PM
Ime domene: Srčno popuščanje
Ime: Matjaž Gams, Anton Gradišek, Jure Šorn
Ime institucije: Institut "Jožef Stefan"
Jamova cesta 39
1000 Ljubljana
Elektronski naslov: ("anton.gradisek@ijs.si","anton.gradisek@ijs.si")

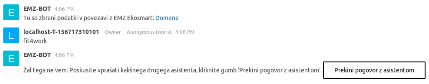
Povezava: http://animal-sounds.ijs.si/

Ime prototipa: Ugotavljanje popuščanja srca
Ime: Mitja Luštrek, Matjaž Gams, Martin Gjoreski, Anton Gradišek
Ime institucije: Institut "Jožef Stefan"
Jamova cesta 39
1000 Ljubljana
Elektronski naslov: ("martin.gjoreski@ijs.simitja.lustrek@ijs.si","mitja.lustrek@ijs.si matjaz.gams@ijs.si martin.gjoreski@ijs.si anton.gradisek@ijs.si")

Povezava: mitja.lustrek@ijs.si

animal-sounds.ijs.si
Animal App

Figure 4: Example of an EMZ-BOT answer when we search for heart failure



EMZ-BOT 4:06 PM
Tu so zbrani podatki v povezavi z EMZ Ekosmart: Domene

localhost-T-156717310101 Owner Anonymous tourist 4:06 PM
fit4work

EMZ-BOT 4:06 PM
Žal tega ne vem. Poskusite vprašati kakšnega drugega asistenta, kliknite gumb 'Prekini pogovor z asistentom'.

Figure 5: Example of an EMZ-BOT answer when looking for something which the assistant does not find

The last Figure 5 shows us the response from the bot when it did not find anything after we have chosen to talk to just one working agent. We have an option to click on the button to talk to all agents.

4. FUTURE DEVELOPMENT OF THE EMZ PLATFORM

We have identified several issues and possible research directions for future work. The most important thing is to shorten the general response time of the platform. One possible solution would be to define functions in Python as coroutines and adding asynchronous behavior which is an abstraction of using threads. The authors of the Summoner [5] showed that using threads can shorten the response time by a factor of four. However, before speeding up the system, we should also run tests to evaluate the system's ability to provide the correct answer. After this, we would be able to provide the assistants that are integrated into the system an improved and broader knowledgeable database. The assistants should learn from past questions and improve the database on the fly.

5. CONCLUSIONS

We have created a comprehensive platform using the latest information-communication technology based on the AS-IT-IC platform. One of the essential parts of this ecosystem is the EMZ-BOT. Users can communicate with it in natural language and obtain structured and informative answers. The system is intended to provide information and integrate services for the healthcare domain. Thus, it can be used by chronic patients and people with disabilities, as well as healthy individuals and the general population, who are just searching trough the platform and seeking information about healthcare, information regarding municipalities, active life, the Eco-SMART project, and other. Soon, there will be a need for advanced ICT solution since our healthcare system is overwhelmed and in need of digitalization to support the need of patients. Our solution to these problems is the EMZ platform.

6. ACKNOWLEDGMENTS

The work and investment is co-financed by the Republic of Slovenia and the European Union from the Regional Development Fund. It is also co-financed by the Eco-SMART project. For help and support of the project we thank Jernej Zupančič and Miha Štravs.

7. REFERENCES

[1] A. Ajanović, J. Konda, A. Peterlin, K. Počivavšek, G. Fele-Žorž, A. Gradišek, M. Gams, and M. Matičič. Application for sexually transmitted infection risk assessment. *Informatica 2017*, 41(2):253–254, June 2017.

[2] A. Ajanović, A. Ulčar, A. Peterlin, K. Počivavšek, G. Fele-Žorž, A. Gradišeka, M. Gams, and M. Matičič. Application for viral hepatitis infection risk assessment - hepy. *Informatica 2018*, 42(2):279–281, June 2018.

[3] M. Drnovšek and M. Gams. Emz in ekosmart-asistent. *Information society 2018*, 1(3):11–14, September 2018.

[4] M. Drnovšek, M. Gams, A. Tavčar, and G. Grasselli. Pregled asistentov ijs e9. *Information society 2018*, 1(3):15–19, September 2018.

[5] A. Glavač, J. Zupančič, and M. Gams. The summoner - 'izbirčnež'. *Information society 2018*, 1(6):28–31, September 2018.

[6] J. Zupančič and M. Gams. Austrian-slovenian intelligent tourist information center:project progress report 2018. *AS-IT-IC Workshop 2018*, E(21):5–9, September 2018.