Development of the text analysis software agent (chat bot) for the library based on the question and answer system TWIN

Anastasia Ivanovskaya¹, Konstantin Aksyonov¹*, Igor Kalinin¹, Yuriy Chyrshev¹, and Olga Aksyonova¹

¹Ural Federal University, 620002 Mira 32, Ekaterinburg, Russia

Abstract. The development of a text analysis software agent is presented for a library based on the TWIN question-response system. A review of modern platforms for creating chat bots. The results of experiments with a trained text analysis software agent are described. The trained agent fully provided correct information during the experiments.

1 Introduction

Information technologies are applied in all spheres of human activity, this allows to reduce the time for processing a large amount of information. For example, in the library system, when the reader asks for help in selecting a book, the librarian often has difficulty quickly providing an answer about the availability of a book or advice on choosing literature. These difficulties give rise to problems of quick help and relevant search. To solve these problems, a text analysis software agent based on the TWIN question-answer system was developed.

2 Overview of question-answer systems for the processing of Russian speech, focused on the voice interface and chat-bots interface

A question-answer system is an information system that is a hybrid of search, reference, and intelligent systems that uses a natural-language interface. The input to such a system is a request formulated in natural language, after which it is processed using NLP methods, and a natural language response is generated. The QA system uses either local storage or a global network, or both, as a source of information [1]. At the moment, there are many cloud solutions for the needs of NLP and the development of software agents for analyzing texts based on the question-answer system. Table 1 provides an overview of several chat bots that can recognize text and voice [2]. TWIN question-answer system was chosen for the development of a text analysis software agent for the library.

* Corresponding author: wiper99@mail.ru
| Chat bot development environment | Cost | Machine learning simulation | Messaging Platform Integration | Notes |
|---------------------------------|------|----------------------------|-------------------------------|-------|
| Wit. ai (Facebook, 2013) https://wit.ai | free | Training with examples | Not. Web Services API | There is a limit in the ability to transmit only one request per second. |
| Api.ai (Google, 2010) https://api.ai | free | Learning is present | Has paid funds for deployment on Heroku | |
| IBM WATSON, (IBM, 2010) https://www.ibm.com/watson | Trial free version 30 days | Learning is present | Integrates with Azure | There is a large functionality that will require skills and a long time to master. |
| TWIN, (TWIN, 2018) https://twin24.ai/ | free | Training with examples | It has integration with various communication systems. | |

### 3 Description of the TWIN system

The TWIN system is an omni channel communication platform for building head-on and chat bots, which is able, in particular, to receive data in one language and transmit it in another. TWIN can keep voice and text recording, display detailed statistics and analytics for each call or dialogue. TWIN system supports and integrates 6 communication channels - SMS, calls, instant messengers, online chat, mail, social networks.

The platform provides the ability to connect to the customer's IP-PBX, use of its telecom operator and integration with virtually any CRM or ERP system. The robot can work with any modern IP-PBX. The basis of the formation of TWIN knowledge is the work with Big Data and neural networks [3].

### 4 Application of TWIN system

It is necessary to distinguish named entities and intentions before developing a software agent. Named entity is a word or phrase intended for a specific, well-defined object or phenomenon, distinguishing this object or phenomenon from a number of similar objects or phenomena. Intention can be defined as the meaning of what has been said, i.e. what the user meant when he said a certain phrase [4].

The main entities of this subject area are the following (Figure 1): book; author name; genre; description.

The following intentions are the ones of this subject area: determine the year of publication; define a book description; identify a book by name; identify a book by ISBN; identify a book by author; identify a book by genre (see. fig. 2).

Composite intentions solve the main problem of understanding human speech, namely, people very often express several intentions in one phrase. Therefore, the markup of compound phrases is carried out after adding entities and intentions for the task of training the agent and thereby the agent’s knowledge base is filled. After the preparatory actions, the agent was trained.
After filling the agent’s knowledge base, the agent is trained using a neural network. The results of processing a phrase by a trained agent are presented in Figure 3.
Fig. 3. An example of processing a phrase by a text analysis program agent.

Figure 4 shows a summary chart that displays the percentage of recognition of intentions.

Fig. 4. Recognized Intent Relation Chart.

The TWIN system is developed using the RASA platform [5]. This is an open and very popular chat bot framework. It consists of two independent components, Rasa NLU and Rasa Core. Rasa NLU (Natural Language Understanding). The main goal of this component is to convert user input in natural language into objects that the program can work with. Rasa Core - this component is responsible for creating a chat bot script based on intentions and entities.

The previously created agent is automatically transferred to the synonyms directory in the NLU section (see figure 5), while determining the text of the users message; the intent which should be associated with the text; the entities are specific parts of the text which need to be identified.
The Stories page defines training examples for the interactive system (see figure 6). A story starts with a name preceded by two hashes **## story1**. Messages sent by the user are shown as lines starting with * in the format intent{"entity1":"value", "entity2": "value"}. Actions executed by the bot are shown as lines starting with - and contain the name of the action.

The Domain defines the universe in which assistant operates (see figure 7). It specifies the intents, entities, slots, and actions bot should know about. Also, phrases patterns that the bot can speak are defined in this section.
Fig. 6. Intentions, entities, slots and actions of the bot.

A chat bot model was created after all the steps taken. The bot is trained during the creation of the model. Further, experiments with the bot were carried out with communication during which the developed bot demonstrated accurate and necessary answers to user replicas.

4 Conclusion

The developed text analysis software agent (chat bot) for the library on the basis of the question-answer system solves the problems posed: it allows you to answer quickly interesting questions from the reader about the availability and selection of books. The trained agent fully provided correct information during the experiments.

This work is supported by Act 211 Government of the Russian Federation, contract № 02.A03.21.0006.

References

1. A. Nikitin, P. Raikov. Question-answer systems (2018) Access mode: https://gigabaza.ru/doc/67598.html
2. Comparative analysis of chatbots API (2018) Access mode: https://activewizards.com/blog/a-comparative-analysis-of-chatbots-apis/
3. Future Bunking. Twin - omni-channel communication platform for building voice and chat bots (2018) Access mode: http://futurebanking.ru/post/3647
4. K. Aksyonov, D. Antipin, T. Afanaseva, I. Kalinin, I. Evdokimov, A. Shevchuk, A. Karavaev, U. Chiryshev, E. Talancev. Testing of the speech recognition systems using Russian language models. CEUR, 2298, 7 (2018), http://ceur-ws.org/Vol-2298/paper13.pdf
5. Opencorpora. Instructions for determining named entities [Electronic resource] - Access Mode: http://opencorpora.org/wiki/Nermanual1