Using of Open-Source Technologies for the Design and Development of a Speech Processing System Based on Stemming Methods

Andrey Tarasiev, Margarita Filippova, Konstantin Aksyonov, Olga Aksyonova, Anna Antonova

Using of Open-source Technologies for the Design and Development of a Speech Processing System Based on Stemming Methods

Andrey Tarasiev[1], Margarita Filippova[1], Konstantin Aksyonov[1], Olga Aksyonova[1], Anna Antonova[1]

[1]Ural Federal University
620000, Yekaterinburg, Russia
antonovaannas@gmail.com

Abstract. This article discusses the idea of developing an intelligent and customizable automated system for real-time text and voice dialogs with the user. This system can be used for almost any subject area, for example, to create an automated robot - a call center operator or smart chat bots, assistants, and so on. This article presents the developed flexible architecture of the proposed system. The system has many independent submodules. These modules work as interacting microservices and use several speech recognition schemes, including a decision support submodule, third-party speech recognition systems and a post-processing subsystem. In this paper, the post-processing module of the recognized text is presented in detail on the example of Russian and English dictionary models. The proposed submodule also uses several processing steps, including the use of various stemming methods, the use of word stop-lists or other lexical structures, the use of stochastic keyword ranking using a weight table, etc.

Keywords: Multi-Agent, Design, Development, System, Decision-Making, Real-Time, Twin, Stemming, Postprocessing, Open-source.

1 Introduction

At a present day technologies related to automated control processing systems are rapidly developing. As a part of these modern trends in the development of information technology the creation of effective and technologically advanced solutions for organizing the processing of incoming requests is required

Data processing technologies are progressively evolving, with more and more systems replacing human resources every day. Automation and the creation of information systems are at the moment the most promising areas of activity of modern society. One of the reasons for the active development of these areas is that automation is the basis for a fundamental change in management processes that play an important role in the activities of companies.

Thus, there is a need to use a control system, whose operation is aimed at maintaining and improving the operation of the all business processes at all.

As an important part of this area an automation of call centers working process require the help of a control device (complex of means for collecting, processing, transmitting information, and generating control signals or commands).

The module of speech recognition in this system uses the two most developed at the moment existing solutions YandexSpeechKit and GoogleSpeech API. But integration with these third-party services and other related solutions can't provide complete accordance with system requirements including features, adequate recognition quality, scalability, flexibility and so on.

As previous part of this work the testing of the speech recognition systems using Russian language models including selecting and testing a speech recognition systems and designing architecture of the "Twin" automation system was conducted [5].

Based on the received information, we can conclude that the Yandex system is good at recognizing short expressive phrases, as well as numerals. On contrary, the Google API is good at recognizing long phrases and terms. The results of this research are being used in development of the automatic system that calls the customers of "Twin" [1, 2, 4]. At the same time, both systems have problems with recognition in terms of noise, uneasy speech tempo and voice defects of the interlocutor.

Development experience of answer and question automated systems are described in some works: based on frame approach [9], based on agent approach [10-11, 15] and conceptual graphs [12-14]. This is due to the fact that both systems better recognize phonemes and phrases, and individual sounds are worse, especially if there are noises and other factors that distort the quality of the transmitted audio message. This observation is confirmed by the findings obtained by other independent researchers [3].

To eliminate such problems, it is necessary to include pre- and post-processing. These actions can provide requirements of recognition quality.

Pre-processing uses the dynamic selection of recognition services provided. Subsequent processing of the received textual information includes the lemmatization and normalization of individual words in the recognized speech and other methods. For example, in the case of processing numbers, this is the application of tokenization technology.

Also in previous part of this work the research and selection of methods for evaluating information about part of speech, word form and statistical relationships between words have been done.

As the conclusion of this research we can claim that the lemmatization method is most effective for solving the problems of dynamic text processing and recognition. However, to the same extent, this approach is extremely dependent on the exact definition of a part of speech (the lexical category of a word) and the language model.

Application of this algorithm leads to the fact that the analyzer module receives an array of words in primary form, due to which it is possible to easily recognize part of speech using the rules of the language model.

However, this information in practice may not be enough for the further work of the decision-making module and generating a response. In addition, there is the possi-

bility of recognition errors associated with incorrect recognition of the lexical category of a word and its initial form.

To obtain a more accurate recognition and processing result, further post-processing mechanisms are used.

The implementation of the lemmatization mechanism in the system was performed through some open-source package solutions of morphological processing and lemmatization for the PHP language.

Within the framework of this work, the design and development of an automated system named "Twin" was carried out and details presents module of recognized text post-processing on the example of the Russian and English languages dictionary models.

2 Integration of Speech Recognition Functionality With Ranking and Database Generation Functionality

Since the approach of reducing the words to normal form and determining the parts of speech is insufficient for the voice robot to understand the interlocutor's speech and for further decision making, the system uses post-processing algorithms for normalized words. This is due to the presence of a number of problems, inaccuracies and recognition errors.

In addition, the lemmatization method can occur recognition errors since the lexical category of a word can be incorrectly determined. There is also a big problem with the correct recognition of numerals, which leads to difficulty to identifying numbers, dates, addresses, etc. To solve these problems, several additional processing algorithms are used.

Many researchers provide several methods to improve quality of natural language processing based on stemming methods.

When building the speech recognition module for voice system, we were guided for the most part not by coming up with conceptually new algorithms, but by focusing on the use and adaptation of existing and well-established approaches.

The most often usable method based on implementation of a large vocabulary speech recognition system for the inflectional languages such as Slavic group languages [8].

Based on this approach the following method for additional processing was proposed.

Conventionally, we can distinguish the following steps of the additional processing of the second stage:

- primary processing;
- introduction of a keyword system and the use of stochastic keyword ranking;
- introduction of a word stop-list system.

2.1 Primary Processing

Decisions on the basic additional processing of normalized words are determined on the basis of the given settings and information on the lexical category of the word (part of speech) as a preparatory stage for more serious text analysis methods.

The essence of the primary processing of the second stage of text analysis is the use of simple algorithms for automated morphological analysis.

To increase the accuracy of recognition of numerals, the information contains numerals goes through the tokenization procedure. Also, the system provides for the adjustment of language models, which allows you to customize the perception of numerals, dates, cases, as well as the need to use keywords. In addition, it is possible to configure the methods used for stemming.

2.2 Application of the Keyword System

The introduction of a keyword system allows for significantly more effective recognition of language units. Firstly, this is due to the fact that keywords can be tied to specific blocks of the dialogue script, which largely solves the problems associated with the context. This can be useful, for example, when processing paronyms.

When creating a script in a special interface, creator of the scenario can set keywords for each person’s answer.

This dialogue scenario consists of various information blocks, most of which involve the interlocutor's response or expectation of a phrase.

Moreover, depending on the frequency of mentioning of certain words and phrases, further training takes place. The system, each time making a decision, increases the weight of a certain keyword in the expert table of stochastic data. Thus, the likelihood of a particular word appearing during further launches of this dialogue scenario increases depending on the frequency of its appearance and correct recognition earlier.

The processed text is transferred to the block of the analyzer of coincidence with keywords.

At the same time, both processed normalized words and initially recognized words are transmitted to the system. Keywords themselves can be set intentionally in various forms. Due to these two conditions, the recognition quality can be significantly improved, and the table of knowledge base weights for the stochastic ranking method will be more trained. A similar approach is widespread in systems using stochastic knowledge bases.

In addition, this makes it easy to solve the potential problem of the loss or distortion of the meaning of either a single word or the phrase itself as a whole.

That is, if several variants of the same word are specified in the system for a certain block of the script, and both coincide with the original and lemmatized recognized words, then the probability weights for both keywords will significantly increase, and the weight of the script branch will also increase by the weight of both keywords.

By default, the conditional weight of each keyword and each branch of the dialogue at the initial moment of execution time of a script is equal to one. When the

lemmatized word coincides with the originally recognized word with the keywords, the weight of each keyword increases by one. In this case, the weight of the branch increases by two units. If only the keyword denoting the initial form and the lemmatized language unit coincide, then the weight of the script dialog branch increases by one.

Based on the dynamically calculated weights of each branch, at every moment a decision is made on the further development of the dialogue scenario taking into account the largest branch weight.

This approach allows the author of the script to work out the dialogue features important for the script, which are influenced by the degree and conjugation of the verb, etc.

As already mentioned, in addition to setting keywords the system provides for the use of tables of stochastic ranking of words by probability depending on the frequency of their appearance and correct recognition in dialogs.

Of course, to ensure the correct operation of such a system, it is necessary to implement and maintain a knowledge base. It is worth noting that this storage takes up much less volume than other knowledge base implementations for stochastic stemming methods would occupy. To ensure dynamic interaction, in the system it is implemented as a base tree.

The use of the stochastic method is primarily due to the fact that an approach that implements only lemmatization algorithms is not able to provide sufficient reliability for correct speech recognition and decision making, since it places great responsibility on the system operator itself, which may be mistaken due to the human factor or not take into account all the options, which can have a big impact on the course of the dialogue.

2.3 Application of the Stop-word System

Similar to using the keyword system in a word processing module, stop words are also used. Their semantic load is similar to keywords. The essence of this idea is to strictly discard branches when the recognized word coincides with the stop word.

This approach allows us to simplify the writing of some dialogue scenarios by providing the user with the opportunity to work with negation and exceptional situations, or to write down fewer keywords and work out less on premediating of potential scenarios. It also allows to speed up the processing of text, due to the fact that some cases can be immediately discarded and not be considered in the future.

According to the implementation method, the process of recognizing stop words is practically no different from keywords. A lemmatized word and a primary word are transmitted to the handler in their original form recognized by an external open-source morphy system. These words are compared with stop words. If a match is found, the weight of the stop word in the knowledge base of stochastic rules increases, and the weight of the branch decreases to the minimum possible. Thus, during the execution of the further procedure for selecting a scenario vector, this branch cannot be selected in any way.

After applying all the rules and treatments, the search for a solution begins in accordance with the rules specified in the script, after which a response is generated.

2.4 Recognition Process Description

Based on everything described earlier, a simplified process of additional processing of a phrase recognized by third-party systems occurs as follows:

At the first step, the text recognized by a third-party system is transmitted to the post-processing module. This module analyzes this text in several stages.

First of all, the text goes through the procedure of splitting into separate linguistic units – words and statistical grouping by phrases. Each word is lemmatized, as a result of which the word in the initial form is returned at the output of the work of the stemming algorithms, and the lexical category of the word is also determined.

Next, the second stage of processing begins. Based on the part of speech defined in the previous step, initial processing is performed. In this case, the rules defined when setting up language models are used, such as processing numerals, tokenization, processing dates, cases for nouns, etc.

The processed words and words in their original form, recognized by third-party systems, are transferred to the next data processor, which deals with their comparison with keywords and stop words, which are pre-introduced to the system for the corresponding blocks of the dialogue script.

If the words transmitted to the input of the algorithm coincide with the ones specified in the system, the statistical weights of the keywords and corresponding branches of the script increase.

The weights of the words specified in the system are recorded in the production knowledge base of stochastic rules. Thus, the ranking of words is provided and the model is trained. The more often these or other words and phrases are used when executing a script during dialogs, the more likely it is the appearance of these words in these forms with other passages of this scenario. Accordingly, over time, these words will be more correctly recognized.

Finally, based on which branch, illustrating the possible choice of further development of the dialogue, is gaining the greatest weight, a decision is made and a response is generated.

This process can be visualized using the following sequence diagram (see Fig. 1).

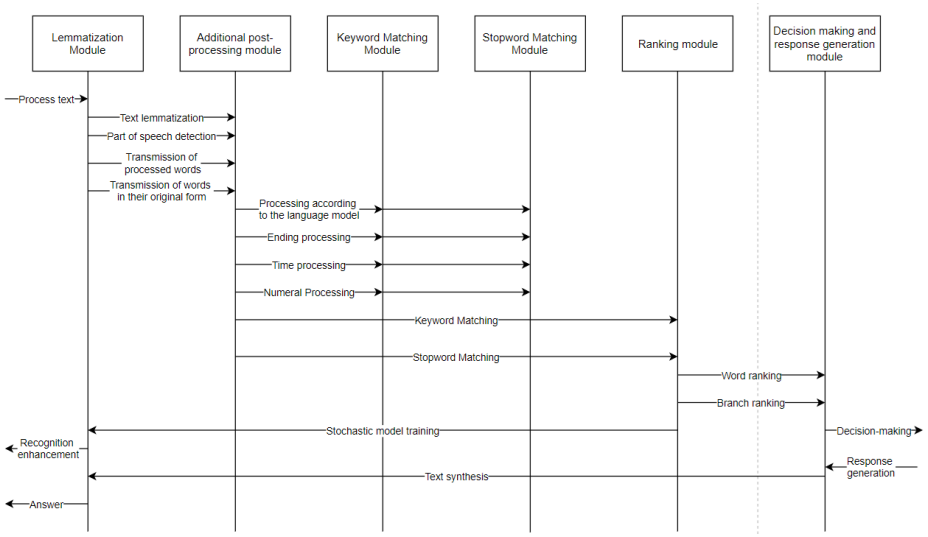

Fig. 1. Sequence diagram for additional post-processing of recognized text.

We are considering options for highlighting our own developments, such as a system of keywords, into separate freely distributed solutions.

3 Using of Open-source Technologies to Development Post-processing System

The implementation of the lemmatization mechanism in the system was performed through an open-source package solution of morphological processing and lemmatization for the PHP language – PHP Morphy, which contains the corresponding knowledge bases - dictionaries for various language models. This solution easily integrates with the used development tool stack.

This library supports AOT and myspell project dictionaries. For the library to work, you need to compile a binary dictionary. When compiling, the dictionary is used in its original form.

Physically, the dictionary is presented in the form of several files, which are a data store in a key-value format, as well as storage of reference lexical information, grammatical information, end tables, grammes, etc.

Other word processing packages for processing declension cases, such as petrovitch library also are used. This package also allows to gender determination and develop and apply own linguistic rules.

It should be noted that using of open-source systems in not limited by mentioned solutions.

Flexibility of proposed application and modern technologies itself allows to simple scaling or replacing third-party modules or adding new.

Moreover, proposed dialogue systems use other open-source packages for different modules that are not limited to post-processing speech recognition module.

For example, we use Rasa Core module for speech-analyze subsystem and chat-bots.

Mentioned speech-analyze subsystem using as a part of customizable script.

4 Conclusion

Open-source solutions provide ability to integrate various implementations of exists methods, such as stemming text recognition methods in proposed case, with each other and own development.

Due to additional processing, the correct perception of speech and the meaning of what has been said and decision-making occurs in developing system.

The development of the system involves the development and implementation of additional functions, such as the sending of statistical data and the creation of hints in the compilation of the script. To do this exists open-source solutions also can be used. Analysis of internal statistics will help determine the priority directions for improving the interface and system features.

The range of use of the system can be extended, due to the initial flexibility of the structure and wide ability to integrate with third-party solutions, packages and so on.

It should be noted that proposed system Twin is used in teaching intellectual technologies at the Ural Federal University on a free basis.

We are considering options for highlighting our own developments, such as a system of keywords, into separate freely distributed solutions.

5 Acknowledgments

The work was supported by Act 211 Government of the Russian Federation, contract no. 02.A03.21.0006.

References

1. Aksyonov, K., Antipin, D., Afanaseva, T., Kalinin, I., Evdokimov, I., Shevchuk, A., Karavaev, A., Aksyonova, O., Chiryshev, U.: Testing of the speech recognition systems using Russian language models. In: Proceedings of the 5th International Young Scientists Conference on Information Technologies, Telecommunications and Control Systems, ITTCS 2018. Yekaterinburg, Russian Federation, December, (2018).
2. Aksyonov, K., Kalinin, I., Tabatchikova, E., Chiryshev, U., Aksyonova, O., Talancev, E., Tarasiev, A., Kanev, V.: Development of decision making software agent for efficiency indicators system of IT-specialists. In: Proceedings of the 5th International Young Scientists Conference on Information Technologies, Telecommunications and Control Systems, ITTCS 2018. Yekaterinburg, Russian Federation, December (2018).

3. The study of the reliability of speech recognition by the system Google Voice Search. Cyberleninka.ru, https://cyberleninka.ru/article/v/issledovanie-nadezh, last accessed 2020/03/10.
4. Features of TWIN, https://twin24.ai/#features, last accessed 2020/03/10.
5. Tarasiev, A., Talancev, E., Aksyonov, K., Kalinin, I., Chiryshev, U., Aksyonova, O.: Development of an Intelligent Automated System for Dialogue and Decision-Making in Real Time. In: Proceedings of the 2nd European Conference on Electrical Engineering& Computer Science (EECS 2018).Bern, Switzerland, December (2018).
6. Kartavenko, M.: On the use of acoustic characteristics of speech for the diagnosis of mental states of a person, https://cyberleninka.ru/article/v/ob-ispolzovanii-akusticheskih-harakteristik-rechi-dlya-diagnostiki-psihicheskih-sostoyaniy-cheloveka, last accessed 2020/03/10.
7. Loseva, E., Lipnitsky, L.: Recognition of human emotions by spoken using intelligent data analysis methods, https://cyberleninka.ru/article/n/raspoznavanie-emotsiy-cheloveka-po-ustnoy-rechi-s-primeneniem-intellektualnyh-metodov-analiza-dannyh, last accessed 2020/03/10.
8. Rotovnik, T., SepesyMaucec, M., Kacic, Z.: Large vocabulary continuous speech recognition of an inflected language using stems, https://hal.archives-ouvertes.fr/hal-00499182/document, last accessed 2020/03/10.
9. Minsky, M.: A framework for Representing Knowledge in The Psychology of Computer Vision, P. H. Winston (ed.), McGraw-Hill (1975).
10. Greenwald, A., Jennings, N., Stone, P.: Guest Editors' Introduction: Agents and Markets. Intelligent Systems, vol.18, pp. 12–14 (2003).
11. Dash, R., Parkes, D., Jennings, N.: Computationals-Mechanism Design: A Call to Arms. Intelligent Systems, vol.18, pp. 40–47 (2003).
12. Sowa, J.F.: Knowledge Representation: Logical, Philosophical, and Computational Foundations, CA: Brooks/ Cole Publishing Co. (2000).
13. Sowa, J.F.: Conceptual graphs for a database interface. IBM Journal of Research and Development 20(4), 336–357 (1976).
14. Sowa, J.F.: Conceptual Structures: Information Processing in Mind and Machine. MA: Addison – Wesley (1984).
15. Vittikh, V.A., Skobelev, P.O.: Multiagent Interaction Models for Constructing the Needs-and-Means Networks in Open Systems. Automation and Remote Control 64, 162–169 (2003).