

METHODS OF SPEECH AND TEXT DATABASES DEVELOPMENT FOR QA-SYSTEMS

A.L. Ronzhin, A.A. Zaytseva, S.V. Kuleshov, K.V. Nenausnikov

*Saint-Petersburg Institute for Informatics and Automation of Russian Academy of Science,
Saint-Petersburg, Russian Federation
E-mail: ronzhin@ias.spb.su*

The paper is devoted to the problems of question-answer systems development (QA-systems). The subject of the study is discussion of approaches to the automatic filling of the database of the QA-system based on the analysis of the unstructured text sources currently available in the public domain of the Internet.

The analysis reveals that the following ways of implementing QA-systems are distinguished: based on inference for ontologies, rules and syntax, using artificial neural networks.

The methods for automatically search of question-answer pairs based on the structure of sentences and on the basis of associative-ontological analysis has been developed and tested in the research.

The method based on the analysis of the structure of sentences is effective for texts such as lists of frequently asked questions (FAQ), as well as literature texts containing dialogs, direct speech, based on preliminary processing of the text, expressed in the form of a heuristic rule.

The method based on associative-ontological analysis is focused to the class of reference and dictionary texts and is based on the assumption that in the descriptive text there is a sentence (or a group of sentences) containing the main idea of the text. In this case, the title of the text can be considered a question, and this sentence (or a group of sentences) is the answer. We need to make the selection of meaning-generating sentences due to the semantic reduction of the text automation. For this purpose, algorithms of self-referencing are applied based on the associative-ontological approach to the processing of texts in natural language.

For the experimental verification of the possibility of creating an open QA-system based on the automatic collection of question-answer pairs from the Internet, a prototype of a collection module for the database of the QA-system has been developed.

Keywords: question-answer pair; associative-ontological analysis; text; automatic text processing; natural language; speech recognition.

Introduction

The task of automatic speech recognition in real conditions is far from its solution, taking into account the variability of the source of the speech signal and the acoustic noise that harbors the initial sequence of audio segments. In recent years, significant progress has been made in this area and there are commercial voice-independent applications that quite successfully recognize speech in the processing of voice commands (Google maps, Yandex maps), in interactive systems (Siri), in stenographic systems [1]. The accuracy of recognition of speech units in these systems has reached the necessary threshold, so that users begin to trust automatic voice input and think about the transition from the usual means of contact input of information to contactless ones.

The reached success in the field of speech recognition is associated with the development of cloud technologies, which made it possible to use: 1) "large" heterogeneous data for teaching a multi-level hierarchical acoustic language model of language and speech; 2) crowdsourcing technologies for manual processing of a huge volume of training and recognizable audio and text data; 3) distributed computing resources for servicing client voice applications.

The advantageous factors, that reduce the complexity of the task, are the possibility of preliminary tuning to a specific speaker and a relatively small size of the dictionary of recognizable speech units.

Among the possible areas of research contributing to the solution of the problem are the methods with the application of: 1) multichannel recording and processing of audio signals using an array of microphones for filtering audio noise; 2) multi-sensory recording of the process of speech formation using different types of datacom (microphones, laryngophones, video cameras, etc.); 3) biometric analysis of the psychophysiological state of the speaker with the evaluation of speech capabilities and the choice of the most accessible communication channel.

The effectiveness of human-machine interaction is also related to the current state of the operator. In the works [2, 3] the technology of personified monitoring of working conditions of the personnel of industrial enterprises and industries, implemented in the interests of ensuring reliable activity and health preservation, is presented. The general scheme of a personified indicator of working conditions is presented. In the works [4, 5] the analysis of domestic patents for methods and devices for diagnosing the functional state of a human operator has been performed, showing a low innovative ability of inventions, and the forecast of the process of scientific and inventive activity indicates a decrease in the number of inventions in this branch of science and equipment for the next years.

The problem of the variability of speech in the various psychophysiological states of the speaker caused by external factors is less studied and represents the greatest complexity. To study it, it is required to create the speech databases necessary for the subsequent learning of the on-board speech recognition system. But first of all it's important to determine the hardware and software resources that can be allocated for the processing of speech audio. This will determine which generation of speech recognition systems (based on comparison of standards, hidden Markov models, artificial neural networks, etc.) can be launched on the client device.

Given the responsibility of the tasks to be solved with the help of on-board client devices, it is difficult to record training voice databases in real operating conditions. The only option for the introduction of speech technologies is the iterative procedure for the gradual modification of speech training databases, recorded primarily in an artificially recreated acoustic environment. The main steps in the formation of speech databases are: 1) classification, analysis of the amplitude-frequency characteristics of audio noise and the creation of appropriate databases; 2) an analysis of the variability of the speaker's speech caused by audio noise.

It is probably possible to organize the implementation of the first step in conditions closed to real operation. Audio recordings in the second study can be carried out in the laboratory, giving the headphones audio speaker with the specified characteristics. The implementation of additional devices for audio signals recording in real conditions, of course, significantly accelerated the process of solving the problem of noise and variability of the speaker's speech filtering.

Automatic text processing is an integral step in the formation of human-machine speech interfaces. For QA-systems, it is important that the equivalent in sense questions can be recognized as the same question, regardless of the words, style, syntactic interconnections and idioms used. To search or generate an answer to a question, a QA system must have access to some knowledge base that contains information allowing you to formulate a response.

There are two main types of QA-systems: closed-domain or specialized (with a limited thematic area) and open-domain (not limited to a particular subject area). The Open-domain QA-systems work with information in all areas of knowledge, which provides the ability to conduct search in related areas. An open-domain QA-system usually works with several sources of knowledge, in which it searches for answers depending on the class of the given question [6, 7].

The following ways of QA-systems implementing can be distinguished: on the basis of inference on ontologies [8], rules, and syntax [9], using artificial neural networks [10]. Also it is worth noting that there is the availability of approaches to improve the quality of QA-systems based on the user satisfaction score [11].

The system's response should be presented in the form of a phrase in natural language. In some cases, the simple search for the data of the copy of the communicative act is enough, that gives the question was ever used and an answer was given to it (a question-answer pair was formed).

The existing database filling technologies for QA-systems include expert filling [12], the use of crowd sourcing technologies [13], methods of procedure generation [14], automatic filling methods using existing anthologies (text corpus).

The growth of the number of public information resources in the Internet, which allows, on the one hand, the completeness of the terminological thesaurus within individual subject areas, and on the other hand, the diversity of thematic areas, has become the basis for making the assumption of the possibility of automatic analysis of texts of various content with the purpose to detect and highlight communicative acts for their subsequent entry into the database of the QA-system in the form of QA-pairs.

The joint use of the voice interface and QA-systems within the framework of human-machine interaction gives the following features:

1) the use of a closed-domain QA-system within the voice interface of interaction with the operator can expand the functionality in cases where the operator's co-command can't be directly executed, in which case the phrase is transmitted as a request to the QA system for issuing recommendations or receiving situational help. In this case, the QA system should be built on the extended thesaurus of the voice interface of a specific board system and include the basic aspects of the functioning of such a system in the base of question-answer pairs.

2) the use of an open-domain QA system operating in the voice assistant mode for issues not directly related to the operation of the on-board or mobile system (the analog of the assistants Siri, Cortana, Google Assistant, etc.) increases the process of satisfaction in communication with the system. In this case, the filling of the system can be made from available open resources, but in addition it is necessary to take into account the variability of speech and the difference in the forms of question phrases constructing.

1. The approach to the QA-system's database development

Consider the functional features of the QA-system that allows creating a database of QA-pairs, extracting knowledge from the publicly available Internet resources and providing a dialog question-answer interface in the form of web service (the block diagram is shown in Figure). As we can see on the figure, the system consists of functionally independent blocks for generating a database and for using this database to respond to user requests.

For filling the database there is a set of web crawlers and a module for collecting QA-pairs, that collect, download and analyze text documents, as well as extracting question-answer pairs from them.

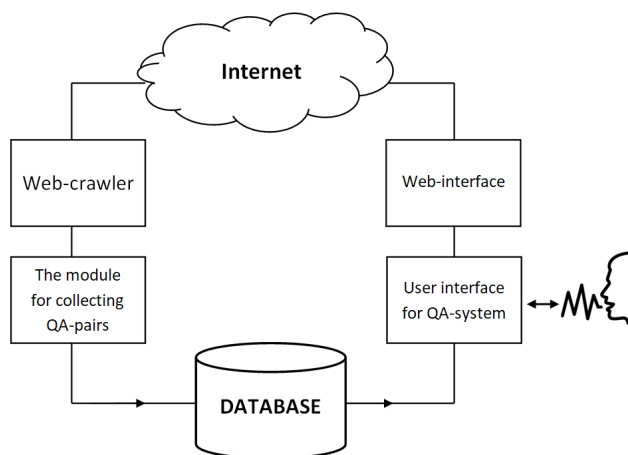
For the analysis of search queries (texts of questions) and the choice of the most relevant answer to this question, among the available question-answer pairs there is the interface search-and-dialogue component, represented on the structural diagram by the interface module of the question-answer system.

The formulation of the final answer is made by the module for responses generating (included in the interface module of the QA-system), so that the result looks syntactically natural and represents exactly what the user was looking for.

The mechanisms of decomposition of the question (user query), search and generation of the answer are considered, for example, in [15]. We will consider only methods of automatic collection of documents for filling the database (DB) of the QA-system based on the analysis of texts available in the web.

The available pages from the Internet are downloaded using web crawler technology [16]. It crawls links in processed documents according to specified algorithms, in conjunction with a headless browser that parses the original format of the downloaded document (PDF, HTML, MS Word, etc.) and converts it to text format. Additionally, the title of the document is retrieved. At this stage, the elements of the document are filtered, containing blocks of information that are not related to the main text: text blocks, navigation bars, etc.

In the work several methods of automatic



The structured scheme of QA-system with speech interface

selection of question-answer pairs were developed and tested based on the structure of sentences and on the basis of an associative-ontological approach to text analysis [16].

Before the direct allocation of question-answer pairs by any of the developed methods, the received texts are subjected to preliminary processing, in this case the graphematic analysis [17], which includes the definition of the boundaries of paragraphs, sentences and words, taking into account the structure of sentences.

The selection of sentences from the text is made by heuristic rules based on the search for the delimiter characters of the sentences: «.», «!», «?», «...» and the line transfer symbol. The words boundaries are delimited characters: « », «,», «;», «-», «(», «)», «:» и «"».

Each word in the text is being lemmatized – normalized using the function m of the morphological analysis (m -function). In this context the normalization means the obtaining the base form of the word (called 'lemma'), $w_i \xrightarrow{m} \bar{w}_i$, $w_i \in \{\mathbf{W}\}$, $i \in \mathbb{N}$, where $\{\mathbf{W}\}$ is the set of word forms, \bar{w}_i is the base form of the word, w_i is a word, \mathbf{W} – set of words, and ω is the set of valid words in the language, $i \in \mathbb{N}$. A valid set of words in the language $\{\omega\}$ is defined by some thesaurus $w_i \in \{\mathbf{W}\} \subset \{\omega\}$. The words belonging to the set of stop-words are ignored.

2. The method of QA-pairs delivering based on the sentence structure analysis

For texts such as lists of frequently asked questions (FAQs), as well as prose texts containing dialogs and direct speech, a method based on the analysis of the sentence structure obtained by preprocessing the text, expressed in the form of the following heuristic rule is effective. A sentence containing a direct speech is a sentence satisfying any of the following conditions:

- the first symbol of the sentence is the symbol «-»;
- within the sentence, a pair of symbols are sequentially located: the first character is the element of the set {«,», «.», «!», «?», «"»}, the second character is the «-» symbol;
- inside the sentence, a pair of «:» and «"» symbols are sequentially located.

From the sentences received, the author's words are deleted. The author's words are the text fragment that satisfies any of the following conditions:

- the text fragment is located after a pair of characters: the first character is the element of the set {«,», «.», «!», «?», «"»}, the second character is the «-» symbol;
- the fragment of the text is separated by the symbols «-»;
- the fragment of the text is located before the sequence of characters: «:» and «"».

Proposals that do not contain direct speech are considered in their original form, because they aren't needed in preprocessing.

Interrogative sentences are allocated from the text.

These sentences satisfy the following condition:

(the sentence contains more than two words) AND (the sentence ends with the symbol «?»).

Immediately after the interrogative sentences within one paragraph, a sentence that satisfies the conditions is selected:

(the sentence must not end with a «?» symbol) AND (the sentence contains at least one word).

Such a proposal will be considered an answer to the question posed. If any sentence in this paragraph doesn't satisfy these conditions, we believe that the question doesn't contain a response and it won't be entered into the database.

These heuristic rules can be written in the form of a generating grammar and implemented as a finite automaton.

3. The method on QA-pairs delivering based on associative-ontological approach

The method based on associative-ontological analysis is primarily focused on the class of reference and dictionary texts and it is based on the assumption that in the descriptive text there is a sentence (or a group of sentences) containing the main idea of the text. In this case, the title of the text (including that indicated through the meta tags of the online document) can be considered as a question, and this sentence (or a group of sentences) is the answer.

The use of abstracting algorithms based on the associative-ontological approach to the processing of texts in natural language [18] makes it possible to automate the selection of meaning-generating sentences through the semantic reduction of the text. The abstracting of texts is based on bi-grams, where a bi-gram is a pair of words found in one sentence. A couple of words that are often found in one sentence are considered associated, and the more often this bi-gram occurs, the stronger the connection. The sentences containing concepts, whose sum of connections is greatest, better than all other reflect the subject area described in the text.

4. The experiments and discussion

For the experimental verification of the possibility of creating an open-domain QA-system based on the automatic collection of question-answer pairs from the Internet, a prototype of the collection module, working in conjunction with the web crawler of the monitoring system for Internet resources is developed [18, 19]. The system processed 310,239 documents with useful volume of the text 1.92 GB (without taking into account the layout of the document and media data). While analyzing the texts, 2,230,325 questions and answers were received, the database size is 710 MB. The quantitative results obtained during the experimental verification of various methods are presented in Table.

Method	QA-pairs quantity
The method based on the structure sentences analysis without the direct speech registration	529117
The method based on the sentence structure analysis for direct speech	1080730
The method of QA-pairs delivering based on associative-ontological approach	310239

The greatest contribution to the formation of the database of question-answer pairs among the texts containing recorded communicative acts, mainly due to the high specific content of question-answer pairs within each document was made:

- by the prose texts containing dialogues of heroes (26 %);
- by the sections of frequently asked questions (FAQ) (17 %);
- by the reference and dictionary sources using the algorithm based on associative-semantic analysis (21 %);
- by the user generated content (UGC): forums, blogs, comments;
- by the documentary texts and news content.

Conclusion

A prototype of a system for collecting question-answer pairs was developed on the basis of the actual material contained in the public domain of the Internet.

Available pages were downloaded using web crawler technology, which crawls links in conjunction with a headless browser that parses the original format of the loaded document.

Two methods were tested for identifying question-answer pairs: a method based on analysis of the structure of sentences, and a method based on an associative-ontological analysis of texts.

Based on the analysis of the results obtained by the developed methods, it can be asserted that for a particular sample the average number of question-answer pairs was 7,9 per 1 document (one question-answer pair per 1 KB of text).

At the same time, an expert evaluation of the quality and completeness of the database, carried out using the interactive prototype, showed the impossibility of obtaining adequate answers for most of the specified search queries that the expert asked the search system without regard to the subject area.

This indicates the limited ability to create an open-domain (not specialized) QA-system only by directly identifying question-answer pairs from unstructured text sources currently available in the public domain of the Internet.

In conclusion, the authors are pleasant to express their sincere gratitude to Professor A.V. Bogomolov for his constructive criticism, a joint discussion of the problems of human-machine interaction in the framework of medical and biological research and congratulate him on the forty-fifth anniversary.

The research is granted by the budget (projects No. 0073-2014-0005 and No. 0073-2018-0002).

References

1. Kipyatkova I.S., Karpov A.A. Automatic Russian Speech Recognition Using Factored Language Models. *Artificial Intelligence and Decision Making*, 2015, no. 3, pp. 62–69. (in Russ.).
2. Bogomolov A.V., Kukushkin Yu.A. Personalized monitoring automation of the labor conditions. *Avtomatizatsiya. Sovremennye tekhnologii* (Automation. Modern technologies), 2015, no. 3, pp. 6–8. (in Russ.).
3. Zinkin V.N., Soldatov S.K., Kukushkin Yu.A., Afanasyev R.V., Bogomolov A.V., Akhmetzyanov I.M., Svidovyi V.I., Pirozhkov M.V. Hygienic evaluation of work conditions for noise-related occupations in aircraft repair plants. *Meditsina truda i promyshlennaya ekologiya* (Occupational Medicine and Industrial Ecology), 2008, no. 4, pp. 40–42. (in Russ.).
4. Goryachkina T.G., Ushakov I.B., Evdokimov V.I., Bogomolov A.V. Methodical and Methodological Recommendations for Inventors of Innovations Aimed at Assessing the Functional State of a Human Operator. *Technologies of Living Systems*, 2006, Vol. 3, no. 3, pp. 33–38. (in Russ.).
5. Kukushkin Ju.A., Bogomolov A.V., Guzij A.G. Principles of Construction of Life Support Systems of Human Controllers of Systems "Man-Machine", Adaptive to Their Functional State. *Mechatronics, Automation, Control*, 2005, no. 3, pp. 50–54. (in Russ.).
6. Lapshin V.A. Voprosno-otvetnye sistemy: razvitiye i perspektivy (Question-answer systems: development and prospects). *Nauchno-tekhnicheskaya informatsiya. Seriya 2. Informatsionnye protsessy i sistemy* (Scientific and technical information. Series 2. Information Processes and Systems), 2012, no. 6, pp. 1–9. (in Russ.).
7. Rodrigo A., Peñas A. A study about the future evaluation of Question-Answering systems. *Knowledge-Based Systems*, 2017, Vol. 137, pp. 83–93. DOI: 10.1016/j.knosys.2017.09.015
8. Zou L., Huang R., Wang H., Yu J.X., He W., Zhao D. Natural Language Question Answering over RDF: A Graph Data Driven Approach. Proc. 2014 ACM SIGMOD International Conference on Management of Data, SIGMOD'14, Snowbird, Utah, USA, June 22–27, 2014, pp. 313–324. DOI: 10.1145/2588555.2610525
9. Fader A., Zettlemoyer L., Etzioni O. Open Question Answering over the Curated and Extracted Knowledge Bases. Proc. 20th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '14, New York, New York, USA, August 24–27, 2014, pp. 1156–1165. DOI: 10.1145/2623330.2623677
10. Li J., Liu H., Zhang Y., Xing C. A Health QA with Enhanced User Interfaces. Proc. 13th Web Information Systems and Applications Conference (WISA), 23–25 Sept. 2016, pp. 173–178. DOI: 10.1109/WISA.2016.43
11. Liu Y., Bian J., Agichtein E. Predicting Information Seeker Satisfaction in Community Question Answering. Proc. 31st annual international ACM SIGIR conference on Research and development in information retrieval, SIGIR '08, Singapore, Singapore, July 20–24, 2008, pp. 483–490. DOI: 10.1145/1390334.1390417
12. Sutyagin I.V. *Molodoy uchenyy* (Young Scientist), 2012, no. 1-1, pp. 151–153. (in Russ.).
13. Fedorkova, G.S. Kraudsorsingovye tekhnologii v rossiyskikh sotsial'nykh media (Crowdsourcing technologies in the Russian social media). *Materialy Vserossiyskoy nauchno-prakticheskoy konferentsii "Kommunikatsiya v sovremennom mire"* (Proc. All-Russian Scientific and Practical Conference "Communication in the Modern World"), Voronezh, May 11–13, 2017, pp. 154–155. (in Russ.).
14. <https://www.wolframalpha.com/>
15. Nikitin A., Raykov P. *Voprosno-otvetnye sistemy* (Question-answer systems): <http://yury.name/internet/06ia-seminar.ppt> (in Russ.).
16. Kuleshov S.V., Zaytseva A.A., Markov V.S. Associative-Ontological Approach to Natural Language Texts Processing. *Intellectual Technologies on Transport*, 2015, no. 4, pp. 40–43. (in Russ.).
17. Pervushin A. Modul' grafematicheskogo analiza v sisteme obrabotki russkoyazychnykh tekstov (Module of graphematic analysis in the system for processing Russian-language texts). *Novye informatzionnye tekhnologii v avtomatizirovannykh sistemakh*, 2012, no. 15, pp. 187–190. (in Russ.).
18. Alexandrov V.V., Kuleshov S.V. Analytical Monitoring of Internet Content. Info Logical Approach. *Kachestvo. Innovatsii. Obrazovanie* (Quality. Innovation. Education), 2008, no. 3, pp. 68–70. (in Russ.).

19. Mikhailov S.N., Kuleshov S.N. Expert monitoring of unstructured content in the interest of information and analytical support of space researches. *Proceedings of the Southwest State University*, 2013, no. 6-2 (51), pp. 40–43. (in Russ).

Received May 16, 2018

УДК 51-7, 004.89

DOI: 10.14529/mmph180307

МЕТОДЫ СОЗДАНИЯ РЕЧЕВЫХ И ТЕКСТОВЫХ БАЗ ДАННЫХ ВОПРОСНО-ОТВЕТНЫХ СИСТЕМ

А.Л. Ронжин, А.А. Зайцева, С.В. Кулешов, К.В. Ненаусников

Санкт-Петербургский институт информатики и автоматизации Российской академии наук,
г. Санкт-Петербург, Россия
E-mail: ronzhin@ias.spb.su

Работа посвящена проблемам построения речевых вопросно-ответных систем (QA-систем). Предметом исследования являются подходы к автоматическому наполнению базы данных вопросно-ответной системы путем анализа неструктурированных текстовых источников, имеющих в настоящий момент времени в открытом доступе в сети Интернет.

В результате анализа выявлено, что выделяют следующие способы реализации QA-систем: на основе логического вывода по онтологиям, правилам и на основе синтаксиса, с использованием искусственных нейронных сетей.

В исследовании разработаны и протестированы методы автоматического выделения вопросно-ответных пар на основе структуры предложений и на основе ассоциативно-онтологического анализа.

Метод на основе анализа структуры предложений эффективен для текстов типа списков часто задаваемых вопросов (FAQ), а также художественных текстов, содержащих диалоги, прямую речь, основан на предварительной обработке текста, выраженный в виде эвристического правила.

Метод на основе ассоциативно-онтологического анализа ориентирован на класс справочных и словарных текстов и основан на предположении о том, что в тексте описательного характера имеется предложение (или группа предложений), содержащее основную мысль текста. В этом случае заголовок текста может считаться вопросом, а это предложение (или группа предложений) – ответом. Для автоматизации выделения смыслообразующих предложений за счет семантической редукции текста применяются алгоритмы реферирования на основе ассоциативно-онтологического подхода к обработке текстов на естественном языке.

Для экспериментальной проверки возможности создания открытой вопросно-ответной системы на базе автоматического сбора вопросно-ответных пар из сети Интернет был разработан прототип модуля сбора базы данных вопросно-ответной системы.

Ключевые слова: вопросно-ответная пара; ассоциативно-онтологический подход; текст на естественном языке; автоматическая обработка текста; распознавание речи.

Литература

1. Кипяткова, И.С. Автоматическое распознавание русской речи с применением факторных языковых моделей / И.С. Кипяткова, А.А. Карпов // Искусственный интеллект и принятие решений. – 2015. – № 3. – С. 62–69.
2. Богомолов, А.В. Автоматизация персонифицированного мониторинга условий труда / А.В. Богомолов, Ю.А. Кукушкин // Автоматизация. Современные технологии. – 2015. – № 3. – С. 6–8.

3. Гигиеническая оценка условий труда работников «шумовых» профессий авиаремонтных заводов / Зинкин В.Н., Солдатов С.К., Кукушкин Ю.А. и др. // Медицина труда и промышленная экология. – 2008. – № 4. – С. 40–42.
4. Методико-методологические рекомендации авторам инноваций по диагностике функционального состояния человека-оператора / Т.Г. Горячкина, И.Б. Ушаков, В.И. Евдокимов, А.В. Богомолов // Технологии живых систем. – 2006. – Т. 3, № 3. – С. 33–38.
5. Кукушкин, Ю.А. Принципы построения системы обеспечения жизнедеятельности операторов систем «человек-машина», адаптивных к их функциональному состоянию / Ю.А. Кукушкин, А.В. Богомолов, А.Г. Гузий // Мехатроника, автоматизация, управление. – 2005. – № 3. – С. 50–54.
6. Лапшин, В.А. Вопросно-ответные системы: развитие и перспективы / В.А. Лапшин // Научно-техническая информация. Серия 2. Информационные процессы и системы. – 2012. – № 6. – С. 1–9.
7. Rodrigo, A. study about the future evaluation of Question-Answering systems / A. Rodrigo, A.A. Peñas // Knowledge-Based Systems. – 2017. – Vol. 137. – P. 83–93.
8. Natural Language Question Answering over RDF: A Graph Data Driven Approach / L. Zou, R. Huang, H. Wang *et al.* // Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data, SIGMOD'14. – Snowbird, Utah, USA, June 22–27, 2014. – P. 313–324.
9. Fader, A. Open Question Answering over the Curated and Extracted Knowledge Bases / A. Fader, L. Zettlemoyer, O. Etzioni // Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '14, New York, New York, USA, August 24–27, 2014, pp. 1156–1165.
10. A Health QA with Enhanced User Interfaces / J. Li, H. Liu, Y. Zhang, C. Xing // Proceedings of the 13th Web Information Systems and Applications Conference (WISA). – September 23–25, 2016. – P. 173–178.
11. Liu, Y. Predicting Information Seeker Satisfaction in Community Question Answering / Y. Liu, J. Bian, E. Agichtein // Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval, SIGIR '08. – 2008. – P. 483–490.
12. Сутягин И.В. Методы формализации экспертных знаний для наполнения базы знаний / И.В. Сутягин // Молодой ученый. – 2012. – № 1-1. – С. 151–153.
13. Федоркова, Г.С. Краудсорсинговые технологии в российских социальных медиа / Г.С. Федоркова // Материалы Всероссийской научно-практической конференции «Коммуникация в современном мире». – Воронеж, 11–13 мая 2017 г. – С. 154–155.
14. <https://www.wolframalpha.com/> (Date of access: 27.12.2017).
15. Никитин, А. Вопросно-ответные системы / А. Никитин, П. Райков // URL: <http://yury.name/internet/06ia-seminar.ppt> (Date of access 27.12.2017)
16. Кулешов, С.В. Ассоциативно-онтологический подход к обработке текстов на естественном языке / С.В. Кулешов, А.А. Зайцева, В.С. Марков // Интеллектуальные технологии на транспорте. – 2015. – № 4. – С. 40–43.
17. Первушин, А. Модуль графематического анализа в системе обработки русскоязычных текстов / А. Первушин // Новые информационные технологии в автоматизированных системах. – 2012. – № 15. – С. 187–190.
18. Александров, В.В. Аналитический мониторинг Internet контента. Инфологический подход / В.В. Александров, С.В. Кулешов // Качество. Инновации. Образование. – 2008. – № 3(34). – С. 68–70.
19. Михайлов, С.Н. Экспертный мониторинг неструктурированных информационных ресурсов в интересах информационно-аналитического обеспечения космических исследований / С.Н. Михайлов, С.В. Кулешов // Известия Юго-Западного государственного университета. – 2013. – № 6–2 (51). – С. 40–43.

Поступила в редакцию 16 мая 2018 г.