Expert Assignment Method Based on Similar Document Retrieval

Denis Zubarev[1], Dmitry Devyatkin[1], Ilia Sochenkov[1], Ilia Tikhomirov[1], and Oleg Grigoriev[1]

[1] Federal Research Center "Computer Science and Control" of Russian Academy of Sciences, Moscow, Russia

Abstract. The paper describes the problem of expert assignment. Based on the analysis of methods that are currently used to solve this problem, the main shortcomings of these methods were identified. These shortcomings can be eliminated by analysing large collections of documents whose authors are potential experts. The article describes the method of compiling a ranked list of experts for a given document, using similar document retrieval. To evaluate the proposed method, we used a collection of grants applications from a science foundation. Experimental studies show that the more documents are available where experts are authors, the better the performance of the proposed method becomes. In conclusion, the current limitations of the proposed method are discussed, and future work is described.

Keywords: Scientific expertise, expert assignment, unstructured data analysis, text analysis, similar document retrieval

1 Introduction

A competent and objective examination of applications for grants and scientific publications is a prerequisite for scientific progress. But it requires a competent and objective selection of experts. Currently, in most cases, the appointment of experts is based on manually assigned codes from manually created classifiers or manually chosen keywords. Experts and authors independently assign codes or keywords to their profiles or documents (application for a grant, report on a grant, an article, etc.), and the appointment of an expert is carried out by comparing the assigned codes or keywords. Classifiers are rarely updated, so they quickly become obsolete, have uneven coverage of the subject area (one code can correspond to thousands of objects, and the other to dozens) and have all the other drawbacks of manual taxonomies. In addition, experts often assign themselves several codes, but their level of competence varies greatly between these codes [1]. If there are several dozen experts who correspond to the same code (which happens quite often), then the further choice will be extremely subjective and non-transparent (in fact, manual selection of an expert is performed). All this leads to insufficient compliance of the competence of the selected expert and the object of examination and, possibly, to the subjective choice of the expert. As a result, there are refusals of examination, or it is conducted incompetently and, possibly, subjectively. Therefore, it is important not to determine the formal coincidence of the expert interests and the expertise subject topic, but to use all possible information for accurate expert ranking.

Information about expert competence is accumulated in the documents in which he participated (scientific articles, scientific and technical reports, patents, etc.). This information is much more precise in determining the expert's knowledge area than the classifier codes or keywords. This article describes the method of searching and ranking experts for a given object of expertise using thematically similar documents retrieval. The method requires a database of experts and a large set of texts associated with experts. It is assumed that this method will become the basis for a whole class of methods that use unstructured information to select experts for the objects of expertise.

2 Related Works

Automating the search for an appropriate expert for examination has long been a subject of research. Researchers often narrow the research scope, for example, limiting themselves only to the appointment of experts to review articles submitted to the conference [2], or to select an expert who will answer user questions on the corporate knowledge base [3].

As a rule, expert assignment methods are divided into two groups [4]. The first group includes methods that require additional actions from experts or authors. For example, one of the methods involves the examination of the submitted abstracts of articles and the self-assessment of his readiness to consider any of the works in question. Another involves the selection by an expert of keywords that describe his competence from the list provided by the conference organizers and comparing these keywords from the expert with the keywords chosen by the authors of the article. These approaches are well-suited for small conferences but are not suitable for events in which several tens of thousands of participants take part. Even with relatively small conferences, the use of keywords is inappropriate if the number of topics for this event is large enough.

The second group includes methods that automatically build an expert's competence model based on his articles and / or other data and compare the resulting model with peer-reviewed articles submitted using the same model [5]. In this work, the name and surname of the expert were sent as a request to Google Scholar and CiteSeer. For the full texts of the articles found and the article submitted, the Euclidean distance was measured. This method does not take into account the namesakes, the dynamics of changes in the expert interests, a possible conflict of interests and requires significant computational resources. Another method [6] uses annotations and titles to classify articles according to topics predetermined by the conference organizers. However, it is not always possible to pre-determine a specific set of topics. The method presented in [7] uses bibliographic data from the reference list of the presented article. First names and surnames of authors are mined from bibliographic references, co-authors are determined for them using external resources (DBLP), etc. Thus, a co-authorship graph is constructed, on which a modification of the page ranking algorithm for identifying experts is performed. In [8] a special similarity measure that compares the reference lists are used to determine the proximity between expert publications and article submission. The comparison takes place under headings and authors, it also takes into account the case when the expert's articles are cited in the presented article. Bibliographic list comparison is a fairly effective operation, but it is difficult to assess the expert's competence only by bibliographic references, without using full texts. In [9] topic modelling is used to represent an object of examination and each document associated with an expert. Topic distribution of an expert is adjusted according to the time factor that is meant to capture the changes of research directions of an expert as time goes on. Cosine measure is used to measure the similarity between the expert's topic distribution and the topic distribution of the object under review. Furthermore, vector space model (with TF-IDF weighting scheme) is used to calculate an additional similarity score between experts and the object of examination. The final score for relevance is calculated using a weighted sum that takes into account the two previous scores. In the experimental studies of this work, the number of topics was chosen to be 100, which, according to the authors, reflects the real number of topics in information technology knowledge area. In this study, words are used as features.

In [10] a hybrid approach is used that combines full-text search (performed using ElasticSearch) over experts' articles and an expert profiling technique, which models experts' competence in the form of a weighted graph drawn from Wikipedia. The vertices of the graph are the concepts extracted from the expert's publications with TagMe tool. Edges represent the semantic relatedness between these concepts computed via textual and graph-based relatedness functions. After that, each vertex is assigned a score corresponding to the competence of the expert. This score is computed employing a random walk method. Concepts with a low score are removed from the expert's profile, due to the assumption that they cannot be used to characterize his competence. Also, the vertices are assigned a vector representation which is learned via structural embeddings techniques on concepts graph. At query time, the object of examination is parsed with TagMe tool, and embeddings are retrieved for extracted concepts, then they are averaged. As a result, a cosine measure is used to measure the similarity between averaged expert's vectors and vector that represents the object of the expertise. The final list of relevant experts is obtained via combining full-text search results and results of semantic profiles matching. It should be noted that impact of semantic profiles is rather small. According to the results of experimental studies conducted in this work, the increase in the quality assessment using the expert's semantic profile was 0.02, compared to the use of full-text search with the BM25 ranking function. This method was tested on a dataset [11], in which short phrases describe areas of knowledge (GT5). These phrases were used as queries (objects of expertise).

Thus, the existing methods for the expert search do not use all available information related to this task. Some methods are limited to processing only bibliographic lists or annotations with titles while ignoring the full texts of articles. Others are based on full text analysis of articles but they use ordinary full-text retrieval tools that apply to simple keyword search and are not effective for thematically similar documents retrieval. In addition, it should be noted that some of the methods described are computationally expensive since they do not use efficient means of indexation, and when selecting experts for each new object, it is necessary to repeat complex computational operations. In the approach proposed in this article, the main part of computationally intensive operations is performed only once.

3 Expert Assignment Method

The first step of the proposed method is the search for thematically similar documents for a given object of examination (application) [12]. Search is made on the collections of scientific and technical texts. These can be scientific articles, patents and other documents related to experts. The collections are pre-indexed. Before indexing the text undergoes a full linguistic analysis: morphological, syntactic and semantic [13, 14]. Indexes store additional features for each word (semantic roles, the syntactic links and so on) [15]. During indexing, several types of indexes are created, including an inverted index of words and phrases, which is used to search for thematically similar documents. Indexing is incremental; that is, after initial indexing, one can add new texts to the collection without re-indexing the entire collection [15].

When searching for thematically similar documents, the given document is represented as a vector, elements of which are TF-IDF weights of keywords and phrases. Phrases are extracted based on syntactic relations between words. This allows extracting phrases consisting of words that are not adjacent to each other but have a syntactic connection. For example, the phrases "images search" and "digital images" will be extracted for the fragment: "search for digital images". The degree of similarity is calculated between the vector of the original document and the documents vectors from the index. Some similarity measure is used to calculate the degree of similarity (we tried cosine and hamming distance). The main parameters of the search method for thematically similar documents are presented in Table 1.

Based on the list of thematically similar documents, a list of candidates for experts is compiled. This is a trivial operation since the documents relate to the expert: the expert is one of the authors, he reviewed this article/application, etc.

After that, if there is the necessary meta-information, the experts are excluded from the list of candidates according to various criteria. For example, if meta-information about belonging to organizations is available for a peer-reviewed document and an expert, some experts could be filtered out because of a conflict of interest. At present, this step depends on the available meta-information, and it is related to the type of reviewed document. The experimental implementation of the method used several filters that are appropriate for grant applications:

1. All experts who are involved as participants in the given application are excluded.
2. All experts working in the same organizations as the head of the given application are excluded.

Table 1. The main parameters for thematically similar documents search method

Description	Name
The percent of words and phrases in the source document that determine the similarity of documents	TOP_PERCENT
The maximum number of words and phrases that are used to determine document similarity	MAX_WORDS_COUNT
The minimum number of words and phrases that are used to determine the similarity of documents	MIN_WORDS_COUNT
Minimum TF-IDF weight of a word or phrase included in the top keywords of the document	MIN_WEIGHT
The minimum value of the similarity score	MIN_SIM
The maximum number of similar documents for the source document	MAX_DOCS_COUNT

After that, the relevance of each expert to the object of expertise is calculated. The calculation takes into account the similarity of the documents (S_{sim}), with which the expert is associated, to the reviewed document, as well as several additional measures. In case if the expert has multiple documents, their ratings of similarity are averaged out. The set of additional measures depends on the type of the reviewed document. In the implemented method, one simple measure (S_{sci}) was used: the equality of the knowledge area code assigned to the expert and to the document under review (0 when the codes are not equal and 1 otherwise). The overall relevance score of the expert is calculated using the following formula:

$$W_{sim}\,S_{sim} + W_{sci}\,S_{sci},$$

where S_{sim}, S_{sci} are values of the measures described above, as W_{sim}, W_{sci} are weights with the condition $W_{sim} + W_{sci} = 1$. S_{sci} criterion was useful in ranking experts who were heads of interdisciplinary projects. An interdisciplinary project can relate to several scientific areas, but the head is an expert in only one area, so he should be ranked lower than the experts who have the same area of knowledge. The score of the relevance of each expert may lie in the interval [0;1]. After evaluation, experts are ranked in descending order of relevance.

4 Description of the Experimental Setup

4.1 Dataset Description

As a result of cooperation with the Russian Foundation for Basic Research, it was possible to conduct a series of experiments on the applications accumulated by the Foundation in various competitions held from 2012 to 2014. The Fund provided an API for indexing the full texts of applications. The application text included:

- summary of the project;
- description of the fundamental scientific problem the project aims to solve;
- goals and objectives of the study;
- proposed methods;
- current state of research in this field of science;
- expected scientific results;
- other substantive sections that are required by the competition rules.

For each application, a meta-information containing the following fields was provided:

- document identifier;
- the identifier of the head (principal investigator);
- identifier of the organization in which the head works;
- a list of the identifiers of participants (co-investigators);
- coded participants full names;
- publication year of the grant application;
- code of the field of knowledge which the application belongs to (Biology, Chemistry, etc.));
- main code and additional application codes;
- keywords of the application.

There was also presented impersonal information about the experts who reviewed the applications:

- expert identifier;
- identifier of the organization which the expert works in;
- expert keywords;
- code of the main area of knowledge of an expert;
- applications which the expert is the head in (list of identifiers);
- applications which the expert is the participant in (list of identifiers);
- applications reviewed by the expert (list of identifiers);
- applications the expert refused to review (list of identifiers).

The size of the collection of applications was about 65 thousand documents. Information was also received about 3 thousand experts, where the share of experts who were the head (principal investigator) of at least one application was 78%. At first, it was supposed to use only the applications of experts, in which they were principal investigators. However, it turned out that the share of such documents was about 9% among all grant applications. Moreover, most of the experts were associated with only one grant project. To increase the number of documents associated with the experts, we took into account applications in which the expert participated as co-investigator. We also used an external collection of scientific papers, which mainly consisted of articles from mathnet.ru and cyberleninka.ru, to search for additional experts publications. First, we looked for documents that confirm the support of grants with the participation of the expert (the grant identifier is usually written in the acknowledgments section). This provided us with about 4,000 additional documents. In addition, we performed a search for similar works for each expert application. To filter documents that are similar, but not related to experts, we compared the full names of the authors of the article with the full names of the applicants. If at least one full name corresponded, then we considered that this document is associated with an expert. Usually, there are no full names of the authors of the article, there are only short names (last name with initials), then there should be at least two matches with the short names of the applicants. We received about 30,000 new documents related to experts, using the search for similar documents.

Since the names of authors of papers are not structured and are presented as text, we parsed names into their individual components. We will briefly describe the parsing method. First, given the input string that contains the name of the author, the type of pattern is identified. Multiple patterns are supported:

1. The Slavic pattern includes several variations:
   a. Last name[,] First name [Patronymic];
   b. Last name Initials;
   c. First name [Patronymic] Last name;
   d. Initials Last name;
2. The Western pattern consists of several variations:
   a. First name [Middle]… Last name
   b. Last name, First name [Middle]…
   c. First name Initial [Initial]… Last name
3. Spanish pattern similar to the western one, except that there may be two last names:
   a. First last name [Second last name], First name [Middle]
   b. First name [Middle]… First last name [Second last name]
4. Asian pattern:
   a. Last name First name [First name]…

This classification is necessary because the parser can match full names with several patterns, e.g. 1.1 and 2.1. As a training set, we used the names of public persons and the country of their citizenship obtained from Wikidata dump, and also added the names with countries obtained from Russian patents (www1.fips.ru). We trained Fasttext classifier on that dataset and obtained 0.96 precision@1 on the test data. When a pattern is identified for the given input text, then all variations available for this pattern are tested. If there is only one matching option, the parsing is complete. If more than one option matches, for example, 1.1 and 1.3; we use the common first names dictionary to determine the right variation.

After performing these procedures, the share of experts with documents increased to 88%. In addition, the number of experts associated with only one document was significantly reduced, as can be seen in Table 2.

Table 2. Distribution of the number of documents associated with an expert including extra documents

Number of documents per expert	Number of experts	Number of documents per expert	Number of experts	Number of documents per expert	Number of experts	Number of documents per expert	Number of experts
1	115	11	50	21	24	31	9
2	101	12	45	22	18	32	14
3	108	13	39	23	26	33	13
4	94	14	39	24	24	34	9
5	83	15	34	25	12	35	7
6	76	16	34	26	19	36	6
7	79	17	36	27	17	37	11
8	64	18	31	28	14	38	3
9	66	19	28	29	16	39	9
10	55	20	19	30	9	40	11

4.2 Evaluation Methodology

To assess the proposed method, data from previous expert selections of applications for participation in the A-2013 competition was used (total of 10,000 applications, an average of 3 experts per application). For every application from this competition, a ranked list of experts (found experts) was compiled using the proposed method. Then this list of experts was compared with the list of experts assigned to the given application.

There are common metrics that are used for evaluation of the experts search methods [16, 17]. Some of these metrics are applicable only if expert assignment goes along with the expert search. Those metrics assess the uniformity of the expert load and the assignment of a certain number of experts to each object of expertise. Each expert search provides a pool of relevant experts for further assignment. Therefore, this task should be evaluated using other metrics. Classical information retrieval metrics are frequently used: MAP, NDCG@100. Using these metrics is justified if the test data contains a large number of relevant experts for each object of expertise. We used a data set of up to 3 relevant experts for the object of expertise. This number of experts is not enough to correctly interpret the assessment results for a large number of selected experts. (like 100). Therefore, recall was used for evaluation in order to determine what total share of relevant experts was in the pool of selected experts. Recall was calculated using the following formula:

$$Recall = \frac{F_{found}}{F_{total}},$$

where F_{found} is the number of found experts from among those that have been assigned to this application; F_{total} is the number of assigned experts that could be found by the method (i.e. only experts that have at least one associated document).

Micro averaging was used to calculate metrics for all applications (i.e. for all applications are summarized F_{found} and F_{total}, and based on this, the required metric was calculated). Also, recall was calculated separately for each knowledge area.

The standard way of measuring precision in this situation is not appropriate since it is not known whether the found expert that has not been assigned to this application is suitable. He might be suitable for this application but was not assigned because he was busy on other projects or for other reasons.

Therefore, to evaluate precision, the information on this application expertise refusals was used. There were about 2 thousand of refusals according to the provided data. The idea is, that the compiled experts list shouldn't contain those who refused to review this application. The precision was calculated using the following formula:

$$Precision\ = 1 - \frac{R_{found}}{R_{total}},$$

where R_{found} is the number of found experts from those who refused to expertise the application; R_{total} is the number of refused experts that could be found by the method (i.e. experts with documents).

4.3 Parameters Optimization

Optimization of the algorithm parameters was performed on a separate sample collection of 700 applications. For optimization, we used a grid search of a single algorithm parameter with a fixed value of the remaining parameters. Optimization was performed to maximize recall. The results are presented in Table 3.

4.4 Experimental Results

We conducted multiple experiments with different similarity measures (cosine, hamming) and a different set of features (only words, words with phrases). Also, we used different datasets: at first, experts were associated with applications, in which they are the head (only-head); then additional documents were added (extra-docs). Table 4 shows the micro-averaged recall and precision on the top 150 for those experiments.

Hamming distance along with adding word phrases result in the best recall on two datasets. New documents addition associated with experts (extra-docs dataset) increased the value of recall, but decreased the precision by almost the same value. As we discussed earlier, MAP is not the best metric for this task. Its value depends on the experts ranking, but this ranking is better when smaller dataset is used (only-head). Using more docs associated with experts (extra-docs) gives greater recall but lesser MAP. It should be borne in mind that experts found should be distributed over several dozens or hundreds of applications, and several experts are usually appointed for each application. Therefore, each application requires a sufficiently large pool of relevant experts. Therefore, recall is a more important metric for this task than MAP.

Table 3. Values of method parameters after optimization

Name	Value
`TOP_PERCENT`	0.4
`MAX_WORDS_COUNT`	200
`MIN_WORDS_COUNT`	15
`MIN_WEIGHT`	0.03
`MIN_SIM`	0.05
`MAX_DOCS_COUNT`	500
W_{sci}	0.1
W_{sim}	0.9

Table 4. Evaluation results

	Only-head			**Extra-docs**		
	Recall	**Precision**	**MAP**	**Recall**	**Precision**	**MAP**
Cosine, only words	0.67	0.52	0.136	0.73	0.43	0.098
Cosine, with phrases	0.69	0.51	0.139	0.75	0.42	0.108
Hamming, only words	0.69	0.52	0.141	0.76	0.4	0.107
Hamming, with phrases	0.7	0.5	0.148	0.77	0.41	0.123

Since the result of the method is a ranked list of experts, it is possible to plot a graph of recall and precision (Hamming, with phrases) shown in Fig. 1.

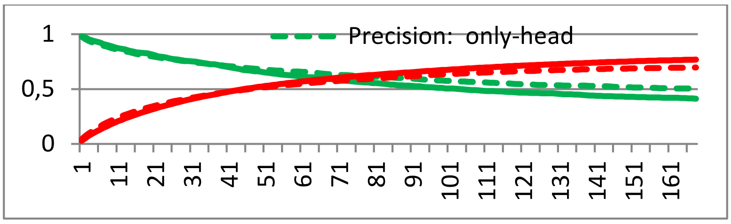

Fig. 1. The dependence of recall and precision on the rank

The graph shows that the maximum recall is achieved at 100–120 rank, after that, recall almost does not change. Also, this graph shows that ranking on smaller dataset (only-head) is better, because recall values on low ranks (1–40) are better than extra-docs dataset.

Recall values were also calculated for each area of knowledge separately, and the results are shown in Table 5.

Table 5. Recall for each knowledge area

	only-head	**extra-docs**
Mathematics informatics and mechanics	0,69	0,79
Physics and astronomy	0,71	0,81
Chemistry	0,74	0,79
Biology and medical science	0,74	0,75
Earth Sciences	0,76	0,87
Human and Social Sciences	0,77	0,92
Information technologies and computer systems	0,55	0,61
Fundamentals of Engineering	0,57	0,64

The table shows that the best results were obtained in the fields of Earth Science (5) and the Human and Social Sciences (6) – recall is about 90%. In other areas, good results were obtained (recall from 70% to 80%). Average results were obtained for the fields of Information technologies (7) and Fundamentals of engineering (8). In addition, the figure shows the increase of the number of documents related to experts has a positive impact on assignment of experts recall in all knowledge areas.

5 Conclusion

In this paper, the expert appointment method based on the analysis of text information was described, and the results of method evaluation experiments were presented. We proposed a new evaluation methodology and conducted experiments on the RFBR data set, which distinguishes this work from the previous ones. The method showed its viability, but it is necessary to improve it in order to increase the recall. Adding more expert-related texts improved the review somewhat, but not as dramatically as expected. According to Fig. 2, there are still many experts that have only one document authored by them. It may be viable to add documents related to these experts in the first place: scientific publications, scientific and technical reports, patents, etc. It is also possible to expand the list of documents indirectly related to the expert, with the exception of authorship, for example, articles that he reviewed. These documents should contribute to the overall score of relevance with a lower ratio since the expert has no direct relationship to the text, however, if he regularly reviews the papers of a certain topic, it should be taken into account when scoring.

In further experiments, it is also proposed to expand the set of criteria that affect the expert's assessment for a given object of examination, for example, add an expert rating calculated using a page ranking algorithm based on quotes from expert works.

Further studies are also expected to improve the methodology for evaluation of the expert’s assignment. In the technique proposed in the article, there are several shortcomings, namely: the dependence of recall on the original expert assignment, which could be subjective; the inability to assess the selected experts, which were not involved in the expertise of this proposal, which makes it impossible to calculate precision of the selection. Precision measurement based on refusals of expertise is also not optimal. Cases of refusal are 15 times less than cases of acceptance, and refusal can occur for other reasons than a mismatch between the competence of the expert and the subject of the application. However, the question of how to evaluate the work of the expert assignment method is currently unresolved. The involvement of external experts can significantly improve the quality of the evaluation, but it will require a large number of experts from different knowledge areas, who should be well acquainted with the expert community.

The proposed method can be used not only when appointing an expert for grant application of a scientific fund but also in the reviewer selection for any text object: articles in scientific journals, conference abstracts, patent applications, etc.

Acknowledgements

The research is supported by Russian Foundation for Basic Research (grant No. 18-29-03087).

References

1. V Rossijskom nauchnom fonde proshlo zasedanie ehkspertnogo soveta po nauchnym proektam [The Russian Scientific Foundation held a meeting of the expert council on scientific projects]. Available at: http://rscf.ru/ru/node/2367 last accessed 2019/08/16.
2. Dumais, Susan T. and Nielsen, Jakob: Automating the assignment of submitted manuscripts to reviewers. Proceedings of the 15th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 233–244 (1992).
3. Balog, Krisztian, Leif Azzopardi, and Maarten De Rijke: Formal models for expert finding in enterprise corpora. Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval. ACM. 43–50 (2006).
4. Kalmukov, Yordan and Rachev, Boris: Comparative analysis of existing methods and algorithms for automatic assignment of reviewers to papers. arXiv preprint Available at: https://arxiv.org/pdf/1012.2019.pdf last accessed: 2019/05/11 (2010)
5. Pesenhofer, Andreas, Mayer, Rudolf, and Rauber, Andreas: Improving scientific conferences by enhancing conference management systems with information mining capabilities.

Digital Information Management, 2006 1st International Conference on. IEEE, 359–366 (2006).
6. Ferilli, Stefano, et al.: Automatic topics identification for reviewer assignment. International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems. Springer, Berlin, Heidelberg, 721–730 (2006).
7. Rodriguez, Marko A., and Bollen, Johan: An algorithm to determine peer-reviewers. Proceedings of the 17th ACM conference on Information and knowledge management. ACM, 319–328 (2008).
8. Li, Xinlian and Watanabe, Toyohide: Automatic paper-to-reviewer assignment, based on the matching degree of the reviewers. Procedia Computer Science **22**, 633–642 (2013).
9. Peng, H. et al. Time-aware and topic-based reviewer assignment. International Conference on Database Systems for Advanced Applications. Springer, Cham, 145–157 (2017).
10. Cifariello, P., Ferragina, P., and Ponza, M.: Wiser: a semantic approach for expert finding in academia based on entity linking. Information Systems **82**, 1–16 (2019).
11. Berendsen, Richard, et al.: On the assessment of expertise profiles. Journal of the American Society for Information Science and Technology **64** (10), 2024–2044 (2013).
12. Sochenkov, I.V., Zubarev, D.V., and Tihomirov, I.A.: Eksplorativnyj patentnyj poisk [Exploratory patent search]. Informatika i ee primeneniya [Informatics and its Applications]. **12** (1), 89–94 (2018).
13. Osipov, Gennady, et al.: Relational-situational method for intelligent search and analysis of scientific publications. Proceedings of the Integrating IR Technologies for Professional Search Workshop, 57–64 (2013).
14. Shelmanov, A.O. and Smirnov, I.V.: Methods for semantic role labeling of Russian texts. Computational Linguistics and Intellectual Technologies. Proceedings of International Conference Dialog **13** (20), 607–620 (2014).
15. Sochenkov, I.V. and Suvorov, R.E.: Servisy polnotekstovogo poiska v informacionno-analiticheskoj sisteme (Chast' 1) [Full-text search in the information-analytical system (Part 1)]. Informacionnye tekhnologii i vychislitel'nye sistemy [Journal of Information Technologies and Computing Systems] **2**, 69–78 (2013).
16. Li, L., Wang, L., Zhang, Y.: A comprehensive survey of evaluation metrics in paper-reviewer assignment. Computer Science and Applications: Proceedings of the 2014 Asia-Pacific Conference on Computer Science and Applications (CSAC 2014), Shanghai, China, 27–28 December 2014. CRC Press, 2015. P. 281.
17. Lin, S. et al. A survey on expert finding techniques. Journal of Intelligent Information Systems **49** (2), 255–279 (2017).