

# Algorithm for Searching Experts in Scientometric Systems

Alexander S. Kozitsin, Sergey A. Afonin, Dmitry A. Shachnev

*Institute of Mechanics Lomonosov Moscow State University, Michurinskiy pr., 1, Moscow, 119192, Russia*

## Abstract

The use of modern methods of thematic analysis for the analytical processing of large amounts of information is currently used in almost all spheres of human activity. Scientometrics is one such area. Research carried out in this area shows that the statistical characteristics of citation rates are highly dependent on the thematic area of papers at hand. In this regard, systems for calculating citation indicators and constructing assessments of the quality of scientific works, ranks of authors or organizations should take into account the thematic area of researchers in order to normalize scientometric indicators.

Many scientometric systems and citation systems use subject headings for information retrieval and processing. The world famous citation systems, including WoS, Scopus, Google Scholar, and others have such headings.

Important practical tasks that can be solved using the methods of thematic classification are: assessment of the dynamics of the development of thematic areas in an organization, in a particular country and in world science as a whole; search for articles on a given topic; search and assessment of the authority of experts; search for journals for publication and other urgent tasks. The authors have created software implementations of algorithms for solving some of the listed problems, and research is being conducted to create new effective mathematical models and algorithms in this area.

## Keywords

thematic search, bibliographic data, expert search, information systems, scientometrics

## 1. Introduction

With the development of communications and remote interaction systems, the task of finding highly qualified specialists and experts to perform work remotely is becoming increasingly important. Such work can be short-term, intermittent, or performed on an ongoing basis. Currently, the search for experts is mostly demanded in the area of media. Journalists and publishers are looking for consultants on the subject of published articles. The owners of news and information portals are looking for specialists to write articles in the given thematic areas. In the manufacturing sector and in software development, experts are also being sought to conduct one-time consultations, to work on an ongoing basis in startups and developing projects. Such searches are usually done through social networks or through task exchanges. An active search for experts through social networks is carried out by publishing announcements and searching among acquaintances or in specialized communities, for example, communities [1, 2], containing twelve thousand members each. For the media, there are also article exchanges where you can find a performer to write an article on a given topic [3–5]. The search for specialists on such exchanges is carried out by placing orders. Customers post assignments, and experts post descriptions of their competencies. Specialists must be registered on the exchange and regularly review orders on their topic. This search method is effective for finding specialists of not very high qualifications who specialize in writing texts for the media and Internet resources in order to earn money. In most cases, the qualifications of specialists are assessed by the number of orders sold. To successfully conduct a search in such systems, it is necessary to actively participate in the search process both for the customers of the work and the experts themselves.

---

SSI-2021: Scientific Services & Internet, September 20–23, 2021, Moscow (online)

EMAIL: alexanderkz@mail.ru (A.S. Kozitsin); serg@msu.ru (S.A. Afonin); mitya57@gmail.com (D.A. Shachnev)

ORCID: 0000-0002-8065-9061 (A.S. Kozitsin); 0000-0003-3058-9269 (S.A. Afonin); 0000-0002-5940-9180 (D.A. Shachnev)



© 2021 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

There are specialized sites that allow experts to exchange their contacts. For example, the *deadline.media* project, designed to find specialists for the media. Each registered user independently marks the topics in which he is an expert, and his contacts will be highlighted when searching by headings. The classifier is single-level, contains about eighty headings, for example “science, innovation and technology”, “incidents”, “culture and art”. The contact database contains about fourteen thousand users of varying degrees of relevance. You can also mention the following English-language services: *www.profnetconnect.com* for finding PR specialists; the currently closed project of searching for experts in arbitrary fields *www.allexperts.com*; service for the analysis of materials by volunteer experts before publication *pressfeed.ru*. The interest of experts in such work may be not only financial. For example, an expert may be interested in additional mentioning of his name in the media.

Similar exchanges are created to find an expert when creating projects. For example, a site for finding experts *expertme.ru*. The organizer of the project indicates the description of the project and the requirements for the expert, while the experts, when searching, indicate their competencies and review the proposals of the customers. An expert's qualifications are assessed only on the basis of the description provided by him and the subsequent personal interview.

A common disadvantage of the above manual and automated methods is the need for active participation in the search for both parties and the lack of criteria for assessing the qualifications of an expert. The assessment is made either on the basis of a subjective assessment of a small number of other people, or on the basis of the activity of work on a search site, which has nothing to do with the real qualifications of an expert.

The task of thematic search for experts in the scientific field differs significantly from the tasks described above for searching for the media and developing software. Such a search is in demand in the selection of reviewers for scientific articles, in the selection of experts in competitions, in the search for highly qualified consultants for scientific work and in the selection of scientific advisors by students. In most cases, with such a search, experts are not ready to take an active part in the search process. A world-renowned scientist may agree to read and give a review of an article at the request of the editors, but he will not independently search for articles sent to the editors for publication, on which he could give reviews. In this regard, the use of exchange systems for thematic search in the scientific field is practically impossible, and it is required to use other principles and search mechanisms.

## 2. Overview of existing systems

The easiest way to find specialists is to search for competency descriptions in an employee's profile. For example, in the project of searching for experts among employees of US and Canadian universities [6], a keyword search is carried out among employees of US and Canadian universities registered in the system. Contextual search is carried out only according to the descriptions of the competencies, which were presented by the employees themselves in the description of their profile. Any assessment of the conformity of the declared subject matter and real scientific activity is not made. Experts are invited to fill out a list of their competencies on their own in their personal profile. The search engine searches for the specified competency names using standard keyword search methods. This approach is easy to implement, but has a very low search recall. In their competencies, employees indicate general concepts, but do not indicate special terms. For example, in the system [6], you can find forty registered experts with the “big data” competence, but you cannot find a single specialist with the “k-means” or “SVM” competencies. In addition, this method does not allow assessing the qualifications of experts and ranking search results.

There are a significant number of researchers' social networks that allow their publications to be uploaded, shared and searched. The *Academia.edu* project allows you to upload your articles to your profile and search for other people's articles (full-fledged full-text search is available only on a paid basis). The *ResearchGate* project additionally organizes joint work on articles. The *www.bibsonomy.org* project allows you to download and index Russian-language articles, however, in a very small volume. For ten years, it contains only fourteen Russian-language articles containing the word “search” in Russian.

Better methods of searching for experts are implemented based on the analysis of publication data. Such data are contained in citation systems and scientometric systems. Projects implementing this search method can be divided into three groups.

Projects of the first group make it possible to carry out a thematic search of publications with subsequent manual analysis of information about the authors. For example, the *CiteSeerX* project of the Uni-

versity of Persylvania [7] contains about one million documents and allows you to search for articles by keywords and authors by last name. Each author's profile contains a selection of their articles. The Semantic Scholar project [8] contains forty million articles and enables keyword searches. Publishers' websites provide the ability to search by keywords and bibliographic data for paid access to the text of articles [9–11]. The disadvantage of such projects is the lack of an assessment of the author's relevance to the search query. The user needs to find articles that match the request, and manually iterate over all co-authors to assess their competence.

Projects of the second group not only allow you to search for articles, but also show aggregated data on the authors of publications. The Digital Library project [12] allows you to search for articles by keywords and shows, when viewing the results, a list of authors of all found articles, indicating the number of articles for each author. This service has a user-friendly search interface, however, the search base contains three million articles only.

The Google Scholar project [13] has the most complete index of articles, including many articles in Russian. When searching by keywords, it shows the number of citations of articles according to its data, which significantly exceeds the number of citations according to other systems. Following the global Google concept of training the system by the users themselves, the service provides the user with the ability to create their own profile and add articles to the profile from the results of a search query or by filling out the article description form manually. Thus, the user can add Russian-language articles to his profile, but the information is not verified. Profiles are missing or incomplete for most Russian-speaking authors. The assessment of the author's compliance with the subject of the request is also done manually by enumerating all the co-authors of the found works.

The Microsoft Academic Graph project [14] allows you to search for experts based on the analysis of scientific publications. The user is provided with a selection of articles relevant to the query based on the keywords of the search query. For the generated collection, the user is shown the distributions: according to the list of the most popular topics for these articles; by type of publication; by magazines; by scientific organizations; by conferences; by sections of classifiers and by authors. Each element of these lists contains an indicator of aggregated relevance to the user's request and can be used for additional filtering. Thus, you can select the author or organization that is closest to the user's request and view a list of their publications. The project is in English, does not process requests in Russian and does not analyze Russian-language articles. Correction by the authors of data on their works is not provided.

The third group of projects includes projects with the ability to search for author profiles by keywords. The most famous project of this group is the Chinese project AMiner [15]. To conduct a thematic analysis, it uses data on publications of authors, received patents and participation in projects. A search query based on the selected keywords can be refined using one of forty headings of an additional thematic classifier. The list of found authors can be ranked both by their correspondence to the text of the request, and by various indicators of the author's authority. For each author, the list of search results additionally shows the total number of registered articles of the author, the number of citations found and a list of automatically built competencies. For detailing and additional acquaintance with the author's competencies, it is possible to go to his automatically built profile, which presents: the distribution of the selected competencies by years of publication; list of co-authors; list of publications, patents and projects of the author. A list of articles, projects and patents of the expert is available for viewing in separate tabs. Presented are neighboring vertices from the co-authorship graph (Ego Network). Additionally, the distribution of the number of articles with given keywords by years is presented. For example, you can see that the popularity of “K-means” and “Elliptic curves” in open scientific articles was quite high between 2001 and 2011, and then began to decline. And the popularity of “SVM”, on the contrary, began to grow after 2008.

The most serious drawback of the above projects from the point of view of Russian-speaking users is the complete absence or extremely insignificant amount of Russian-language articles used for the analysis. This fact is especially important for experts in the humanitarian field, since the results of the scientific activities of many well-known Russian scientists in the humanitarian field are poorly represented in English-language journals indexed by Web of Science.

In addition, problems with the support of the Russian language entail difficulties with identifying names and surnames in English-language articles, which leads to incorrect creation of profiles and incorrect linking of articles to authors. For example, for article [16], the system defines the authors “V. Valery, Z. Maxim, K. Alexander, K. Maxim, S. Dmitry” [17]. The bibliographic data processing system has mixed up first and last names. The system has left the names in full and shortened the last names to one letter. Also, the definition of organizations does not work correctly in many cases. For example, for article [18] out of five working at the Moscow State University. M.V. Lomonosov co-authors of the article,

three authors are assigned to the organization “Moscow State University”, and two to “Michigan State University” [19]. Such errors in the identification of authors lead to a significant deterioration in the quality of search among Russian-speaking authors. Another disadvantage is the lack of convenient filtering by territory or language. In many cases, the user needs to find not the best expert in the world on the chosen topic, but a good enough expert with whom he can consult.

It should be noted that the AMiner and Microsoft Academic Graph projects provide users with the opportunity to download their data for independent analysis: bibliographic descriptions of articles (authors, title, journal, pages, and others); keywords; links to the pdf-version and annotations (if any), as well as descriptions of the authors. Data are provided for 2018 only. The data of recent years, which are most in demand for analysis, are closed.

Separately, the Russian-language project RSCI should be highlighted [20]. The project does not support search for authors by keywords. However, the project implements a search for authors by last name, by various identifiers (ORCID and others), by city or organization. Additionally, the project provides for filtering by one of seventy thematic headings. Search results are ranked by citation. The main drawback of the search module in this project is the impossibility of specifying the thematic search area, since the available headings give too general a formulation of the topic (for example, “mathematics”, “geology”). In addition, the system does not have the ability to assess the relevance of a search query. When searching in the “mathematics” section, an expert with thirty citations will have a lower rank than an expert with forty citations, even if the first of them has all works devoted to mathematics, and the second only ten percent belongs to the “mathematics” section, and the rest – to other sections of the classifier.

It should be noted that there is an active development of scientometric systems for the effective management of scientific organizations and incentives for their employees at the present time. The need for permanent analysis of the results of scientific activities of organizations using scientometric methods requires the collection of various quantitative indicators of scientific and technical and innovative and technological activities. Traditionally, one of the main quantitative indicators for assessing the results of such activities is the citation of publications. Impact factors of journals, Hirsch index and g-index for authors of scientific publications, i-index for organizations and other important scientometric indicators are based on the measurement of this indicator. However, modern scientometric systems are not limited to collecting indicators only for publications, but also collect indicators for other results of scientific and pedagogical activity. Such indicators are presentations at conferences, supervising dissertations and theses, participation in editorial boards and dissertation councils, obtaining patents and certificates, teaching training courses, carrying out scientific projects, and others. In this regard, modern scientometric systems are a good supplier of data for various types of thematic searches, including the search for experts in a given subject area. The data of scientometric systems can be combined with the data of traditional systems for calculating the citations of publications, such as Web of Science, Google Scholar, Scopus, RSCI and others.

On the basis of the above review of existing systems and an assessment of their strengths and weaknesses, it is possible to formulate the requirements that are imposed on thematic search systems for experts in the scientific field. The system should search automatically without participation of the experts themselves in the search process. To ensure the completeness of the search and a more objective assessment of the author's qualifications, articles and other information about the scientific activity of the expert should be used. The system intended for use in Russia must be able to work with Russian-language materials, including recognizing Russian-language names and surnames. The system should provide the ability to perform search queries on general thematic sections of the classifier and in special thematic areas, for example, using user-specified keywords to search [21]. Search results should be automatically ranked according to the authority of the specialists found in the given subject area. Initial data for such a search can be obtained from scientometric systems.

## **2. Description of the method**

Using information from scientometric systems [22, 23] allows automatic search for experts and rank search results. For thematic search of authors, the following data placed in the scientometric system can be used: bibliographic descriptions of articles, abstracts and monographs; reports at conferences; dissertation; descriptions of projects; patents and software registration certificates; appearances in the media; description of the taught courses and others. For each of the listed types of scientific products, the scientometric system may contain a full-text description (title, abstract, keywords) and a list of authors speci-

fied manually or using automatic recognition algorithms [24, 25]. For some types of work, the description may contain additional information that reflects the significance of the work: citation, membership in the WoS and VAK (Higher Attestation Commission at the Ministry of Science and Higher Education of the Russian Federation) lists, the number of publications made on the project, the type of courses taught, the amount of project funding, and others.

The task of thematic search for employees may be formally specified as follows. Let  $Q$  be a set of possible queries,  $A$  be a set of authors, and  $S=S_1U_1S_2U_2\dots U_N$ , be a set of object, divided into disjoint subsets, where  $N$  is the number of types of scientific activities. Given the function of the relevance of the text description to the search query by keywords, the function of the significance of the work and the function of the belonging of the work to the author are given, it is required to build a function of matching the author to the search query, which for each pair “author, query” defines a real number (match rank for the user's search query).

The function of the relevance of the text description to the search query by keywords is implemented using standard contextual search tools, taking into account morphological analysis. To improve the completeness of the search, automatic expansion of the set of keywords with related terms and translations can also be used [26].

**Table 1**  
Significance coefficient for novelty

Year	Coefficient
0	1
1	0.96
2	0.92
3	0.89
4	0.85
5	0.81
6	0.78
7	0.74
8	0.7
9	0.66
10	0.62
11	0.59
12	0.55
13	0.51
14	0.48
15	0.44
16	0.4
17	0.36
18	0.33
19	0.29
>19	0.25

The implementation of the result's significance function is different for different types of work. For articles, significance value depends on the journal membership in the lists of the Higher Attestation Commission, Scopus, or RSCI, the impact factor of the journal, and the citation rate of the article itself. For projects, the value of the function is determined by the amount of funding, as well as the number and importance of publications made on the project. For the courses taught, the course category is taken into account: streaming lectures; seminars; special courses at the choice of the department; special courses at the student's choice. For book, the value of the function is determined by the number of pages and size of publication circulation. For other types of scientific results, the value of the function is identically equal to 1. Further, the significance of the work, determined by its type, is multiplied by the significance of the novelty (Table 1). This coefficient takes into account that a researcher can change his field of scientific activity over time and his knowledge will become obsolete.

The function of the belonging of the work to the author describes the degree of participation of the author in the creation of the work. As shown in [27], when defining a function of the belonging of the work to the author as binary function (1 for all coauthors of the work and 0 in other cases), the search accuracy

significantly decreases, since coauthors can only indirectly participate in writing an article and not be experts on the topic considered in the article.

In many cases, articles are written by only one author, whose surname is put in the first place in its bibliographic description. The rest of the co-authors may participate in the work on the article insignificantly, and their main direction of scientific activity may not coincide with its subject matter.

To test the hypothesis about the importance of the order of authors in the thematic analysis, an assessment was made of the proportion of articles in which the order of authors is determined by the lexicographic order, and not by the importance in the work on the article. All articles in journals for three years with the number of authors from two to seven were selected from the scientometric system of Moscow State University for analysis.

For each of the specified number of authors, the percentage of articles for which the correct set of authors is determined by lexicographic order was calculated. The calculation results for a different number of authors are shown in Table 2.

**Table 2**  
Percentage of articles with lexicographic order of authors

Number of authors	Percentage of articles
2	24%
3	16%
4	9%
5	6%
6	6%
7	3%

Therefore, to set the connection of authors with publications and reports, a formula from [25] is used. The function of the belonging of the work to the author is defined as follows. The first author gets half the rating per article. The second half of the rating is evenly distributed among all authors (including the first one).

At the first stage of the query execution, a search is carried out for objects of all types. For objects of each type, the relevance of the query is determined in accordance with the function of the relevance of the text description to the query. Based on the ranking results, the coefficient of compliance with the query is calculated for each type of object and each author follows. For each element of the author's scientific production, the product of the function of the relevance of the text description to the search query by keywords, the function of the significance of the work and the function of the belonging of the work to the author is considered, after which the results are summed up. Thus, the correspondence of each author to the search query for each type of scientific work is calculated.

Using the calculated coefficients, the authors are ranked for each type of work, and for each author his rank is determined. A value of 1 corresponds to the maximum value for all authors for this type of work, 0 – to the minimum. The total rating of each author is calculated as the sum of its ranks for each type of scientific work, multiplied by the constants of the algorithm. The constants of the algorithm are the coefficients of significance of each type, calculated at the training stage. The result of this step of the algorithm is the value of the author's match to the search query without taking into account its authority.

At the next step, to determine the authority, the co-authorship graph built on the data of the scientometric system is used. To build this graph, information about publications, scientific reports, projects and other sources is used. The full column of co-authorship is used in scientometric systems to distinguish research teams, resolve ambiguities in defining authors [28], and in other cases. The co-authorship subgraph is constructed as the intersection of the full co-authorship graph and the set of authors, found at the previous step of the algorithm. So co-authorship subgraph is created for all selected authors. The PageRank algorithm [29] is used to determine the authority of the authors. The final function of matching the author to the request, taking into account the authority, is calculated as the product the value of the author's match to the search query without taking into account its authority and the square root value of the authors authority.

The developed algorithm can use the data of any scientometric system. However, when developing a software implementation of the algorithm, it is necessary to take into account the data format used in a

particular scientometric system. The software implementation of the algorithm developed by the authors was tested on the data of the ISTINA scientometric information-analytical system [22], which contains data on 27 organizations, including Lomonosov Moscow State University, Russian University of Chemical Technology them. D. Mendeleev, A.N. Nesmeyanov Institute of Organoelement Compounds of Russian Academy of Sciences, Frumkin Institute of Physical chemistry and Electrochemistry Russian academy of sciences, National Medical Research Center for Neurosurgery named after Academician N.N. Burdenko (Ministry of Health of Russia), Moscow State Pedagogical University, Schmidt institute of physics of the earth of the Russian academy of sciences, Topchiev Institute of Petrochemical Synthesis of the Russian academy of sciences, Institute of Geography of the Russian academy of sciences and others.

### 3. Search interface

To view the search results and test the algorithm, an interface has been implemented that allows you to search by keywords with the ability to filter results by one of seventy general thematic headings. Figure 1 shows the result of a search query for the keywords “thematic analysis”.

The screenshot shows a search interface with the following components:

- Search Criteria:**
  - Ключевые слова (через запятую): тематический анализ
  - Область знаний: Все
  - Подразделение: Все
- Buttons:** Отправить
- Table Settings:** Show by 10 items
- Search Results Table:**

V	NAME	DEP
1.139	д.ф.-м.н. проф. Васенин Валерий Александрович	404 Лаборатория автоматизации экспериментальных исследований(Совм.)(МГУ); Межфакультетская кафедра математического моделирования и компьютерных исследований(МГУ); Кафедра вычислительной математики(Совм.)(МГУ) Заведующий лабораторией, Заведующий кафедрой, Профессор
1.012	к.ф.-м.н. Афонин Сергей Александрович	404 Лаборатория автоматизации экспериментальных исследований(МГУ); Кафедра вычислительной математики(Совм.)(МГУ); Межфакультетская кафедра математического моделирования и компьютерных исследований(Совм.)(МГУ) Ведущий научный сотрудник, Доцент, Доцент
0.957	к.ф.-м.н. Козицын Александр Сергеевич	404 Лаборатория автоматизации экспериментальных исследований(МГУ); Кафедра вычислительной математики(Совм.)(МГУ) Ведущий научный сотрудник, Доцент
0.542	Шачнев Дмитрий Алексеевич	404 Лаборатория автоматизации экспериментальных исследований(Совм.)(МГУ) Специалист (зоолог, программист, геолог, инженер и т.п.)
0.417	к.ф.-м.н. Голомазов Денис Дмитриевич	None
0.391	д.ф.-м.н. проф. Мальковский Михаил Георгиевич	Кафедра алгоритмических языков(МГУ) Профессор
0.314	к.т.н. Коршунов Андрей Александрович	404 Лаборатория автоматизации экспериментальных исследований(МГУ) Ведущий научный сотрудник
0.307	Зачурин Максим Анатольевич	404 Лаборатория автоматизации экспериментальных исследований(МГУ); Межфакультетская кафедра математического моделирования и компьютерных исследований(Совм.)(МГУ) Младший научный сотрудник, Ассистент

**Figure 1:** Search result for the query “Thematic analysis”

The system is intended for Russian-speaking users and uses Russian to present information. In the search window, three fields are filled in “keywords (separated by commas)”, “Knowledge area” and “Department”. The example in Fig. 1 use the keywords «thematic analysis», knowledge areas «all», and department “all”. Additional filtering by knowledge area allows you to refine your query in the case of multivalued keywords, and a department filter allows you to restrict searches within an organization. As a result of the search query, the user is shown a table with three columns. The column “Name” contains the last name, first name, patronymic and academic title of the employee. Column “Dep” contains a description of the position and place of work of the found employee. Column “V” contains the degree to which the user's scientific results match the search query. The main positions in the list of search results for the query “thematic analysis” are occupied by employees specializing in the development of systems for thematic analysis of information.

For a detailed study of the results of search results, you can follow to the author's profile or to the list of scientific results found upon request.

To view the profile of each found employee, you can click on the employee's last name in the table shown in Fig. 1. In the employee's profile, you can see aggregated and detailed data on all scientific re-

sults of the employee, list of his most frequent co-authors, work experience and scientific degree, citation indices (Web of Science, Scopus) and his ORCID. Also, from the employee's profile page, you can go to the pages of his publications to view annotations or full texts of works (if they are uploaded to the system).

To view the list of found works (Fig. 2), you can click on the numerical value of the degree of compliance of the author with the request in column V.

Show by  items Search:

Тип работы	Название работы	Релевантность
Публикация	Автоматизированная система тематического анализа информации	13
Публикация	Методы и средства тематического анализа данных в больших системах на основе ключевых слов и косвенных связей между ними	13
Доклад	Методы и средства тематического анализа данных в больших системах на основе ключевых слов и косвенных связей между ними	13
Публикация	Интеллектуальная Система Тематического Исследования НАучно-технической информации (ИСТИНА)	10
Достежение	Развитие и поддержка ИАС «Наука МГУ» («ИСТИНА»)	10
Проект	Развитие и сопровождение системы подготовки принятия решений на основе анализа информации о результатах научно-исследовательской, педагогической и инновационной деятельности ИАС «ИСТИНА»	10
Проект	Математическое и программное обеспечение сложных информационно-вычислительных систем	10

**Figure 2:** Search result for the query “Thematic analysis”

In the list of scientific papers (Fig. 2), the first column indicates the type of scientific work (article, book, report, dissertation management, etc.), the second column indicates the name of the scientific work, and the third column indicates the relevance to the search query.

The interface of the module for thematic search for experts is developed using the module Sqlreports [30]. This module has built-in functions for generating and visualizing tables, filtering and sorting results, building cross-references, convenient markup of results and other useful functions. This allows developers to focus on writing data mining code.

## 4. Conclusion

The use of a large amount of bibliographic data accumulated in scientometric systems and citation systems allows a thematic search of specialists for all authors of publications, patents, research projects and other objects of scientific activity. This method significantly expands the range of search, and allows you to find leading specialists in a given subject area who did not take active steps to search for part-time jobs with registration in project exchanges.

At the moment, the services of thematic search for experts in the Russian-speaking segment of the Internet are not widely represented. The use of algorithms for thematic analysis of data from scientometric systems makes it possible to create high-quality services for solving the problem of finding experts. The algorithm presented in this paper for thematic search and ranking uses a combination of full-text and graph analysis methods. The software implementation of the algorithm was tested on the data of the IAS ISTINA scientometric system.

In the interface developed for testing the software implementation of the algorithm, you can specify keywords, heading and department when searching for an employee. To check the results for each employee, drill down is available with a list of jobs that affect a given thematic search area.

The results of the work of the module can be used, among other things, in editorial offices to search for reviewers for scientific articles, in high-tech industries to search for consultants, research workers to organize research teams in order to solve certain practical problems.



## 5. References

- [1] Gruppa “Pomogite zhurnalistu”. URL: <http://www.facebook.com/groups/pomogitej>
- [2] Gruppa “Geroy i eksperty dlia SMP”. URL: <http://www.facebook.com/groups/1792146334361293>
- [3] Birzha kopiraitinga. URL: <http://www.textsale.ru>
- [4] Birzha kopiraitinga Advego. URL: <http://Advego.ru>
- [5] Birzha kontenta eTXT. URL: <http://eTXT.ru>
- [6] Poiskovaia sistema Search Engine to Find Experts from Universities. URL: <http://network.expertisefinder.com>
- [7] Poiskovaia sistema CiteSerX. URL: <http://citeseerx.ist.psu.edu>
- [8] Poiskovaia sistema Semantic Scholar. URL: <http://www.semanticscholar.org>
- [9] Poiskovaia sistema ScienceDirect. URL: <http://www.sciencedirect.com>
- [10] Izdatelstvo Shpringer. URL: <http://link.springer.com>
- [11] Poiskovaia sistema izdatelstva IEEE. URL: <http://ieeexplore.ieee.org>
- [12] Poiskovaia sistema Digital Library. URL: <http://dl.acm.org>
- [13] Sistema tsitirovaniia Google Scholar. URL: <http://scholar.google.com>
- [14] Sistema poiska avtorov Microsoft Academic. URL: <http://academic.microsoft.com>
- [15] Sistema poiska avtorov AMiner. URL: <http://www.aminer.cn>
- [16] V. Vasenin, M. Zanchurin, A. Kozitsyn et al., Architectural and technological aspects of the cloud data analysis system development, case of istina system. 5th International Conference on Actual Problems of System and Software Engineering, APSSE 2017. CEUR Workshop Proceedings 1989 (2017) 90–96.
- [17] Opisanie stat’i “Architectural and technological aspects of the cloud data analysis system development, case of istina system” v sisteme Microsoft Academic. URL: <http://academic.microsoft.com/paper/2771441762>
- [18] V. A. Vasenin, S. A. Afonin, A. A. Zenzinov i dr. Mekhanizmy sistemy ISTINA dlia intellektualnogo analiza sostoiianiia i stimulirovaniia khoda vypolneniia proektov v sfere nauki i vysshego obrazovaniia. Nauchnyi servis v seti Internet: trudy XXI Vserossiiskoi nauchnoi konferentsii (23–28 sentiabria 2019 g., g. Novorossiisk). Pod red. V. V. Voevodina. IPM im. M.V. Keldysha, Moskva, 2019. S. 210–221.
- [19] Opisanie stat’i “Mechanisms of <ISTINA> system for intelligent state analysis and progress stimulation for projects in the sphere of science and higher education” v sisteme Microsoft Academic. URL: <http://academic.microsoft.com/paper/2986189703>
- [20] Sistema tsitirovaniia RINTs. URL: <http://elibrary.ru>
- [21] V. Vasenin, K. Lunev, S. Afonin, D. Shachnev, Methods for intelligent data analysis based on keywords and implicit relations: The case of “istina” data analysis system. In Proc. of the International Conference “Actual Problems of Systems and Software Engineering” (APSSE 2019), IEEE Conference Proceedings. United States, 2019. P. 151–155,
- [22] V. A. Vasenin, Arkhitekturno-tekhnologicheskie aspekty razrabotki i soprovozhdeniia bolshikh informatsionno-analiticheskikh sistem v sfere nauki i obrazovaniia. Programmnaia inzheneriia. 8(10) (2017) 448–455.
- [23] S. A. Afonin, D. D. Golomazov, A. S. Kozitsyn, Ispolzovanie sistem semanticheskogo analiza dlia organizatsii poiska nauchno-tekhnicheskoi informatsii. Programmnaia inzheneriia (2) (2012) 29–34.
- [24] A. S. Kozitsyn, S. A. Afonin, Razreshenie neodnoznachnostei pri opredelenii avtorov publikatsii s ispolzovaniem grafov soavtorstva v bolshikh kollektiiakh bibliograficheskikh dannykh. Programmnaia inzheneriia 8(12) (2017) 556–562.
- [25] A. S. Kozitsyn, S. A. Afonin, Algoritm razresheniia neodnoznachnosti imen avtorov v IAS ISTINA. Sovremennye informatsionnye tekhnologii i IT-obrazovanie. 16(1) (2020) 108–117.
- [26] D. A. Shachnev, Searching for activity results and experts in a given subject area, taking results significance into account. Programmnaia Inzheneriia. 12(5) (2021) 260–266.
- [27] A. S. Kozitsyn, S. A. Afonin., D. A. Shachnev, Metod otsenki tematicheskoi blizosti nauchnykh zhurnalov. Programmnaia Inzheneriia (6) (2020) 335–341.
- [28] A. Kozitsin, S. Afonin, D. Shachnev, The use of thematic analysis methods in scientometric systems. Proceedings of the 22nd Conference on Scientific Services & Internet (SSI-2020). CEUR Workshop Proceedings 2784 (2020) 178–188.

- [29] D. F. Gleich, PageRank Beyond the Web. Proc. Society for Industrial and Applied Mathematics (2015) 321–363.
- [30] S. Afonin, A. Kozitsyn, I. Astapov, Sqlreports: Yet another relational database reporting system. Proceedings of the 9th International Conference on Software Engineering and Applications, 2014. P. 529–534.