Design and Development of Information System of Scientific Activity Indicators

Aleksandr Spivakovsky, Maksym Vinnyk, Yulia Tarasich and Maksym Poltoratskiy

Kherson State University, 27, 40 rokiv Zhovtnya St., 73000 Kherson, Ukraine

{Spivakovsky, Vinnik, YuTarasich}@kspu.edu,
max1993poltorackii@gmail.com

Abstract. The article provides a brief overview of the most popular information systems of evaluation of scientific activity of scientists. The vision of functional capabilities of processing of the system scientometric indicators of the scientific team, the organization and its business units on the basis of scientific profiles of existing scientometric and bibliometric systems are described. The example of the implemented solutions with the authors description of its components, basic algorithms and used technologies is presented.

Keywords: scientific activity, information systems, scientometric systems, bibliometric systems, scientometric indicators.

Key Terms: ICTCInfrastructure, ICTComponent, InformationTechnology, WebService

1 Introduction

Scientific information is a special kind of information that affects the development of any and all sectors of modern society. Analysis of scientific information can be divided into polysyllabic such as information on the research teams, scientific collections, scientist, scientific works and more. Elementary but the objective component, in our opinion, is the scientific activities of the scientist. Today there are many information systems that attempt to create methods and technologies of processing and saving information on the activities of scientists.

The most outstanding services with rapidly growing impact are Google Scholar, Scopus, Orcid, Academia.edu, Research Gate, Mendeley, arXiv.org, cs2n, Epernicus, Myexperiment, Network.nature, Science community.

These services contribute to satisfying the needs of the scientific com-munity. In fact, this positively influences scientific and technical progress and creates a new paradigm of scientific research. A big number of the recently created scientometric services allow assessing the relevance of the research results by a scientist. Having these measurements at hand opens up new opportunities and prospects. In this article we consider the existing information systems for the processing of scientific activities

ICTERI 2016, Kyiv, Ukraine, June 21-24, 2016

(section 2), describe your own vision and capabilities to design and develop our system (section 3), as well as the basic methods and technologies (section 4) used for its implementation.

2 Related works

After analyzing the information systems that run on the activities of scientists, scientific groups, publishers, etc..., we offer to look for the most interesting projects.

Bibliometrics of Ukrainian Science. The pilot project of information-analytical system "Bibliometrics of Ukrainian Science", is implemented by the Department of bibliometric and scientometrics of information and analytical support of Vernadsky National Library [1].

The system "Bibliometrics of Ukrainian Science" is representation of in-formation of Ukraine scientists' profiles who provided information about their publication in the Internet; national component of the project Ranking of Scientists (Cybermetrics Lab).

Information resources of systems are formed by processing: created by scientists on the platform of Google Scholar bibliometric profiles containing information of their publication activity results, bibliometric indicators of Scopus, Web of Science, Ranking Web of Research Centers. Updating of information on value of Hirsch index in bibliometric profiles of scientists is executing on monthly, the value of other indicators is updated quarterly (Hirsch index of scholar is h, if he has h publications, each of which is cited at least h times) [1].

Scopus. Scopus is a single the world's largest abstract database, which indexes more than 17 000 items of scientific, technical and medical journals about 4,000 international publishers [2].

Scopus system is designed to maintain efficient workflow of researchers, helping them to: find new articles from the area of their specialization; find information about the author; analyze the publication activity in the subject area; track citation; view the h-index; identify the most cited articles and authors; assess the relevance of the study.

Scopus enables researchers to combine their articles under a single profile [2].

Google Scholar. Google Scholar is freely accessible search system, which indexed the full text of the scientific publications all formats and disciplines.

Google Scholar executes not only informational, but scientometric function. From the list of results on a hyperlink Search Cited by we can obtain the information how many and what documents are linked on the publication in data base Google Scholar. The number in Cited by reflects the degree of authoritativeness and publicity of publication [3].

Web of Science. Web of Science – International established database of Scientific Citation, it is presented by company Thomson Reuters. Web of Science gives possibility to search among 12 000 magazines and 148 000 materials of conferences in the field of natural, social, human sciences and arts, which allows to obtain the most relevant information for your questions. In addition to search, Web of Science establishes a reference link between the specific research using the cited materials and thematic

links between articles established reputable re-searchers working in this field. It is the most extensive database of abstracts. It is available by subscription [4].

Russian Science Citation Index (RSCI). RSCI is a national information-analytical system, accumulating more than 2 million publications of Russian authors, as well as information about the citation of these publications from more than 3,000 Russian magazines. It is designed not only for the operational support of research to date reference and bibliographic information, but is also a powerful tool to carry out evaluation of the impact and effectiveness of research organizations, scientists, the level of scientific journals, etc. [4].

Earlier research team of Kherson State University (KSU), which included the authors of the article, took part in a number of international and national projects whose aim was the development and implementation of scientific and management processes of analytical information systems and services [10].

In addition, this article is a continuation of the previous work of the authors [5] which addressed the issue of openness of scientific activities of Ukrainian scientists, as well as the construction of an open scientific training system, one of the main elements of which are the scientometric information processing system.

The authors also conducted a study of the technical component of the implementation of feedback services in the KSU [6], as well as the formation of the ICT infrastructure at higher education institution [7, 8].

3 Vision of the system. Criteria

Analysis of information systems described in Section 2 (Table 1) confirms once again the need to implement a system that would allow build the consolidated ratings of scientists, scientific groups and organizations in the automatic mode.

Why consolidated? A significant part of scientometric databases and systems, which are presented in the scientific world are closed, and accordingly assess only the academic publications that are indexed by them, while the rest of the scientific work in this assessment are not included. For example, Scopus indexes, indicators of the other part of scientometric databases are not always accounted for as tangible.

In addition, for the analysis of the scientific activities of scientists' group, or a specific organization, it should be carried out manually. The only option of its partial automation is now rating the organization's profile in Google Scholar (which makes the system "Bibliometrics of Ukrainian Science"). But what should do if this profile is not created? Or if not all scientists working in the organization or are part of the research team, and their articles are incorporated in the profile?

Thus, the main task of building our system is the realization of the possibility of automatic processing of scientometric and bibliometric indicators of scientific groups and organizations on the basis of analysis of scientific profiles of known scientometric databases and systems, including automatic search and its analysis.

Table 1. Compare features considered information systems

Information Systems / criteria	Scopus	Google Scholar	RSCI	"Bibliometrics of Ukrainian Science"	Our System
Scientist profile	+	+	+	-	+
Scientific institution Profile	-	+	-	-	+
Profile of structural units of scientific institutions	-	+	-	-	+
Construction ratings	-	-	-	+	+
Scientometric and bibliometric indicators	+	+	+	+	+
Personal notifications and reports	-	-	-	-	+
The openness of the system	+ -	+	+	+	+
Possibility of automatic comparison of the scientific work of several scientists, organizations, groups, etc.	+ -	-	-	-	+

3.1 Concept of solutions

Goals and objectives.

Estimated system should have the following features: parsing pages of scientists in scientometric systems and databases; processing and display scientometric indicators of the author, the scientific team, the organization; formation of a library of publications of scientists and the ability to sort by the specified criteria: the author's name; the name of the organization of the university; the department, etc.; statistical processing of the obtained information; the ability to compare the quality performance of universities, research groups, etc.; the possibility of multi-threaded data processing; presence feedback.

Allocation objects of system: parser; scientometric system and the database works with resource; data store; Web- site of resource; reporting.

Parser and data warehouse include the following attributes: author's name; link to profile scientometric databases; number of publications; scientometric indicators (Hirsch index, citation index, I-code, etc.); publications; links to the publications; publication description.

Web-site of resource. Select the following attributes: rating of universities in Ukraine; rating of Ukrainian scientists; rating of university scientists; rating of the university departments; rating of the university faculties; profiles of the scientists of the university.

Reporting includes the following attributes: scientometric indicators; graphics.

The interaction of key system components shown in Fig.1.

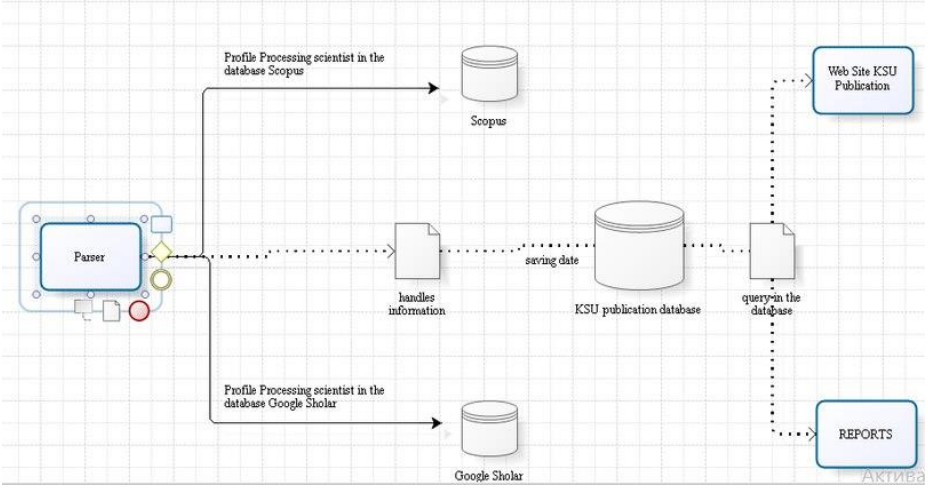


Fig. 1. The interaction of key system components

Thus, the user can view general information of scientific activity of scientist, scientific group, certain university or scientific organization, as well as the consolidated rating. Scientist, registered in the system is able to receive notifications about changes in their scientometric indicators. The system administrator can generate a general statistical report of their organization.

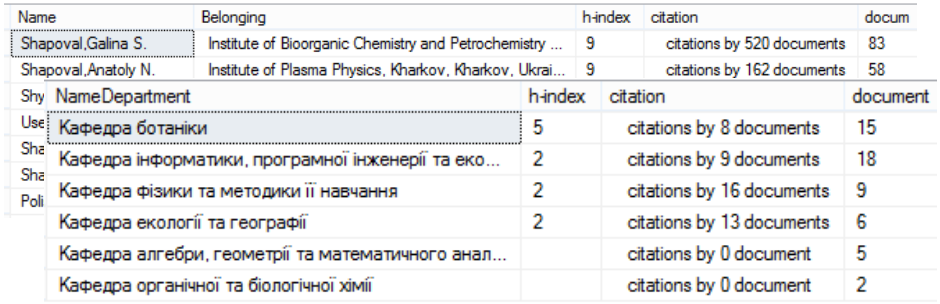
Assumptions and Constraints

In the current version of the system it is implemented the ability to handle scientist indicators on Scopus data and Google Scholar. The algorithm of automatically search for links to profiles of Ukrainian scientists is developed, the algorithm of automatic distribution profiles of scientists on the name of the organization in which they work is implemented, the automatic generating of department ratings, faculties and research teams is implemented, the ability to send messages to e-mail scientists about changes of academic indexes.

Scientometric indices on which ratings are based in the system are:

1. h-index (Scopus&Goggle Scholar). The h-index is based on the highest number of papers included that have had at least the same number of citations;
2. citations (Scopus& Goggle Scholar). Numbers of total citations of documents that are indexed by the system;
3. i10-index (Goggle Scholar). Numbers of total citations by documents that have ten or more citations;

At present, about 3,000 profiles of scientists in Scopus has been processed by the system, of which 680 have been identified as the profiles of Ukrainian scientists. Automatic processing of the found profiles allowed constructing the rating of Ukrainian scientists on their indices in Scopus. By sorting the results of belonging the scientists to the university (e.g. KSU), it was implemented the ability automatically generate ratings of chairs, faculties and scientific researches of the university groups (Fig. 2).



Name	Belonging	h-index	citation	docum
Shapoval,Galina S.	Institute of Bioorganic Chemistry and Petrochemistry ...	9	citations by 520 documents	83
Shapoval,Anatoly N.	Institute of Plasma Physics, Kharkov, Kharkov, Ukrai...	9	citations by 162 documents	58

NameDepartment	h-index	citation	document
Кафедра ботаніки	5	citations by 8 documents	15
Кафедра інформатики, програмної інженерії та еко...	2	citations by 9 documents	18
Кафедра фізики та методики її навчання	2	citations by 16 documents	9
Кафедра екології та географії	2	citations by 13 documents	6
Кафедра алгебри, геометрії та математичного анал...		citations by 0 document	5
Кафедра органічної та біологічної хімії		citations by 0 document	2

Fig. 2. Example of system work

The highest number of publications (on 10.02.16) has such scholars as - Oleg Shishkin (581), Leonid Levchuk (463) and Vladimir Gun'ko (322).

The analysis of the scientific activity of KSU scientists' shows the greatest number of publications has the teachers of Chair of Informatics, Software Engineering and Economic Cybernetics (98). And the most h-index has the teachers of the Chair of Botany (5).

The construction of similar ratings according Goggle Scholar, it is currently possible only in the presence of links on it's, as distinct from Scopus, the author himself should register in the system. There is more complicated the ability to search scientists. Thus, we have been processed the records, links have been provided by the University scientists. Now for viewing and analyzing there is available indicators of scientists of Faculties of Physics, Mathematics and Informatics of KSU, Faculty of Pre-School and Elementary Education of KSU, general Chair of Philosophy and Social and Humanities Sciences.

The next stage of development and improvement of the system will:

- automatic integration and analysis of information on scientometric indicators scientist in the case of duplication of its profile in these scientometric database;
- improving the algorithm of processing information on scientometric indicators of organizations, scientific collectives in case of misspelling or change their names;
- improving the algorithm of finding links on the profiles of scientists according their belonging to the country;
- the ability to automatically compare the indexes of scientific activities of scientists, research groups, organizations and the structural divisions.

Analysis of the use

There are two user groups allocated in the system: the administrator of the system on the part of the establishment; user.

The category of "user" is the staff of institutions, scientists, as well as the rest of Internet users, who can view the information provided on the Web-site of the system.

As example, Consider the algorithm of the system work with Scopus in details:

The parser takes a reference to the scientific profile from the database system and loads the appropriate page of Scopus. After that, two parallel streams are run – processing of scientometric indicators of scientist and processing of information about

his articles. Once when processing of the whole page is over, there is an inquiry about the presence records under consideration "name" in the database system. If the name is, it updates the information about scientometric indicators and publications of the scientist. Otherwise - in the database record is created about the author by assigning a unique identifier to him, and information about his articles and scientometric indexes is entered into the appropriate tables. After the upgrading all the database system the administrator and scientists registered in the database get e-mail with information about changes of their indexes.

4 Tools and Technologies

Developing of solutions requires the use of certain products and technologies:

- JSON. It is used in the system for the exchange of data for third-party systems. Thus, our system can be a source of data for other resources. It implements the data exchange via json requests.
- asp.net and framework Entity. It is used to implement Web- Site of System.
- Library of html align pack. This library is used for processing of Scopus pages. It uses PATCH requests and then adds the results to the database. In the previous version of the system the regular expressions were used. The use of html align pack is greatly affected on her productivity.

One of the most important algorithms used in the system is Levenstein algorithm [9].

This algorithm is used for solving the problem of determining belonging the scientist to a particular organization, which arises at changing of the organization's name, its spelling errors in the article, the change of scientists their place of work, etc.

Let’s consider the algorithm in detail:

Algorithms of fuzzy search are also known as similarity search or fuzzy string search are the basis of the spelling checker systems and full of search engines like Google or Yandex, This algorithm is an extremely useful feature of any search engine. However, its effective implementation is much more complex than implementing of a simple search by exact match. The most commonly used metric is the Levenshtein distance or edit distance, its algorithms calculation can be found at every turn.

Thus, we compare the author's field of membership of the organization specified in Scopus with many possible names of organizations in the system database. This takes into account the possibility of errors.

Conclusions

The work is developed by processing system of scientometric indicators of scientist on the basis of its profile in Scopus and Google Scholar systems. The main difference between the systems developed by us from others is the ability to automatically build research teams rankings, organizations and entities to which the scientist applies. In

addition, the algorithm of automatically search and group profiles of scientists for their attitude to this or that state, organization, is already have developed.

The personal profile of each scientist collected information about his scientometric and bibliometric indicators is a list of his publications, displays statistics of scientific work - change the number of publications, citations, h-index, etc. Graphical display of the dynamics of scientific work was implemented for the research teams, organizations and their subdivisions.

Today the system is used to calculate indicators of scientific activity of Kherson State University and its structural units - departments, faculties, Specialized Academic Council, etc.

The next stage in the development of the system, we see in the realization of its interaction with other scientometrics systems and databases. Also, one of the most important and necessary features, we consider the need for implementation of the comparison options of several organizations, research groups and scientists.

The implementation of the algorithm of automatic search of references to the scientific profile of the membership of a particular country and the organization, and improve the efficiency of the algorithm allows us to speak about the possibility of sampling and processing of large amounts of information. Thus, in the next version of the system it is supposed to build a data warehouse on the principles of Big Data and Map Reduce. That, in turn, will generate ratings of the scientific activities of scientists, scientific groups and organizations with minimal resources and time-consuming.

References

1. Bibliometrics of Ukrainian Science, Http://nbuviap.gov.ua/bpnu/index.php?page_sites=pro_proect
2. Abstracts database Scopus, http://health.elsevier.ru/electronic/product_scopus/
3. Scientometric database, http://www.nbuv.gov.ua/node/1367
4. Science Citation Index for scientists, http://index.petrsu.ru
5. Spivakovsky, A., Vinnik, M., Tarasich, Y.: Web Indicators of ICT Use in the Work of Ukrainian Dissertation Committees and Graduate Schools as Element of Open Science. In: Yakovyna, V., Mayr, H.C., Nikitchenko, M., Zholtkevych, G., Spivakovsky, A., Batsakis, S. (Eds.) ICTERI 2015. CCIS, vol. 594, pp. 3–19. Springer, Heidelberg (2016)
6. Spivakovsky, A., Klymenko, N., Litvinenko, A.: The problem of architecture design in a context of partially known requirements of complex web based application “KSU Feedback”. Inf. Technol. Educ. **15**, 83–95 (2013)
7. Spivakovsky, A., Vinnik, M., Tarasich, Y.: To the Problem of ICT Management in Higher Educational Institutions. Inf. Technol. Learn. Tools **39**, 99–116 (2014). (In Ukrainian)
8. Spivakovska, E., Osipova, N., Vinnik, M., Tarasich, Y.: Information competence of univer-sity students in Ukraine: development status and prospects. In: Ermolayev, V., Mayr, H.C., Nikitchenko, M., Spivakovsky, A., Zholtkevych, G. (eds.) ICTERI 2014. CCIS, vol. 469, pp. 194–216. Springer, Heidelberg (2014)
9. Lowenstein, V .: Binary codes with correction for deletions, insertions and substitutions of character. Reports, USSR Academy of Sciences 163.4 (1965).
10. International Projects, http://www.kspu.edu/About/DepartmentAndServices/DSAICI/internationalprojects