

DEVELOPMENT OF RATING SYSTEMS FOR SCIENTOMETRIC INDICES OF UNIVERSITIES

Aleksandr Spivakovsky^[0000-0001-7574-4133], Maksym Vinnyk^[0000-0002-2475-7169], Maksym Poltoratskiy^[0000-0001-9861-4438], Yulia Tarasich^[0000-0002-6201-4569], Kateryna Panova and Anton Melnychenko

Kherson State University, 27, Universytets'ka St., 73000 Kherson, Ukraine
{Spivakovsky, Vinnik, YuTarasich, KPanova,
AMelnychenko}@kspu.edu, mpoltorackiy@gmail.com

Abstract. The article provides a brief overview of the most popular scientometric systems. In our opinion scientific papers are one of the main objects of evaluation of researcher's scientific activity, scientific journals and research organizations. The key idea of the article is description of authors' viewing to develop the automatic analysis system of scientometric indicators, which are various in structure, systems' features and construction on their basis of consolidated ratings. The philosophy of the system is providing open data of different scientometric systems.

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

1 Introduction

Our time is characterized by the phenomenal acceleration of knowledge accumulation and the complication of its structure. According to Dell-EMC [9], the amount of produced data is growing more than twice every two years. Based on IDC (International Data Corporation) report prediction, the global data volume will grow exponentially to 44 trillion gigabytes by 2020. This tendency is inherent to all areas of human life. The same problem of storage of information exists in scientometric systems.

Human consciousness is objectively incompetent in processing and storing of large volume of complex accurate data. Today information technologies are one of main ways to arrange and create effective tools for organizing the interaction and processing large amount of information. In our opinion, nowadays in Ukraine the archaic methods of collection, processing and presentation of information and scientific activity are adopted.

Today many information systems attempt to create methods and technologies of processing and saving information on the activities of scientists. In our opinion, presentation of information on the university's scientific activity should be in the rating form. The rating accumulates several aspects and provides an opportunity to analyze development in different directions and its changes.

It is important to note that rating is subjective concept, and based on the principles of rating, it is possible to model the development of scientific activity at the university according to its goals, and for the same purpose the influence of a certain element of the rating can be changed at any time.

Our system allows automating the processing of information and its presentation, so we get more accurate result much faster. It is important that result can be obtained at any moment, and this allows us to get a dynamic picture, that helps to make decisions related to the scientific activity.

The availability of information system that would collect, processes and presents the scientific indicators of organizations is the actual. Therefore, the aim of our work is to present our experience in developing system of automatic construction of ratings of scientific organizations based on their scientometric indicators.

In the article we consider the existing information systems for the processing of scientific activities (2), describe the key components of our system and basic principles of its work (3), as well as the basic methods and technologies (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 (in our opinion)[6].

1. Scopus

Scopus is the largest abstract and citation database of peerreviewed literature, which indexes more than 7 000 items of scientific, technical and medical journals and about 4,000 international publishers [2].

Scopus is designed to serve the research information needs of researchers, educators, administrators, students and librarians across the academic community.

Scopus enables researchers to combine their articles under a single profile [3].

Our system gets the following attributes from Scopus:

- author's name;
- number of publications;
- scientometric indicators (Hirsch index, citation index, etc.);
- links to the publications;
- publication description.

2. Google Scholar

Google Scholar is freely accessible search system, which indexed the full text of the scientific publications of 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, 4].

Google Scholar classifies articles in the same way as scientists, by evaluating the entire text of each article, its author, the publication in which the article appeared, and the frequency of citing this work in the scientific literature. The most relevant results are always displayed on the first page.

Our system gets the following attributes from Google Scholar:

- scientometric indicators (Hirsch index, citation index, etc.);
- articles in Google Scholar.

3. Web of Science

Web of Science is an International established data base of Scientific Citation, and a search platform that combines abstract databases of publications in scientific journals and patents, including databases of the mutual citation of publications. 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 researchers working in this field. It is the most extensive database of abstracts. It is available by subscription [3, 4].

4. ORCID

ORCID (Open Researcher and Contributor ID) is a nonproprietary alphanumeric code to uniquely identify scientific and other academic authors and contributors [14]. This addresses the problem that a particular author's contributions to the scientific literature or publications in the humanities can be hard to recognize as most personal names are not unique, they can change (such as with marriage), have cultural differences in name order, contain inconsistent use of first-name abbreviations and employ different writing systems.

The ORCID offers an open and independent registry intended to be the de facto standard for contributor identification in research and academic publishing.

In our system ORCID id is used for unique scientist identification in different scientific databases and systems.

5. Tutor Network

Tutor Network is a web service, developed in Kherson State University. It was developed using such technologies, as ASP.NET MVC, C#, ADO.NET Entity Framework, JavaScript Framework – JQuery, AJAX and Microsoft SQL Server.

This service provides an opportunity to take into account all types of scientific works and publications, which can not be considered by other systems.

A distinctive feature of this system is displaying information about such scientific works as manuals, monographs etc.

Technical details will be considered in a subsequent article.

As mentioned in the previous article [6], the 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. These projects were - Tempus TACIS CP No 20069-1998 "Information Infra-structure of Higher Education Institutions"; Tempus TACIS MP JEP 23010-2002 "UniT-Net: Information Technologies in the University Management Network"; US Department of State Freedom Grant SECAAS-03-GR-214 (DD) "Northern New York and Southern Ukraine: New Partnership of University for Business and Economics Development", etc.

During the scientific activity we faced with such problem, as the absence of a clear mechanism of evaluation of personal contribution in the work of the University, and incomprehension of the construction of university decisions related to scientometric. In addition, for the analysis of the scientific indicators of scientists' group, or a specific organization, it should be carried out manually. The only option of its partial automation is rating the organization's profile in Google Scholar. These reasons motivated us to implement the system of automatic rating construction, its basic principles were described in the previous article [6].

Based on scientometric analysis of the systems described in section 2, we decided to implement the interaction of our system with such scientometric systems as:

- Google Scholar, as the most commonly used bibliographic database and it is easy in using;
- Scopus, as the largest and the most authoritative abstract database.

On the next step we consider the system architecture.

3 Structure of the system

The high-level system architecture is shown in Fig.1.

Parser is search, receiving and transfer the open information of scientometric indicators of authors and journals provided by Scopus and Google Scholar.

Parser is made by using xpath queries and regular expressions. Each xpath query turns to the page of resource. We developed multipletstream parser; each its stream initializes the parsing of particular resource.

For interaction with Scopus and Google Scholar parser uses xpath queries, and API is used for getting information from Web of Science and Tutor Network.

All data received by parser is stored in the system database. DB of system is distributed by the data storage. Individual entities of DB are database of scientometric indicators of researcher and scientific publications.

Information processing is realized by performing a set of predefined SQL queries.

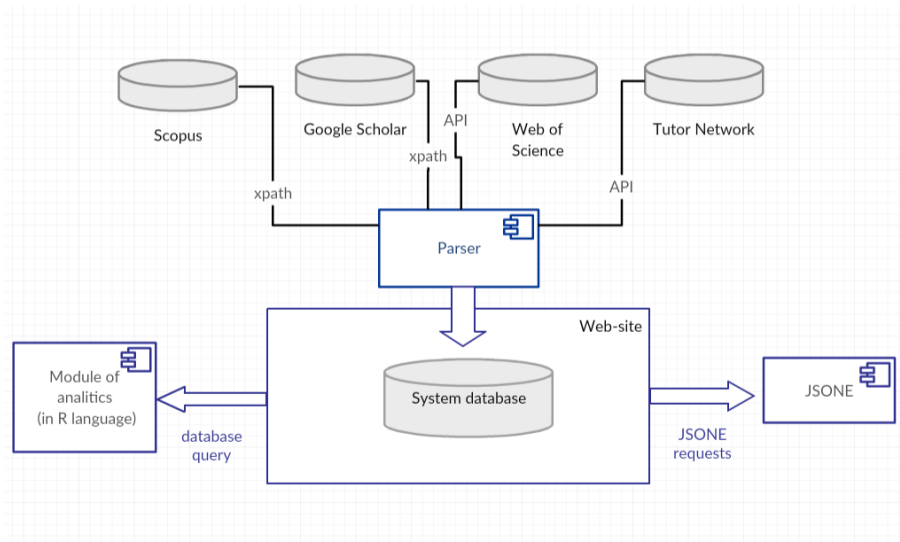


Fig. 1. The interaction of key system components

4 System description

The main task of the system is automatic construction of consolidated rating of scientists, research groups, organizations according to indicators of processed scientometric systems. These indicators are:

1. h-index (Scopus & Google 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 & Google Scholar). Numbers of total citations of documents that are indexed by the system;
3. publications (Scopus). Total number of documents that are indexed by the system.

Thus, the main task of the new version was the realization of an automatic construction of consolidated rating, which allows building a rating for any scientists, research groups, organizations.

We determined types of presentation of the results for scientists:

- profiles of the scientists of the university with generalized information of scientometric indicators for each database;
- the rating list of all the scientists of organization;
- the rating list of all the scientists of organization's structural subdivision (faculty or department);
- the general scientometric information about the university.

The example of system work is shown in the Fig.2.

The system allows to search a scientist by:

- ORCID id;
- Scopus ID;
- Google Scholar ID;
- full name of a scientist.

Інформаційно-аналітична система оцінки наукової діяльності університету

Остання ревізія: 17 лютого 2018 р.

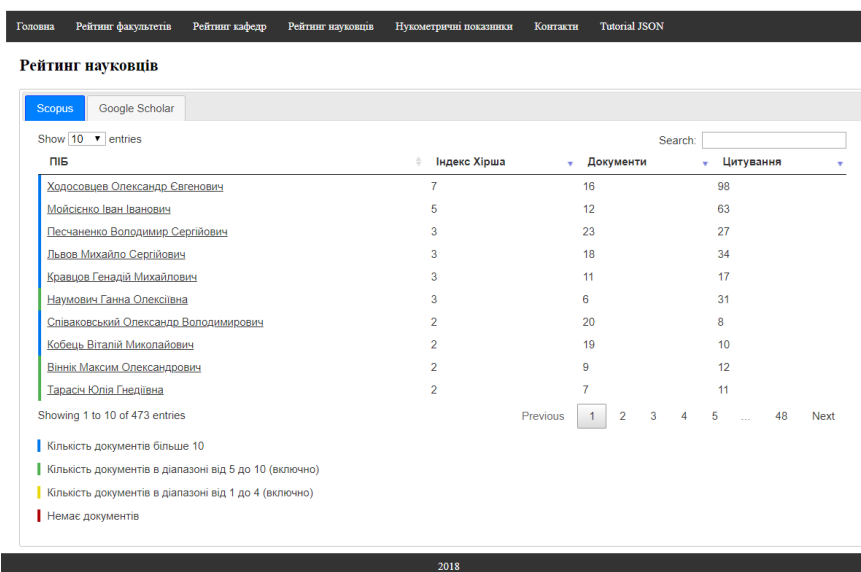


Fig. 2. Example of system work

In response to a search query, the system will return the list of references to scholars, information about which is in the database, or the message that nothing is found. After clicking on the link, the personal profile of scientist with generalized information of scientometric indicators for each database will open.

On the tab "Rating of Faculties" we get the list of faculties and the highest H-index on the faculty. There is an option to select the number of list items that will be displayed on the page. You can also sort the list of faculties by increasing or decreasing in alphabetical order or by the value of the H-index. Depending on the selected tab you can view information from Scopus or Google Scholar. Information on the tab "Rating of Departments" is displayed similarly.

On the tab "Rating of Scientists" we see the list of all scientists of the university, sorted in descending order of such scientific metric indicators as the h-index, the number of documents in the chosen scientometric database, and the number of citations.

Based on the number of scientist's papers (publications) in scientometric system, we offer to divide scientists into 4 categories and we defined an equivalent color for each of them (Fig. 2.):

- blue — the number of documents is over 10;
- green — the number of documents is in the range of 5 to 10 (inclusive);
- yellow — the number of documents is in the range from 1 to 4 (inclusive);
- red — no documents.

On the tab "Scientific metrics", you can find general university statistics, grouped by scientometric systems, such as:

- the total number of university staff, registered in selected systems, as well as their distribution by departments;
- maximum H-index ;
- maximum number of documents in selected system;
- maximum citation index.

The analysis of the scientific activity of KSU scientists' shows the highest number of publications in Scopus has such scholar as Volodymyr Peschanenko (23), and the maximum number of citations has Alexander Khodosovtsev (98). And the most h-index has the teachers of the Chair of Botany (7).

5 Module of analytics

As a separate part of the system, module of analytics (written in R language) was added in the current version. In our system, we use it for processing a scientometric data array, and graphical representations of statistical data.

In particular, we used this language opportunity as a time series analysis that allows the theoretical opportunity to apply models to predict the growth of dynamics of university rating.

The diagrams present data showing relation between such scientometric indicators, as the value of h-index in Scopus and the number of papers in Scopus or Google Scholar. The diagram can be displayed within the university or faculty, grouped by departments. The example of diagram is shown in Fig.3.

As it is evident from the diagrams, the best result is one, located most right and high.

Another option for using module in R is possibility to get a variety of statistical tables, such as the table of universities and the number of the scientists, whose articles were written in co-authorship with the KSU's scientists; the table that shows the general information about set of scientometric indicators of scientists, and total number of registered users on each department and faculty, etc.

Thus, the graphical and table representation of statistical data makes the process of perception of information easier. Moreover, the R language provides an opportunity to analyze the relations between indicators.

The main task of developing system was the realization of the possibility of automated construction of the rating of scientometric indicators for the evaluation of scientific activity not only in Kherson State University, but in any university. Thus, our system allows constructing a rating of scientists, research groups and organizations

(as well as their structural subdivisions) by using the API (Application Programming Interface).

The system provides access by request in such form:

http://publication.kspu.edu/api/v1/teacher?option=[orcid_id/scopus_id/google_scholar_id/name] &value=[search_value].

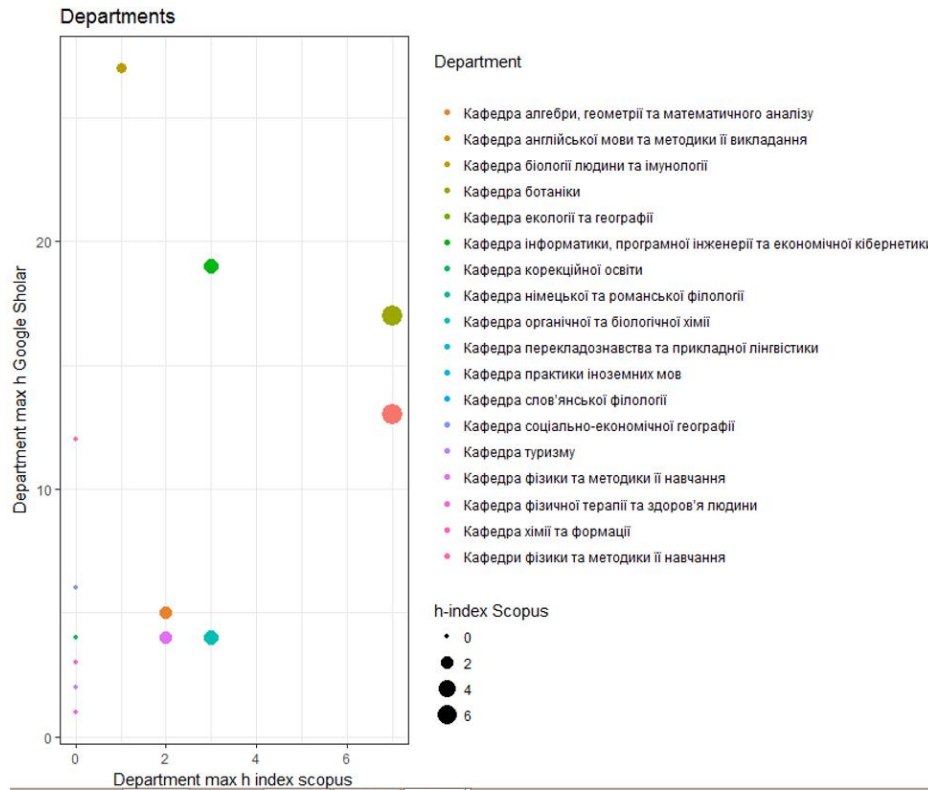


Fig. 3. Diagram of relation between department Google Scholar max h-index and Scopus with in university

By specifying the search parameters in the request (some scientometric system and scientist's id in it), we will get a list in the json format that looks like this:

Conclusions

The article presents our experience in developing system of automatic construction of ratings of scientific organizations based on their scientometric indicators in Scopus and Google Scholar, based on the algorithm of constructing ratings of scientific publications by the presence/absence in various ratings based on their scientometric indicators, proposed in our previous article.

In the current version of the system the collection, statistical processing and presentation of rating of scientometric indicators of scientists, research groups and organizations are realized. Data source of the system is open information provided by such scientometric systems as Scopus and Google Scholar.

Today the system is used to build a consolidated rating of scientists of Kherson State University and its structural units. It was designed in such a way that it is possible to deploy it in other universities, and to customize it for their specific individual goals and tasks. Our system enables to build an automatic rating based on scientometric indicators by using Application Programming Interface (API).

The next stage in the development of the system, we see in the realization of its interaction with Web of Science, as the second in authority international database. The system and technical details will be considered in more detail in the next article.

References

1. Bibliometrics of Ukrainian Science. Social Communications Research Center, <http://nbuviap.gov.ua/>
2. The largest database of peer-reviewed literature. Elsevier, <https://www.elsevier.com/solutions/scopus>
3. Scientometric databases. National library of Ukraine., <http://www.nbuv.gov.ua/node/1367>
4. Science Citation Index for scientists. Regional Center of New Information Technologies, <http://index.petrstu.ru/>
5. Spivakovsky, A., Vinnyk, 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., Nikitchenko M., Zholtkevych G., Spivakovsky A., Batsakis S. (eds) Information and Communication Technologies in Education, Research, and Industrial Applications. ICTERI 2015. Communications in Computer and Information Science, vol 594., pp. 3-19. Springer, Cham (2015).
6. Spivakovsky, A., Vinnyk, M., Tarasich, Y., Poltoratskiy M.: Design and development of information system of scientific activity indicators. . Ermolayev, V., Spivakovsky, A., Nikitchenko, M., Ginige, A., Mayr, H. C., Plexousakis, D., Zholtkevych, G., Burov, O., Kharchenko, V., and Kobets, V. (Eds.): ICT in Education, Research and Industrial Applications: Integration, Harmonization and Knowledge Transfer. Proc. 12th Int. Conf. ICTERI 2016, vol. 1614, pp. 103-110. CEUR-WS.org (2016).
7. Lowenstein, V. : Binary codes with correction for deletions, insertions and substitutions of character. Reports, USSR Academy of Sciences 163.4 (1965).
8. International Projects. Kherson State University, <http://www.kspu.edu/>
9. World's Data More Than Doubling Every Two Years. Dell-EMC, <https://www.emc.com/about/news/press/2011/20110628-01.htm>

10. Ukrainian scientific citation index. Ukrainian Research and Academic Network, <http://uincit.urau.ua/scientists/fronts/about>
11. Semantic Scholar. Allen Institute for Artificial Intelligence, <https://www.semanticscholar.org/#subscribe>
12. ICI Journals Master List. Index Copernicus, <http://jml2012.indexcopernicus.com/page.php?page=2>
13. Introducing JSON. ECMA International, <http://www.json.org/>
14. What is ORCID?, <https://orcid.org/about/what-is-orcid/mission>