

MEASURING THE RESEARCH CAPACITY OF A UNIVERSITY: USE OF WEB OF SCIENCE AND SCOPUS

Head of Unit, Dr. Sergey Shashnov¹

Junior Research Fellow, Maxim Kotsemir^{1,*}

¹Institute for Statistical Studies and Economics of Knowledge, National Research University – Higher School of Economics, **Russia**

ABSTRACT

This paper discusses the possibilities and limitations of the use of publication databases such as Web of Science and Scopus to determine the research capabilities and prospective areas of research and development of universities. It also analyses major problems related with the analysis of universities' publication activities in Scopus and Web of Science databases such as author surname variations, identification of author profile among authors with the same surname, author and organization profile merging, identification of author affiliation etc.

This paper proposes a list of bibliometric indicators for the analysis of publication activities of individual researchers, university departments and universities as a whole. Furthermore, it describes the methodological approaches for interpreting these indicators.

Finally, the paper reviews the possibilities of VOSviewer software for analysis of different aspect of publication activities at individual and department level such as international collaboration networks, detection of the hot topics of research activity and co-citation networks.

Keywords: university research capacity, bibliometric analysis, research evaluation, publication activity, international citation databases

INTRODUCTION

Russia is in the process of implementing a number of initiatives that are recognized to stimulate research activity in universities, direct them towards solving some of the most important socio-economic problems and significantly increase the effectiveness of use of R&D results. The most important of these initiatives include developing innovation infrastructure at universities, promoting collaboration between universities and business sector, involvement of universities in complex high-tech production projects, awarding grants to universities to attract leading professors and researchers, supporting and developing national research universities, and the implementation of the 5/100 Programme which funds 15 universities in their ambition to develop into the top 100 ranked universities globally. It's targeting to empower 5 Russian universities to enter this prestigious league.

These initiatives involve the selection of key development priorities and concentrating resources and efforts on those areas where Russia can enter a leading global position and where development is greatly needed in the modern Russian economy (Shashnov and Poznyak, 2011 [1]). Thus, many universities are faced with the task of defining a system of mid- and long-term priorities for their R&D and teaching activities. The chosen priorities are recognised to achieve the highest possible level of results, but only

if they fully take into account the existing research potential of a university and its departments. This task is therefore of crucial interest to many universities in Russia.

This article describes an approach to defining the research potential of a university and its departments based on a bibliometric analysis of its publication activity in Web of Science and Scopus database.

Tools of bibliometric analysis are widely used for research evaluation on a country (e.g. Falagas et al., 2006 [2]; Tian et al. 2008 [3]; Zhang et al., 2009 [4]; Kotsemir, 2012 [5], Zyoud et al., 2013 [6]), organization (eg. Liu et al., 2005 [7]; Sevukan and Sharma, 2008 [8]; Van Raan, 2008 [9]) or individual researcher level (Lewison, 2001 [10]; Van Raan, 2006 [11]). The results of this approach make it possible to rank countries, organizations and individual researchers according to indicators of their publication activity, to identify the impact of certain scientific results on other research, and to determine the intensity of national and international scientific collaboration of organisations and individual researchers.

Analysis of a university's research potential

A bibliometric analysis of the research potential of a university (and its departments) in international scientific citation databases like Web of Science and Scopus can be examined as aggregation of publication activity of its staff members.

Web of Science was built on the basis of the first global citation database, the Science Citation Index, which was developed by E. Garfield and the Institute for Scientific Information (USA) in 1964. This database was launched in early 1990s and is now owned by Thomson Reuters. As of July 2015, Web of Science contained roughly 60.0 million document records. The Scopus database was formed by the publishing corporation Elsevier in 2004. As of July 2015, almost 57.6 million documents had been indexed in Scopus.

We propose the following approach for evaluation the research potential of a university:

- I. Selection of key fields of science (areas of research) for analysis according to the university's (department's) scientific specialisation.
- II. Formation of the population of staff members (university teachers and researchers) to be analysed.
- III. Deep analysis of the university's (department's) corpus of publications in different aspects.

Below is a short overview of the stages of this approach and the main problems which arise when following this method.

I. Selection the key areas of research according to the university's (department's) specialisation

This stage involves defining the key areas of the university's (department's) research within some classification of fields of science. When selecting these areas, it is worth to take into account certain restrictions which exist in these databases (Jacso, 2005 [12]; Yang and Meho, 2006 [13]).

One of the advantages of Web of Science is that it has a detailed classification of research areas. The 'research areas' search field (SU search filed in advanced search mode in Web of Science) classifies publications under 151 different areas. The 'Web of

Science categories' search field classifies (WC search filed in advanced search mode in Web of Science) scientific journals into 263 research areas. In general, we can ignore these differences and consider both 'research areas' and 'Web of Science categories' as classifiers of publications, since the share of publications 'outside the journal scope' are few in number. 'Web of Science categories' classification scheme is more detailed than 'research areas'. E.g., according to the 'research areas' classification, physics corresponds to a single scientific field: 'Physics'. While under the 'Web of Science categories' classification, physics is broken down into 8 scientific fields (Physics, Applied; Physics, Atomic, etc.). This classification is more convenient when analysing the thematic structure of publication activity of universities since it proposes more precise identification of publication thematic (e.g. "physics atomic" instead of more general "physics").

At the same time, Web of Science database does not have more general classifications of fields of science. However, Thompson Reuters developed a scheme of matching Web of Science categories with OECD fields of science classification¹. Despite this, the problem is that users have to manually enter all of the 'Web of Science categories' that they want to aggregate into one of the OECD fields of Science.

In Scopus, the situation is reversed. In Scopus fields of sciences are classified only on 27 board research areas (like "physics and Astronomy", "Mathematics", "Medicine" etc.) and there is no detailed classification of fields of science. A more detailed classification of scientific fields (314 'Subject Categories') grouped in 27 'Subject Areas' can be found on the electronic analytical resource SCImago Journal and Country Rank, developed on the basis of Scopus², but not within Scopus itself. Obviously that research areas classification is very general and board for description the thematic structure of publications of university, its departments or individual researchers.

II. Formation the population of university staff members to be analysed.

Research potential of the university can be analysed both on the level of the university as a whole and on the level of individual staff members. The main problem which should be solved at this stage is how to take into account all publications of university (departmental) staff members in Scopus and Web of Science. Unfortunately, each of the databases mentioned have their own difficulties in solving this issue.

The main restriction of Web of Science database when searching for publications of specific authors is the lack of unique author identifier system. Accordingly, Web of Science does not have author profiles either. For any corpus of publications Web of Science offers users only a list of authors' names and the number of publications of each author without any additional information. Users therefore have to identify authors and their affiliations themselves. Due to the lack of a unique author identifier system, users do not know how many authors have been grouped together under records such as 'Ivanov A', 'Chan Y', 'Smith J', 'Kumar C', 'Nakamura T' etc. Unfortunately, there is no satisfactory solution to this problem on Web of Science.

¹ This scheme can be accessed at http://incites.isiknowledge.com/common/help/h_field_category_oecd_wos.html

² For example, a classification of journals can be accessed at <http://www.scimagojr.com/journalrank.php>. The SCImago Journal and Country Rank resource is free. Users can find basic information on the publication activity of various countries (for the period 1996 to 2013) and key bibliometric indicators of scientific publications indexed in Scopus (for the period 1999 to 2013).

Meanwhile Web of Science allows users to search for specific authors through the 'Author search' menu. However, when using this function, certain difficulties again arise which prevent users from fully taking into account all publications of a specific author.

The main difficulty is the loss of publications where the author's affiliation is not indicated. In Web of Science (as well as in Scopus), many publications (especially publications in Russian journals) do not have author affiliations or sometimes even an abstract. Thus, when searching for an author through their association with a particular organization, these publications are omitted. We cannot know in advance what share of a specific author publications will be omitted due to lack of affiliation (for some authors, the share of publications without affiliations is up to 80%). Using the 'Include records that do not contain organization information' function and taking into account publications without affiliations helps in part to solve this problem. Meanwhile this approach works only if the user is convinced that all publications without affiliations are publications of this specific author. This approach is more or less applicable only for authors with very rare surnames and forename-patronymic combinations.

Another serious problem is the lack of standardised organization names and unique research institution identifiers in Web of Science. Users have to manually select all variants of names of affiliations (organisations) where the given author works from the drop-down list of affiliations in the 'Author search' menu. In addition, users have to manually compile a list of all variations on an organization's name in advance.

The main advantage of Scopus when searching for individual researchers is the fact that it has a unique author identifier system. Each author has a unique identification code. As a result, Scopus makes it possible to collect all publications of a given author in his or her profile, including publications, which do not have any affiliation.

However, even when searching for individual authors on Scopus users still face a range of difficulties. Since Scopus is an English-language system, problems arise when transliterating the surnames of Russian (as well as Chinese, Spanish, and Indian etc.) authors. For authors whose surnames include letters which can be transliterated in different ways, users have to search for all possible spelling variants of the given author's surname. The other, more significant problem is the existence of namesakes without affiliations. For common surnames (Ivanov, Yang, Saenz, Carlos, Wright, Singh, Kumar, Lee, etc.), it is impossible to take the full list of publications of a specific author since in Scopus exists dozens of Kumar Ps (Ivanov As, Chang Ys, Smith Js etc.) without any affiliation.

In addition, mechanisms of automatic author affiliating in Scopus in many cases work with bugs. These bugs occur for all researchers and are not country specific. For example, these mechanisms can affiliate a specific author with another organization or add publications of namesakes into his or her author profile in Scopus. Manual correction of affiliation errors is possible only if the number of 'namesakes' (i.e. authors with the identical surnames and forename-patronymic) is no more than 5-6 and only if the user knows precisely where this specific author works. In case of common surnames (Ivanov, Yang, Saenz, Carlos, Wright, Singh, Kumar, Lee, etc.) the full detection and correction of errors in author affiliating procedure is impossible.

Since Web of Science has no unique organization and author identifiers, it is extremely difficult to analyse publications of specific departments. In essence, Web of Science

takes into account only those publications of organizations where the author has indicated his or her affiliation. The publication activity of an organization's department can be analysed on Web of Science using the following algorithm. First, the user determine the set (corpus) of publications of a given organisation using the search fields OG (Organization-Enhanced, the relatively standardized names of organizations), and OO (Organization, the short names of organizations with greater name variations). Further user should extract all authors from this set of publications. We should take into account, that this list of authors will include not only authors of a given organisation but also their co-authors from other organizations. Since for this list of authors only author surnames (forename and patronymic initials) and the number of his or her publications in this set of publication are available author name disambiguation is impossible in case of common surnames. The user cannot determine automatically if 16 publications of "Ivanov A" are papers of one Ivanov A. or this author name aggregates 5 publications by several authors.

Scopus offers far more capabilities in terms of analysing the publication activity of an organization as a whole and of its departments. In addition to the unique author identifiers in Scopus, there are unique organization identifiers which aggregate into a single profile the majority of publications affiliated with the specific organization³. In organization's profile user will find the list of authors that was automatically linked to the organization by Scopus author affiliating mechanisms⁴. Therefore, the user can select members of the required department from the full list of authors within the profile of organisation in Scopus and add those department members who were not added into organization profile automatically. For each staff member, it is also possible to restrict the search by years of his or her working in a particular department or by affiliation with the given organization.

In view of these capabilities and restrictions, we have developed an approach to analyse the research potential of a university department. The Scopus database should be used to analyse the publication activity of individual researchers as well as the department as a whole. The Web of Science database should be used to for analysis the thematic structure of the set of publications of this department.

III. The analysis of the university's (department's) research potential can be carried out in the following order:

- Analysis of the dynamics of publication activity of university (departmental) staff members;
- Analysis of national and international scientific collaboration (detection of key national and foreign partners for a university (and its departments));
- Analysis of co-authorship networks on the level of individual authors;
- Characteristics of journals (conference proceedings, book series etc.) where the set of publications of university (department) was issued

³ Clearly, in all cases to some extent publications affiliated with the given organization do not link up with its profile in Scopus. Nevertheless, the share of such "homeless" publications is quite small (in general no more than 25%, ignoring organizations with extremely low numbers of publications). However, users can extract all variants of the organization's name from its list of publications and then take into account all of its publications which were not automatically linked to its profile. Moreover, Scopus team step-by-step add "homeless" publications to an organization's profile in coordination with members of the organization.

⁴Lists of staff members in an organization's profile on Scopus are far from complete, as Scopus, in many cases, incorrectly identifies the affiliation of authors. Moreover, an author's profile in Scopus shows only one affiliation and authors themselves cannot correct their own affiliation.

- Analysis of the citation indicators;
- Analysis of the thematic structure of publication activity and its dynamics;
- Defining of hot topics of research activity of university (its departments) using the capabilities of VOSviewer software;
- Ranking of departments within the university based on their research potential.

The analysis of the university's (department's) research potential is based on indicators for individual researchers. The following key indicators were selected:

- Total number of publications (in Scopus);
- Total number of citations;
- Hirsch index;
- Average number of citations per publication;
- Share of self-citation and citation by co-authors;
- Share of publications without citations.

To see the whole picture of the level of research of a particular author, we should to look at the number of his or her publications from various angles of view: the co-authors of publications; the countries which the co-authors of the publications represent; the organizations which the co-authors of the publications represent; the types of publication (article, conference paper, review, book chapter, book, etc.) and its quality; the thematic fields of publications; the quality of journals (conf. proceedings, book series etc.) where author's works are published. The choice of a particular aspect of analysis is determined by the research objectives.

Various relative indicators, such as the 'share of publications in foreign scientific journals' and 'share of publications in co-authorship with foreign researchers' can be calculated based on absolute figures.

The key indicator of quality of journals where the publications of a given researcher are issued is its impact-factor. We also should take into account not only the value of impact factor itself but also the position of journal in the ranking on impact factor within some thematic area.

When interpreting citation indicators, we need to take into account that citation practices vary considerably across disciplines, and even within a single discipline. For a more comprehensive analysis of citation levels for an author's work, additional citation indicators are needed: the share of self-citation and citation by co-authors, the share of citation from publications by foreign authors (without Russian co-authors), the share of publications which have never been cited, and indicators of the distribution of citing publications by journals, organisations and countries. Based on the distribution of citing publications by countries it is possible to assess the popularity of publications of a given authors in global scientific community. Indicators such as the 'share of publications by foreign authors in the total number of citing publications' can be good proxy for measuring the popularity of publications of a given author abroad.

For analysis of publication activity of university or its departments, we can take indicators used for assessment of individual authors: total number of publications; total number of citations; average number of citations per publication etc.

Capabilities of the VOSviewer software can also be used for visualization of different aspects of publication activity of university as a whole its department and individual staff membersthe. VOSviewer was developed in 2010 by specialists at Leiden

University⁵. This software is one of the most highly developed and at the same time easy-to-use software for so-called bibliometric mapping. VOSviewer software works with standardised descriptions of publications downloaded from Scopus or Web of Science and with corpuses of simple texts⁶. VOSviewer extracts items from the publication descriptions and plots them on a two-dimensional map. These items can be: terms from abstracts and titles, the authors of the publications, affiliations of the authors, journal names. Assessments of the similarity (closeness) of the objects under investigation are calculated based on their co-occurrence in the given set of objects (publications)⁷. These objects are further placed on the map in based on the following principle: the more frequently two terms co-occurred in titles and abstracts of publications the closer they are to each other on the map.

By working with standardized descriptions of publications extracted from Scopus and Web of Science, VOSviewer can map terms extracted from the abstracts and titles of publications, as well as the authors' affiliations, author names when building a co-authorship map. When building a co-citation map, there is the possibility of mapping not only the authors and their affiliations, but also the journals where the given set of publications is issued.

Author mapping makes it possible to identify existing co-author groups and to identify those university staff members who, for some reason, have not been involved in these co-authorship networks. Mapping the organizations extracted from the set of publications of specific organization (or department), allows to identify key national and foreign partners for this organization (department) and to develop the collaboration policy.

The results of the analysis of university (department) research potential using the proposed approach can be used in a different ways:

- Correction of the university (department) research and publication activity strategy;
- Identification of priority research areas and adjusting the thematic work plan,
- Opening of new research lab within departments for research in priority research areas;
- Formation of research groups around the most productive staff members;
- Invitation of leading national and foreign researchers and professors;
- Development of programmes of national and international collaboration.

CONCLUSION

The proposed approach is not only of interest for Russia, but could also be used for other countries which have comparatively little background in assessing the research potential of universities using international scientific citation databases. In addition, the problems of taking into analysis all publications of a given researcher in international citation databases are fairly important for author from any country especially for researchers with hard-to-transliterate surnames. This approach can also be used in foresight studies to identify a university's (department's) scientific research priorities. The following approach can be used for this purpose:

⁵ The portal dedicated to this software can be found at <http://www.vosviewer.com/Home>

⁶ An article describing the main technical characteristics of the VOSviewer programme can be found at <http://repub.eur.nl/pub/14841>

⁷ An article describing in detail the 'visualization of similarities' method can be found at <http://repub.eur.nl/pub/7654>

- carrying out a pre-foresight study, including defining the key fields of research activity of a university department and the methods to be used;
- analysis the publication activity of university departments in these fields of research activity to determine its research potential;
- selection of preliminary Research priorities for the university departments based on materials from national and foreign foresight studies and the results of the publication activity analysis;
- carrying out surveys and formation a summary list of medium- and long-term research priorities.

In the context of such a study, prospective research areas can be determined using a bibliometric analysis, a department's research potential can be analysed, potential national and foreign scientific partners can be identified, etc., which can all be further refined through additional review procedures.

REFERENCES

- [1] Shashnov S., Poznyak A. (2011) Nauchno-tehnologicheskie priority dlya modernizatsii rossiyskoy ekonomiki [S&T Priorities for Modernization of Russian Economy]. Foresight-Russia, vol. 5, no 2, pp. 48-56 (in Russian).
- [2] Falagas, M. E., Papastamataki, P. A., & Bliziotis, I. A. (2006). A bibliometric analysis of research productivity in Parasitology by different world regions during a 9-year period (1995–2003). BMC infectious diseases, 6(1), 56.
- [3] Tian, Y., Wen, C., & Hong, S. (2008). Global scientific production on GIS research by bibliometric analysis from 1997 to 2006. Journal of Informetrics, 2(1), 65-74.
- [4] Zhang, G., Xie, S., & Ho, Y. S. (2009). A bibliometric analysis of world volatile organic compounds research trends. Scientometrics, 83(2), 477-492.
- [5] Kotsemir M. (2012) Dinamika rossiyskoy i mirovoy nauki skvoz' prizmu mezhdunarodnykh publikatsiy [Dynamics of Russian and World Science through the Prism of International Publications]. Foresight-Russia, vol. 6, no 1, pp. 38-58 (in Russian).
- [6] Zyoud, S. E. H., Al-Jabi, S. W., Sweileh, W. M., & Awang, R. (2013). A bibliometric analysis of toxicology research productivity in Middle Eastern Arab countries during a 10-year period (2003-2012). Health research policy and systems/BioMed Central, 12, 4-4.
- [7] Liu, N. C., Cheng, Y., & Liu, L. (2005). Academic ranking of world universities using scientometrics-A comment to the "Fatal Attraction". Scientometrics, 64(1), 101-109.
- [8] Sevukan, R., & Sharma, J. (2008). Bibliometric analysis of research output of biotechnology faculties in some Indian central universities. DESIDOC Journal of Library & Information Technology, 28(6), 11-20.
- [9] Van Raan, A. F. (2008). Bibliometric statistical properties of the 100 largest European research universities: Prevalent scaling rules in the science system. Journal of the American Society for Information Science and Technology, 59(3), 461-475.
- [10] Lewison, G. (2001). The quantity and quality of female researchers: a bibliometric study of Iceland. Scientometrics, 52(1), 29-43.
- [11] Van Raan, A. F. (2006). Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. scientometrics, 67(3), 491-502.
- [12] Jacso P. Testing the calculation of a realistic h-index in Google Scholar, Scopus, and Web of Science for FW Lancaster //Library trends. – 2008. – T. 56. – №. 4. – C. 784-815.
- [13] Meho L. I., Yang K. Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar //Journal of the American society for information science and technology. – 2007. – T. 58. – №. 13. – C. 2105-2125.