

Webliometric Indicators as Elements of the AI Technique of Estimation of the Language Teacher's Net Proficiency

Matukhin PAVEL^{a,1}, Provotorova ELENA^a, Petrova MARINA^a, Gracheva OLGA^b,
Rybakova IRINA^a and Ismailova KHOLISAKHON^c

^aPeoples' Friendship University of Russia (RUDN University), Moscow, Russia

^bInstitute of International Educational Programs of the Moscow State Linguistic
University (IMOP MGLU), Moscow, Russia

^cTchaikovsky Moscow State Conservatory, Moscow, Russia

Abstract. The article deals with the issues related to the scientometric indexing of the net resources of the group of language teachers and scientists. A detailed accounting of all types of publications allows us to obtain initial information about the level of their Internet engagement. The analysis was carried out including a wide range of genres of publications, the structure of publications, the format and language of publications, the corpus of academic subjects, the language aspects of the publications under study, the composition of the authors of publications, based on their position. The authors' use of wide opportunities to present their developments on various Internet resources provides them with the opportunity to be detected by search systems based on artificial intelligence (AI) as well as Big Data techniques and to be most fully characterized by existing and prospective scientometric systems.

Keywords. monitoring of publication activity, scientometric indexing, Big Data, artificial intelligence systems.

1. Introduction

Currently, we can observe an intensive growth in the number of theoretical and applied research in the field of development and application of Big Data and Artificial Intelligence (AI) systems for evaluating the performance indicators of scientific and pedagogical staff of higher education, particularly in the field of foreign language teaching (FLT). An integral part of the work of teachers are development of teaching resources, participation in scientific conferences and seminars, regular publication of scientific papers. At the same time, there is publication methods' redistribution in favor of the network placement of the results of the intellectual activity of teachers. One of the tasks of Big Data as well as AI systems is to monitor publication activity. To solve the problems of indexing the plenty of works of authors and their teams, it is necessary to develop a set of indicators based on which the selection of publications for inclusion in a particular database is carried out, and the corresponding indexes are calculated. Also,

¹ Corresponding Author, Peoples' Friendship University of Russia (RUDN University), Moscow, Russia;
E-mail: m-pg@mail.ru.

it is necessary to develop the information and technological competence of the authors themselves who create electronic educational resources. In this case, there is a need to create tools that allow authors to search for information posted on the global network and carry out statistical analysis of the obtained data sets [1].

The interest in monitoring and analyzing the publication activity of various groups of authors engaged in the study of specific scientific issues or who are members of the same teaching staff is essential for evaluating the organization's effectiveness and management of the work of the department, faculty.

One example of indexing systems is the Scopus system of ELSEVIER B. V. Corporation. [<https://www.scopus.com>], which is currently the primary source of quantitative information about the publication activity of employees of scientific and educational institutions of the Russian Federation. The main quantitative characteristic of this scientometric system is the Hirsch index which depends on the technical means of distributing scientific publications. It can undergo significant changes over time as a particular work becomes available to different groups of researchers. Such considerations allow us to suggest that the AI systems being developed today will soon form and calculate more complex indicators.

Another example of accounting and monitoring of publication activity is the Digital Object Identifier (DOI) system by the CrossRef. A big plus is a possibility of as-signing such a number not only to the entire scientific work but also to each part. DOI eliminates an error in identifying an information product and facilitates its search in the network, facilitating access to it.

Russia has created its electronic scientometric database – the National Electronic Library [elibrary.ru], combined with the RSCI (Russian Science Citation Index) service, which repeats the principles of the Scopus system in several parameters.

Considering the central systems of scientometric indexing, we can conclude that there is an active development of such services. In addition, the strategic direction of their development is the search for forms of application of AI elements and systems that allow monitoring and evaluating the publication activity of individual authors and author groups to determine their scientific productivity and level.

2. Methods

This study aimed to develop a set of characteristics that describe the level of information and technical competence and Internet engagement of the authors of electronic language educational resources. It is based on a statistical analysis of those information products that they created for use at different levels of language education in the humanities, natural science, and technical direction. At the same time, only published works in electronic format and placed in open access in the databases of electronic publications were studied. It should be clarified that speaking about the Internet engagement of the authors; we mean their focus on the use of tools and resources of the Internet in their work. It also includes the authors' competence in network communications with the editorial offices of electronic publications of various profiles, knowledge of the principles of the device, and operation of the main scientometric systems, which allows authors to develop and implement a targeted policy of promoting their publications. In our article, we rely on the results of a study conducted using the collector_stat v. 1. 0 programs package, which was developed by the staff of the Joint Electronic Resources Fund "Science and Education" (OFERNiO) [2]. This instrument allows selecting from

the database of the fund containing huge electronic resources. Unfortunately, this tool does not allow making complete statistical analysis of data, which reduces the possible potential of the program. Because of this, it was decided to try to combine the features offered by collector_stat v. 1.0 and the MS EXCEL table processor. An array of 561 papers was taken as a base for the study. The team of authors consisted of 90 authors who created these works for 11 years – from 2005 to 2020.

The set of initial research data is further called in work as the webliography of a group of authors. This term is introduced by analogy with the traditional concept of bibliography, which can be considered a subset of webliography, which includes printed works that are not presented in electronic versions on the Internet resources. It should be noted that the webliography contains the full texts of manuscripts and basic information about them, given according to the MARK IV standard. They can be obtained using information search engines when the full texts of publications are not available.

Similarly, we use the term "webliometry" to describe a set of techniques and quantitative characteristics of a webliographic massive obtained by mathematical and statistical methods.

During the research, special attention was paid to the manuscripts of FLT in the field of natural and computer science. The studied webliography corpus of teachers of general and language disciplines includes information products on Latin, Russian as a foreign language, anatomy, and physics for medical students and future information technology specialists. The works included in the webliography of the study were created for use at preparatory faculties for students from Asia, Africa, and Latin America of different departments including medical and agricultural institutes.

Almost all project participants have prepared and applied many new forms of presentation of educational content (electronic notes, presentations, videos, tests, etc.). To provide the complete presentation of the works of each participant, not only traditional types of publications, such as monographs, articles, teaching aids, were included. We use the more general concept of "Information product" to refer to all types of materials created by teachers. By accumulating electronic methodological materials on teaching language and general education disciplines, the teacher can create his educational blogs. Unfortunately, access to such information products is sometimes difficult, and artificial intelligence systems should play a saving role in solving this problem.

It was necessary to form a set of appropriate webliometric parameters based on the obtained statistical data. It aimed to get a general landscape characterizing the level of Internet engagement and the richness of the tools' range used by the authors of electronic teaching and scientific publications. We have selected such indicators that allow us to get the most general idea of various forms of publication activity of project participants. Here we present a few of the main ones from among these characteristics.

3. Results and discussions

We took a corpus of 561 publications for scientific, methodological, and educational purposes to conduct a statistical study. The authors of these works are 90 employees of educational institutions. First, the structure of publications was considered, describing the number of publications with each participant's participation in the project. We have analyzed separately Russian-language and English-language works. At the same time,

all the works of the study participants that are contained in the OFERNiO database were considered [3].

These collections have a structure resembling a cone. The analyzed array of webliography is based on the works of 5-6 authors, which are characterized by the largest number of publications included in the study sample. This group of authors makes up no more than 10 percent of members of the sample, but the number of publications of these authors per person is the largest. It can be concluded that these authors show the greatest creative activity and can be considered the initiators and organizers of creative and educational projects. Above them are the names of authors who participated in three and two creative projects. The number of these authors is also quite small. The most significant group consists of authors whose works are presented in the database only in the singular. It does not mean that these authors have no other manuscripts at all. It indicates that other works of these authors were not included in the studied corpus of publications (which again indicates the need to create comprehensive AI-based search algorithms that will allow, in the absence of a single depository, to build the complete array of webliography as a set of publications from different sources containing the necessary information about the publications of a particular author).

Table 1. Genres of IP.

Genre	Number	Percentage
article	107	19,1%
talk abstracts	88	15,7%
annotation	45	8,0%
IC (information card)	45	8,0%
report	45	8,0%
ATD USPD)	45	8,0%
ATD copy	45	8,0%
ATD text	45	8,0%
ERR certificate)	45	8,0%
TA (teaching aid)	28	5,0%
EMM	18	3,2%
contributions	3	0,5%
guide	2	0,4%

The next stage of the analysis was structuring the webliography and getting an opportunity to determine which type of publication is the most common. Analyzing the sample available to us, we determined that the project participants used a wide range of genres of electronic scientific and methodological publications presented by them on the website of the OFERNiO resource. The list of genres is given in table 1. ATD USPD means advertising and technical description, a unified system of program documentation, ERR - electronic resource registration, EMM - educational and methodical manual.

The analysis of academic subjects presented in the studied corpus of scientific and methodological IP is of interest. The list of 22 items includes Language disciplines (English, Latin, Russian, and Russian as a foreign language), Information disciplines (programming, computer science), Natural science disciplines (biology, physiology).

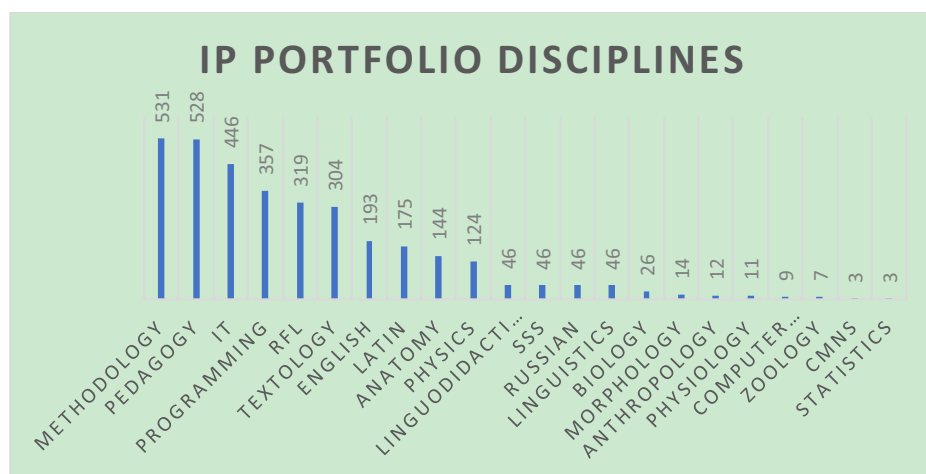


Figure 1. The structure of the IP corpus academic disciplines.

Most of the scientific and methodological papers related to general pedagogy and teaching methods are presented. These works contain various examples of the author's techniques. Works on information technologies and programming represent the second largest group. A significant part of the studied works belongs to the field of language disciplines. These are both languages themselves (Russian and foreign), as well as linguistics and linguodidactics. Here SSS means the scientific style of speech, RFL - Russian as a foreign language, CMNS - concepts of modern natural science).

The diagram shown in Figure 1 gives a visual representation of the structure of the webliographic array. There are several main groups in the spectrum. These general methods and pedagogics form the basis of the pyramid. It characterizes the group of developers as specialists focused not only on creating manuals but also on publicly presenting their methods and pedagogical technologies for the study, analysis, and application by their colleagues. Information technologies and Programming form the next tier of the structure. It reflects the orientation of IP developers to use modern tools in their manuals. This group also includes Testology, the achievements of which are considered in the preparation of electronic forms of training and control-measuring components of educational and methodological complexes. A particular group is formed by language disciplines (RFL, Russian and Latin) [4]. The scientific publications on the Russian language are included in the same group in a separate line. At the same time, these publications can also be attributed to the group of disciplines of the scientific and methodological profile, which includes Linguodidactics, Linguistics. Publications on these disciplines, in turn, form the basis for academic subjects that are taught at the preparatory faculty for foreign citizens and the initial bachelor's degree courses. The latter also includes such disciplines as Morphology and Anthropology.

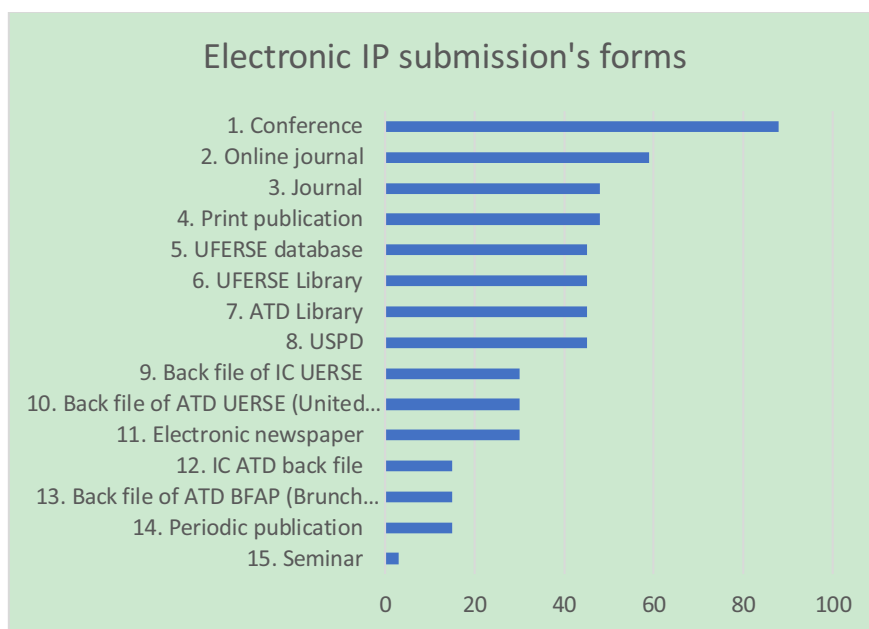


Figure 2. Forms of submission of electronic IP.

The most popular format of publications was conference materials in abstracts and articles, which occupy 18 % of publications. The second most popular is articles in scientific journals, which make up 11 %. The next ones are publications directly in OFERNiO. In addition, the materials of scientific seminars are the least represented. An analysis of the types of educational products created by the project participants was carried out. It should be noted that in most cases, the publication could be assigned to different categories. All of them were identified as informational, most of them as language and profile-specific. A significant part can be considered as soft-ware and algorithmic materials and electronic means of educational and methodological support of the learning process. Forms of the submission of the electronic IP are listed in diagram shown in Figure 2.

The works are mainly presented in two languages – Russian and English. Out of the total number of authors (90), 50 people took part in writing joint works in English, that is, more than half, while the number of English-language works is only 9 percent. Many joint works explain this discrepancy in indicators. In addition, these works are intended mainly for students of Russian universities who are studying in Russian. Accordingly, the demand for Russian-language works is much higher than for English-language ones. The presence of English-language publications is probably due to the need to be published in foreign journals, which can also confirm a sufficiently high level of authors of publications.

4. Conclusion

Summing up the above analysis of the overall picture and details of the study of several indicators of Internet engagement by a group of authors-developers of the considered corpus of webliography, we can draw a crucial key conclusion.

A detailed account of publications allows us to identify the main nuances and shades of the spectrum of special and general information technology tools that were used by the developers of electronic IP, which allows us to obtain initial information about the level of their Internet engagement. The authors' use of wide opportunities to present their developments on various Internet resources provides them with a chance to be detected by AI-based search systems and to be most fully characterized by new scientometric systems based on an integrated approach to evaluating the webliography of individual authors and author groups. The systematic registration of the results of developments in structures like OFERNIO not only allows each of the authors to present their results in a wide range of genres of scientific and pedagogical publications. When forming indicators of the publication activity of teachers, presented in the form of electronic IP, it is advisable to combine the data array of collector_stat-type systems based on a sample from the OFERNIO database with the works of authors posted on other platforms. The use of complex characteristics and statistical methods of their analysis makes it possible to establish the presence and determine the relationship between different types of information products and different groups of authors [5, 6].

References

- [1] Galkina AI, Matukhin PG, Matyash GA, Provotorova EA. State and cooperative resources as a means of access to information and methodological support for educational communication. Proceedings of the 13th International Conference-Exhibition Information Technologies in Education; Moscow: LLC NPP "BIT pro"; 2013. p.30-31.
- [2] Galkina AI, Grishan IA. The program of Collector_stat v. 1.0 - No. AAAA-A17-117041810029-2 dated 23.10.2017; Moscow: CITIS; 2017.
- [3] Portal of the United Fund for Electronic Resources of Science and Education, URL <http://ofernio.ru/portal/newspaper.php>, last accessed 2021/06/25.
- [4] Matukhin PG, Galkina AI, Gracheva OA, Komissarova NV, Zaporozhets VV, Saltykova OV. Reflexive portfolio of a team of authors - developers of information products of the FTP "Russian language 2016-2020". Educational Technologies and Society. 2018; 4:360-372.
- [5] Galkina A and Grishan I. The electronic portfolio of a research university in the mirror of webliometrics. Bulletin of Science and Practice. 2021; 7(5):441-450.
- [6] Galkina AI, Burnasheva EA, Grishan IA, Kadyrova EA. Electronic reflective portfolio of a university teacher in diagrams, tables, graphs. Educational Technology and Society. 2018; 21(2):500-514.