ISSN 0147-6882, Scientific and Technical Information Processing, 2013, Vol. 40, No. 1, pp. 11–16. © Allerton Press, Inc., 2013.
Original Russian Text © V.M. Moskovkin, 2013, published in Nauchno-Technicheskaya Informatsiya, Seriya 1, 2013, No. 1, pp. 26–31.

The Construction of Academic Publishing and Terminological Structures Using the Google Scholar Search Engine: an Example of Environmental Terms in Publications at the Classical Universities of Kharkiv and Skopje[1]

V. M. Moskovkin[a, b]

[a] *Belgorod State University, Belgorod, Russia*
[b] *Karazin Kharkiv National University, Kharkiv, Ukraine*
e-mail: Moskovkin@bsu.edu.ru

Received October 8, 2012

Abstract—A methodology for constructing academic terminological and publishing structures using the Google Scholar search engine is presented. These structures are formed for the classical universities of Kharkiv and Skopje based on the example of basic environmental terms of a general nature that are distinguished in English-language publications.
The environmental theme is more actively studied at Kharkiv National University. The first publications produced by the researchers of this university in the considered area of science were written in the early 1980s.
An analysis of the most-frequently cited publications that contain the selected environmental terms shows that such publications are often a result of the work of an international team of authors. This is of great importance for the development of journal strategies and policies. By using Google Scholar, one can find significantly more publications on the website of Kharkiv University compared to that of the University of Skopje, due to the electronic open access archive system of publications at Kharkiv University. The frequency of occurrence for publications that contain selected environmental terms is higher in the case of the University of Skopje due to the fact that the publications of this university are poorly represented on the web and also because of their predominantly English-language character.

Keywords: publication structures, clusters of publications, Karazin Kharkiv National University, University of Skopje (FYR Macedonia), Google Scholar, ecology, environment, biodiversity

DOI: 10.3103/S0147688213010024

INTRODUCTION

It is convenient to use the Google Scholar search engine for analyzing publication activity trends and for constructing publication clusters and structures. Without doubt, the clusters and structures that are created using this engine are less accurate than those built using information analytical systems, such as Web of Science and SCOPUS; however, the advantages of Google Scholar consist in the ability to use it free of charge and to promptly obtain the results of the conducted search.

A publication cluster, or more precisely a cluster of published articles, which is created for one random scientific term, is understood as a set of scientific papers that refer to this particular term. These clusters are viewed both in the broad and narrow sense. In the former case, one can assume that a specific term is used in the text of a scientific paper, while in the latter case this term is used in publication titles [1]. Essentially, these kinds of publication clusters are created by researchers who analyze the output that was obtained from enquiries about specific scientific terms using Google Scholar [2–7]. A sound and comprehensive approach to the construction of such clusters was first suggested in [1].

Publication structures can, for example, be viewed as publications that are distributed across various fields of science (subject areas of Google Scholar) for different universities. The construction and analysis of such academic publication structures has been conducted based on the example of various universities in the study [8].

It is possible to create narrower academic publication structures on the basis of germane publication clusters that are related to various universities. Such

[1] This paper was prepared on the basis of a report that was presented by the author at a conference entitled The Impact of Modern Technologies on the Environment and Environmental Protection Measures. This conference was held in the framework of the Days of Science dedicated to the Former Yugoslav Republic of Macedonia and organized in Ukraine on September 10–14, 2012.

structures can be described as academic terminological publication structures. They can be constructed within a single field of knowledge by using Google Scholar for both a general search on the Internet and the websites of universities under consideration conducted using the following operator: site: university domain.

This paper reviews the ability to build simple academic terminological publication structures by using Google Scholar based on the example of two classical universities in Kharkiv (Ukraine) and Skopje (FYR Macedonia). Building such structures using any search engine is associated with the problem of establishing a full list of the official names of universities due to the possibility of changes in their names.

RESEARCH METHODOLOGY

In order to study convertible English-language academic publications, it is necessary to assign English language names to the selected universities.

For V. N. Karazin Kharkiv National University, one can distinguish the following minimal list of English names: Kharkov State University, Kharkiv State University, Kharkov National University, and Kharkiv National University. A search for publications using these names will naturally include all those publications that are referring to a university's names that contain its own title, for example, V. N. Karazin Kharkov National University.

For the University of Skopje, two English-language names were distinguished: University of Skopje and Skopje University. The search for publications conducted by using the first name will also result in publications that refer to the full name of this university, which is the Ss. Cyril and Methodius University of Skopje.

It is appropriate to test the names of the identified universities with Google Scholar, by using one of the functions in Advanced Search called "with exact phrase." Our experience shows that when conducting a search "anywhere in the article," the Google Scholar search engine first of all retrieves those articles in which a university's name is included in the metadata of these articles and is assigned to the workplaces of their authors [8]. If the metadata of the articles do not contain the university's name, it can be found by conducting a search in the entire document, which will result in irrelevant references. Generally, the share of the identified papers whose authors are not related to the university is insignificant.

It should be emphasized that although the search for academic publications is conducted by using the option "with exact phrase," Google Scholar will also provide the responses to other universities if their titles contain some parts of the names of the universities, for example, Kharkov National University of Agriculture (Economics, etc.) when Kharkov National University is sought. Our study resulted in an insignificant number of such responses.

Let's assume that we want to know which publications that are related to some title of a university refer to some term. For this purpose, there is an entry called "with at least one of the words" under the "with exact phrase" entry. The term has to be specified in this particular place. In addition, in order to test this term with an exact phrase, it should be put in parentheses.

Next, we will test basic environmental terms, such as "ecology," "biodiversity," and "environment," taking account of the fact that the latter can refer to any kind of environment. Unless we put this term in parentheses, we will also obtain references to papers that contain the terms "environmental" and so on. It is also to be noted that if the places of entries for university names and terms are changed, the results of the search will be the same. For example, the term "ecology" and the name of the university are tested in the entries "with exact phrase" and "with at least one of the words," respectively.

We will conduct a series of experiments with the above-mentioned universities and terms using Google Scholar both without and with time limits. In the latter case, the trends of dynamic publication structures can be defined. In parallel, such experiments allow one to identify the most frequently cited and the earliest indexed papers.

At the last stage of our experiments, the academic terminological publication structure constructed on the basis of publications that was found on the websites of the universities will be analyzed. As mentioned in the introduction, the Google Scholar operator: site: university domain will be used for this purpose.

RESULTS AND DISCUSSION

Let us test three selected terms for all distinguished names of the universities. The first series of experiments will be conducted with no time limit for the "anywhere in the article" search and two specific cases: (1) "include citation" and (2) "at least summaries." To reduce the information noise and to exclude patents, we will tag all seven subject categories. The results of these experiments are presented in Table 1. It can be seen that the number of publications of Kharkiv University is three times larger than that of the University of Skopje for all selected terms. The term "environment" is the most frequently used one, followed by the terms "ecology" and "biodiversity." Differences in the English spelling of Kharkiv (Kharkov and Kharkiv are the Russian and Ukrainian ways of spelling it, respectively) reflect the specific ethnic self-identification of the authors of these papers, as well as their preferences for one of the languages at different periods of time. In the soviet period and during the 1990s, when the classical university of Kharkiv was called Kharkov State University, the English-language publications of

Table 1. The occurrence of the terms "ecology," "biodiversity," and "environment" in the publications of two classical universities in Kharkiv and Skopje (Google Scholar, August 8, 2012)

University name	Term					
	Ecology		Biodiversity		Environment	
	Include citation	At least summaries	Include citation	At least summaries	Include citation	At least summaries
Kharkov State University	42	34	8	8	311	285
Kharkiv State University	9	7	3	2	56	54
Kharkov National University	60	50	19	17	596	566
Kharkiv National University	83	78	18	17	490	423
Total	194	169	48	44	1453	1328
University of Skopje	33	28	13	12	420	378
Skopje University	17	17	5	5	131	122
Total	50	45	18	17	551	500

Table 2. The occurrence of the terms "ecology" and "biodiversity" in the publications of two classical universities in Kharkiv and Skopje in different 5-year intervals (Google Scholar, August 8, 2012)*

University name	Time intervals					
	1986–1990	1991–1995	1996–2000	2001–2005	2006–2010	Total
Kharkov State University	2/11	2/15	5/89	9/53	6/80	24/248
Kharkiv State University	0/0	0/1	4/16	3/17	0/12	7/46
Kharkov National University	0/1	0/4	0/11	15/178	20/263	35/457
Kharkiv National University	0/0	0/1	1/3	24/83	33/196	58/283
Total	2/12	2/21	10/119	51/331	59/551	124/1034
University of Skopje	0/7	1/21	2/32	5/132	15/113	23/305
Skopje University	0/0	0/4	1/14	3/31	6/40	10/89
Total	0/7	1/25	3/46	8/163	21/153	33/394

* "Ecology" is the numerator; "environment" is the denominator; "at least summaries" is the option.

the authors from this university were dominated by the Russian name of their workplace (i.e. Kharkov State University). At the end of the 1990s, this university was granted national status. As a result, in the case of the most frequently used term, "environment," the above-described dominance has significantly shrunk, while in case of the term "ecology," the Ukrainian name, viz., Kharkiv National University, prevailed. It is possible to calculate the frequency of occurrence for the above terms by relating it to the total number of responses.

Thus, for the most frequently used term "environment" (see Table 1) the share of publications that refer to this term was 17.5% of the total number of publications for the name Kharkov National University, 21.7% for the name Kharkiv National University, and about 31% for both English-language names of the University of Skopje. In these experiments, we used the option "at least summaries."

Let's analyze the publication activity of the researched universities by the two most frequently used terms for 5-year intervals (Table 2). An increase in the publication activity by all names of the universities (for Kharkov State University this trend that could be observed until the early 2000s) is associated with global trends in environmental research. However, the factual growth in the publication activity is less significant because of a smaller scale of digitalization and indexation of journal papers at the end of the 20th century compared to the first decade of the 21st century.

The existence of publications that refer to the old name of Kharkiv University in the period between 2001 and 2010 is due to three factors: (1) a considerable time lag between the submission of manuscripts to journal editors and their actual publication; (2) errors made by the Google Scholar search engine in identifying the time of publication due to the improperly structured metadata of the articles; and (3) some authors still mention the old name of their university

by custom (we came across papers that were published in the period between 2006 and 2008 in which the authors referred to their workplace as Kharkov State University).

Google Scholar allows one to identify the earliest indexed and the most highly cited papers for random terms [1].

As can be seen from Table 2, the first indexed papers that refer to the selected terms were published in the period from the early to mid-1980s. Google Scholar also allows one to identify the most highly cited papers by using the option "cited by," which enables one to review all publications that cite the selected paper. Meanwhile, one can be convinced visually that the most highly cited paper belongs to the university. As mentioned above, a paper can refer to a specific university, but at the same time it can be produced by an author from another organization.

Table 3 presents the most highly cited articles for the terms "ecology" and "environment." In the case of the term "ecology" and the old name of the university in Kharkiv, the citation indexes of the articles we found were not higher than three or four; this is the reason that we did not include in table 3 the rarely cited papers that contain this term.

It should be noted that Google Scholar sometimes borrows digitalized and indexed books from Google Books. In the case of the term "ecology," two such books were identified: *Monotone Random Systems. Theory and Application* by Igor Cheshuev from Kharkiv National University and *Alpine Biodiversity in Europe* (a group of authors, including one from the University of Skopje). Both publications were published 10 years ago and collected 87 and 69 citations, respectively. An article published in series B of Proceedings of the Royal Society (Great Britain) obtained more than 60 citations in 5 years. One of the coauthors of this paper was Sergei Ultevskii from the Zoology and Animal Ecology Chair of Kharkiv National University. In addition, two other papers in which Sergei Ultevskii was one of the coauthors that obtained 12 and 19 citations, respectively, are distinguished. The minimal number of citations for the term "ecology" was obtained by a paper that was coauthored by an author from the University of Skopje and dedicated to the endangered species that populate the Balkan region (see Table 3).

For the term "environment" a number of highly cited papers, which were published by the authors from the two universities, was identified in the field of chemistry, physics, astronomy, and medicine. Thus, A. Doroshenko from the Institute of Chemistry at Kharkiv National University was mentioned as the coauthor of two papers. As well, I. Bel'skaya from the Astronomy Observatory at the same university published a paper on ArXiv.com as part of an international team of authors.

The most impressive paper was produced by a large team of authors (about 150 scientists) from 27 universities and research institutes around the world. This paper, which was published in *Nature Photonics* in 2007, has already been cited 469 times. Yu. Ivanisenko from Kharkiv National University is one of the authors in this team. This paper presented the results of measuring an unprecedented peak and the average strength of a coherent powerful ultraviolet source of radiation using a free-electron laser (13.7 nm).

Among two papers that were published by Macedonian authors that refer to the term "environment," one can distinguish a paper on power energy that was coauthored by L. Greev from the Faculty of Electrical Engineering at the University of Skopje. This article was cited 213 times in 22 years (see Table 3). Table 3 shows that the most highly cited papers are published by collective international teams of authors. Therefore, by increasing citation rates, international coauthorship also increases the impact factors of journals that publish these articles.

While conducting these experiments, we noticed that the Faculties of Ecology and Geology and Geography of Kharkiv National University are practically excluded from presenting research findings in English. This finding can be a signal for the scientific management of these two faculties pointing to the need to stimulate English-language publications, as has, for instance, been done at the Belgorod State University.

It is important to mention that the research on the use of such broad terms ("ecology" is a field of knowledge; "environment" is a setting in various fields of knowledge) in the stream of academic publications is not effective for the study of detailed publication structures and dynamic trends. However, this research was important for us in order to test the methodology of constructing academic terminological publication structures by using the Google Scholar search engine. In order to conduct an in-depth study of publication structures, it is necessary to test narrower terms, which cannot always be described by one word. Our experiments on testing two compound terms in the field of environmental pollution and nature conservation ("water pollution," "air pollution," "soil pollution," and "nature conservation") for the two universities only resulted in 10 to 11 responses by Google Scholar.

There was practically no response to the queries made on the basis of ecological and economic terms, such as "ecological management," "ecological marketing," "ecological audit," and several others. Obviously, the classical universities analyzed here conduct large-scale research in a wide range of topics in the area of nature conservation, as well as ecological and economic problems; however, the research findings are primarily published in their national languages.

Let us analyze the frequency of occurrence for the three initially distinguished terms in the publications that are made available on the websites of the universi-

Table 3. The most highly cited publications that were produced by authors from two classical universities in Kharkiv and Skopje, which contain the terms "ecology" and "environment" (Google Scholar, August 11, 2012)

The names of universities	The authors of a paper	Output data of a publication	The number of citations obtained by a paper
		Ecology	
Kharkov National University	I. Chyeshov	Monotone Random Systems. Theory and Application (Lecture Notes in Mathematics). Springer, 2002, books.google.com (Book)	87
Kharkiv National University	Mark E. Sindall, Peter Troutelj, Serge Y. Ultevsky, Mary Nkamany, Kenneth S. Mackdonald	Diverse molecular data demonstrate that commercially available medicinal leeches are not turido medicinalic // Proc. R. Soc. B, 2007. – Vol. 274, No. 1617. – P. 1481–1487.	67
Skopje University	Laszlo Nagy, Des Thomson, Georg Grabherr, Christian Korner	Alpine biodiversity in Europe, 2003, books.google.com (Book)	69
University of Skopje	D. Lakusic, F. Conti	Aseneuma pichlery (lampanulaceae), a neglected species of the Balkan// Plant Systematics and Evolution, 2004. –Vol. 247, No. 1–2. – P. 23–36.	12
		Environment	
Kharkov State University	A. O. Doroshenko, A. V. Kirichenko, V. G. Mitina, O.A. Ponomaryov	Spectral properties and dynamics of the excided state structural relaxation of the ortho analogues of POPOP. Effective abnormally large Stokes shift Cuminophores // Journal of Photochemistry and Photobiology. A: Chemistry, 1996. – Vol. 94, No. 1. – P. 15–26.	43
Kharkiv State University	V. G. Pivovarenko, A. V. Klueva, A. O. Doroshenko, A. P. Demchenko	Bands separation in fluorescence spectra of ketocyanine dyes: evidence for their complex formation with monodydric alcohos // Chemical Physics Letters, 2000. – Vol. 325, No. 4. – P. 389–398.	32
Kharkov National University	N. Peixinho, A. Doressoundiram, A. Delsanti, H. Boehnhardt, M. A. Barucci, I . Belskaya	Reopening the TNOs Color Controversy: Centaurs Bimodality and TNOs Unimodality // ArXiviastro-ph/030942v1, 2003	51
Kharkiv National University	Y. Ivanisenko, et al.	Operation of a free-electron laser from the extreme ultraviole to the water window // Nature Photonics1, 2007. – P. 336–342.	469
University of Skopje	L. Grcev F. Dawalibi	An electromagnetic model for transients in grounding systems // Power Delivery, IEEE Transaction on Power Delivery, 1990. – Vol. 5, No. 4. – P. 1773–1782.	213
Skopje University	Nada Pop-Jordanova	Psychological characteristics and biofeedback mitigation in preadolescents with eating disorders // Pediatrics International, 2000. – Vol. 42, No. 1. – P. 76–81	24

ties. For this purpose, one should first indicate the following operators site:univer.kharkov.ua and site:ukim.edu.mk in the top entry of the Advanced Search function offered by Google Scholar. In the former and latter case, we obtained 7190 and 504 references, respectively, as of August 15, 2012. In the sec-

Table 4. The frequency of occurrence for the selected terms in publications that were posted on the websites of two classical universities in Kharkiv and Skopje (Google Scholar, August 15, 2012)

Term	Kharkiv National University		University of Skopje	
	The number of responses		The number of responses	
	abs.	%	abs.	%
Ecology	46	0.6	20	4.0
Biodiversity	20	0.3	12	2.4
Environment	151	2.1	129	25.6

ond stage, the selected terms have to be indicated in the entry "with exact phrase." By relating the responses that were obtained from these terms to their total number, we will obtain the respective frequency rates. The results of this study are presented in Table 4.

As can be seen from Table 4, the frequency of the occurrence of ecological terms is higher by an order of magnitude on the website of the University of Skopje compared to that of Kharkiv National University. This situation is due to a very fair representation of publications on the web in the case of the University of Skopje (overall, 500 publications compared to about 7000 publications that were uploaded on the website of Kharkiv National University), as well as the better English-language representation of these publications. The availability of publications published by the authors from Kharkiv National University is related to the DSpace OA (open access) repository that was launched at this university in 2009 in the framework of the 2008 Belgorod Declaration on Open Access to Scientific Knowledge and Cultural Heritage. Neither the University of Skopje nor other universities of FYR Macedonia have access to such a repository. The availability of such repositories has a positive impact on the webometric performance of universities, which has been calculated twice a year by the Spanish Cybermetric Laboratory since 2004. In this context, one should note that the operator site: "university domain" had until recently been used by this laboratory for the calculation of the SCHOLAR indicator, which was part of an aggregate webometric indicator that provided the basis for global university rankings (since 2012, university publication activity has been calculated using the database of the Spanish SCIMAGO group).

CONCLUSIONS

We elaborated and presented in this paper a methodology for construction of academic terminological publication structures using the Google Scholar search engine. Such a publication structure was constructed for the classical universities of Kharkiv and Skopje based on the example of simple environmental terms that are contained in the English-language body of publications.

It has been shown that the environmental research we analyzed is conducted more actively at Kharkiv National University. One can consider that this research was started at the university in the early 1980s.

The analysis of the most highly cited publications in the selected environmental area has shown that such publications are produced by international teams of authors. Google Scholar identifies significantly more publications on the website of Kharkiv University compared to the University of Skopje, which is due to the existence of the OA repository at the former university. At the same time, the frequency of occurrence for publications that refer to the selected environmental terms was higher in the case of the University of Skopje because of the poor availability of publications on the website of this university and their primarily English-language character.

REFERENCES

1. Moskovkin, V.M., Constructions of clusters of research results with the help of Googl'e tools, *Nauchn.-Techn. Inf., Ser.2.,* 2012, no. 8, pp. 9–13.
2. Noruzi, A., Google scholar: the new generation of citation indexes, *Libri,* 2005, vol. 55, pp. 170–180.
3. Robinson, M.L. and Wusteman, J., Putting Google scholar to the test: a preliminary study, *Progr.: Electron. Lib. Inf. Sys.*, 2007, vol. 41, no. 1, pp. 71–80.
4. Aalst, J., Using Google Scholar to estimate the impact of journal articles in education, *educ. Res.*, 2010, vol. 39, no. 5, pp. 387–400.
5. Mastrangelo, G., et al., Literature search on risk factors for sarcoma: Pub. Med. and Google Scholar may be complementary sources, *BMC Res. Notes*, 2010, no. 3, p. 131. http://www.biomedcentral.com/1756-0500/3/131
6. Walters, W.H., Comparative recall and precision of simple and expert searches in Google Scholar and eight other databases, *Portal: Lib. Acad.,* 2011, vol. 11, no. 4, pp. 971–1006.
7. Moskovkin, V.M., Simulation expert system for making students' college decision, *Autom. Docum. Mathem. Ling.,* 2009, vol. 43, pp. 292–295. http://www.springerlink.com/content/14788m08u7q7745x/fulltext.pdf
8. Moskovkin, V.M., Delux, T., and Moskovkina, M.V., Comparative analysis of university publication activity by Google Scholar (by example of leading Chech and Germany universities, *Cybermetrics,* 2012, vol. 16, no. 2, paper 2. http://cybermetrics.cindoc.csic.es/articles/v16i1p2.html