

УДК 004:378.4:[002.2:001.4:502]

В. М. Московкин

## Построение университетских публикационно-терминологических структур с помощью поисковой машины Google Scholar: на примере экологических терминов и классических университетов Харькова и Скопье\*

*Развита методика построения университетско-терминологических публикационных структур с помощью поисковой машины Google Scholar. Такие структуры построены для классических университетов Харькова и Скопье на примере простейших экологических терминов общего характера в англоязычном сегменте публикаций.*

*В Харьковском национальном университете имени В. Н. Каразина рассматриваемая экологическая тематика ведется более активно, и первые публикации ученых этого университета в рассматриваемой области исследований можно отнести к началу 80-х гг. XX века.*

*Анализ наиболее цитируемых публикаций с избранными экологическими терминами показал, что такие публикации представлены международными коллективами авторов. Это обстоятельство имеет большое значение при разработке журнальных стратегий и политики. Google Scholar находит на сайте Харьковского университета намного больше публикаций по сравнению с сайтом университета Скопье, что связано с наличием электронного архива публикаций открытого доступа в первом университете. Частота встречаемости публикаций с избранными экологическими терминами для второго университета была больше, что связано с плохим web-представлением публикаций университета Скопье и их преимущественно англоязычным характером.*

**Ключевые слова:** публикационные структуры, кластеры публикаций, Харьковский национальный университет имени В. Н. Каразина, университет Скопье (Македония), Google Scholar, ecology, environment, biodiversity

### ВВЕДЕНИЕ

Для изучения трендов в публикационной активности и быстрого построения публикационных кластеров и структур удобно использовать поисковую машину Google Scholar. Конечно, построенные с ее помощью публикационные кластеры и структуры будут уступать по точности таковым, построенным с помощью информационно-аналитических поисковых систем Web of Science и SCOPUS, но бесплатность и

быстрота получения результатов дают преимущества Google Scholar.

Под публикационным кластером, а точнее, кластером научных статей, порожденным произвольным научным термином, мы будем понимать совокупность научных статей, использующих данный термин. Такие кластеры будем понимать, как в широком, так и узком смысле. В первом случае предполагается, что данный термин используется в тексте научных статей, а во втором – в их заголовках [1]. Фактически, такого рода кластеры публикаций строятся разными исследователями, когда они анализируют отклики на запросы научных терминов, с помощью Google Scholar [2-7]. Но систематический и строгий подход к построению таких кластеров впервые предложен в работе [1].

\* Статья подготовлена на основе доклада прочитанного автором на конференции “Влияние современных технологий на окружающую среду и меры по ее защите”, проводимой в рамках Дней науки Македонии в Украине (Харьковский национальный университет имени В.Н.Каразина, 10-14 сентября 2012 г.)

Под публикационными структурами можно понимать, например, распределения публикаций по разным областям знания (предметные категории Google Scholar) для разных университетов. Построение и анализ таких университетских публикационных структур на примере различных университетов были проделаны в работе [8].

Более узкие университетские публикационные структуры можно строить на основе родственных кластеров публикаций, относящихся к разным университетам. Такие структуры можно назвать университетско-терминологическими публикационными структурами. Их можно строить в рамках одной области знания, с помощью Google Scholar как при поиске по всему Интернету, так и по сайтам изучаемых университетов с помощью оператора: `site: университетский домен`.

Здесь мы рассмотрим возможность построения простейших университетско-терминологических публикационных структур с помощью Google Scholar на примере двух классических университетов из Харькова (Украина) и Скопье (Македония). При построении таких структур с помощью любой поисковой системы проблема состоит в подборе полного перечня официальных названий университетов с учетом возможного изменения их названий в течение времени.

## МЕТОДИКА ИССЛЕДОВАНИЙ

Если мы хотим изучать конвертируемые англоязычные университетские публикации, то нам при поиске следует задавать англоязычные названия выбранных университетов.

Для Харьковского национального университета имени В.Н. Каразина можно выделить следующий минимальный перечень таких англоязычных названий: *Kharkov State University, Kharkiv State University, Kharkov National University, Kharkiv National University*. Поиск публикаций по этим названиям, естественно, включит и все публикации, в которые входят названия университета с собственным его именем, например, *V.N. Karazin Kharkov National University*.

Для университета Скопье мы выделили два англоязычных названия: *University of Skopje, Skopje University*. При поиске публикаций по первому названию мы уловим также статьи, в которых этот университет фигурирует под полным названием: *Ss. Cyril and Methodius University of Skopje*.

Выделенные названия университетов, с помощью Google Scholar, целесообразно тестировать в строке «with exact phrase» («с точной фразой») в расширенном поиске (Advanced Search). Наш опыт показывает, что при поиске по всей статье («anywhere in the article») поисковый механизм Google Scholar в первую очередь возвращает статьи, в которых название университета входит в метаданные статей и приурочено к месту работы их авторов [8]. Если в метаданных статей запрашиваемое название университета отсутствует, то оно может быть найдено при поиске по всему документу, и тогда это будет нерелевантный отклик. В целом, доля найденных статей, авторы которых не относятся к запрашиваемому университету, незначительна.

Следует также обратить внимание на то, что, несмотря на поиск университетских публикаций с помощью опции «с точной фразой», Google Scholar будет давать отклики на другие университеты, в случае если в их названия входят составной частью названия запрашиваемых университетов. Например, *Kharkov National University of Agriculture (Economics и др.)* при запросе *Kharkov National University*. В наших экспериментах доля таких откликов была незначительной.

Допустим, мы хотим посмотреть, в каких публикациях того или иного названия университета упоминается тот или иной термин. Для этого ниже строки «with exact phrase» имеется строка «with at least one of the words». В эту строку и следует записывать изучаемый термин. При этом, чтобы этот термин тестировался также с точной фразой, его следует брать в кавычки.

Мы будем тестировать простейшие экологические термины «ecology», «biodiversity» и «environment», понимая при этом, что последний термин может относиться к любым средам, а не только к окружающей среде. Если мы не поставим последний термин в кавычки, то дополнительно будем получать отклики на статьи, в которые входят термины «environmental» и др. Отметим, что от перемены строк названий университета и термина результаты поиска не изменятся. Например, термин «ecology» тестируется в строке «with exact phrase», а название университета идет в кавычках в строке «with at least one of the words».

С помощью Google Scholar нами будут проделаны эксперименты с вышеуказанными названиями университетов и терминов как без ограничения, так и с ограничением времени. В последнем случае определяются тренды в динамике публикационных структур. Попутно в этих экспериментах удастся идентифицировать наиболее цитируемые, а также наиболее ранние индексируемые публикации.

В последней серии экспериментов будет изучена университетско-терминологическая публикационная структура, построенная на основе публикаций, найденных на сайтах изучаемых университетов. Для этого в Google Scholar, как отмечалось во Введении, будет использоваться оператор: `site: университетский домен`.

## РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЙ И ИХ ОБСУЖДЕНИЕ

Итак, протестируем три выбранных термина для всех выделенных названий университетов. Первую серию экспериментов проведем без ограничения времени при поиске по всей статье («anywhere in the article») для двух случаев, включая: 1) цитирование («include citation»); 2) по крайней мере, резюме («at least summaries»). Для уменьшения информационного шума и исключения патентов, поставим метки перед всеми семью предметными категориями. Результаты экспериментов приведены в табл. 1. Как видим, количество публикаций Харьковского университета более чем в три раза превышало количество публикаций университета Скопье по всем избранным терминам. Наиболее часто встречался термин «environment», за ним следовали термины «ecology» и «biodiversity». Различия в написании города Харьков на английском языке (*Kharkov* – русифицированное название, *Kharkiv* – украинизиро-

ванное название) дает некоторые представления об этнической самоидентификации авторов публикаций и их приверженности к тому ли иному языку в различные периоды. В советский период и в 90-е годы XX в., когда классический университет Харькова назывался Харьковским государственным университетом, в англоязычных статьях ученых из этого университета преобладало русифицированное название - Kharkov State University в качестве места их работы. В конце 1999 г. этот университет получил статус национального, и для наиболее распространенного термина «environment» такое преобладание стало на много меньше, а по термину «ecology» преимущественно перешло к украинизированному названию Харьковского национального университета имени В.Н. Каразина. Можно подсчитать частоту встречаемости вышеуказанных терминов, относя её к общему количеству откликов.

Так, для наиболее часто встречаемого термина «environment» (см. табл. 1) доля публикаций, в которых встречается этот термин, составила 17,5 % от общего количества публикаций для названия Kharkov National University, 21,7 % – для названия Kharkiv National University и около 31% для обоих англоязычных названий университета Скопье. Отсюда следует, что рассматриваемый термин в относительных единицах в 1,4 – 1,7 раза встречался чаще в публикациях университета Скопье. В этих экспериментах использовалась опция «at least summaries».

Таблица 1

**Встречаемость терминов «ecology», «biodiversity» и «environment» в публикациях двух классических университетов Харькова и Скопье (Google Scholar, 8 августа 2012 г.)**

Название университета	Термин					
	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
Всего	194	169	48	44	1453	1328
University of Skopje	33	28	13	12	420	378
Skopje University	17	17	5	5	131	122
Всего	50	45	18	17	551	500

Рассмотрим теперь публикационную активность рассматриваемых университетов по двум наиболее часто встречающимся терминам на пятилетних интервалах (табл. 2). Возрастающая по времени публикационная активность по всем названиям университетов (для Харьковского госуниверситета это справедливо до начала 2000-х гг.) связана с глобальными трендами в экологических исследованиях. Но фактический рост этой публикационной активности будет меньше, учитывая меньший охват оцифрования и индексирования журнальных публикаций в конце XX в., по сравнению с первым десятилетием XXI в.

Таблица 2

**Встречаемость терминов «ecology» и «environment» в публикациях двух классических университетов Харькова и Скопье на различных пятилетних интервалах (Google Scholar, 8 августа 2012 г.)\***

Название университета	Интервалы времени					Всего
	1986–1990	1991–1995	1996–2000	2001–2005	2006–2010	
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
Всего	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
Всего	0/7	1/25	3/46	8/163	21/153	33/394

\* Числитель – «ecology», знаменатель – «environment»; опция – «at least summaries»

Наличие публикаций, подписанных их авторами старым названием Харьковского университета в период с 2001 по 2010 гг., связано с тремя факторами: 1) большим временем запаздывания между представлением рукописей в редакцию журнала и выходом их в печать; 2) ошибками поискового механизма Google Scholar в идентификации времени публикаций, связанными с плохо структурированными метаданными статей; 3) некоторые авторы по привычке дают старое название университета (мы видели статьи 2006 – 2008 гг., авторы которых указывали название места своей работы как Kharkov State University).

Google Scholar позволяет идентифицировать наиболее ранние индексируемые и наиболее цитируемые статьи для произвольных терминов [1].

Из табл. 2 следует, что первые индексируемые статьи по избранным нами терминам появились в начале и середине 80-х гг. XX в. Google Scholar также позволяет идентифицировать наиболее цитируемые статьи с помощью опции «cited by», которая позволяет просматривать все публикации, цитирующие избранную статью. Но при этом нужно визуально убедиться, что наиболее цитируемая статья принадлежит тестируемому университету. Как мы отмечали ранее, в статье может упоминаться конкретный университет, но сама статья может принадлежать автору из другой организации.

Для терминов «ecology» и «environment» наиболее цитируемые статьи приведены в табл. 3. Для термина «ecology» и старого названия рассматриваемого университета из Харькова цитируемость найденных статей не превышала 3 – 4, поэтому мы не стали приводить низкоцитируемые статьи для этого термина.

Отметим, что иногда Google Scholar заимствует оцифрованные и индексируемые книги из Google Books. В нашем случае для термина «ecology» это были «Монотонные случайные системы. Теория и приложения» Игоря Чуешова из ХНУ имени В. Н. Каразина и «Альпийское биоразнообразие в Европе» (коллектив авторов, один из которых из университета Скопье). Обе работы, опубликованные десять лет назад, набрали, соответственно, 87 и 69 ссылок. Более 60 ссылок за пять лет набрала статья, опубликованная в одной из серий Докладов Королевского общества (Великобритания), соавтором которой был Сергей Ултевский с кафедры зоологии и экологии животных ХНУ имени В. Н. Каразина. Отмечены также две другие англоязычные статьи, в которых соавтором был Сергей Ултевский, и набравшие 12 и 19 ссылок. Минимальное количество ссылок для термина «ecology» набрала статья, подписанная одним из авторов университета Скопье, и посвященная изучению исчезающих видов Балкан (см. табл. 3).

Для термина «environment» были идентифицированы высокоцитируемые статьи в области химии, физики, астрономии и медицины, принадлежащие ученым из двух рассматриваемых университетов. Здесь А. Дорошенко из Института химии ХНУ имени В. Н. Каразина фигурировал в двух коллективных статьях, а И. Бельская из Астрономической обсерватории этого же университета с международным коллективом авторов опубликовала статью в ArXiv.com.

Наиболее впечатляет статья огромного коллектива авторов (около 150 ученых) из 27 университетов и НИИ мира, опубликованная в «Nature Photonics» в 2007 г. и уже процитированная 469 раз. В составе этого коллектива был Ю. Иванисенко из ХНУ имени В. Н. Каразина. В этой работе с помощью лазера на свободных электронах (13,7 nm) были измерены беспрецедентный пик и средняя мощность для когерентного мощного ультрафиолетового радиационного источника.

Среди двух статей македонских ученых, в которых встречался термин «environment», отличается статья

по электроэнергетике, в которой соавтором является L. Grcev с факультета электротехники университета Скопье. Она за 22 года была процитирована 213 раз (см. табл. 3). Из табл. 3 следует, что в целом наиболее цитируемые статьи написаны международными коллективами авторов, таким образом международное соавторство, повышая цитируемость работ, повышает также и импакт-факторы журналов, в которых эти статьи были опубликованы.

При проведении этих экспериментов мы заметили, что экологический и геолого-географический факультеты ХНУ имени В. Н. Каразина были практически исключены из процесса представления результатов своих исследований на английском языке. Это может послужить сигналом для научного менеджмента вышеуказанных факультетов в деле стимулирования англоязычных публикаций, как это сделано, например, в Белгородском государственном национальном исследовательском университете.

Важно отметить, что изучение встречаемости такого рода широких терминов («ecology» – область знания, «environment» – среды в различных областях знаний) в потоке университетских публикаций мало что даёт при изучении детализированных публикационных структур и трендов в их динамике. Для нас такое изучение было важно с точки зрения отработки методики построения университетско-терминологических публикационных структур с помощью поисковой машины Google Scholar. Для более детального изучения публикационных структур необходимо тестировать более узкие термины, которые не всегда представлены одним словом. Наши эксперименты по тестированию двухсложных терминов в области загрязнения окружающей среды и охраны природы («water pollution», «air pollution», «soil pollution», «nature conservation») для двух изучаемых университетов давали не более 10 – 11 откликов от Google Scholar.

Практически отсутствовали отклики на запросы «ecological management», «ecological marketing», «ecological audit» и другие подобные эколого-экономические термины. Очевидно, что рассматриваемые классические университеты проводят широкие исследования по всему спектру вышеуказанной природоохранной и эколого-экономической проблематики, но публикуют их результаты, в основном, на национальных языках.

Посмотрим теперь, как часто встречаются выделенные нами, первоначально, три термина в публикациях, размещенных на сайтах рассматриваемых университетов. Для этого, на первом шаге, в самой верхней строке расширенного поиска Google Scholar следует записать операторы site:univer.kharkov.ua и site:ukim.edu.mk. В первом случае, на уровень 15 августа 2012 г., мы получили 7190 откликов, во втором – 504. На втором шаге избранные термины следует записывать в строке «with exact phrase». Относя отклики от этих терминов к общему их числу, получим их частоты встречаемости. Результаты проделанных экспериментов приведены в табл. 4.

Наиболее цитируемые статьи ученых из классических университетов Харькова и Скопье, в которых встречаются термины «ecology» и «environment» (Google Scholar, 11 августа 2012 г.)

Название терминов и университетов	Авторы статьи	Выходные данные публикации	Количество цитирований статьи
<b>Ecology</b>			
Kharkov National University	I. Chyeshov	Monotone Random Systems . Theory and Application (Lecture Notes in Mathematics) . Springer, 2002, books.google.com (книга)	87
Kharkiv National University	Mark E. Sindall, Peter Troutelj, Serge Y. Ultevsy, 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, № 1617. – P. 1481 – 1487.	67
Skopje University	Laszlo Nagy, Des Thomson, Georg Grabherr, Christian Körner	Alpine biodiversity in Europe, 2003, books.google.com (книга)	69
University of Skopje	D. Lakušic, F. Conti	Aseneuma pichlery (lampanulaceae), a neglected species of the Balkan// Plant Systematics and Evolution, 2004. –Vol. 247, № 1-2. – P. 23 – 36.	12
<b>Environment</b>			
Kharkov State University	A. O. Doroshenko, A. V. Kirichenko, V. G. Mitina, O.A. Ponomaryov	Spectral properties and dynamics of the excited 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, № 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 monodyric alcohols// Chemical Physics Letters, 2000. – Vol. 325, № 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 // ArXivastro-ph/030942v1, 2003	51
Kharkiv National University	Y. Ivanisenko, et al	Operation of a free-electron laser from the extreme ultraviolet 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, № 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, № 1. – P. 76 – 81.	24

Частоты встречаемости избранных терминов в публикациях, размещенных на сайтах классических университетов Харькова и Скопье (Google Scholar, 15 августа 2012 г.)

Термин	ХНУ имени В. Н. Каразина		Университет Скопье	
	Количество откликов		Количество откликов	
	абс.	%	абс.	%
Ecology	46	0,6	20	4,0
Biodiversity	20	0,3	12	2,4
Environment	151	2,1	129	25,6

Из табл. 4 видим, что частоты встречаемости экологических терминов на сайте университета Скопье на порядок выше, чем на сайте Харьковского университета, что связано с очень слабым web-представлением публикаций первого университета (около 500 публикаций на фоне более 7 тыс. публикаций, размещенных на сайте ХНУ имени В.Н. Каразина) и их лучшим англоязычным представлением, по сравнению с Харьковским университетом. Большинство публикаций ученых ХНУ имени В. Н. Каразина связано с функционированием DSpace OA (open access) – репозитория в этом университете, который был запущен в 2009 г. в рамках принятия в 2008 г. Белгородской декларации об открытом доступе к научному знанию и культурному наследию. Университет Скопье, как и все остальные университеты Македонии, не имеет такого репозитория. Наличие таких репозитариев положительно влияет на вебметрические рейтинги университетов, которые два раза в год, начиная с 2004 г. рассчитывает Испанская киберметрическая лаборатория. В связи с этим отметим, что оператор site:”домен университета” до последнего времени использовался этой лабораторией в расчёте индикатора SCHOLAR, входящего в интегральный вебметрический показатель, согласно значениям которого назначались глобальные рейтинги университетов (с 2012 г. публикационная университетская активность берётся из базы данных испанской SCIMAGO group).

## ЗАКЛЮЧЕНИЕ

Таким образом, в нашей работе была разработана методика построения университетско-терминологических публикационных структур с помощью поисковой машины Google Scholar. Такого рода публикационные структуры построены для классических университетов Харькова и Скопье на примере простейших экологических терминов в англоязычном поле публикаций.

Показано, что в Харьковском национальном университете имени В. Н. Каразина рассматриваемая экологическая тематика ведётся более активно. Её зарождение можно отнести к началу 80-х гг. XX в.

Анализ наиболее цитируемых публикаций избранной экологической тематики показал, что такие публикации представлены международными коллекти-

вами авторов. Google Scholar находит на сайте Харьковского университета на порядок больше публикаций по сравнению с сайтом университета Скопье, что связано с наличием ОА-репозитория в первом университете. При этом частота встречаемости публикаций с избранными экологическими терминами для второго университета была больше, что связано с плохим web-представлением публикаций университета Скопье и их преимущественно англоязычным характером.

## СПИСОК ЛИТЕРАТУРЫ

1. Московкин В.М. Построение кластеров результатов исследований с помощью инструментов Google // НТИ. Сер.2. - 2012. - № 8. – С.9 – 13.
2. Norouzi A. Google Scholar: The New Generation of Citation Indexes // Libri. - 2005.- Vol. 55. - P. 170-180.
3. Robinson M.L., Wusteman J. Putting Google Scholar to the test: a preliminary study // Program: electronic library and information systems. - 2007. - Vol.41, № 1.- P. 71-80.
4. Aalst J. Using Google Scholar to estimate the impact of journal articles in education // Educational Researcher.- 2010. - Vol. 39, № 5. - P. 387-400.
5. Mastrangelo G. et al. Literature search on risk factors for sarcoma: Pub. Med and Google Scholar may be complementary sources // BMC Research Notes.- 2010. - N 3. - P. 131. - URL: <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: Libraries and the Academy. - 2011. - Vol. 11, № 4.- P. 971 - 1006.
7. Московкин В.М. Имитационная экспертная система выбора университетов для обучения // НТИ. Сер. 2. - 2009. – № 10. – С. 19-21; Moskovkin V.M. Simulation expert system for making students' college decisions // Automatic Documentation and Mathematical Linguistics. – 2009. – Vol. 43, № 5. – P. 292-295. – URL: <http://www.springerlink.com/content/14788m08u7q7745x/fulltext.pdf>.
8. Moskovkin V. M., Delux T., Moskovkina M.V. Comparative Analysis of University Publication Activity by Google Scholar (On Example of Leading Czech and Germany Universities) // Cybermetrics - 2012. - № 16. - Issue (1), Paper 2. – URL: <http://cybermetrics.cindoc.csic.es/articles/v16i1p2.html>.

*Материал поступил в редакцию 08.10.12.*

## Сведения об авторе

**МОСКОВКИН Владимир Михайлович** – доктор географических наук, профессор кафедры мировой экономики Белгородского государственного университета, профессор кафедры экологии и неэкологии Харьковского национального университета им. В. Н. Каразина

E-mail: [Moskovkin@bsu.edu.ru](mailto:Moskovkin@bsu.edu.ru)