

В. М. Московкин

МЕТОДОЛОГИЯ ОЦЕНКИ И СТРУКТУРНОГО АНАЛИЗА УНИВЕРСИТЕТСКОЙ ПУБЛИКАЦИОННОЙ АКТИВНОСТИ С ПОМОЩЬЮ ПОИСКОВОЙ МАШИНЫ GOOGLE SCHOLAR (на примере Приграничного белорусско-российско-украинского университетского консорциума)

Предложена методология оценки и структурного анализа университетской публикационной активности с помощью поисковой машины Google Scholar на примере Приграничного белорусско-российско-украинского университетского консорциума. Она основана на испытаниях всего спектра англоязычных названий университетов для семи областей знаний и пятилетних временных интервалов с помощью запросов в вышеуказанной поисковой машине. Построены детализированные, укрупненные и упорядоченные обобщенные публикационные структуры для университетов консорциума. Предложенный методологический инструментарий может быть полезен для проведения вебметрического университетского бенчмаркинга (мониторинга, аудита) на всех уровнях – международном, национальном, региональном, корпоративном (ассоциации университетов), внутриуниверситетском.

Ключевые слова: университетская публикационная активность, поисковая машина Google Scholar, структурный анализ публикаций, вебметрический университетский бенчмаркинг, Приграничный белорусско-российско-украинский университетский консорциум, методология оценки и анализа.

V. M. Moskovkin

The methodology of an estimation and structural analysis university publishing activity by means of search engine Google Scholar (on an example of the Border Byelorussian-Russian-Ukrainian university consortium)

The methodology of an estimation and a structural analysis university publishing activity by means of search engine Google Scholar is suggested in the work on an example of Border Byelorussian-Russian-Ukrainian university consortium. It is based on tests of all spectrum of English-language names of universities for seven fields of knowledge and five-years time intervals by means of requests in the above mentioned search engine. The detailed, aggregated and ordered generalized publishing structures for consortium's universities were constructed. The offered in the article methodological tools can be useful to realization webometrical university benchmarking (monitoring, audit) at all levels - international, national, regional, corporative (associations of universities), intra-university.

Key words: university publishing activity, search engine Google Scholar, structural analysis of publications, webometrical university benchmarking, Border Byelorussian-Russian-Ukrainian university consortium, methodology of estimation and analysis.

В настоящее время в зарубежной научной литературе, посвященной наукометрическим методам исследования, возник кластер публикаций, связанный с использованием поисковой машины Google Scholar при проведении таких исследований. Отмечается, что до последнего времени единственным источником данных при цитировании были базы данных Института научной информации США (ISI Citation Indexes) [2].

Несмотря на критическое отношение к этим

базам данных, они уже давно широко используются в научном менеджменте за рубежом.

Относительно недавно возникли две альтернативы этим базам данных – коммерческая поисковая система «Scopus», созданная крупнейшим издательством научной периодики «Elsevier», и свободно доступная поисковая машина Google Scholar [2].

В работах [4, 5] показано, что Google Scholar покрывает гораздо большее количество научных

документов по сравнению с базами данных Института научной информации США.

В работе [6] отмечается, что поисковая машина Google Scholar использует новый метод обнаружения потенциально релевантных статей по данной проблематике за счет идентификации статей, которые были процитированы в других работах. Поэтому важной особенностью этой поисковой машины является то, что исследователи могут использовать ее для отслеживания взаимных связей между авторами, цитирующих статьи на одну тему, а также для определения частоты, с которой другие авторы цитируют конкретную статью (за счет опции «cited by»). В вышеуказанной работе сделано заключение, что поисковая машина Google Scholar обеспечивает свободную альтернативу и дополнения к другим индексам цитирования.

В работе [3] отмечается, что базы данных Института научной информации США индексируют около одной трети от общего количества рецензируемых научных журналов, которых сейчас насчитывается около 25 тыс. Google Scholar и Google Books индексируют гораздо больше научных документов, но пока не в состоянии достичь полного их покрытия, так как только 15 % текущего ежегодного научного выхода представлено публикациями открытого доступа (OA-publications) [3].

Наш обзор наукометрических исследований показал отсутствие работ, которые бы изучали публикационную структуру и вебметрические оценки университетского научного выхода с помощью поисковой машины Google Scholar.

При изучении этого выхода мы обратили внимание на то, что его нельзя качественно получить с помощью измерения откликов на URL-адреса сайтов университетов. Для постсоветских университетов характерно большое количество нерелевантных откликов в виде различной административной информации (решения ученого совета, ректората и др.). Для западных университетов характерна ситуация, когда, например, при приблизительно одинаковой публикационной активности ученых американских (Гарвардский и Чикагский университеты) и британских (Кембриджский и Оксфордский университеты) университетов первые имели на порядок больше откликов на запросы их URL-адресов¹, хотя запросы на названия этих уни-

верситетов дали преимущества британским университетам².

Причина этого, на наш взгляд, — в качестве организации информации на сайте. Так, множество откликов на сайт Гарвардского университета связано с наличием научного интернет-магазина (sciencemag.org), а при запросе сайта Чикагского университета (site: uchicago.edu) первая тысяча откликов, разрешенная для показа поисковой машиной Google Scholar, идет на статьи превосходно составленной коллекции журналов университета («Chicago Journals»), размещенной на платформе uchicago.press.

Далее, очевидно, что на сайте университетов представлены далеко не все публикации их ученых, а в постсоветских университетах вообще отсутствует практика размещения опубликованных научных статей на их сайтах.

В связи с вышеизложенным мы решили испытывать с помощью поисковой машины Google Scholar не URL-адреса сайтов университетов, как это делается испанской киберметрической лабораторией при расчете вебметрического рейтинга университетов мира (www.webometrics.info), а их общепринятые англоязычные названия. Эксперименты с ведущими университетами мира показали хорошую релевантность такого поиска. В первую очередь Google Scholar находит статьи, размещенные на онлайн-платформах крупнейших издательств, таких как «Elsevier», «Springer», «Blackwell», «Wiley» и др., т. е. «конвертируемые» статьи, входящие в базы данных Института научной информации США. Кроме того, эта поисковая машина хорошо находит статьи из онлайн-журналов и университетских репозитариев открытого доступа.

Отметим также, что Google Scholar в результате своего поиска включает дополнительно небольшой процент научных монографий, предоставляемых Google Books.

Нам удалось показать, что релевантность расширенного поиска с точной фразой возрастает в следующем направлении: отсутствие ограничений на области наук и интервалы времени → задание областей наук → задание одновременно областей наук и временных интервалов поиска.

Помимо общего количества статей в данной области знаний (7 областей), полученных на запрос англоязычного названия конкретного университета, Google Scholar также дает значения общего числа цитирований на каждую статью с

¹ Site: harvard.edu дает 1 310 000 документов, site: uchicago.edu — 60 400, site: ox.ac.uk — 8090, site: cam.ac.uk — 9330, измерения проводились нами в начале января 2009 г.

² Расширенный поиск с точной фразой (advance search, with exact phrase).

возможностью просмотра названий научных работ, цитирующих данную статью (с помощью опции «by cited»). Наши контакты с командой Google Scholar показали, что пока отсутствует процедура, которая позволяла бы суммировать цитирования по всей совокупности найденных статей, но команда Google Scholar с интересом восприняла идею разработки такой процедуры. При ее реализации появляется возможность рассчитывать полноценный вебметрический рейтинг научно-публикационной активности университетов мира. При расчете такого рейтинга возникает проблема идентификации всех общепринятых названий университетов. Например, для университетов франкоговорящих провинций Канады необходимо использовать их общепринятые франко- и англоязычные названия, для европейских университетов неанглоговорящих стран, помимо англоязычных названий университетов, необходимо использовать все их основные иностранноязычные названия в соответствии с принятыми в этих странах языками. Для постсоветских стран следует учитывать перманентный процесс переименования классических университетов, часть из которых вышла из пединститутов, а другие сменили названия в угоду иных обстоятельств. Для части университетов Приграничного белорусско-российско-украинского университетского консорциума (ПБРУУК), которые ранее были пединститутами, запросы на старые англоязычные их названия дают пренебрежимо малое количество откликов (низкий статус вузов в прошлом; снижение вероятности оцифровки статей при увеличении их возраста). Мы учитывали старое название Южного федерального университета (Southern (South) Federal University) — Rostov State University. Для всех украинских университетов ПБРУУК приходилось делать запросы на несколько названий, учитывая, что эти университеты после распада СССР получили статус национальных (Харьковский, Симферопольский, Запорожский, Донецкий, Восточно-Украинский). В Украине в целом ситуация осложняется еще и тем, что англоязычные названия университетов стали трансформироваться в связи с украинофикацией названий городов, где они расположены (например, Kharkov → Kharkiv). При работе с поисковой машиной GS мы отметили флуктуацию откликов на запросы названий университетов, что связано с возможным временным отсутствием доступа, исключением дублирующих или нерелевантных откликов и др. Поэтому при расчете итогового кумулятивного вебметрического по-

казателя целесообразно, на наш взгляд, использовать сглаживающие процедуры (рассчитывать усредненный временной тренд).

Мы полагаем, что со временем, по мере активизации создания университетских репозитариев открытого доступа, будет возрастать вероятность дублирования откликов на запросы URL-адресов сайтов университетов, так как в таких репозитариях будут размещаться (самоархивироваться) ранее опубликованные статьи (в основном в виде авторских PDF-файлов). Трудно заранее сказать, насколько эффективно поисковая машина Google Scholar будет справляться с нарастающим масштабом дублирования статей.

Некоторый опыт использования поисковой машины Google Scholar для оценки университетской публикационной активности стран MEA изложен нами в работе [1]. Для проведения систематических экспериментов с поисковой машиной Google Scholar нами выбраны десять университетов Приграничного белорусско-российско-украинского университетского консорциума (ПБРУУК). Структурная динамика научных статей, опубликованных учеными из молодых университетов ПБРУУК на начало января 2009 г., показана в табл. 1.

Первые отклики на англоязычные названия украинских университетов, получивших статус национальных, появились в 1999 г. (на название Восточно-Украинского национального университета, Луганск) и в 2005 г. (на название Запорожского национального университета). Для сравнения в табл. 2 и 3 приведены аналогичные данные для старых университетов ПБРУУК — Харьковского национального университета им. В. Н. Каразина и Воронежского государственного университета.

В 1990–1994 гг. появились отклики на украинофицированное англоязычное название первого университета — Kharkiv, и в последующие интервалы времени учитывался национальный статус этого университета (всего четыре варианта названий, отклики на которые прослеживаются вплоть до последнего временного интервала). Из этих таблиц видим, что публикационная активность Харьковского университета в два раза выше, чем Воронежского, и на один-два порядка превышает публикационную активность молодых университетов ПБРУУК. Отметим, что публикационная активность Южного федерального университета, оцененная с помощью поисковой машины Google Scholar, сопоставима с аналогичной активностью Харьковского национального университета им. В. Н. Каразина.

Таблица 1

Структурная динамика научных статей, опубликованных учеными из молодых университетов ПБРУУК
(Google Scholar, 12 января 2009 г.)

Области наук	БелГУ			КурГУ			ВУНУ			ЗНУ						
	1995-1999	2000-2004	2005-2009	2000-2004	2005-2009	Всего	1995-1999	2000-2004	2005-2009	Всего	1985-1989	1990-1994	1995-1999	2000-2004	2005-2009	Всего
1. Biology, Life Sciences, Environmental Science	0	7	17	24	2	1	3	0	0	0	0	0	0	2	1	3
2. Business, Administration, Finance, Economics	0	1	0	1	1	1	2	00	00	03	00	00	00	30	00	30
3. Chemistry, Materials Science	2	25	37	64	3	17	20	3	23	5	31	0	2	7	2	13
4. Engineering, Computer Science, Mathematics	2	6	28	36	2	11	13	8	0	0	8	0	3	9	3	30
5. Medicine, Pharmacology, Veterinary Science	0	0	1	1	0	1	1	0	0	0	0	0	0	0	0	0
6. Physics, Astronomy, Planetary Science	14	67	79	160	0	12	12	5	9	1	15	1	0	7	2	15
7. Social Sciences, Arts, Humanities	0	1	10	11	1	1	2	0	0	0	0	0	1	4	0	5
Всего	18	107	172	297	9	44	53	16	32	6	54	1	5	19	11	69
								1	33	69	103	0	0	0	20	20

Примечания: 1. Здесь и в других табл.: ГГУ – Гомельский государственный университет, БелГУ – Белгородский государственный университет, ВГУ – Воронежский государственный университет, КурГУ – Курский государственный университет, ЮФУ – Южный федеральный университет, ВУНУ – Восточно-Украинский национальный университет (Луганск), ДонНУ – Донецкий национальный университет, ЗНУ – Запорожский национальный университет, ТНУ – Таврический национальный университет, ХНУ – Харьковский национальный университет.

2. Второй ряд цифр для украинских университетов соответствует публикации статусу национального университета.

Таблица 2

Структурная динамика научных статей, опубликованных учеными Харьковского национального университета им. В. Н. Каразина (Scholar Google, 3–6 января 2009 г.)

Области наук	До 1950	1950–														Всего
		1950–1954	1955–1959	1960–1964	1965–1969	1970–1974	1975–1979	1980–1984	1985–1989	1990–1994	1995–1999	2000–2004	2005–2009			
1. Biology, Life Sciences, Environmental Science	1	0	0	0	0	0	1	0	1	8	13	57	41	122		
2. Business, Administration, Finance, Economics	0	0	0	1	0	0	0	0	0	0	2	6	6	15		
3. Chemistry, Materials Science	0	0	2	6	11	25	31	55	31	53	110	205	200	729		
4. Engineering, Computer Science, Mathematics	0	0	0	18	3	3	19	22	20	70	258	653	595	1661		
5. Medicine, Pharmacology, Veterinary Science	0	0	0	1	0	0	0	0	0	0	4	14	14	33		
6. Physics, Astronomy, Planetary Science	0	0	3	3	8	21	23	19	16	100	442	714	633	1982		
7. Social Sciences, Arts, Humanities	2	0	0	0	0	0	1	1	1	5	14	36	30	90		
<i>Всего</i>	3	0	5	28	23	49	75	97	69	236	843	1685	1519	4632		

Таблица 3

Структурная динамика научных статей, опубликованных учеными Воронежского государственного университета (Scholar Google, 9 января 2009 г.)

Области наук	До 1950	1950-1954	1955-1959	1960-1964	1965-1969	1970-1974	1975-1979	1980-1984	1985-1989	1990-1994	1995-1999	2000-2004	2005-2009	Всего
1. Biology, Life Sciences, Environmental Science	1	0	0	1	0	0	1	0	4	8	29	86	64	194
2. Business, Administration, Finance, Economics						0	0	0	0	2	0	15	3	20
3. Chemistry, Materials Science		0	0	2	7	10	10	11	16	14	36	239	214	559
4. Engineering, Computer Science, Mathematics		0	0	1	15	34	34	21	18	50	84	206	171	634
5. Medicine, Pharmacology, Veterinary Science						0	0	0	1	1	0	5	7	14
6. Physics, Astronomy, Planetary Science			0	0	21	26	42	48	33	36	123	299	247	875
7. Social Sciences, Arts, Humanities	0	0	1	0	1	2	4	1	1	2	4	16	12	44
<i>Всего</i>	1	0	1	4	44	72	91	81	73	113	276	866	718	2340

Процентные распределения статей, опубликованных учеными университетов ПБРУУК, полученные с помощью поисковой машины Google Scholar на конец декабря 2008 г., без ограничения на временной интервал поиска, приведены в табл. 4. На основе этой таблицы построена укрупненная публикационная структура (табл. 5), а также упорядоченные обобщенные публикационные структуры для университетов ПБРУУК (табл. 6).

В табл. 5 показано полное превалирование «конвертируемых» англоязычных публикаций в области естественных и технических наук во всех университетах (ситуация, характерная для всех

постсоветских университетов). В отличие от мировых тенденций развития научных областей отмечается небольшая доля публикаций в области медико-биологических исследований (лучшие позиции здесь имеют Таврический, Воронежский и Белгородский университеты). Среди университетов ПБРУУК социально-экономическая и гуманитарная составляющая в публикационной структуре «конвертируемых» публикаций преобладает в молодых Курском и Запорожском университетах, в которых в советское время не успели сложиться традиции проведения фундаментальных естественно-научных исследований.

Таблица 4

Распределение статей, опубликованных учеными из университетов ПБРУУК, по областям наук без ограничений на временной интервал поиска, % (Scholar Google, 30.12.2008 г.)

Области наук	ГГУ	БелГУ	ВГУ	КурГУ	ЮФУ	ВУНУ	ДонНУ	ЗНУ	ТНУ	ХНУ
1. Biology, Life Sciences, Environmental Science	1,9	8,3	8,1	4,9	3,4	0	1,7	3,3	5,9	2,6
2. Business, Administration, Finance, Economics	0,6	1,1	1,1	4,9	0,7	1,6	2,4	4,2	0,6	1,1
3. Chemistry, Materials Science	22,2	20,5	25,0	34,4	39,8	63,8	32,8	11,7	33,0	15,2
4. Engineering, Computer Science, Mathematics	32,1	13,4	26,4	26,3	14,2	15,4	28,0	55,8	13,2	34,1
5. Medicine, Pharmacology, Veterinary Science	0	0,3	0,7	1,6	2,6	0	0,3	0	3,8	1,9
6. Physics, Astronomy, Planetary Science	41,3	51,6	35,8	19,7	37,5	16,5	28,9	16,7	40,0	41,9
7. Social Sciences, Arts, Humanities	1,9	4,8	2,9	8,2	1,8	2,7	5,9	8,3	3,5	3,2
<i>Всего</i>	100	100	100	100	100	100	100	100	100	100

Таблица 5

Укрупненная публикационная структура университетов ПБРУУК, %

Укрупненные области научных исследований	ГГУ	БелГУ	ВГУ	КурГУ	ЮФУ	ВУНУ	ДонНУ	ЗНУ	ТНУ	ХНУ
Все естественные и технические науки, кроме наук о жизни	95,6	85,5	87,2	80,4	91,5	95,7	89,7	84,2	86,2	91,2
Науки о жизни (биология, экология, медицина)	1,9	8,6	8,8	6,5	6,0	0	2,0	3,3	9,7	4,5
Социально-экономические и гуманитарные науки, искусство	2,5	5,9	4,0	13,1	2,5	4,3	8,3	12,5	4,1	4,3
Кумулятивное количество статей	361	351	2690	61	5397	188	911	120	340	5821

**Упорядоченные обобщенные публикационные структуры университетов ПБРУУК
в трех ведущих научных областях**

Упорядоченные количества публикаций	Университеты ПБРУУК	Примечания
1. $N_{Ph.} > N_{Eng.} > N_{Ch.}$	ГГУ, ВГУ, ХНУ	Исторически сложившиеся по значимости приоритетные научные школы
2. $N_{Ph.} > N_{Ch.} > N_{Eng.}$	ТНУ, БелГУ	Исторически сложившиеся по значимости приоритетные научные школы в ТНУ и сформировавшиеся в течение последних десяти лет на основе приезжих коллективов ученых научные школы в БелГУ
3. $N_{Ch.} > N_{Ph.} > N_{Eng.}$	ВУНУ, ДонНУ, ЮФУ	Все университеты из донбасского углехимического региона
4. $N_{Ch.} > N_{Eng.} > N_{Ph.}$	КурГУ	Регион Курской магнитной аномалии
5. $N_{Eng.} > N_{Ph.} > N_{Ch.}$	ЗНУ	Металлургический регион

Примечание. N – количество англоязычных статей, найденных поисковой машиной Google Scholar; Ph. – Physics, Astronomy, Planetary Science; Eng. – Engineering, Computer Science, Mathematics; Ch. – Chemistry, Material Science.

Даты выхода первых статей, обнаруживаемых Google Scholar, относятся к 50–60-м гг. XX в. Согласно кумулятивному количеству статей (табл. 5), все университеты ПБРУУК можно разбить на три условных кластера: университеты с высоким научным выходом (2,5–5,5 тыс. статей) – ХНУ, ЮФУ, ВГУ; университеты с низким научным выходом (300–1000 статей) – ДонНУ, ГГУ, БелГУ, ТНУ; университеты с очень низким научным выходом (до 300 статей) – ВУНУ, ЗНУ, КурГУ.

Упорядоченные обобщенные публикационные структуры «конвертируемых» англоязычных статей хорошо соответствуют исторически сложившимся по значимости приоритетным научным школам и промышленной специализации регионов, в которых расположены рассматриваемые университеты (табл. 6). Приведем примеры ряда конкретных упорядоченных публикационных структур: ХНУ: 2320 > 1989 > 841; ВГУ: 964 > 711 > 673; БелГУ: 181 > 72 > 47; ГГУ: 149 > 116 > 80.

Для университетов ПБРУУК можно построить дисциплинарно-временные таблицы публикационной активности на ежегодной основе, которые очень удобны в качестве инструмента университетского вебометрического бенчмаркинга. В целях проведения такого бенчмаркинга их целесообразно составлять в виде электронных гиперссылочных таблиц.

Следует отметить, что поисковая машина Google Scholar хорошо индексирует статьи из

отечественных ОА-репозитариев и ОА-журналов (например, журнал «Исследовано в России»), что может быть интересно для всех ученых, формирующих свои публикационные стратегии.

Предложенный в статье методологический инструментальный может быть полезен для проведения вебометрического университетского бенчмаркинга (мониторинга, аудита) на всех уровнях – международном, национальном, региональном, корпоративном (ассоциации университетов), внутриуниверситетском.

1. *Московкин В. М., Бадер Э. А.* Университетские и научно-исследовательские системы стран MEDA // Университетское управление: практика и анализ. 2008. № 3. С. 50–56.

2. *Bar-Ilan J.* Which h-index? A Comparison of WoS, Scopus and Google Scholar // Scientometrics. 2008. Vol. 74, № 2. P. 257–271.

3. *Brody T., Carr L, Gingras Y. et al.* Incentivizing the Open Access Research Web Publication – Archiving, Data-Archiving and Scientometrics // CTWatch Quarterly. 2007/ Aug.

4. *Kousha K., Thelwall M.* Google Scholar citations and Google Web/URL citations: A multi-discipline exploratory analysis // Journal of the American Society for Information Science and Technology. 2007. Vol. 58, № 7. P. 1055–1065.

5. *Kousha K., Thelwall M.* Sources of Google Scholar citation outside the Science Citation Index: A comparison between four science disciplines // Scientometrics. 2008. Vol. 74, № 2. P. 273–294.

6. *Noruzi A.* Google Scholar: The New Generation of Citation Indexes // Libri. 2005. Vol. 55, № 4. P. 170–180.