

**Моделирование комплектования научной периодики
на основе объективной тематики публикаций**

**Modeling scientific periodicals collection development
based on publication intrinsic subject scope**

**Моделювання комплектування наукової періодики
на основі об'єктивної тематики публікацій**

В. Н. Гуреев

*Государственный научный центр вирусологии и биотехнологии «Вектор»,
Кольцово Новосибирской обл., Россия*

Н. А. Мазов

*ИБЦ Института нефтегазовой геологии и геофизики
им. академика А. А. Трофимука СО РАН, Новосибирск, Россия*

Vadim Gureev

*VECTOR State Scientific Center for Virology and Biotechnology,
Koltsovo, Novosibirsk Region, Russia*

Nikolay Mazov

*A. Trofimuk Institute for oil and Gas Geology and Geophysics,
Siberian Division of the Russian Academy of Sciences,
Novosibirsk, Russia*

В. М. Гуреев

*Державний науковий центр вірусології та біотехнології «Вектор»,
Кольцово Новосибірської обл., Росія*

М. А. Мазов

*ДБЦ Інституту нафтегазової геології та геофізики
ім. академіка А. О. Трофімука ВО РАН, Новосибірськ, Росія*

Авторы настоящей работы предлагают к тестированию метод оценки научных журналов, основанный на результатах запроса из ключевых слов и индикаторов предметных категорий, которые присваиваются статьям организации в различных базах данных. Используя инструментарий Web of Science (WoS), авторам удалось в автоматическом режиме извлечь и обработать ключевые слова KeyWords Plus и индикаторы WoS Categories из статей сотрудников двух научных институтов различной направленности. На основе полученного списка был сформирован запрос в базу данных (БД) WoS и получен результат по всем источникам с теми же ключевыми словами и предметными категориями. Последующий анализ позволил создать ранжированные списки научных журналов, которые можно использовать для редактирования подписного репертуара научной библиотеки.

Ключевые слова: научная периодика, ключевые слова, KeyWords Plus, Web of Science, Scopus, комплектование, репертуар подписки.

The authors make a proposal for testing an assessment method for scientific journals based on the results of keyword, subject indicators assigned to the articles in various databases. Using Web of Science (WoS) instruments, the authors succeeded in producing and processing KeyWords Plus and WoS Categories out of the papers written by the researchers in their two research institutes. The WoS database query was then formulated, and the result for every source was obtained for the same keywords and subject categories. Further analysis enables to rank scientific journals in the lists to be used to update subscription repertoire of scientific libraries.

Keywords: scientific periodicals, keywords, KeyWords Plus, Web of Science, Scopus, collection development, subscription repertoire.

Автори цієї роботи пропонують для тестування метод оцінки наукових журналів, який базується на результатах запиту із ключових слів та індикаторів предметних категорій, які присвоюються статтям організації в різних базах даних. Використовуючи інструментарій Web of Science (WoS), авторам вдалося в автоматичному режимі вибрати та опрацювати ключові слова KeyWords Plus та індикато-

ри WoS Categories із статей співробітників двох наукових інститутів різного спрямування. На основі отриманого списку було сформовано запит до бази даних (БД) WoS та отримано результат із усіх джерел з такими ж ключовими словами та предметними категоріями. Подальший аналіз дозволив створити ранжовані списки наукових журналів, які можна використовувати для редагування передплатного репертуару наукової бібліотеки.

Ключові слова: наукова періодика, ключові слова, KeyWords Plus, Web of Science, Scopus, комплектування, репертуар передплати.

Время компьютерных технологий продолжает вносить свои коррективы в различные сферы библиотечной деятельности. Растущий объем научных публикаций, увеличение количества электронных журналов требуют новых методов и принципов отбора научных журналов. Особая актуальность разработки современных подходов обусловлена изменением форм чтения, когда читатель не просматривает оглавления журнала целиком, как было прежде, но пользуется реферативными базами данных или переходит к новому тексту по гиперссылкам. Как результат, современного читателя в меньшей степени интересует, в каком именно журнале опубликована статья, – на первое место выходит сам текст. Задачей специалиста-комплектатора, таким образом, выступает создание такого оптимального набора журнальных названий, выписываемых научной библиотекой, которые содержали бы максимальное количество текстов, потенциально интересных читателям.

Ранее мы уже обращались к теме формирования журнального фонда научной библиотеки с использованием библиометрического анализа. Были проанализированы все ссылки за последние пять лет из работ научных сотрудников двух институтов. Анализ цитируемых статей позволил нам в полуавтоматическом режиме создать рейтинговые списки журналов, которые мы предлагали к использованию при комплектовании библиотеки научной периодикой. Подробные результаты и методику проведения анализа можно найти в работах [1, 2, 3].

Между тем анализ пристатейных списков литературы, хотя он и выигрывает в сравнении с экспертной оценкой, является не вполне объективным методом, поскольку субъективно само цитирование. Автор работы [4], рассматривая мотивации цитирующего автора, выделил 15 функций ссылок, среди которых, например, отрицательные отзывы о научной работе, дань уважения первооткрывателям и ссылки на родственные работы знакомых коллег. Журналы, выявленные на основе таких ссылок, будут давать погрешность в общем массиве полученных журналов, не отвечая реальным информационным потребностям ученых, а измерить эту погрешность достаточно трудно!

При разработке нового, более объективного, на наш взгляд, метода оценки научных журналов мы руководствовались тем, что современный метод должен быть максимально автоматизированным, объективным и простым в исполнении. Ручная обработка данных уже не позволяет библиотечным работникам отследить в обширном многообразии названий научных журналов новые, но ценные издания. Например, недавно организованное американское издательство Scientific and Academic Publishing [5] за три последних года открыло более 120 реферируемых журналов по 16 тематикам, а созданный пять лет назад журнал Microbial Biotechnology (издательство Wiley) уже индексируется в Web of Science и набрал относительно высокий импакт-фактор 2,5.

Предлагаемый нами метод отвечает означенным требованиям и позволяет получить ранжированные списки научных журналов, способные помочь специалисту-комплектатору оптимизировать фонд научной периодики. Ядром метода выступает использование в качестве поискового запроса ключевых слов из статей сотрудников научной организации. Насколько нам известно, это первый опыт формирования подписки по такому принципу. Мы исходим из того, что публикации научного института наиболее широко и достоверно отражают сферу научных интересов авторов и реальные достижения организации. Дополнительный список ключевых слов, присвоенный каждой публикации системой (**KeyWords Plus** – в Web of Science или **Indexed keywords** – в Scopus), позволяет сформировать сложный запрос, результаты которого являются прямым отражением в базе данных тех научных интересов, которые авторы конкретной организации выразили своими статьями. Особо обращаем внимание на то, что для достижения большей объективности мы выбрали для обработки именно машинные ключевые слова, а не авторские. Их присваивает система, разработанная экспертами, они присутствуют во всех статьях, тогда как авторские ключевые слова в ряде

случаев могут отсутствовать. Подробнее об алгоритмах присвоения KeyWords Plus статьям можно узнать в оригинальной работе Ю. Гарфилда [6], а также в статье Qin J. [7].

Были проанализированы статьи сотрудников ГНЦ ВБ «Вектор» и ИНГГ СО РАН за период 2008–2012 гг., отраженные в БД Web of Science. Выбор этой базы данных был произвольным, и все описанные ниже алгоритмы в равной мере применимы к базе данных Scopus, с соответствующими поправками на формирование запроса. Всего нами было рассмотрено 190 публикаций «Вектора» и 408 – ИНГГ СО РАН. Экспорт статей производился в формате WinTab, впоследствии данные обрабатывались в редакторах Microsoft Excel и Word. Ключевые слова KeyWords Plus каждой из статей были объединены в группы оператором AND, затем к этим группам посредством оператора AND были добавлены индикаторы WoS Categories из тех же соответствующих статей, а сами группы связаны оператором OR. Таким образом, в запросе участвовало столько же групп, сколько было и статей. В итоге получился массивный запрос, содержащий более чем из 3000 терминов. Схематично структура запроса показан на рис. 1.

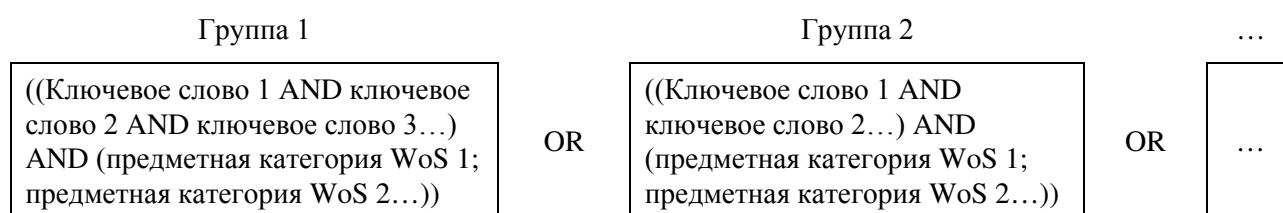


Рис. 1. Схематичная структура запроса для расширенного поиска в Web of Science

Результатом расширенного поиска стали 68 тысяч статей по биологии и 47 тысяч – по геологии без учета не интересующих нас материалов конференций, книг, статей из области общественных наук и всех языков кроме английского и русского. Временной интервал нашей подписки WoS – с 1980 года. Затем мы воспользовались опцией анализа результатов по параметру Source Titles и получили ранжированный список из 800 названий в области биологии и 700 – в геологических дисциплинах, в котором явно прослеживается ядро. Ведущие журналы, охватывающие 30 % всех статей с интересующими нас ключевыми словами, показаны в табл. 1.

Таблица 1

Ведущие журналы, включающие 30 % всех статей с заданными KeyWords Plus

Журналы биологического профиля	%	Всего %	Журналы геологического профиля	%	Всего %
1. Aids	6,007	6,007	1. Geophysical Research Letters	6,85	6,85
2. Vaccine	5,416	11,423	2. Precambrian Research	2,989	9,839
3. Aids Research and Human Retroviruses	3,852	15,275	3. Palaeogeography Palaeoclimatology Palaeoecology	2,951	12,79
4. Journal of Infectious Diseases	3,51	18,785	4. Journal of Volcanology and Geothermal Research	2,641	15,431
5. Applied and Environmental Microbiology	2,859	21,644	5. Geological Society of America Bulletin	2,158	17,589
6. Clinical Infectious Diseases	2,5	24,444	6. Canadian Journal of Earth Sciences	2,152	19,741
7. Genomics	2,456	26,6	7. Journal of Hydrology	2,051	21,792
8. Journal of General Virology	2,047	28,647	8. Advances in Space Research	1,956	23,748
9. Gene Therapy	1,98	30,627	9. Journal of Structural Geology	1,935	25,683
			10. Annales Geophysicae	1,863	27,546
			11. Quaternary Science Reviews	1,773	29,319
			12. Journal of the Geological Society	1,669	30,988

Данный список может помочь комплектатором при формировании подписки научной периодики. В то же время мы имеем возможность сопоставить эти результаты с данными, полученными на основе анализа ссылок из тех же статей сотрудников наших НИИ. В табл. 2 показаны наиболее активно цитируемые журналы, выявленные на основе анализа списков пристатейной литературы. При анализе авторы использовали программное обеспечение SciMAT [8].

Таблица 2

**Ведущие журналы, включающие 30% всех статей, процитированных авторами наших НИИ.
Жирным помечены журналы, общие с данными предыдущей таблицы**

Журналы биологического профиля	%	Журналы геологического профиля	%
1. J VIROL	4,12	1. RUSS GEOL GEOPHYS	6,41
2. P NATL ACAD SCI USA	2,8	2. DOKL EARTH SCI	2,56
3. J BIOL CHEM	2,51	3. J GEOPHYS RES-SOL EA	1,69
4. NATURE	2	4. EARTH PLANET SC LETT	1,68
5. SCIENCE	1,96	5. TECTONOPHYSICS	1,55
6. VIROLOGY	1,59	6. GEOPHYS J INT	1,45
7. J GEN VIROL	1,45	7. GEOPHYSICS	1,37
8. J MOL BIOL	1,43	8. GEOLOGY	1,18
9. VACCINE	1,4	9. PRECAMBRIAN RES	1,13
10. NUCLEIC ACIDS RES	1,25	10. NATURE	1
11. J INFECT DIS	0,82	11. GEOCHIM COSMOCHIM AC	0,99
12. EMERG INFECT DIS	0,79	12. STRATIGR GEOL CORREL	0,92
13. APPL ENVIRON MICROB	0,78	13. GEOPHYS RES LETT	0,84
14. CELL	0,73	14. J GEOPHYS RES	0,71
15. J IMMUNOL	0,7	15. SCIENCE	0,64
16. J MED VIROL	0,66	16. PALAEOGEOGR PALAEOCL	0,58
17. CANCER RES	0,66	17. GEOKHIMIYA	0,5
18. VOP VIRUSOL+	0,63	18. CHEM GEOL	0,49
19. J CLIN MICROBIOL	0,58	19. CONTRIB MINERAL PETR	0,48
20. LANCET	0,58	20. PHYS EARTH PLANET IN	0,47
21. NAT MED	0,58	21. ORG GEOCHEM	0,46
22. VIRUS RES	0,52	22. GEOL SOC AM BULL	0,4
23. NEW ENGL J MED	0,51	23. GEOTEKTONIKA	0,37
24. BIOCHEMISTRY-US	0,49	24. J PETROL	0,35
25. ATMOS ENVIRON	0,47	25. J VOLCANOL GEOTH RES	0,34
26. ARCH VIROL	30,01	26. LITHOS	0,33
		27. J GEOL SOC LONDON	0,32
		28. TECTONICS	0,31
		29. APPL GEOCHEM	0,29

Применив два различных метода для получения ранжированных списков научных журналов, мы получили данные, совпадающие лишь частично. Анализ ключевых слов не выявил мультидисциплинарных журналов, таких как Nature и Science, которые активно цитировались сотрудниками обоих институтов и, безусловно, должны выписываться каждой научной библиотекой. Запрос из ключевых слов KeyWords Plus не выявил и более узких междисциплинарных журналов, каковыми является, например, Journal of Virology в биологии и Russian Geology and Geophysics – в области наук о Земле. Очень малая (менее 1 % от общего количества журналов) представленность рус-

скоязычной литературы в Web of Science не позволила выявить высокоцитируемые российские журналы. Из этого мы делаем вывод, что в его настоящем виде метод построения ранжированных списков журналов на основе анализа ключевых слов наилучшим образом подходит для выявления именно предметного ядра периодики, а также позволяет выявить наиболее активные направления научной деятельности внутри института. Из данных табл. 1 видно, что в области биологии на «Векторе» лидируют исследования по ретровирусам и СПИДу, а также вакцинология. Для таких наиболее сильных научных групп могут оказаться полезными новые журналы в подписном репертуаре научной библиотеки.

Предложенный нами метод полностью автоматизированный, достаточно прост в исполнении и не требует больших временных затрат от работников библиотек. Имеется возможность работы как в базе данных Web of Science (KeyWords Plus), так и в Scopus (Indexed keywords). Хотя метод и нуждается в доработке, уже сейчас он позволяет получить такой список научных журналов, который может оказаться весьма ценным при комплектовании фонда научной периодики. Остаются требующими решения вопросы работы с мультидисциплинарными журналами и журналами на других языках, кроме английского. Авторы ведут активную работу по усовершенствованию методологии построения ранжированных списков научных журналов, в частности – по объединению двух методов на основе ключевых слов и пристатейного списка литературы. Впоследствии это позволит выработать более достоверные и доступные принципы отбора журналов в фонды научных библиотек, а также решить все обозначенные в настоящей работе проблемные моменты.

Литература:

1. Мазов Н.А. Оценка потока научных публикаций академического института на основе библиометрического анализа цитирования // Информационные технологии в гуманитарных исследованиях. – 2011. – № 16. – С. 25–30.
2. Гуреев В.Н., Мазов Н.А. Изучение информационных потребностей ученых с использованием библиометрического анализа для оптимизации комплектования // Библиосфера. – 2012. – № 4. – С. 57–66.
3. Гуреев В.Н., Мазов Н.А. Практическое применение библиометрического анализа при формировании журнального фонда // Информационные технологии в гуманитарных исследованиях. – 2012. – № 17. – С. 81–87.
4. Weinstock M. Citation indexes. Essays of an Information Scientist, 1. – Philadelphia: ISI Press, 1977. – P. 188–216.
5. Scientific and Academic Publishing. <http://www.sapub.org/>
6. Garfield E. KeyWords Plus – ISI's breakthrough retrieval method. 1. Expanding your searching power on current-contents on diskette // Current Contents. – 1990. – V. 32. – P. 295–299.
7. Qun J. Semantic similarities between a keyword database and a controlled vocabulary database: An investigation in the antibiotic resistance literature // Journal of the American Society for Information Science. – 2000. – V. 51(2) – P. 166–180.
8. Cobo M.J., López-Herrera A.G., Herrera-Viedma E., and Herrera F. SciMAT: A New Science Mapping Analysis Software Tool // Journal of the American Society for Information Science and Technology. – 2012. – V. 63(8). – P. 1609–1630.