

¹ Государственный научный центр вирусологии
и биотехнологии «Вектор»
р. п. Кольцово, 630559, Новосибирская обл., Россия

² Институт нефтегазовой геологии и геофизики
им. акад. А. А. Трофимука СО РАН
пр. Акад. Коптюга, 3, Новосибирск, 630090, Россия

E-mail: gureyev@vector.nsc.ru; MazovNA@ipgg.sbras.ru

МОДЕЛИРОВАНИЕ ОПТИМАЛЬНОГО РЕПЕРТУАРА НАУЧНОЙ ПЕРИОДИКИ ОРГАНИЗАЦИИ С ИСПОЛЬЗОВАНИЕМ БИБЛИОМЕТРИЧЕСКИХ МЕТОДОВ

Динамичное развитие современного рынка научной периодики ставит перед всеми его участниками – издателями, библиотекарями и учеными – ряд новых нетривиальных задач. Практически завершившийся переход журналов в электронную форму изменил парадигму получения информации, способствовал открытию сотен новых изданий, а библиографические базы данных уравнивали публикации в плане их поиска и доступности. В таких условиях первостепенное значение приобретает выявление ядра журналов, необходимых ученому / научной группе / организации, причем обязательным условием выступает автоматизация процессов поиска. На примере собственных разработок получения ранжированных списков научных журналов на основе ссылок, ключевых слов и предметных рубрик мы показываем потенциальные возможности современного библиометрического инструментария, доступного для использования каждому заинтересованному человеку.

Ключевые слова: Web of Science, Scopus, библиометрический анализ, анализ цитирования, научная периодика.

Введение

С конца XX в. в жизни современного общества четко наметились две, на первый взгляд, взаимоисключающие тенденции. Во-первых, это глобализация, проникающая во все сферы жизнедеятельности человека, которая проявляется в унификации разных культур, вкусов, стремлений, вымирании части языков и завоевании английским языком лидирующих позиций, например, в бизнесе и науке. Во-вторых, это дифференциация общества, которая находит проявление в создании различных групп по интересам. Эти группы, как правило, наднациональны, не имеют языковых и географических границ; участники объединены исключительно проявленным интересом к той или иной сфере жизни или области знаний. Таким образом, наблюдается перестройка общества: частично теряя традиционные устоявшиеся ценности, члены современного общества, пользуясь информационно-коммуникационными технологиями, создают новые.

Эти тенденции отчетливо прослеживаются и в науке: исследования все чаще становятся плодом коллективного труда, в проведение экспериментов вовлекаются ученые разных стран, проводятся междисциплинарные работы. Одновременно происходит дальнейшее дробление дисциплин на более узкие, специальные, что подтверждается постоянным введением новых рубрик в классификаторы научной информации.

Гуреев В. Н., Мазов Н. А. Моделирование идеального репертуара научной периодики организации с использованием библиометрических методов // Вестн. Новосиб. гос. ун-та. Серия: Информационные технологии. 2013. Т. 11, вып. 4. С. 38–45.

Обе тенденции получили свое отражение в издательском деле. С одной стороны, появляется большое количество междисциплинарных журналов, как общенаучных, так и охватывающих отдельные дисциплины, а с другой – узкоспециальные, отвечающие требованиям небольших групп специалистов.

Ситуация усложняется изменившейся за последнее время парадигмой чтения научных статей, новой моделью получения информации. Читатели все реже ходят в библиотеку, не просматривают оглавления журнала целиком, как было прежде, а работают с реферативными базами данных или переходят к новому тексту по гиперссылкам. Таким образом, на первое место выходит непосредственно текст, а в каком именно журнале он опубликован, читателю зачастую уже не столь важно. Безусловно, продолжают существовать классические научные журналы с многолетней и богатой историей. Однако мы находимся лишь в начале процесса изменения публикационной модели, и вполне возможно, что само понятие научного журнала с утратой физического носителя в скором времени изменит свой смысл.

Уже сейчас мы можем наблюдать, как издательства пытаются найти новые подходы, которые отвечали бы требованиям современных пользователей. Публикуются так называемые статьи «в печати», которые своим существованием вступают в противоречие с главным журнальным атрибутом – периодичностью. Повсеместно внедряемое цитирование по идентификатору цифрового объекта (DOI) также заменяет собой традиционные журнальные элементы – том, номер и страницы. Сравнительно новой подписной моделью становится подписка организации на заданный набор статей (Article choice), безотносительно журналов. Особо следует упомянуть об открытии множества сайтов с препринтами, которые противопоставляют свою модель журнальной и решают проблему временного лага, длящегося от срока подачи статьи до ее опубликования.

Методы отбора журналов для информационного обеспечения исследований

В описанных условиях перед потребителями научной информации встает трудная задача – из тысяч названий журналов выделить необходимое ядро, отвечающее их информационным потребностям. Отработанные десятилетиями методы ручной обработки данных уже не позволяют отслеживать в обширном многообразии названий научных журналов новые, но ценные издания.

Время поставило дополнительные условия перед читателями, но оно же предоставило уникальные инструменты обработки и анализа данных, которые необходимо найти и задействовать самим пользователям.

Основные инструменты, доступные для использования, можно распределить по четырем большим группам: формальные признаки, экспертная оценка, анализ используемости журналов и библиометрический анализ.

Формальные признаки. Группа формальных признаков представляет собой первичный, достаточно грубый фильтр, который нельзя использовать в отрыве от других, более тонких методов. Эта группа включает в себя такие общие категории, как

- соответствие тематике исследований организации;
- тип журнала;
- стоимость журнала;
- научная ценность;
- наличие или отсутствие реферирования;
- авторитет издающей организации;
- представленность журнала в индексирующих и реферативных службах;
- наличие у издательства статуса национального научного общества;
- распространенность издания;
- язык статей;
- наличие и язык рефератов;
- наличие ключевых слов;
- полнота пристатейной библиографии.

Сейчас к этим традиционным категориям добавились новые элементы, среди которых DOI, «резюме второго порядка» Highlights, отдельно индексируемые подписи к таблицам и рисункам и проч. Все эти элементы упрощают навигацию в информационной среде научных статей и могут использоваться при оценке изданий. Так или иначе, метод формальных признаков присутствует в каждом из прочих методов и является первоосновой для последующих принципов отбора.

Экспертная оценка, или анкетирование, представляет собой наиболее консервативный критерий ранжирования научных журналов. Данный метод основан на допущении, что ни библиотекари, ни компьютерные программы не могут выявить потребностей ученых лучше них самих. Однако в современной ситуации экспертную оценку можно с определенной долей уверенности считать устаревшим методом, поскольку накопленные им недостатки вряд ли найдут свое разрешение:

- учитывая рост количества научных журналов, которые нередко быстро набирают рейтинг, специалисты не могут охватить все разнообразие периодики и тем более дать ей оценку;
- случается, что отмеченные экспертами журналы впоследствии пользуются небольшим спросом или вообще остаются невостребованными;
- в методе экспертной оценки не предусмотрено ответственности за возможную неэффективность подписки, поэтому специалисты, скорее всего, не станут вдумчиво и всесторонне оценивать предлагаемые им варианты.

Анализ спроса литературы. Что касается данных об использовании литературы, то, хотя из них и можно извлечь некоторую полезную информацию, они не вполне достоверны. Отметим, что даже современные инструменты отслеживания числа загрузок не дают точных показателей по следующим причинам:

- ряд издательств не предоставляет прав на просмотр статистики использования ресурсов;
- нет возможности отследить количество загрузок статей из журналов открытого доступа;
- даже для тех ресурсов, которые предоставляют статистику использования, она не будет точной из-за возможности получить один и тот же текст из нескольких источников.

Библиометрический анализ в настоящее время является наиболее прогрессивным и многообещающим методом, который в отличие от описанных выше позволяет быстрее и объективнее выявлять уровень удовлетворенности ученых текущей подпиской и их информационные потребности.

Все чаще делается акцент на объективности библиометрического анализа даже при его сравнении с научным рецензированием, что подчеркивает его возросшую значимость [1]. Базы данных, предоставляющие необходимую для анализа информацию, доступны в большинстве научных организаций, причем зарегистрированные пользователи получают к ним удаленный доступ. Компании «Томсон» и «Эльзевир» популяризируют свои продукты Web of Science и Scopus, проводят вебинары и тем самым расширяют круг технически грамотных пользователей.

Библиометрический анализ основан на различных исходных данных и производится специальными программами по заданным алгоритмам. Такими данными могут выступать импакт-фактор журнала, нормированный импакт-фактор SNIP, Эйгенфактор, спрос на издания, индекс Хирша и проч. Используются математические формулы и методы статистического анализа.

Разработано множество подходов к этому виду анализа. Далее на примере двух методов, посвященных моделированию журнального фонда, мы продемонстрируем уникальные возможности современных инструментов, доступных в каждой организации.

Анализ цитирования как основа определения информационных потребностей ученых

Информационные потребности пользователей и тенденции их развития, которые можно применять при моделировании репертуара периодики, прекрасно выявляются незаслуженно

редко используемым методом цитат-анализа, который мы применили для анализа пристатейной литературы на примере публикаций авторов из наших организаций [2–4]. Тематические профили ГНЦ ВБ «Вектор» и ИНГТ СО РАН – медико-биологические исследования и геофизика – занимают значительное место в журнальной периодике, а обслуживающие эти отрасли журналы обладают наибольшими импакт-факторами, наблюдается их количественный рост и конкуренция между ними. Поэтому наши методы и результаты вполне могут использоваться в иных организациях естественно-научного профиля.

Для анализа мы рассмотрели все статьи наших сотрудников за последние пять лет. Из пристатейных списков литературы были выгружены 1 300 и 1 000 наименований научных журналов в области биологии и геологии соответственно. Доля отечественных изданий составила четверть. В совокупности ученые каждого из институтов сделали порядка 8 000 ссылок. По каждой из дисциплин мы получили списки журналов, которые затем разбили на три равные части. Ядро цитируемых российских журналов показано в табл. 1.

Выяснилось, что лишь 9 % от общего числа цитируемых российских журналов медико-биологического профиля обеспечивают $\frac{2}{3}$ всех ссылок и всего на 6 % журналов приходится 80 % всех цитирований в области геологии и геофизики. Данные двух институтов по цитируемости хорошо согласовывались друг с другом, с той лишь разницей, что в статьях в области геологии больше цитировались отечественные журналы, тогда как в области биологических наук предпочтение отдавалось зарубежной литературе.

Полученные данные по цитируемости зарубежных журналов приводятся в табл. 2.

Таблица 1

Ядро отечественных журналов, выявленное на основе цитирования

Название журнала	Количество ссылок, %	Всего, %
Журналы биологического профиля		
Вопросы вирусологии	16,94	16,94
Молекулярная биология	6,44	23,38
Биотехнология	6,07	29,45
Доклады Академии наук	5,09	34,54
Журналы геологического профиля		
Геология и геофизика	35	35
Доклады РАН	19	54
Геохимия	7	61
Петрология	4	65

Таблица 2

Ядро зарубежных журналов, выявленное на основе цитирования

Название журнала	Количество ссылок, %	Всего, %
Журналы биологического профиля		
Journal of Virology	6,03	6,03
Proceedings of the National Academy of Sciences of the United States of America	3,36	9,39
Virology	3,15	12,54
Vaccine	2,77	15,31
Journal of Biological Chemistry	2,4	17,71
Journal of General Virology	2,4	20,11
Nature	2,04	22,15
Science	1,94	24,09
Journal of Clinical Microbiology	1,94	26,03
Emerging Infectious Diseases	1,89	27,92

Окончание табл. 2

Название журнала	Количество ссылок, %	Всего, %
Nucleic Acids Research	1,59	29,51
Journal of Infectious Diseases	1,38	30,89
Journal of Molecular Biology	1,35	32,24
Journal of Immunology	1,24	33,48
Journal of Medical Virology	1,19	34,67
Журналы геологического профиля		
Earth Planetary Science Letters	7	7
Contributions to Mineralogy and Petrology	6	13
Geochimica et Cosmochimica Acta	5	18
Journal of Geophysical Research	5	23
Nature	4	27
American Mineralogist	3	30
Lithos	3,5	33,5
Journal of Petrology	3,5	37
Chemical Geology	2,5	39,5
Geology	2,5	42
Tectonophysics	2	44
Science	1,5	45,5
Economic Geology	1,5	47
European Journal of Mineralogy	1,5	48,5
Canadian Mineralogist	1,5	50

Ссылки на иностранные журналы показывают распределение, схожее с отечественными: бóльшая концентрация ссылок представлена в журналах геологического профиля, притом что на 25 журналов приходится $\frac{2}{3}$ всех ссылок. В области же биомедицины $\frac{2}{3}$ всех ссылок представлены в 100 журналах. Лишь 1,3 % процитированных журналов обеспечивает треть всех ссылок, сделанных из статей медико-биологического профиля; для наук о Земле этот показатель составляет 0,9 %.

Данные по цитируемости зарубежных журналов распределяются более гладко в сравнении с отечественной периодикой, где счет ядерных журналов идет на единицы. Это можно связать с меньшим количеством российских журналов в целом. С помощью полученных на основе цитирования данных мы быстро выявили ядро журнального фонда, в первую очередь необходимого для информационного обеспечения научных исследований. В области биологии и медицины в первые две зоны (66 % всех ссылок) вошли 16 отечественных и 100 зарубежных журналов; а в области наук о Земле – 4 российских и 25 иностранных. Примечательно, что в области геологических наук ссылки как на отечественную, так и на зарубежную периодику приходятся на количество журналов, в четыре раза меньшее, чем в области биологии.

Тематический анализ публикаций организации как основа формирования репертуара научной периодики

Основой второго метода выступает использование в качестве поискового запроса ключевых слов из статей сотрудников научной организации. По нашим данным, это первый опыт составления ранжированных списков журналов 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 (в нашем случае – KeyWords Plus [5]), который присваивается каждой публика-

ции в реферативной базе данных, позволяет сформировать сложный запрос, адресуемый той же базе данных. Список журналов, полученных на основе такого запроса, составленного из машинных ключевых слов, будет напрямую отражать те научные интересы, которые авторы конкретной организации выразили в своих статьях.

Мы отдали предпочтение машинным, а не авторским ключевым словам, для достижения большей объективности. Во-первых, их присваивает система, разработанная экспертами, во-вторых, в отличие от авторских ключевых слов машинные ключевые слова присутствуют во всех статьях.

Используя данный метод, как и в предыдущем случае, мы проанализировали статьи сотрудников двух научных институтов за последние 5 лет. Из статей организаций, проиндексированных в базе данных Web of Science, выбор которой был произвольным, были выгружены машинные ключевые слова, которые впоследствии были объединены в группы оператором AND. Затем для конкретизации результатов запроса к этим группам с помощью оператора AND были добавлены индикаторы Web of Science 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 более чем из 3 000 терминов, структура которого описана ниже.

Пусть $C = \{D_1, \dots, D_n\}$ – множество документов организации, отраженных в базе данных. Запросный профиль для поиска документов, релевантных множеству C , должен выглядеть следующим образом:

$$Q = P(D_1) + P(D_2) + \dots + P(D_n),$$

где «+» – это операция логического «или» / «OR».

Каждый документ D_i , в свою очередь, может быть представлен терминами KeyWords Plus (KWP) и WoS Categories (WC) следующим образом:

$$D_i = \{KWP_1^i, \dots, KWP_{m_i}^i\} \& \{WC_1^i, \dots, WC_{l_i}^i\}.$$

Запросный профиль для документа D_i имеет вид

$$P(D_i) = (KWP_1^i \& \dots \& KWP_{m_i}^i) \& (WC_1^i \& \dots \& WC_{l_i}^i),$$

где «&» – это операция логического «и» / «AND».

Итак, полный запросный профиль выглядит следующим образом:

$$Q = (KWP_1^1 \& \dots \& KWP_{m_1}^1) \& (WC_1^1 \& \dots \& WC_{l_1}^1) \text{ OR } \dots \text{ OR } (KWP_1^n \& \dots \& KWP_{m_n}^n) \& (WC_1^n \& \dots \& WC_{l_n}^n).$$

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. Затем мы воспользовались опцией анализа результатов по параметру «Источники публикаций» и получили ранжированные списки из 800 названий в области биологии и 700 – в геологических дисциплинах, в которых явно прослеживалось ядро. Данные по двум институтам, как и в предыдущем анализе, хорошо коррелировали друг с другом. Ведущие журналы, охватывающие 30 % всех статей с интересующими нас ключевыми словами, показаны в табл. 3.

Таблица 3

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

Название журнала	Количество ссылок, %	Всего, %
Журналы биологического профиля		
Aids	6,00	6,00
Vaccine	5,41	11,42
Aids Research and Human Retroviruses	3,85	15,27
Journal of Infectious Diseases	3,51	18,78

Окончание табл. 3

Название журнала	Количество ссылок, %	Всего, %
Applied and Environmental Microbiology	2,85	21,64
Clinical Infectious Diseases	2,5	24,44
Genomics	2,45	26,6
Journal of General Virology	2,04	28,64
Gene Therapy	1,98	30,62
Журналы геологического профиля		
Geophysical Research Letters	6,85	6,85
Precambrian Research	2,98	9,83
Palaeogeography Palaeoclimatology Palaeoecology	2,95	12,79
Journal of Volcanology and Geothermal Research	2,64	15,43
Geological Society of America Bulletin	2,15	17,58
Canadian Journal of Earth Sciences	2,15	19,74
Journal of Hydrology	2,05	21,79
Advances in Space Research	1,95	23,74
Journal of Structural Geology	1,93	25,68
Annales Geophysicae	1,86	27,54
Quaternary Science Reviews	1,77	29,31
Journal of the Geological Society	1,66	30,98

Применив два различных метода получения ранжированных списков научных журналов, мы обнаружили, что данные, полученные разными способами, совпадают лишь частично, хотя в рамках каждого из методов данные по разным дисциплинам хорошо коррелируют друг с другом. Анализ на основе запроса из ключевых слов не выявил мультидисциплинарных журналов, таких как Nature и Science, которые в то же время активно цитировались сотрудниками обоих институтов. Небольшая представленность русскоязычной литературы в Web of Science не позволила выявить высокоцитируемые российские журналы. Разница в данных также может объясняться частой недоступностью ряда издательств на территории России: и для чтения, и, как следствие, для цитирования. Особое внимание мы обратили на осторожное цитирование в России журналов открытого доступа.

В настоящем виде метод построения ранжированных списков журналов на основе анализа ключевых слов наилучшим образом подходит для выявления именно предметного ядра периодики, а также позволяет определить наиболее активные направления научной деятельности внутри института.

Заключение

Предложенные нами методы полностью автоматизированы, достаточно просты в исполнении и не требуют больших временных затрат. На выбор имеется возможность работы в Web of Science или Scopus – двух крупнейших инструментах для научно- и библиометрических изысканий. Свободно распространяется множество профессиональных программ для проведения подобных исследований, как, например, SciMAT [6], который авторы использовали в своей работе. Описанные методы позволяют получить такие списки научных журналов, которые могут оказаться весьма ценными, например, при комплектовании фонда периодики или выборе журнала для опубликования статей.

В последнее время активно проводятся библиометрические работы и ведется поиск новых областей применения библиометрии к различным практическим случаям. На примере предложенных нами методов видно, какие широкие возможности предоставляют научным сотрудникам и работникам библиотек современные коммерческие инструменты, внедряемые в реферативные базы данных. Компьютеризация библиотечных и издательских процессов

предъявила новые требования к аналитической обработке информации, но она же и предложила новые способы решения этих трудных задач, многие из которых еще предстоит найти.

Список литературы

1. Raan A. F. J. van. The Use of Bibliometric Analysis in Research Performance Assessment and Monitoring of Interdisciplinary Scientific Developments // *Technikfolgenabschätzung – Theorie und Praxis*. 2003. Vol. 1 (12). P. 20–29.
2. Мазов Н. А. Оценка потока научных публикаций академического института на основе библиометрического анализа цитирования // *Информационные технологии в гуманитарных исследованиях*. 2011. № 16. С. 25–30.
3. Гуреев В. Н., Мазов Н. А. Изучение информационных потребностей ученых с использованием библиометрического анализа для оптимизации комплектования // *Библиосфера*. 2012. № 4. С. 57–66.
4. Гуреев В. Н., Мазов Н. А. Практическое применение библиометрического анализа при формировании журнального фонда // *Информационные технологии в гуманитарных исследованиях*. 2012. № 17. С. 81–87.
5. Garfield E. KeyWords Plus – ISI's Breakthrough Retrieval Method. 1. Expanding your Searching Power on Current-Contents on Diskette // *Current Contents*. 1990. Vol. 32. P. 295–299.
6. Cobo M. J., López-Herrera A. G., Herrera-Viedma E., Herrera F. SciMAT: A New Science Mapping Analysis Software Tool // *Journal of the American Society for Information Science and Technology*. 2012. Vol. 63 (8). P. 1609–1630.

Материал поступил в редколлегию 30.07.2013

V. N. Gureyev, N. A. Mazov

MODELING OF EFFICIENT SET OF ACADEMIC PERIODICALS IN RESEARCH INSTITUTE USING BIBLIOMETRIC APPROACHES

Dynamic development of contemporary market of academic periodicals offers new challenges for all participants, namely publishers, librarians, and researchers. Almost complete journal conversion into electronic format changed approaches of receiving information and provided foundation of hundreds of new journal titles. As for bibliographic databases, they equalized publications in terms of their search and availability. In the circumstances, detection of core journals that are necessary for researcher / science team / research institute is of high priority. Furthermore, automation of search processes is essential condition. On the base of our own approaches to obtain ranked lists of academic journals using references, keywords, and subject areas we demonstrate possible modern bibliometric tools available for usage for every interested person.

Keywords: Web of Science, Scopus, bibliometric analysis, citation analysis, academic periodical.

References

1. Raan A. F. J. van. The use of Bibliometric Analysis in Research Performance Assessment and Monitoring of Interdisciplinary Scientific Developments // *Technikfolgenabschätzung – Theorie und Praxis*. 2003. Vol. 1 (12). P. 20–29.
2. Mazov N. A. Estimation of a Flow of Scientific Publications of the Academic Institute on a Basis Bibliometric Citation Analysis // *Information Technologies in Social Researches*. 2011. Vol. 16. P. 25–30.
3. Gureyev V. N., Mazov N. A. Studying the Information Needs of Scientists Using Bibliometric Analysis for Acquisition Optimization // *Bibliosphere*. 2012. Vol. 4. P. 55–66.
4. Gureyev V. N., Mazov N. A. Practical Application of Bibliometric Analysis when Development of Periodical Holding // *Information Technologies in Social Researches*. 2012. Vol. 17. P. 81–87.
5. Garfield E. Keywords Plus – ISI's Breakthrough Retrieval Method. 1. Expanding your Searching Power on Current-Contents on Diskette // *Current Contents*. 1990. Vol. 32. P. 295–299.
6. Cobo M. J., López-Herrera A. G., Herrera-Viedma E., Herrera F. SciMAT: A New Science Mapping Analysis Software Tool // *Journal of the American Society for Information Science and Technology*. 2012. Vol. 63 (8). P. 1609–1630.