

**Библиометрический анализ как основа формирования
подписки на журналы для научной библиотеки**

В.Н. Гуреев

Институт нефтегазовой геологии и геофизики СО РАН, Новосибирск, Россия
Государственный научный центр вирусологии и биотехнологии «Вектор», Кольцово, Россия
GureyevVN@ipgg.sbras.ru

Н.А. Мазов

Институт нефтегазовой геологии и геофизики СО РАН, Новосибирск, Россия
MazovNA@ipgg.sbras.ru

Аннотация

Продолжающийся рост числа научных публикаций и журналов, появление новых направлений, возникающих на стыках различных дисциплин, постановка новых научных проблем, требующих информационного обеспечения, а также появление ряда недобросовестных издателей, выпускающих журналы открытого доступа, требуют разработки более точных моделей поиска информации, необходимой для решения последующих прикладных задач. В библиотечном деле одной из основных задач является отбор журналов в фонды научной периодики. В условиях, когда прежние методы оценки журналов утрачивают свою эффективность, особое внимание стоит уделять более объективным автоматизированным подходам, основывающимся на методах библиометрии.

Введение

Проблема выявления журнального ядра, необходимого к подписке в научных библиотеках, а также оценки уже подписанных изданий для выявления журналов – кандидатов на списание, не утрачивает своей актуальности. В частности, это связано со следующими обстоятельствами.

- В последние годы появляется большое количество междисциплинарных журналов, как общенаучных, так и охватывающих отдельные дисциплины, а также узкоспециальных, отвечающих требованиям небольших групп специалистов. Средние темпы роста числа научных журналов составляют 3,3 % в год [11, 12], причем большинство из них приходится на второстепенные по значимости издания.
- Не останавливается рост стоимости подписки на научные журналы. Следует отметить, что платные журналы пока продолжают занимать лидирующие позиции в

удовлетворении информационных потребностей ученых. Темпы увеличения стоимости подписки для журналов коммерческих издательств составляют от 6 до 8 % в год [9, 14], тогда как для журналов, издаваемых научными обществами – 5,5 % [10, 16]. Примечательно, что модели открытого доступа еще не оказывают существенного влияния на ценообразование в издательском деле.

- Трансформируется само понятие «научный журнал», которое постепенно утрачивает традиционные атрибуты (например, периодичность, нумерацию страниц) и в то же время приобретает новые (например, статьи в форме видеолекций, цитирование по doi). Изменения в форме представления научной информации требуют и новых подходов к ее поиску.
- Открываются и успешно функционируют недобросовестные издательства, выпускающие журналы открытого доступа с высокой оплатой статей авторами, но с низким уровнем рецензирования [7, 8].
- Происходит смена парадигмы чтения научных статей с опорой на поиск необходимых текстов по реферативным базам данных в удаленном режиме и на систему гиперссылок, что в итоге затрудняет выявление информационных потребностей ученых [15].
- Снижается эффективность основных используемых методов комплектования, в первую очередь экспертной оценки, поскольку эксперты не в состоянии оценить все многообразие современных журналов, а также статистики используемости журналов, которая в основном недоступна для журналов открытого доступа и дает искажение из-за возможности получить одну и ту же публикацию из разных источников.

Использование в отделах комплектования научных библиотек библиометрических методов во многом может способствовать решению обозначенных задач. Библиометрический анализ с появлением специализированных ресурсов и программ стал более доступным и менее трудоемким, что делает разработку методов на его основе оправданными и необходимыми.

Сбор библиографической информации о публикациях организации

Сбор информации об исходящих из организации публикациях – первый этап библиометрического анализа. Выбор в качестве фактологической базы данных публикаций сотрудников организации обусловлен тем, что каждая научная организация уникальна: в них работают сложившиеся и вновь образующиеся научные группы и школы, проводятся особенные научные исследования, разрабатываются собственные стратегии развития. В этих условиях составление репертуара периодики на основе более широкой базы данных (в литературе

предлагалось использовать ссылки из отдельных журналов, их совокупности, а также всего мирового потока публикаций) не будет учитывать существенные особенности различных организаций, хотя многие из них и проводят исследования в одинаковых областях.

Для проведения качественного библиометрического анализа необходим сбор исчерпывающей информации о публикациях сотрудников организации, которую обслуживает библиотека. Используемые в большинстве случаев запросы по названию организации дают лишь половинный результат, что в основном связано с проблемами транслитерации, с частой сменой названий организации и с указанием авторами головной организации в качестве места работы. В ИНГГ СО РАН и ГНЦ ВБ «Вектор» разработана и успешно применяется технология сбора библиографических метаданных о публикациях на уровне авторов и публикаций, описанная нами прежде [3].

На запросы по авторам, состоящие из всех возможных вариантов транслитерации фамилий сотрудников – в Web of Science и всех авторских идентификаторов – в Scopus, в базах данных устанавливается режим оповещения, по которому на электронный адрес ответственного за работу сотрудника библиотеки поступает информация о новых индексируемых публикациях. Идентификаторы новых публикаций впоследствии заносятся в общий список идентификаторов публикаций организации, который может включать несколько тысяч записей. Запрос из этих идентификаторов публикаций используется для последующего поиска журналов для подписки на основе библиометрического анализа.

Поисковая модель, основанная на библиометрических подходах, для формирования подписки в научной библиотеке

Разработанная в ИНГГ СО РАН и впоследствии апробированная в ГНЦ ВБ «Вектор» модель автоматизированного поиска основана на двух основных библиометрических методах – анализе цитирования и контент-анализе. По стандартной методике, описанной в предыдущих работах [2, 13], были рассмотрены цитирования из публикаций научных сотрудников двух организаций за пятилетний период, наиболее часто используемый для оценки текущих информационных потребностей ученых. Каждому журналу был присвоен коэффициент значимости по цитируемости C_{cit}^i , который вычислялся по формуле:

$$C_{cit}^i = \frac{C_i}{N} \cdot 100$$

где C_i – число ссылок из работ сотрудников организации на i -ый журнал ($i = 1 \dots L$), N – общее число ссылок из публикаций сотрудников организации на все журналы [4]. Множитель 100 был

принят для того, чтобы коэффициенты были более различимы и не нужно было повышать точность до 5-го знака.

В процитированных за этот срок журналах четко обозначалось ядро, доказывающее действенность закона рассеяния информации С. Брэдфорда. Учитывая ряд справедливых критических замечаний к анализу цитирования, а также то, что использование одного метода не может служить достаточной основой для принятия решений о подписке журналов, мы дополнительно использовали библиометрический метод контент-анализа, впервые предлагаемый для использования в комплектовании библиотек.

Особенность предлагаемого подхода к поиску и оценке научной информации заключается в том, что он основан на создании комплексного запроса в БД и состоит из автоматически извлеченных из публикаций организации ключевых слов. Относясь к количественным библиометрическим методам, подход предполагает использование естественного языка и по классификации отечественного исследователя С. Д. Хайтуна относится к тезаурусному методу [6]. Теоретически возможно использование ключевых слов трех типов: ключевые слова KeyWords Plus, присваиваемые публикациям в Web of Science, ключевые слова из тезаурусных словарей предметных баз данных (например, MeSH), а также ключевые слова, извлекаемые из полных текстов публикаций. В нашей работе использовались KeyWords Plus ввиду их наибольшей доступности для библиотечных работников.

Основное преимущество предлагаемого метода заключается в предельной тематической точности поиска, релевантности результатов, отсутствии информационного шума, автоматизации извлечения ключевых слов и построения запроса, а также в объективности подхода, поскольку в создании запроса не требуется участия экспертов.

Полученные по запросу из ключевых слов (KeyWords Plus) публикации содержали точно те же последовательности ключевых слов, которые были присвоены публикациям сотрудников наших организаций, что прямо указывало на их соответствие информационным потребностям ученых. Заключительным этапом в применении данного метода стал анализ журналов, содержащих наибольшее число релевантных статей. Для каждого журнала был вычислен ранговый коэффициент значимости по ключевым словам C_{kwp}^i согласно следующей формуле:

$$C_{kwp}^i = \frac{C_i}{M} \cdot 100$$

где C_i – число статей из i -го журнала ($i = 1 \dots L$), найденных по запросу из ключевых слов, M – общее число статей во всей базе данных цитирований, найденных по запросу из ключевых слов KeyWords Plus.

В полученном списке журналов так же, как и в списке, полученном при анализе цитирования, четко пролеживалось ядро, необходимое к подписке.

Ядерные журналы обоих списков в обеих организациях хорошо коррелировали друг с другом. Для получения универсального (единого) списка журналов был вычислен универсальный коэффициент значимости журнала:

$$C_u^i = \frac{C_{cit}^i \cdot C_{kwp}^i}{C_{vol}^i}$$

Для более тонкой оценки полученных значений C_u^i было проведено их нормирование коэффициентом C_{vol}^i , который учитывает объем i -го издания. Данная операция была вызвана необходимостью уравнивать журналы с бóльшим количеством выпусков (или статей) в году с журналами с меньшей периодичностью (или объемом издания), поскольку журналы первой группы с большей вероятностью могли оказаться в верхних позициях списка.

Коэффициенты объема журналов C_{vol}^i были получены по данным Journal Citation Reports (см. подробнее [1]). Отметим, что аналогичным образом можно при необходимости ввести поправку на стоимость журналов.

Результаты ранжирования журналов представлены в табл. 1–2.

Таблица 1. Первые 15 журналов в области биомедицины, ГНЦ ВБ «Вектор».

Название журнала	C_{cit}	C_{kwp}	C_{vol}	C_u
1. Journal of Virology	4,13	3,02	0,095	131,29
2. Virology	1,59	1,53	0,027	90,10
3. Journal of General Virology	1,45	0,91	0,023	57,36
4. Journal of Biological Chemistry	2,52	2,82	0,317	22,41
5. Journal of Molecular Biology	1,43	0,76	0,051	21,30
6. Nucleic Acids Research	1,26	1,49	0,089	21,09
7. Proceedings of the National Academy of Sciences of the USA	2,84	1,88	0,262	20,37
8. Journal of Medical Virology	0,67	0,49	0,021	15,63
9. Vaccine	1,40	0,83	0,091	12,76
10. Gene	0,37	0,54	0,016	12,48
11. AIDS	0,29	0,76	0,018	12,24
12. Science	1,96	0,37	0,063	11,51
13. Nature	2,01	0,34	0,061	11,20
14. Archives of Virology	0,45	0,52	0,021	11,14
15. Virus Research	0,53	0,42	0,020	11,13

Таблица. 2. Первые 15 журналов в области наук о Земле, ИНГГ СО РАН.

Название журнала	C_{cit}	C_{kwp}	C_{vol}	C_u
1. Геология и геофизика	6,46	0,13	0,040	20,99
2. Precambrian Research	1,14	0,44	0,030	16,72
3. Tectonophysics	1,57	0,59	0,065	14,25
4. Earth and Planetary Science Letters	1,70	1,34	0,165	13,80
5. Geology	1,20	0,75	0,090	10,00
6. Geochimica et Cosmochimica Acta	1,00	1,24	0,146	8,49
7. Geophysical Journal International	1,46	0,66	0,124	7,77
8. Стратиграфия. Геологическая корреляция	0,93	0,07	0,013	5,00
9. Contributions to Mineralogy and Petrology	0,49	0,37	0,040	4,53
10. Journal of Petrology	0,36	0,28	0,025	4,27
11. Доклады Академии наук	2,59	0,17	0,103	4,27
12. Chemical Geology	0,50	0,60	0,073	4,10
13. Organic Geochemistry	0,46	0,39	0,040	4,03
14. Geophysics	1,38	0,22	0,079	3,84
15. Canadian Journal of Earth Sciences	0,24	0,34	0,025	3,26

Данные списки включают в себя как платные журналы, так и журналы открытого доступа, в них представлено все разнообразие издательств, имеется возможность определить ранговые места российских журналов среди международной периодики.

Заключение

Накопленные объемы знаний, особенно в точных и естественных науках, требуют новых подходов к их извлечению и последующей обработке. Устоявшиеся методы часто не дают возможности полного и своевременного охвата научных знаний, отчего требуется разработка новых автоматизированных методов на основе существующего инструментария, внедренного в наукометрические базы данных. Примечательно, что наборы доступных инструментов можно комбинировать, получая новые оригинальные продукты для последующего проведения библиометрического анализа.

В настоящей работе библиометрические методы были внедрены в технологические процессы отбора периодики в фонд научной библиотеки. Была показана самодостаточность использования библиометрических методов применительно к подписке журналов, относительная доступность подходов и их объективность, поскольку ни один из процессов не требовал привлечения предметных экспертов. Учитывая эти преимущества библиометрических методов в применении к комплектованию, можно говорить о предпочтительности их использования в современных научных библиотеках.

Литература

1. Гуреев В.Н., Мазов Н.А. Тематика публикаций организации как основа формирования объективного и оптимального репертуара научной периодики // Научно-техническая информация. Серия 1: Организация и методика информационной работы. – 2013. – № 10. – С. 30–39.
2. Дубров А.П., Красикова О.Л. Использование цитат-анализа для изучения и формирования фонда иностранных журналов научных академических библиотек // Научные и технические библиотеки. – 1998. – № 6. – С. 26–34.
3. Мазов Н.А., Гуреев В.Н. Новые методы формирования публикационного профиля научной организации в сети науки // Научные и технические библиотеки. – 2013. – № 12. – С. 42–48.
4. Мазов Н.А., Гуреев В.Н. Изучение информационных потребностей ученых с использованием библиометрического анализа для оптимизации комплектования // Библиосфера. – 2012. – № 4. – С. 57–66.
5. Мазов Н.А. Оценка потока научных публикаций академического института на основе библиометрического анализа цитирования // Информационные технологии в гуманитарных исследованиях. – 2011. – № 16. – С. 25–30.
6. Хайтун С.Д. Наукометрия. Состояние и перспективы. – Москва: Наука, 1983. – 344 с.
7. Beall J. Criteria for Determining Predatory Open-Access Publishers. – 2012. – [Электронный ресурс]. – Режим доступа: <http://scholarlyoa.com/2012/11/30/criteria-for-determining-predatory-open-access-publishers-2nd-edition> (Дата обращения: 29.10.2014).
8. Beall J. Beall's List: Potential, possible, or probable predatory scholarly open-access publishers. – 2014. – [Электронный ресурс]. – Режим доступа: <http://scholarlyoa.com/publishers> (Дата обращения: 29.10.2014).
9. Bergstrom T.C., Courant P.N., McAfee R.P., Williams M.A. Evaluating big deal journal bundles // Proceedings of the National Academy of Sciences of the United States of America. – 2014. – V. 111. – № 26. – P. 9425–9430.
10. Jurski D. 2013 Study of Subscription Prices for Scholarly Society Journals: Society Journal Pricing Trends and Industry Overview. Allen Press, Inc., 2013. – 18 p.
11. Mabe M. The growth and number of journals // Serials. – 2003. – V. 16. – № 2. – P. 191–197.
12. Mabe M., Amin M. Growth dynamics of scholarly and scientific journals // Scientometrics. – 2001. – V. 51. – № 1. – P. 147–162.

13. McCain K.W., Bobick J.E. Patterns of Journal Use in a Departmental Library - a Citation Analysis // Journal of the American Society for Information Science. – 1981. – V. 32. – № 4. – P. 257–267.
14. Poynder R. The Big Deal: Not Price But Cost // Information Today. – 2011. – V. 28. – № 8. – P. 1–3.
15. Tenopir C., King D.W., Edwards S., Wu L. Electronic journals and changes in scholarly article seeking and reading patterns // Aslib Proceedings. – 2009. – V. 61. – № 1. – P. 5–32.
16. Tillery K. 2012 Study of Subscription Prices for Scholarly Society Journals: Society Journal Pricing Trends and Industry Overview. Allen Press, Inc., 2012. – 20 p.