

7. Дубин Б. В., Зоркая Н. А. Чтение–2008 // Кн. дело. – 2008. – № 3. – С. 70–76.
8. Книгоиздание в России. Состояние, тенденции и перспективы развития : доклад. – М. : Федер. агентство по печати и массовым коммуникациям : Упр. период. печати, книгоизд. и полиграфии, 2009. – С. 98–100.
9. Книжный рынок России: текущее состояние и перспективы развития // MarketReport: маркетинговые исследования и бизнес-планы. – URL: <http://market-report.ru/item.php?id=104820> (дата обращения: 22.01.2010).
10. Маркова Т. Б. Социальная и культурная значимость чтения в информационном обществе // Библиотековедение. – 2010. – № 3. – С. 61–65.
11. Россияне и чтение. – URL: http://www.bal-con.ru/news/2009-06-17/rossiyane_i_chtenie/ (дата обращения: 17.06.2009).
12. Рунет читает в цифре // Книжное дело : проф. журн. : [сайт]. 11 февр. 2011. – URL: http://www.knigdelo.ru/default.asp?id=7&news_id=3711 (дата обращения: 15.06.2010).

УДК 001.891
ББК 78.606.1

РОССИЙСКИЙ ИНДЕКС НАУЧНОГО ЦИТИРОВАНИЯ – ИНСТРУМЕНТ ДЛЯ АНАЛИЗА НАУКИ

© П. Г. Арефьев*, Г. О. Еременко*, В. А. Глухов**, 2012

* Научная электронная библиотека
119421, г. Москва, ул. Новаторов, д. 7а, корп. 1, оф. 107

** Институт научной информации по общественным наукам Российской академии наук
117997, г. Москва, Нахимовский пр., 51/21

Рассматривается текущее состояние проекта создания Российского индекса научного цитирования (РИНЦ). Представлена информация о внедренных и перспективных разработках, выполняемых в рамках развития проекта.

Ключевые слова: Российский индекс научного цитирования, РИНЦ, наукометрия, международные индексы научного цитирования, публикационная активность.

Current situation with the project on the Russian Science Citation Index (RSCI) is considered. Information on implemented and perspective developments is presented.

Key words: Russian Science Citation Index, RSCI, scientometrics, the international indexes of scientific citation; publication activities.

РИНЦ: современное состояние

В рамках комплексной реформы отечественной науки Федеральное агентство по науке и инновациям в 2005 г. инициировало конкурс «Разработка системы статистического анализа российской науки на основе данных Российского индекса цитирования». Конкурс проводился в рамках Федеральной целевой научно-технической программы «Исследования и разработки по приоритетным направлениям развития науки и техники».

Научная электронная библиотека (НЭБ) стала головным исполнителем проекта по созданию РИНЦ. Соисполнителями проекта были также Государственная публичная научно-техническая библиотека Сибирского отделения Российской академии наук (ГПНТБ СО РАН), Библиотека по естественным наукам РАН (Пушинский научный центр), Институт научной информации по общественным наукам РАН, Федеральное государственное унитарное предприятие (ФГУП) «Информрегистр».

Российский индекс научного цитирования – это многофункциональная информационная система, в которой обрабатывается полная библиографическая информация, аннотации и пристатейные списки цитирования из российских научных журналов. Поисковые и информационные сервисы базы данных реализуют различные виды поиска информации, анализируют и рассчитывают количественные показатели по объемам опубликованных статей и цитированию отдельных авторов, организаций, административно-географических регионов, тематических направлений, журналов.

Помимо известных показателей, таких как число публикаций, число ссылок (цитируемость), среднее число ссылок на публикацию, индекс Хирша и импакт-фактор, для оценки производительности, результативности и эффективности науки в РИНЦ рассчитывается около 50 различных метрик. Используя аналитический инструментарий, РИНЦ дает возможность корректно оценить результаты деятельности всей вертикали социального инсти-

туда науки, начиная с отдельного ученого-автора и заканчивая страной в целом.

На сегодняшний день в базе данных (БД) РИНЦ обрабатывается более 3,5 тыс. российских научных журналов, представлена информация по 2,3 млн публикаций, 600 тыс. авторов и 6 тыс. организаций. Для каждого объекта исследования и анализа – автора, журнала, организации – создан профиль, в котором представлена информация по библиометрическому анализу публикационного потока этого объекта.

В авторском профиле и профиле организации есть возможность получить не только полные списки статей, опубликованных автором или организацией, но и показатели их цитируемости, а также списки статей, которые цитируют данные работы. Глубина представленных материалов сейчас в основном составляет более 7 лет – с 2005 по 2012 г. включительно.

Основная задача РИНЦ – собрать информацию о всех публикациях российских авторов и их цитировании. Как уже отмечалось, многие российские ученые публикуются в зарубежных журналах. Поскольку в РИНЦ обрабатываются в основном российские журналы, то этой информации недостаточно для полного анализа. В связи с этим в РИНЦ были добавлены данные из базы Scopus. Это более 500 тыс. библиографических записей о публикациях в зарубежных журналах с участием хотя бы одного российского автора и свыше 1 млн статей, которые цитируют эти произведения. Глубина архива заимствованных данных из базы Scopus превышает 15 лет – с 1996 г. по настоящее время.

Активно развиваются контакты между НЭБ и компанией Thomson Reuters. По соглашению с производителями Web of Knowledge в онлайн-режиме для каждой статьи в РИНЦ отображаются показатели ее цитирования в Web of Science (WoS), разумеется, в случае наличия этого произведения в самой WoS.

В результате интеграции данных из отечественных и зарубежных журналов в РИНЦ получен единый массив документов, отражающий полный поток российских научных публикаций и их цитирований. Уже сегодня это позволяет проводить анализ и оценку результатов научной деятельности отечественных ученых и организаций с применением библиометрических методов.

Проблемы национальных и международных индексов цитирования

Одна из наиболее сложных задач создания индексов цитирования – проблема идентификации авторов научных произведений и организаций, в которых работают ученые, а также привязки ссылок на

оригинальные работы в пристатейных списках литературы.

В РИНЦ разработана полностью автоматическая система идентификации библиографических ссылок. Для этого каждая ссылка, поступающая в БД, программно разбивается на элементы (автор, название работы, источник, где была опубликована статья, выходные данные). Отметим, что задачу усложняет наличие большого количества вариантов оформления пристатейных списков литературы в российских журналах, несоблюдение действующего ГОСТ 7.0.5-2008 на библиографические ссылки. Для выполнения программного разбиения ссылок создано более 500 так называемых шаблонов, которые используются для решения задачи.

В процессе идентификации авторов и организаций основными проблемами являются: наличие однофамильцев, неполное написание имен, различная транслитерация русских фамилий в зарубежных журналах, а также ошибки при оцифровке печатных версий журналов. Сведения о местах работы авторов отсутствуют в ряде журналов. Особенно это относится к архивным выпускам.

Зачастую сведения об организациях, где работают авторы, приводятся без четкой привязки к списку авторов в статье или содержат общеупотребительные сокращения. Например, аббревиатуру КГТУ имеют несколько университетов в стране – Калининградский технический, Костромской технический и Казанский технологический университеты, и таких примеров много.

Кроме того, наименования мест работы авторов в журналах не унифицированы. В результате одной организации в базе РИНЦ может соответствовать от 10 до 100 и более вариантов названий. Это не позволяет выполнить полноценную обработку данных и заметно снижает показатели публикационной активности и индексы цитирования авторов и научно-образовательных организаций.

С подобными проблемами сталкивается любой индекс цитирования. Например, в WoS Национальный исследовательский технологический университет МИСиС представлен 21 различным наименованием.

В системе Scopus публикации известного российского специалиста, доктора медицинских наук, профессора В. А. Горбуновой (возглавляет Отделение химиотерапии в Российском онкологическом научном центре им. Н. Н. Блохина), распределены между двумя разными профилями. В поиске по автору Scopus находит только 88 публикаций В. А. Горбуновой, но, используя интерфейс расширенного запроса, можно найти 102 статьи этого автора.

Примеров такого рода в международных системах цитирования насчитываются тысячи. Для

решения этих проблем специалисты РИНЦ разрабатывают специальные инструменты, которые составляют основу аналитической надстройки над национальным индексом.

Система Science Index как расширение аналитических возможностей РИНЦ

Аналитическая система Science Index – это набор инструментов, позволяющий повысить качество БД и расширить ее аналитические возможности для авторов научных статей, научно-образовательных организаций, издательств, а также государственных органов, использующих РИНЦ для принятия решений в области научной политики.

Science Index состоит из трех основных систем: для авторов, организаций и издательств.

В настоящий момент программные средства РИНЦ производят автоматический отбор сведений о публикациях автора и ссылок на эти публикации. Выше уже отмечались сложности, возникающие на пути программной идентификации авторских публикаций и цитирований. Проблему в большой степени решает разработанный интерфейс Science Index [автор], который через систему уникальной идентификации автора (SPIN-код) дает возможность каждому зарегистрированному ученому провести анализ публикаций и ссылок, автоматически «приписанных» ему системой, отобрать (или удалить) из этого списка публикации и ссылки, провести поиск во всем массиве РИНЦ и включить публикации и ссылки, программно по тем или иным причинам не идентифицированные.

Отметим при этом, что сервисы Science Index [автор], включая регистрацию и работу в системе, для пользователей абсолютно бесплатны. За прошедшие с начала эксплуатации полтора года в Science Index [автор] зарегистрировалось свыше 85 тыс. российских ученых (общий массив авторов в РИНЦ составляет 603 тыс. человек, а действующих авторов, которые за последние 5 лет публиковались хотя бы единожды – около 250 тыс.).

В аналогичной системе ResearcherID, разработанной для БД WoS в 2008 г., за 4 года работы зарегистрировалось около 150 тыс. авторов, из которых всего 2,5 тыс. человек представляют Россию [1].

В августе 2012 г. запущена в эксплуатацию система Science Index для организаций. Интерфейс Science Index [организация] позволяет уполномоченному представителю провести выверку публикаций авторов своей организации и их цитирований. Для каждого сотрудника вносится информация о периоде его работы в данном учреждении, что позволяет, в частности, увеличить показатели организации за счет включения тех статей авторов (и их цитирований), в которых не указаны или ошибочно указаны места их работы.

Важный элемент системы Science Index [организация] – возможность ввода структуры своей организации. Это позволяет администрации производить сравнение по показателям публикационной активности и цитируемости внутри самого учреждения – среди различных подразделений, временных творческих коллективов и т. д.

Однако самым значительным сервисом в Science Index [организация] следует считать добавление описаний тех публикаций, которые отсутствуют в РИНЦ. Сейчас основной массив индекса составляют статьи из журналов. В 2011–2012 гг. в базу были импортированы описания российских авторефератов, диссертаций и патентов.

Сведения о других типах публикаций (монографиях, статьях в научных сборниках, тезисах конференций, научных отчетов) также попадают в БД, но исключительно из пристатейных списков литературы. Идентификация этих данных (включая сведение значительно различающихся между собой дублетных описаний) весьма затруднительна.

В интерфейсе Science Index [организация] научно-образовательные организации могут в онлайн-режиме вносить в базу описания публикаций и размещать, при необходимости и возможности, их полные тексты. Для организаций, решивших принять участие в этом проекте, доступны более детальные таблицы с разнообразными статистическими распределениями и новые формы аналитических отчетов, которые закрыты для остальных пользователей. В международных индексах научного цитирования на сегодняшний день аналогичных инструментов нет.

Перспективной разработкой является интерфейс Science Index для издательств. Отдельные элементы издательского интерфейса созданы еще несколько лет назад, в частности:

- опции по редактированию профиля издательства, что позволяет издателям оперативно вносить изменения в описания своих журналов в РИНЦ;
- просмотр статистики по доступу пользователей к журналу;
- система приема и контроля загрузки выпусков журналов в БД.

В новом интерфейсе будут добавлены онлайн-сервисы для редактирования метаданных оглавлений журнала, заключения договоров и их выполнения, а также средства подготовки выпусков журналов в принятом в РИНЦ формате XML (новая онлайн-система XML разметки Articulus уже разработана и проходит тестирование).

Частью системы Science Index [издательство] станет так называемая электронная редакция – комплекс средств, позволяющий выполнить в Интернете все работы по приему рукописей от авторов, рецензированию, формированию выпусков

и оперативной публикации электронной версии журнала.

В рамках электронной редакции разработан онлайн-редактор для набора авторами научных текстов, который содержит, например, средства для ввода математических и химических формул, иллюстраций, аудио- и видеоматериалов.

Новая методика оценки научной периодики в РИНЦ: интегральный показатель Science Index для журналов

Для оценки качества и авторитетности научной периодики в РИНЦ, как и в Journal Citation Reports, используется ряд метрик, важнейшей из которых является импакт-фактор. Главный недостаток импакт-фактора – возможность искусственного увеличения этого показателя за счет определенных манипуляций с количеством публикуемых статей и списками использованной литературы или за счет роста самоцитирования¹.

Для того чтобы преодолеть проблемы импакт-фактора и более достоверно анализировать российскую науку было принято решение к 2013 г. выполнить работу по нормализации, выверке и расширению архивного представления библиографических данных РИНЦ. На основе анализа общей цитируемости периодических изданий в РИНЦ установлено, что около 1,5 тыс. журналов получают 95% ссылок, которые приходятся на всю научную периодику России. По результатам этого исследования решено отобрать ведущие 1,5 тыс. журналов и обработать их выпуски за 10 лет, начиная с 2003 г.

Однако встает вопрос: а как определить, ведущих это журнал или не ведущий, каков уровень авторитетности издания? Можно использовать проверенный временем показатель импакт-фактора и получить все те искусственные смещения, о которых упоминалось выше. Другой способ: взять общие показатели цитируемости изданий в абсолютном исчислении, но тогда перечень журналов будет непропорционально представлен разными дисциплинами.

Так, журналы по экспериментальным наукам – физика, химия, медицина, биология – цитируются на порядок выше, чем издания по математике, социологии или истории. Кроме того, влияние имеют маркетинговая раскрученность и узнаваемость издания (журналы, давно издающиеся, или вкладывающие дополнительные ресурсы в маркетинг и рекламу, имеют преимущество в позиционировании на рынке научной периодики, порой чрезмерно завышенное и не соответствующее реальному качеству содержания журнала).

Специалистами РИНЦ в результате проведенных исследований был обобщен мировой опыт в разработке методов и алгоритмов оценки научных журналов и разработан собственный показатель для решения этой проблемы – Интегральный показатель Science Index для журналов. Он рассчитывается по следующей методике:

За основу расчетов интегрального показателя берется общее число ссылок, которые получили статьи, опубликованные в журнале за предыдущие 5 лет, в течение года, для которого показатель рассчитывается.

Например, при расчете показателя за 2010 г. за основу берется суммарное число ссылок, сделанных в 2010 г. на статьи, опубликованные в журнале в 2005–2009 гг. Расширенное с 2 до 5 лет окно цитирования позволяет в некоторой степени сгладить различия в хронологическом распределении ссылок для журналов различных тематических направлений. В то же время, безусловно, журналы, статьи в которых начинают цитироваться раньше, получают определенные преимущества.

Показатель не нормируется на количество статей, опубликованных в журнале в течение расчетного периода, как это делается, например, при определении импакт-фактора. Соответственно, он отражает не среднее количество цитирований, полученных одной статьей, а совокупный объем цитирований всех статей в журнале, т. е. вклад данного журнала в общую массу цитирований, сделанных в течение года. Это означает, что при равном среднем уровне цитирования публикаций преимущество получают более крупные журналы. Учитываются только цитирования, сделанные в научных статьях, научных обзорах и кратких сообщениях. Не рассматриваются цитирования из публикаций, у которых нет авторов.

Выбор в качестве основного оценочного показателя общего количества цитирований журнала позволяет отчасти скорректировать различия между обзорными журналами и журналами, печатающими оригинальные исследовательские статьи. Преимущество журналов, содержащих в основном обзорные статьи, которые в среднем цитируются чаще, нивелируется за счет относительно небольшого числа обзоров, выпускаемых журналом за год.

Интегральный показатель Science Index рассчитывается только для российских научных журналов на русском языке, или российских журналов, выходящих сразу на английском языке и не имеющих оригинальной русскоязычной версии. Не рассматриваются реферативные журналы и журналы, не выходящие в настоящее время.

Если журнал переименовывался в течение последних 5 лет либо имело место слияние нескольких журналов в один или, наоборот, разделение журнала на отдельные серии, то учитывается

¹ Как известно, самоцитирование статей в журнале при расчете импакт-фактора не исключается из расчетов.

суммарное количество цитирований данного журнала и его предыдущей версии. Если журнал имеет переводную английскую версию, учитывается суммарное количество цитирований оригинальной русскоязычной и переводной версий. В случае цитирования из российского журнала, имеющего переводную версию на английском языке, которая также представлена в РИНЦ, засчитывается только ссылка из оригинальной статьи на русском языке. Это позволяет избежать дублирования ссылок из оригинальной статьи и ее переводной версии.

Расчет показателя Science Index осуществляется в две стадии. На первой стадии сначала для совокупности статей из журнала, опубликованных в течение последних 7 лет, определяется список статей в других журналах, процитировавших их в течение этого же периода. Затем для массива этих цитирующих статей определяется медианное значение количества ссылок в списке цитируемой литературы на журналы, обрабатываемые в РИНЦ (так называемый потенциал цитирования). Эта величина показывает, сколько ссылок обычно имеют цитирующие данный журнал статьи.

Полученные значения потенциала цитирования далее нормируются путем деления на медианное значение потенциала цитирования по всей совокупности журналов, обрабатываемых в РИНЦ (т. е. такое значение, при котором половина журналов в БД имеют потенциал цитирования выше, а другая половина – ниже). Количество цитирований для каждого журнала корректируется путем деления на рассчитанный таким образом нормированный потенциал цитирования. Чем больше потенциал цитирования (или среднее число ссылок в расчете на одну цитирующую данный журнал статью), тем меньший относительный вес имеет каждая ссылка на данный журнал. При этом самоцитирование не засчитывается. Такой подход позволяет автоматически учесть специфику цитирования в том кластере журналов, к которому относится данный оцениваемый журнал. Причем учитываются не только тематические различия (например, в математических журналах в среднем в несколько раз меньше ссылок, чем в химических), но и другие факторы, влияющие на практику цитирования в конкретной референтной группе журналов.

Например, в одном и том же тематическом направлении теоретические и прикладные журналы могут иметь существенные различия по величине и составу списка цитируемой литературы. Референтная группа, к которой относится данный журнал, определяется для каждого журнала индивидуально: в эту группу включаются журналы, процитировавшие данный журнал в течение расчетного периода. Это позволяет учитывать специфику цитирования не только для узкоспециализированных журналов, но и для мультидисциплинарных.

Еще один важный момент, на который необходимо обратить внимание: среднее число ссылок в расчете на одну цитирующую данный журнал статью зависит от того, насколько хорошо охвачено данное направление в БД РИНЦ.

Например, если в БД более полно представлены журналы по экономике, чем по лингвистике, то даже при одинаковом среднем количестве ссылок в статье экономические журналы получают преимущество, поскольку часть ссылок на лингвистические статьи из необрабатываемых в РИНЦ журналов будет потеряна.

С целью компенсировать в определенной степени возможную неравномерность наполнения БД по различным научным направлениям, при расчете потенциала цитирования учитываются только ссылки на журналы, обрабатываемые в БД РИНЦ.

Кроме того, при расчете потенциала цитирования не учитываются ссылки на монографии, материалы конференций, патенты, сборники статей и другие типы научных публикаций. Это позволяет в определенной степени сгладить различия между естественными и гуманитарными науками, где доля ссылок на публикации в нежурнальных источниках, как правило, выше. Также не учитываются ссылки, сделанные из этого же журнала, что устраняет влияние самоцитирования на определение потенциала цитирования для журнала.

Таким образом, рассчитанный на первой стадии показатель журнала по сути представляет собой количество ссылок в расчетном году, сделанных из других журналов на статьи в данном журнале за 5 предыдущих лет, скорректированное с учетом особенностей цитирования в референтной группе данного журнала.

Вторая стадия – учет авторитетности источника ссылки и уровня самоцитирования журнала. В качестве меры авторитетности источника цитирования используются рассчитанные на первом этапе показатели журналов. Предварительно эти показатели нормируются относительно медианного значения. Полученные относительные значения показателей цитирующих журналов используются в качестве весовых коэффициентов при расчете суммарного цитирования журнала: ссылки, пришедшие из журналов с высоким показателем, имеют больший вес, чем ссылки из малоцитируемых журналов.

При расчете числа цитирований журнала, взвешенного с учетом авторитетности источника, ссылки из этого же журнала (самоцитирование) обрабатываются отдельно. Для этих ссылок вводится дополнительный весовой коэффициент, который зависит от коэффициента самоцитирования журнала, т. е. от отношения числа самоцитирований журнала к общему числу ссылок, исходящих из журнала. Чем меньше коэффициент самоцити-

рования, тем ценнее ссылки, полученные журналом из этого же журнала.

Наконец, рассчитанное с учетом авторитетности источника количество цитирований журнала делится на определенное ранее значение нормированного потенциала цитирования для данного журнала. Полученное в результате значение используется в качестве интегрального показателя при построении ранжированного списка российских научных журналов в системе Science Index.

База данных РИНЦ постоянно растет и пополняется как новыми журналами, так и новыми выпусками уже обрабатываемых журналов, в том числе архивными. Чтобы учесть эти обновления, интегральный показатель журналов следует пересчитывать с интервалом минимум 1 раз в год.

* * *

История развития проекта создания РИНЦ насчитывает 6 лет. Тем не менее уже сейчас уровень разработки и функциональность поискового и аналитического инструментария позволяют считать РИНЦ национальным информационно-аналитическим продуктом, дающим возможность проводить анализ публикационной активности российских ученых и научных организаций.

Несмотря на незначительный срок (для сравнения Science Citation Index приближается к 50-летию юбилею), РИНЦ все более активно используется и как крупнейшая библиографическая БД по российской научной периодике (30 тыс. уникальных посетителей и 400 тыс. запросов в день), и как механизм для анализа результативности научной деятельности.

Начиная с 2009 г. Российский индекс научного цитирования и его показатели включены в типовые методики оценки результативности деятельности научных организаций, выполняющих научно-исследовательские, опытно-конструкторские и технологические работы гражданского назначения, разработанные Министерством образования и науки Российской Федерации (РФ), Министерством здравоохранения и социального развития РФ, Федерального медико-биологического агентства РФ и другими федеральными министерствами и ведомствами, имеющими, согласно Постановлению Правительства РФ № 312 от 08.04.2009 г., в своих структурах научно-исследовательские и образовательные организации. Растет и библиография опубликованных работ, посвященных как самому национальному индексу, так и его использованию в прикладной деятельности для оценки ученых, журналов и организаций [2–4].

Литература

1. ResearcherID. Thomson Reuters, 2012. – URL: <http://www.researcherid.com> (дата обращения: 25.09.2012).
2. Свирюкова В. Г., Ремизова Т. В. Информационное обеспечение библиометрических исследований в регионе: роль справочно-библиографического отдела ГПНТБ СО РАН // Библиосфера. – 2009. – № 4. – С. 76–77.
3. Писляков В. В. Зачем создавать национальные индексы цитирования? // Науч. и техн. б-ки. – 2007. – № 2. – С. 65–71.
4. Костюкова М. В. Современное состояние и развитие Российского индекса научного цитирования // Проф. образование. Столица. – 2011. – № 2. – С. 38–42.

УДК 02:001.8+002.55
ББК 78.606

БИБЛИОМЕТРИЯ – НОВОЕ НАПРАВЛЕНИЕ РАБОТЫ БИБЛИОТЕК УНИВЕРСИТЕТОВ ЕВРОПЫ

© М. С. Галявиева, 2012

*Казанский государственный университет культуры и искусств
420059, г. Казань, Оренбургский тракт, 3*

Рассматривается деятельность библиотек университетов Европы по организации библиометрических исследований и предоставлению библиометрических услуг.

Ключевые слова: библиометрия, наукометрия, университеты, библиотеки, библиометрические исследования, библиометрические услуги, Европа.

Activity of libraries of universities in Europe on the organization of bibliometric researches and to bibliometric servicing is considered.

Key words: bibliometrics, scientometrics, universities, libraries, bibliometric researches, bibliometric services, Europe.