

ИННОВАЦИОННЫЕ ТЕХНОЛОГИИ В ИНФОРМАЦИОННОМ ОБЕСПЕЧЕНИИ ОБРАЗОВАНИЯ, НАУКИ, КУЛЬТУРЫ*

УДК 061.6:025.4.036

Зибарева Инна Владимировна

*Институт катализа им. Г.К. Борескова СО РАН
Новосибирск, Россия*

Оценка научной результативности организации библиометрическими методами: проблемы и решения

Рассмотрены проблемы количественной оценки научной результативности исследовательской / образовательной организации с использованием библиометрических методов и обсуждены пути их решения.

Ключевые слова и словосочетания: *научная результативность, автор, организация, поиск информации, базы данных, Российский индекс научного цитирования, Web of Science, Scopus.*

Введение

В соответствии с постановлением правительства [1] в настоящее время в стране проводится количественная оценка результативности организаций, занимающихся научной деятельностью, в первую очередь институтов Российской академии наук (РАН), которая предполагает серьезные организационные выводы. Нормативные документы (типовые методики) для оценки результативности за определенный период предусматривают использование библиометрических показателей [2]. Последние, в частности, включают: количество научных публикаций и цитируемость сотрудников организации по базе данных (БД) Российский индекс научного цитирования (РИНЦ), БД Web of Science (WoS) или другим авторитетным источни-

* Рубрика подготовлена по материалам 7-й научно-практической конференции «Инновационные технологии в информационном обеспечении образования, науки, культуры», ВГУЭС, Владивосток, 17 – 25 сентября 2012 г.

кам реферативной информации; средний импакт-фактор журналов, опубликовавших эти работы, по БД WoS.

В настоящее время в мире имеется около 20 авторитетных библиографических БД, учитывающих цитирование [3]. Для институтов РАН, разделенных с учетом их научной специализации на референтные группы, решение о том, какие дополнительные БД помимо РИНЦ и WoS могут использоваться в данной группе, принимает соответствующее Отделение РАН. Среди них наиболее востребованной, скорее всего, окажется политематическая БД Scopus.

Количественная оценка результативности научной деятельности с использованием библиометрических показателей важна, конечно, не только для институтов РАН, но и других организаций, выполняющих фундаментальные и прикладные исследования, в первую очередь университетов, например, в контексте активно обсуждаемого сейчас их национального и международного ранжирования. Вместе с тем такая оценка связана с рядом сложных технических проблем, основные из которых, вместе с возможными путями их решения, обсуждены в настоящей статье на примере трех названных БД – РИНЦ, WoS и Scopus.

Получение библиометрических показателей организации

В настоящее время для получения величин библиометрических показателей, особенно в целях сравнений и оценок, во всем мире используются глобальные БД с онлайн-доступом. Основная проблема – максимально полный и точный отбор из БД релевантных публикаций. В случае организации такой отбор можно провести двумя способами – по ее названию и по всем работающим в ней авторам.

Первый способ более быстрый, но, как правило, требующий составления сложных поисковых запросов, учитывающих различные варианты названия организации, встречающиеся в БД, взятые из оригинальных публикаций [4]. Второй – более длительный, также требующий учета различных написаний имен авторов в БД, но, возможно, предпочтительный в случае относительно небольших организаций.

В обоих случаях наибольшие трудности связаны с тем, что в БД имена авторов и названия организаций вводятся так, как они представлены в публикациях – без какой-либо стандартизации. Кроме того, отечественные публикации могут реферироваться как в оригинальной, так и в переводной версии. При этом в некоторых БД, например РИНЦ и Scopus, обе версии нередко присутствуют одновременно [5, 6].

Поиск по организации

Названия организаций порождают серьезные библиометрические проблемы, в частности, в связи с переводом на другой язык в зарубежных БД, а также из-за неоднозначности, возникающей при использовании со-

кращений и аббревиатур. При библиометрическом поиске необходимо учитывать все названия организации и, если имеются, ее филиалов, отделений и т.д. Организации могут изменять названия, объединяться или разделяться. Встречается ошибочная атрибуция производителем БД публикаций организациям. Сравнительно часто названия организаций, в которых выполнена работа, в БД просто отсутствуют. Все это осложняет поиск публикаций организации по ее названию.

В этих условиях полезными дополнительными идентификаторами могут быть географическое положение (геолокация / адрес) и (или) ведомственная принадлежность организации. Однако если в одних БД для журнальных публикаций приводится место работы всех авторов, то в других – лишь первого из них. Кроме того, как и название организации, написание адреса в БД может быть поливариантным, а сам он изменяться со временем. Ведомственная принадлежность организации в публикациях указывается не всегда, к тому же в БД в этом отношении встречается путаница (например, между РАН и РАМН). В публикациях НИИ, входящих в состав университетов, иногда не упоминается сам университет.

Таким образом, при поиске публикаций исследовательской / образовательной организации следует предусмотреть все мыслимые варианты ее названия и адреса, включая лингвистические вариации. Непредвиденные способы написания и неполные, а также отсутствующие адреса могут повлиять на полноту списка найденных публикаций. В каждом случае полученные результаты следует проверять сравнением с официальными списками трудов организации (собственной БД публикаций, если имеется).

Поиск по авторам

С представлением имен авторов в БД связано множество проблем. В одних они указываются так, как приведены в первоисточнике – без какой-либо стандартизации, в других сокращаются до инициалов. Разные журналы имеют разные правила, предлагая, например, указывать первое имя автора полностью или лишь в виде инициала, при этом сокращенная форма «фамилия, инициалы» может быть идентичной для нескольких авторов.

Сами авторы могут использовать разные формы написания своего имени в разных публикациях. В целом невозможно предугадать, как будет представлено российским автором собственное имя в англоязычной литературе. Кроме того, один и тот же автор публикации на русском языке может быть представлен в БД в нормативной транслитерации, а в качестве автора публикаций в зарубежных журналах – в неидентичной ей собственной. Для российских авторов написание имени зависит и от языка зарубежной публикации, например, английского, немецкого или французского [7]. Широко встречаются ошибки в инициалах и фамилиях. При

реферировании в БД переводных версий фамилия и имя автора могут быть перепутаны. В целом можно выделить 10 видов типичных ошибок, связанных с транслитерацией, добавлением лишних знаков, пропусками, перестановками, опечатками, пробелами, знаками препинания, капитализацией букв, различными вариантами сложных имен, а также комбинациями всего вышеперечисленного [8].

Таким образом, для корректного отбора публикаций организации по ее отдельным сотрудникам требуется тщательная предварительная подготовка поисковой стратегии. Полноту отбора в каждом случае следует проверять сравнением поисковых результатов со списками трудов ученых.

Поиск цитирования

Существует два способа поиска цитируемых публикаций: первый – поиск публикаций в реферируемых в БД изданиях и получение для них статистики цитирования (процедура View citation overview); второй – прямой поиск цитируемых ссылок (аналогичный процедуре Cited reference search в БД WoS), при котором находятся и т. н. «некорректные» (содержащие технические ошибки) цитирования.

При поиске цитирования организаций используется первый способ. Результаты могут существенно зависеть от использованной БД, например из-за различного охвата источников, но не только.

Поиск цитирования порождает, возможно, наибольшие библиометрические проблемы. Все перечисленные выше ошибки, связанные с названиями организаций и именами авторов, относятся и к цитированию. При этом следует иметь в виду, что к возможным ошибкам, сделанным авторами при цитировании публикаций, добавляются ошибки, внесенные на стадии ввода ссылки в БД. Осложняющим обстоятельством являются различные правила оформления ссылок в журналах. Цитирующие авторы могут указывать полные имена цитируемых или использовать различные комбинации инициалов. Если написание имени в библиографической БД и в ссылке не совпадают, такая ссылка не будет найдена.

Предлагаемые решения

Хотя БД научно-технической информации создавались отнюдь не для библиометрических оценок результативности профессиональной деятельности организаций и (или) авторов, они все чаще используются в этих целях, что производители ресурсов хорошо понимают. Серьезным шагом с их стороны навстречу сформировавшемуся запросу стало создание в некоторых БД профилей авторов и организаций – в ряде случаев с возможностью их редактирования самими авторами / организациями. Этой возможностью следует активно пользоваться, хотя имеются чисто технологические проблемы. Автоматическая обработка данных не позволяет во всех случаях точно идентифицировать организации / авторов, что

приводит (может привести) к созданию для них нескольких профилей и, тем самым, уменьшению величин библиометрических показателей. Это мотивирует ручную идентификацию организаций / авторов для профилей, позволяющую решить ряд проблем. Ручная работа, однако, требует сравнительно больших затрат времени и (или) денег и не гарантирует полного избавления от ошибок. Более того, устранение одних ошибок может сопровождаться внесением других.

Средства и методы внешнего редактирования профилей авторов / организаций в различных БД различаются [9].

БД WoS. На платформе Web of Knowledge (WoK), где размещена БД WoS, имеется надстройка открытого доступа для авторов ResearcherID (RID) – личная карточка исследователя, обеспечивающая: однозначную идентификацию автора, полную интеграцию с платформой WoK, включающую добавление записей в RID из нее и любых других источников, получение статистических данных по цитированию. Самое главное заключается в том, что RID автора автоматически добавляется в описание статьи в БД WoS и по нему можно проводить поиск публикаций. Таким образом, для отбора всех публикаций организации можно использовать поиск по RID в закладке Advanced search, перечислив все RID сотрудников организации. Наряду с другими элементами информационного массива RID также используются в БД для формирования с помощью специального алгоритма индивидуальных авторских наборов (Distinct author sets), к которым можно перейти из опций Author finder, General search или Advanced search по имени автора. Они содержат полезные в обсуждаемом контексте списки статей, сопоставленных в БД конкретному имени. В ближайшее время корпорация Thomson Reuters, производящая БД, планирует использовать данные RID для уточнения информации в WoK [9].

БД РИНЦ. Здесь каждая ученная организация имеет свой профиль. В принципе, это позволяет получить всю необходимую для оценки результативности информацию – в карточке организации представлено распределение публикаций и цитирований по годам в удобном табличном формате. Точность данных РИНЦ, однако, вызывает критику. В значительной степени это связано с тем, что в БД РИНЦ сопоставление публикаций и ссылок авторам, организациям и журналам производится в автоматическом режиме. При этом встречаются все уже отмеченные выше проблемы. Так, не все публикации могут быть автоматически сопоставлены конкретному автору, поскольку часто имеющейся информации недостаточно для его однозначной идентификации среди однофамильцев. Кроме того, написание фамилии автора на английском языке может отличаться от стандартной транслитерации с русского или быть поливариантным. Еще более сложная задача – автоматическая привязка публикаций к организациям ввиду множества вариантов написания их

полных и сокращенных названий. В связи с этим планируется привлечение к работе по уточнению информации, исправлению ошибок и привязке публикаций и ссылок к авторам самих ученых, научных организаций и издательств. Поскольку авторы, и организации, и журналы заинтересованы в максимальной полноте и точности релевантной им информации, подход представляется вполне реалистичным.

Более детальную библиометрическую информацию и величины ряда важных показателей позволяет получить аналитическая надстройка БД РИНЦ – система Science Index (SI).

Система SI для авторов. После регистрации автора в БД РИНЦ его анкета поступает в службу поддержки, производящую идентификацию автора в этом качестве и поиск по всему информационному массиву БД его публикаций и цитирований. Полученные результаты проверяются. По завершении этих операций, занимающих до 10 рабочих дней, на электронный адрес автора отправляется сообщение о присвоении ему персонального идентификационного кода (SPIN-кода). Обладателю кода автоматически открывается доступ к новым сервисам, предоставляемым системой SI для авторов публикаций. Они могут добавить найденные в БД РИНЦ публикации – в список своих работ, найденные ссылки – в список своих цитирований, удалить из списка своих работ или цитирований ошибочно попавшие туда чужие публикации или ссылки, идентифицировать организации, указанные в качестве места работы, получить актуальные показатели цитируемости не только по БД РИНЦ, но и по БД WoS и БД Scopus с возможностью перехода на список цитирующих статей в них при наличии подписки. Авторам предоставляется возможность идентифицировать свои публикации, в том числе отсутствующие в БД РИНЦ, если на них есть ссылки в списках цитируемой литературы. Планируется предоставить авторам возможность добавлять новые публикации в БД РИНЦ в том случае, если их достоверность можно независимо проверить: например, если статья зарегистрирована в CrossRef и имеет идентификатор DOI.

Система SI для организаций содержит средства идентификации, уточнения и дополнения информации в БД РИНЦ с участием научных организаций, издательств и авторов публикаций. Планируется, что возможность включения новых публикаций будет распространена на авторизованных представителей организаций на основании специальных соглашений. При этом в БД РИНЦ можно будет добавлять не только статьи, опубликованные в не учитываемых в ней журналах, но и другие типы научных публикаций: статьи в сборниках, монографии, материалы конференций, диссертации, патенты и т.д. Достоверность информации обеспечивает организация, заключившая соглашение. В рамках приобретаемой лицензии для организации будут доступны: ввод в систему SI иерархиче-

ской структуры организации; поиск ее сотрудников – авторов публикаций и их привязка к организации; регистрация новых авторов – сотрудников организации; некоторые другие операции.

БД Scopus. Специализированной надстройки для авторов здесь нет, однако они имеют профили и могут: объединять несколько авторских профилей; уточнять информацию, например, отмечать свои и исключать чужие публикации; посылать текстовые запросы на изменение информации по электронной почте.

Авторский идентификатор БД Scopus (Scopus author identifier – SAI) – функция, различающая авторов путем присвоения каждому из них уникального номера и группирующая публикации по этим номерам, что особенно полезно в случае широко распространенных имен. В SAI учитывается возможность поливариантного написания имени автора – варианты сопоставляются и публикации группируются вместе, даже если автор был процитирован при различном написании имени. В этих случаях для однозначного отнесения автора к конкретному номеру и соответствующей группировки публикаций используется алгоритм, учитывающий аффилиацию, адрес, предметную область, название источника, даты публикации и цитирований, а также соавторов. Использование SAI возможно в таких видах поиска, как Author search и Advanced search.

Кроме того, в БД Scopus имеется идентификатор организации Scopus affiliation identifier – первый в мире инструмент, помогающий идентифицировать и сгруппировать полный набор работ организации, т.е. существенно упрощающий этот затратный по времени процесс. Он различает организации (их подразделения) с похожими названиями на основании присвоенных им уникальных номеров. Публикации, аффилированные с организацией, группируются по ее номеру. Алгоритм учитывает поливариантность названия организации в публикациях, что позволяет сгруппировать релевантные ссылки, даже если в них использовались различные названия организации. Использование этого идентификатора возможно в таких видах поиска, как Affiliation search, Author search и Advanced search.

Заключение

Рассмотренные технические проблемы количественной оценки научной результативности отечественной исследовательской / образовательной организации с применением библиометрических методов касаются как оцениваемых и оценивающих структур, так и производителей используемых информационных ресурсов. В макромасштабе, например на уровне страны, многие из обсужденных проблем, ограничений и ошибок не столь существенны. Однако на уровне отдельной организации (и тем более ученого) они способны приобрести решающее значение.

Рассмотренные проблемы, конечно, не могут быть решены отдельными организациями или (и) учеными. Прежде всего, они требуют максимальной и, желательно, единой стандартизации БД, что зависит от их производителей. В любом случае, для улучшения ситуации нужен постоянный диалог пользователей (организаций, а также их сотрудников) и производителей БД, направленный на создание предельно точных ресурсов. Организациям совершенно необходима целенаправленная работа с производителями БД, во всяком случае, наиболее важными в контексте их оценки и ранжирования. Это, в первую очередь, информирование об обнаруженных ошибках. Для организаций очень полезно, если все сотрудники, ведущие научную работу, регистрируются в соответствующих БД и будут контролировать свои показатели – корректное отражение показателей сотрудников во многом обеспечивает корректное отражение показателей организации. Прежде всего, можно рекомендовать получение RID (БД WoS) и SPIN-кодов (БД РИНЦ) каждым публикующимся сотрудником исследовательской / образовательной организации.

Кроме того, руководству организаций следует целенаправленно работать с сотрудниками – авторами публикаций над культурой оформления статей с особым акцентом на корректное / стандартизованное (утвержденное / официальное) указание организации – места работы. Для каждого автора и тем самым для организации в целом важно унифицированное (так или иначе) написание своего имени в публикациях.

Наконец, для организаций необходимо создание и поддержание в актуальном состоянии собственных БД научных трудов сотрудников. В них можно хранить авторские RID из WoS, Spin-коды из РИНЦ, идентификаторы из Scopus, а также номера публикаций (accession numbers) в соответствующих БД. Однажды найденные, эти номера могут использоваться в последующих поисках информации для обеспечения полноты списка публикаций организации.

1. Постановление Правительства Российской Федерации № 312 от 8 апреля 2009 г. «Об оценке результативности деятельности научных организаций, выполняющих научно-исследовательские, опытно-конструкторские и технологические работы гражданского назначения».

2. Оценка результативности деятельности научных организаций РАН (РИД РАН) [Электронный ресурс]. Режим доступа: <http://www.issras.ru/activity/committee/doc/indicator.xls> (дата обращения 06.10.2011).

3. Neuhaus C. Data sources for performing citation analysis: an overview / C. Neuhaus, H.-D. Daniel // Journal of Documentation. – 2008. – V. 64. – №2. – P. 193 – 210.

4. Бузник В.М. Наукометрические показатели химических институтов Новосибирского научного центра СО РАН в 1995 – 2003 гг. по данным Science Citation Index и Chemical Abstracts / В.М. Бузник, И.В. Зибарева, Н.И. Сорокин, Л.С. Филатова // Химия в интересах устойчивого развития. – 2005. – Т. 13. – № 5. – С. 677 – 692.
5. Зибарева И.В. Ранжирование институтов Российской академии наук с помощью Российского индекса научного цитирования на примере институтов химического профиля / И.В. Зибарева, В.Н. Пармон // Вестн. РАН. – 2012. – Т. 82. – № 9. – С. 779 – 789.
6. Зибарева И.В. Российские научные публикации 2005 – 2009 гг. в зарубежных базах данных Science Citation Index, Scopus и Chemical Abstracts / И.В. Зибарева, Н.С. Солошенко // Научно-техническая информация. Сер. 1: Организация и методика информационной работы. – 2011. – № 9. – С. 18 – 29.
7. Зибарева И.В. Опыт использования библиографических БД для наукометрических исследований российской химической науки / И.В. Зибарева, Н.В. Круковская // Образовательные технологии и общество. – 2007. – № 1. – С. 297 – 303.
8. Stefaniak B. Use of bibliometric data bases for scientometric studies / B. Stefaniak // Scientometrics. – 1987. – V. 12. – № 1 – 3. – P. 149 – 161.
9. Мазов Н.А. Проблемы идентификации метаданных в наукометрических базах данных Web of Knowledge, Scopus и РИНЦ на примере профилей авторов [Электронный ресурс] / Н.А. Мазов, В.Н. Гуреев // Библиотеки и информационные ресурсы в современном мире науки, культуры, образования и бизнеса: материалы конф. – М.: ГПНТБ России, 2012. Электрон. опт. диск (CD-ROM).