

Проблемы идентификации данных в индексах цитирования и пути их решения

А. В. Скалабан,
директор,

И. В. Юрик,
заместитель директора;

Научная библиотека Белорусского
национального технического университета

В настоящее время в Республике Беларусь для анализа и оценки результатов научной деятельности наряду с традиционными методами – экспертными оценками (отзывами, рецензиями, мнениями экспертов, рецензентов, научных редакторов и др.) все активнее используются библиометрические индикаторы, связанные с учетом числа публикаций и их цитируемостью. Библиометрические данные используются в научной политике и управлении финансированием науки, в национальных программах развития науки и национальных системах оценки результатов научных исследований, при составлении международных и национальных рейтингов вузов и др.

Данные библиометрического анализа позволяют сделать значимые (обоснованные) сравнения научных организаций, дисциплин и стран, но лишь в том случае, если они используются корректно. О необходимости грамотной и аккуратной трактовки наукометрических индикаторов, о проблеме «злоупотреблений цитированием» еще в 60-е гг. XX в. писал один из основоположников библиометрии и наукометрии Ю. Гарфилд. Указанная проблема не теряет своей актуальности и сегодня. «С одной стороны, это связано с легкостью манипулирования массмедиа этими показателями, с другой – с безответственностью малокомпетентных “специалистов” и администраторов науки, с легкостью берущихся за выполнение такой оценки» [1].

Именно в ответ на «игры с цитированием» в 2010 г. появился манифест альметрики, предлагающий способы оценки публикаций, отличные от анализа цитирований, а в декабре 2012 г. вышла Декларация по оценке науки – Declaration on Research Assessment (DORA), принятая Американским обществом клеточной биологии (The American Society for Cell Biology (ASCB)). К настоящему времени под ней уже поставили свои подписи 10 668 ученых и 467 организаций [2]. И наконец, в апреле 2015 г. представителями наукометрического сообщества был разработан Лейденский манифест по наукометрии, провозглашающий 10 принципов «в оценке исследовательской деятельности, основанной на наукометрии, с тем, чтобы ученые могли бы проверять

тех, кто их оценивает, а “оценщики” могли бы проверять свои индикаторы» [3]. Знакомство с данными документами будет полезно всем представителям отечественного научного сообщества.

Сегодня для библиометрического анализа используются наукометрические базы данных (индексы цитирования) и веб-сервисы, позволяющие в автоматизированном режиме проводить комплексную обработку данных. Наиболее известными среди них являются Web of Science (далее – WoS) (производитель Thomson Reuters), Scopus (Elsevier), Google Scholar (Google), Российский индекс научного цитирования (далее – РИНЦ) (Научная электронная библиотека eLIBRARY.RU), а также веб-инструменты InCites (на основе WoS) и SciVal (на основе Scopus).

Индексы цитирования позволяют осуществлять автоматизированный сбор и анализ библиометрических показателей и автоматический расчет соответствующих индексов. «Востребованность их объясняется, с одной стороны, желанием самих ученых вывести “магическую цифирь”, сводящую в одном или нескольких числах все таланты и заслуги соратников по цеху, а с другой – стремлением администраторов от науки получить объективные показатели ценности результатов деятельности ученого, научной или образовательной организации или научного направления. Но не так все просто» [4].

За кажущейся легкостью расчета показателей научной результативности с помощью баз данных научного цитирования скрываются многочисленные подводные камни. По мнению специалистов в области наукометрии, «...по таким стандартным показателям, как среднее цитирование публикаций или индекс Хирша и его вариации, проводить сравнение результативности научной деятельности научных коллективов или отдельных авторов, работающих в разных научных направлениях, представляется весьма затруднительным» [1]. Рассмотрим некоторые проблемы идентификации данных научных публикаций в индексах цитирования и определим пути их решения.

Проблемы идентификации данных в индексах цитирования и пути их решения. Поскольку различные источники информации о публикациях для оценки научной деятельности формируются по различным принципам, необходимо указывать, на основе какого инструмента производится библиометрический анализ. Следует иметь в виду, что показатели, рассчитанные с использованием одной и той же методики, в различных источниках будут иметь разные значения, поскольку расчет в каждом конкретном ресурсе осуществляется на основе информации (источников), содержащейся именно в данном ресурсе (таблица 1).

Сравнительный анализ наукометрических инструментов

Инструменты	Массив индексируемых данных	Вид доступа	Глубина архива	Отличительная характеристика	Способы анализа и оценки
Международный индекс цитирования Web of Science (WoS) (Thomson Reuters, 2002)	Свыше 12 тыс. – наиболее значимых журналов в каждой области знания, трудов конференций, незначительное количество книг	Подписная база данных, для доступа к которой необходима платная подписка той организации, к которой вы относитесь	С 1900 г.	Ретроспектива баз данных в конкретной организации зависит от суммы оплаченных средств, что напрямую влияет на отображаемые показатели	Учет числа публикаций, цитирований
Международный индекс цитирования Scopus (Elsevier, 2004)	Более 20 тыс. научных и отраслевых журналов, труды конференций, книжные серии, монографии	Подписная база данных; для доступа необходима платная подписка той организации, к которой вы относитесь	С 1823 г.	Доступ ко всему архиву; показатели рассчитываются, учитывая журналы с 1996 г., для некоторых журналов крупных издательств – с 1970 г.	Учет числа публикаций, цитирований
Российский индекс научного цитирования (Научная электронная библиотека eLIBRARY.RU, 2005)	Научные публикации России и стран СНГ	Доступ бесплатный	С 2000 г.	Наличие полных текстов	Учет числа публикаций, цитирований
Google Scholar (Google, 2004)	Научно-образовательные ресурсы, доступные в Интернете в репозиториях, журналах открытого доступа, списках процитированной и найденной в статьях литературы и др.	Доступ бесплатный	Нет данных	Наличие полных текстов	Учет числа публикаций, цитирований

Так, WoS и Scopus индексируют наиболее значимые в мире журналы в каждой области знаний, труды конференций и, с недавнего времени, книги; РИНЦ – российскую научную периодику и ряд научных журналов стран СНГ; Google Scholar – данные из открытых источников информации (репозитории, журналы открытого доступа и др.). Отличается и глубина архивов (с 1900 г. – в WoS, с 1823 г. – в Scopus и приблизительно с 2000-го – в РИНЦ).

Основное отличие WoS и Scopus заключается в том, что в Scopus при подписке доступ дается ко всему архиву, тогда как ретроспектива баз данных WoS зависит от суммы оплаченных средств, что напрямую влияет на отображаемые показатели.

Международные индексы цитирования WoS и Scopus – это библиографические/реферативные базы данных, в которых собирается и обрабатывается полная библиографическая информация о научных статьях, аннотации/рефераты и приставочные списки цитируемой в статьях литературы. Полных текстов статей в указанных базах данных нет. В качестве отличительной характеристики РИНЦ и Google Scholar можно привести наличие полных текстов.

В Беларуси лицензии на доступ к ним имеют: WoS – Центральная научная библиотека имени Я. Ко-

ласа НАН Беларуси; Scopus – Центральная научная библиотека имени Я. Коласа НАН Беларуси, БелСХБ, Научная библиотека БНТУ.

Существенной проблемой, затрудняющей корректный подсчет публикаций и цитирований отдельного автора или научного коллектива в WoS и Scopus, является отсутствие однозначной идентификации объектов. Данные могут быть потеряны:

- при множественности написания фамилии автора (использование различных схем транслитерации; написание фамилии автора с различным количеством инициалов, смена женских фамилий после замужества, наличие опечаток и пр. [5]);

- при отсутствии в статье данных о месте работы автора либо при наличии нескольких вариантов наименования организации, к которой относится автор (в поле «affiliation» название организации, ее адрес переносятся автоматически).

Как следствие, возникают погрешности в количестве публикаций, цитирований и, соответственно, значениях индекса Хирша и др., некорректные итоговые сведения о научных достижениях автора/научного коллектива/региона.

Проблема идентификации данных существует также в РИНЦ и Google Scholar. Они несколько отличаются от зарубежных баз данных.

Во многом эта проблема решается с помощью уникальных идентификаторов, существующих в виде отдельных инструментов или встроенных в базы данных.

Имеется некоторое количество таких систем. Например, Thomson Reuters создан как специальный инструмент объединения и корректировки данных автора, связанный с WoS, – авторский профиль ResearcherID. В Scopus авторский профиль Author ID формируется автоматически для авторов, опубликовавших хотя бы одну статью. ORCID (Open Researcher and Contributor ID) – самостоятельная система – реестр уникальных идентификаторов ученых, независимый от научных дисциплин и национальных границ, синхронизируется с другими системами идентификации, например, с ResearcherID и Author ID. В РИНЦ создана специальная аналитическая надстройка SCIENCE INDEX, в Google Scholar – персональный профиль автора.

Однозначно определить все свои публикации может только сам автор, поэтому ученым настоятельно рекомендуется регистрироваться в системах идентификации и работать со своими персональными профилями. Кроме этого, регистрация авторов в системах идентификации значительно улучшает видимость их работ в мировом научно-информационном пространстве.

ResearcherID – это свободное, общедоступное интерактивное пространство для создания индивидуального номера и персонального профиля. Зарегистрироваться в системе можно на одноименном сайте <http://www.researcherid.com>. Введенные сведения не подвергаются модерации. Не требуется регистрации в системе в том случае, если автор относится к организации, имеющей лицензию на доступ к WoS, и ранее был зарегистрирован в ней. Преимущества регистрации в ResearcherID:

- создание полного списка своих публикаций тремя способами: путем поиска в WoS и сохранения результата; посредством загрузки из библиографического менеджера EndNote; импорт внешних файлов в формате RIS, содержащих статьи, отсутствующие в WoS;
- исключение из своего списка публикаций однофамильцев;

- поиск в WoS по номеру профиля ResearcherID и нахождение всех индексированных публикаций автора даже при наличии ошибок и разных вариантов написания его фамилии;

- определение наукометрических показателей (анализ цитирований автора будет проводиться только по статьям, проиндексированным в WoS);

- поиск партнеров и др.

Потенциал ResearcherID пока не оценен в должной мере белорусскими исследователями, о чем свидетельствуют статистические данные, полученные нами в результате анализа этой системы. Так, по состоянию на 01.12.2014 в ResearcherID зарегистрировано 155 пользователей из Беларуси (для сравнения: Россия представлена 32 330 пользователями, Украина – 1130, Казахстан – 1869, Польша – 7690, США – 66 614). Статистика по научным организациям представлена в таблице 2.

Уникальный идентификатор автора **Scopus Author ID** присваивается при опубликовании даже одной статьи, индексированной Scopus. Он выполняет функцию независимого от статьи источника информации, аккумулируя информацию о местах работы автора, количестве его публикаций и их цитируемости, годах публикационной активности, области исследований, соавторах, индексе Хирша, списке использованных в работах литературных источников и пр. Прямого доступа к редактированию своего профиля автор не имеет, вместе с тем уточнить свою библиографическую информацию он может. При наличии доступа к базе данных через свой профиль в Scopus, а при его отсутствии – через ссылку <http://www.scopusfeedback.com/> автор может самостоятельно:

- объединить несколько своих профилей в один;
- уточнить информацию на уровне статей, отмечая свои и исключая чужие публикации в процессе работы со списком;
- выделить все свои статьи и послать запрос по электронной почте в техническую службу Scopus.

Подобную работу может провести не только сам автор. Например, в БНТУ сотрудники Научной библиотеки осуществляли поиск и анализ множественных профилей университета и работающих в нем авторов.

Таблица 2

Статистика зарегистрированных пользователей в ResearcherID по научным организациям Беларуси

Название организации	Количество пользователей
Belarus State University	29
Belarus State Technological University, BSTU	9
B. I. Stepanov Institute of Physics	8
Belarusian National Technical University	7
Belarusian State University of Informatics and Radioelectronics	6
Brest State Technical University	5
Belarusian Medical Academy of Postgraduate Education	4

Хорошим подспорьем в этой работе стала генерируемая библиотекой база данных трудов научных работников БНТУ. В результате анализа данных в Scopus было обнаружено 27 вариантов названий БНТУ и 170 статей авторов, работающих в университете, но не связанных с его профилем. Был отправлен запрос в службу поддержки. Профили были объединены. В итоге количество статей БНТУ в Scopus увеличилось на 700.

ORCID (Open Researcher and Contributor ID) – совместная разработка нескольких издательств, университетов, научных сообществ. Адрес входа в систему – <http://orcid.org/>.

Основная цель этой системы – решить проблему идентификации ученых с одинаковыми именами и фамилиями. ORCID обеспечивает следующие функции:

- получение уникального идентификатора и ведение соответствующей учетной записи об исследовательской деятельности;
- программное обеспечение для межсистемной коммуникации.

Следует отметить, что данный номер от авторов в последнее время часто требуют издатели во время отправки рукописи в журнал. Таким образом, этот номер, отсылающий к отредактированному списку публикаций автора, является его своеобразной визитной карточкой, поскольку информация видна редактору и (иногда) рецензентам.

Учетная запись ORCID включает: имя ученого, его электронный адрес, название организации и информацию о исследовательской деятельности. ORCID учитывает необходимость контроля за распространением этих данных и предоставляет соответствующие инструменты для управления уровнем приватности данных.

ID представляет собой 16-значный числовой код, согласованный со стандартом ISO (ISO 27729). ORCID ID – это URI, поэтому отображается как адрес вида <http://orcid.org/xxxx-xxxx-xxxx-xxxx>.

Записи в ORCID хорошо синхронизируются с записями в базе данных Scopus: во-первых, можно

осуществлять перенос всех публикаций автора из Scopus в ORCID, указав номер авторского профиля Scopus Author ID; во-вторых, если работы автора в Scopus присоединены к нескольким профилям, то при включении их в ORCID служба технической поддержки Scopus отредактирует профиль автора в базе данных.

Согласно принципам организации ORCID, участником этого проекта может стать любая организация научного сообщества. Для индивидуальных исследователей доступ к реестру ORCID предоставляется бесплатно. По состоянию на 1 декабря 2014 г. в системе был зарегистрирован 1 млн пользователей и содержалось 5 359 470 работ.

SCIENCE INDEX – специальная аналитическая надстройка в РИНЦ, с помощью которой автором производится редактирование своих личных данных, анализируется публикационная активность и цитируемость. По состоянию на 1 декабря 2014 г. всего в РИНЦ зарегистрирован 276 491 автор: 266 632 – из России, 3671 – из Украины, 3397 – из Беларуси, 1096 – из Казахстана. При этом поиск по автору и стране (Беларусь) в РИНЦ дает результат 9037.

В отличие от WoS и Scopus, в РИНЦ нет проблемы с авторским множеством, однако есть другая – часть статей может быть не привязана к профилю автора. Анализ библиометрических показателей 10 самых цитируемых в РИНЦ белорусских авторов показал, что из 10 только 6, отмеченных символом «*», зарегистрированы в SCIENCE INDEX. Вполне вероятно, что сведения о публикациях и цитировании авторов, не прошедших регистрацию в РИНЦ, не полные.

Пищу для размышления сотрудникам и администрации УВО Беларуси дает и анализ количества зарегистрированных в SCIENCE INDEX авторов по вузам (таблица 3).

Улучшению представления научных результатов белорусских научных и образовательных организаций в РИНЦ могли бы способствовать следующие тактические шаги:

Таблица 3

Топ-10 учреждений высшего образования Беларуси по числу зарегистрированных в SCIENCE INDEX авторов

Название организации	Количество зарегистрированных авторов
Белорусский государственный медицинский университет	635
Гродненский государственный медицинский университет	443
Витебский государственный университет имени П. М. Машерова	294
Белорусская медицинская академия последипломного образования	207
Витебский государственный медицинский университет	199
Белорусский государственный университет	196
Витебская государственная академия ветеринарной медицины	131
Белорусский государственный технологический университет	64
Гродненский государственный университет имени Янки Купалы	57
Белорусский национальный технический университет	48

1) обязательная регистрация каждого публикуемого сотрудника в SCIENCE INDEX;

2) активная работа авторов, прошедших регистрацию, со своими списками – выявление «непривязанных» статей, поиск своих статей в пристатейной литературе и пр.;

3) расширение ассортимента белорусских научных журналов на платформе eLibrary.ru (как по названию, так и по глубине архивов).

С точки зрения реальных показателей цитируемости для авторов из белорусских научных и образовательных учреждений Google Scholar представляет больший интерес, чем WoS или Scopus, поскольку система индексирует массив публикаций, размещенных в институциональных репозиториях (на 01.06.2015 в OpenDOAR зарегистрировано 15 репозиториях УВО Беларуси, объем документов – более 146 тыс.), а также другие источники на русском языке (научные журналы открытого доступа, материалы конференций, препринты и др.). Вместе с тем следует отметить, что эта система до сих пор находится в стадии бета-версии, поэтому практически не может использоваться для официальных отчетов (лишь иногда ее включают как факультативную).

С целью получения данных о публикациях для оценки научной деятельности с помощью Google Scholar авторам необходимо создавать личный профиль в этой системе, который позволит также сохранять результаты поиска, следить за цитированием интересующих работ, отслеживать отдельные работы или творчество отдельных авторов.

Адрес системы – <http://scholar.google.ru/>. Сначала необходимо создать новый аккаунт в Google или войти в уже существующий. Заполнение формы регистрации в Google Scholar осуществляется в три шага:

- первый шаг – заполнение полей – имя, место работы, электронная почта для подтверждения, области интересов;
- второй шаг – работа со статьями – необходимо указать именно свои статьи в списке, автоматически предложенном системой, а также добавлять публикации самостоятельно;
- третий шаг – заполнение страницы профиля – установка параметров обновления профиля, загрузка фотографии и др.

Особо следует отметить функцию экспорта списка своих работ в другие системы идентификации авторов, как, например, ResearcherID и ORCID. В связи с этим авторам из белорусских научных и образовательных учреждений, чьи публикации размещены в институциональных репозиториях, мы рекомендуем сначала создавать персональные профили в Google Scholar, добавляя в них список собственных статей, а затем экспортировать их в другие системы идентификации авторов.

Преимущество Google Scholar как наукометрического инструмента заключается в максимальной

широте представленных материалов. К недостаткам можно отнести следующее: в нее не попадают сведения из закрытых источников и публикации, отсутствующие в электронном виде (или упоминания о таковых); нет гарантии в достоверности данных, добавленных в свой личный профиль самим автором.

Таким образом, обязательным условием при измерении и оценке научной деятельности библиометрическими методами являются корректность и полнота исходных данных. Необходимо учитывать существующие в наукометрических инструментах проблемы с идентификацией авторов и организаций. Игнорирование может привести к занижению реальных показателей публикационной активности и цитируемости как отдельных авторов, так и научных коллективов в целом.

Избегать ошибок и неточностей при индексировании публикаций научных и образовательных организаций, возможно, поможет строгая регламентация указания названия организации как места работы на административном уровне.

Регистрация в системах идентификации и регулярная работа автора со своим профилем позволят не только контролировать корректность представленных данных в различных наукометрических инструментах, но и улучшат видимость научных работ.

Особенностью современной научной коммуникации является формирование электронного научного пространства, а ее эффективность в большей мере зависит от активного освоения и использования учеными разных форм электронной научной коммуникации. В связи с появлением альтернативных цитированию способов оценки научных результатов с помощью альметрик научным организациям и их сотрудникам следует принимать во внимание важность качественного присутствия: в блогосфере, социальных сетях, в системах идентификации авторов, библиографических менеджерах.

Список литературы

1. Руководство по наукометрии: индикаторы развития науки и технологии: монография / М. А. Акоев [и др.]; под ред. А. М. Акоева. – Екатеринбург: Изд-во Урал. ун-та, 2014. – 250 с.
2. Обновление инструкции для авторов научных журналов: метод. материалы / пер. с англ. А. В. Бажанова; под ред. А. Ю. Гаспаряна, О. В. Кирилловой. – СПб.: Сев.-Зап. ин-т упр. – фил. РАНХиГС, 2015. – 48 с.
3. Лейденский манифест для наукометрии [Электронный ресурс] / Ин-т всеобщ. истории РАН; пер. А. А. Исэрова. – Режим доступа: <http://www.igh.ru/about/news/1053/>.
4. Новиков, Д. А. Померяемся «Хиршами»? (Размышления о наукометрии) / Д. А. Новиков // Высшее образование в России. – 2015. – № 2. – С. 5–13.
5. Мазов, Н. А. Роль единых идентификаторов в информационно-библиографических системах / Н. А. Мазов, В. Н. Гуреев // Научно-техническая информация. Серия 1: Организация и методика информационной работы. – 2014. – № 9. – С. 32–37.