

## **Взаимодействие баз данных научного цитирования и корпоративных репозиториев научных публикаций**

*Ковязина Елена Васильевна, научный сотрудник, Институт вычислительного моделирования Сибирского отделения Российской академии наук, [elena@icm.krasn.ru](mailto:elena@icm.krasn.ru)*

*Ранжирование научных и образовательных организаций обострило проблемы учета публикационной и иной научной активности сотрудников. Базы данных, с помощью которых ведется учет научной деятельности, перестали быть только библиографическими. Они содержат расширенный набор данных о документах самого разного вида – публикациях, отчетах, патентах, конференциях и т.д. Доклад посвящен проблемам ведения таких баз данных: устранению дублирования ввода, пополнение записей данными, необходимыми для статистических расчетов, организационным проблемам и т.д.*

Политические и организационные процессы, связанные с ранжированием научных и образовательных организаций, породили острую необходимость в учете публикационной активности их сотрудников. Резко возросла роль баз данных научного цитирования. Организации прикладывают максимальные усилия для обеспечения полноты и адекватности представленных в них данных о публикациях, используя для этого как встроенные сервисы, так и личные контакты. В результате прилагаемых усилий данные организаций приобретают законченный и корректный вид. С появлением в РИНЦ облачного сервиса Science Index организации получили инструмент для создания собственной удаленной базы трудов сотрудников организации. Однако, в силу ряда причин, таких как платность данного сервиса и ограничения контроля над данными, ни одна из организаций не может позволить себе ограничиться для оценки публикационной активности только отчуждаемыми данными в РИНЦ. Поэтому внутри организаций огромное внимание уделяется развитию корпоративных репозиториев научных публикаций, совершенствуется их технологическая и организационная составляющие.

По сравнению с данными, представленными в индексах научного цитирования корпоративный репозиторий научных публикаций представляет собой более сложное образование, включающее базы метаданных документов и собственно собрание полных текстов публикаций в электронном виде. Для эффективного функционирования репозитория необходимо оформить его организационно – подготовить пакет документов, предназначенный для обеспечения его пополняемости и легитимности доступа к документам в свете закона об авторском праве. Документы в репозитории в терминах закона понимаются как служебные произведения, имущественные права на которые принадлежат работодателю. Для оформления репозитория необходимо принять Положение о репозитории, регламентирующее порядок поступления и размещения документов, правила доступа к ним и вопросы взаимодействия с авторами. Технические службы организации, библиотека или информационный отдел, берут на себя следующие функции:

1. Организация метаданных документов как базы данных, записи которой включают расширенное библиографическое описание и данные по цитированию.
2. Организация хранилища полных текстов в виде архива научных публикаций организации (репозитория).
3. Организация дифференцированного онлайн-доступа к документам в соответствии с Положением о репозитории.
4. Сбор документов и обслуживание хранилища полных текстов в соответствии с Положением о репозитории.

Работа в течение ряда лет по организации и функционированию репозитория выявила ряд проблемных мест, связанных как с организационной, так и с технологической составляющей. Организационная часть работы осложняется правовой неопределенностью в отношении доступа к электронным документам. Издательства, заключая договора с авторами, как правило, не интересуются, принадлежит ли автору имущественное право на документ, вследствие чего договор с издательством включает целый ряд ограничений доступа. В свою очередь, организация, заинтересованная в появлении публикации, не предъявляет каких-либо претензий к издательству, принимая ограничительные условия. Еще сложнее обстоит дело с монографией - служебным произведением. Договор заключается с институтом только для монографий, напечатанных при поддержке РФФИ. Монографии, вышедшие в сторонних издательствах из-за хронического недофинансирования, вообще бывают потеряны для организации в правовом смысле.

Наиболее трудоемкой частью работы является формирование описательных метаданных документов. Метаданные, предназначенные для оценки публикационной активности, имеют отличия от тех, к которым мы привыкли в электронных каталогах библиотек. Структура метаданных дополняется следующими полями:

- а) полем идентификаторов публикации в индексах научного цитирования (WoS, Scopus, РИНЦ и т.д.);
- б) полем текущего импакт-фактора источника публикации;
- в) полем списка библиографии;
- г) полем разночтения фамилии автора;
- д) показателем текущего цитирования.

Как видно из перечня полей, работа по заполнению их содержимого существенно повышает трудоемкость работы. Если учесть, к тому же, множественность баз данных трудов сотрудников – это и библиотечная база данных, база ученого секретаря, распределенные корпоративные базы данных, личные базы сотрудников и т.д. – проблема обмена данными между различными базами стоит очень остро, что побуждает специалистов искать пути ее решения [1-3].

Международным сообществом разработаны и используются базовые стандарты для обмена библиографическими данными между различными системами. Стандарты определяют содержание полей и правила их заполнения. Если система позволяет выгрузить содержимое полей в одном из таких стандартов, то перенос данных в другую базу уже не является неразрешимой задачей. На-

бор сервисов Web of Science и Scopus позволяет заимствовать данные этих индексов и интегрировать их в систему автоматизации библиотеки. Технология заимствования записей из WoS и Scopus широко применяется многими библиотеками научных организаций. Последовательность заимствования:

- а) поиск с помощью текущих или хранимых запросов – по автору, месту работы, выделение нужных публикаций;
- б) экспорт описаний в табличном или текстовом виде;
- в) табличное преобразование данных в форму пригодную для интеграции в БД репозитория.

Определенные проблемы возникают при слиянии данных Web of Science и Scopus. В конечном итоге процесс слияния превращается в трехэтапную процедуру:

- а) слияние по DOI;
- б) слияние по свертке САБ;
- в) проверка вручную ошибочных данных.

Кроме возможности экспорта описаний, международные индексы научного цитирования предоставляют подписчикам дополнительные сервисы, такие как:

- а) оповещение по электронной почте о цитировании заказанной статьи;
- б) оповещение по электронной почте о выходе статьи заказанного автора;
- в) хранимые запросы, которые можно выполнять по расписанию;
- г) динамическое извлечение данных о цитировании с помощью сервиса ARM.

Набор указанных сервисов существенно облегчает работу ведения базы метаданных репозитория и позволяет отслеживать показатели публикуемости организации в зарубежных источниках.

Хотя политика Министерства образования и науки направлена, главным образом, на учет показателей публикационной активности в зарубежных изданиях, российские публикации все еще играют не меньшую роль в рейтинговых показателях организации. Данные для оценки получаются из Российского индекса научного цитирования. (РИНЦ). К сожалению, разработчики РИНЦ не развивают вышеуказанные сервисы для российских пользователей. Функция экспорта в табличный вид отсутствует, а заимствование с помощью открытых API требует длительных личных переговоров. Сервисы хранимых запросов и оповещение также отсутствуют. Однако, справедливости ради следует отметить, что разработчики РИНЦ сосредоточили усилия на несколько ином направлении. Уже примерно год работает сервис Science Index, который постоянно совершенствуется и позволяет многим организациям существенно повысить свои показатели цитирования путем ввода данных о публикациях, отсутствующих в РИНЦ. Существенным плюсом сервиса является возможность добавления не только статей в научных журналах, но и монографий, сборников статей, материалов конференций, патентов, отчетов и других типов научных публикаций, не обрабатываемых в РИНЦ, а также возможность размещения в РИНЦ ссылок на полные тексты публикаций. Существует возможность добавления публикаций различными путями: можно заполнить все поля вручную, использовать библиографическую ссылку на публикацию в произвольном формате,

или получить библиографическое описание из системы CrossRef, указав код DOI. Отрадно заметить, что появилась и выгрузка записей, но почему-то только в формате xml. Для облегчения и исключения дублирования работы не помешала бы и функция импорта данных в каком-либо из стандартных форматов. Все добавленные и измененные описания публикаций, внесенные представителем организации, проходят контроль администраторами системы, что улучшает качество вводимых данных. Удобный интерфейс позволяет просматривать уже добавленные и исправленные публикации, а также комментарии администраторов к публикациям, отправленным на доработку. Однако проверка данных ведется очень долго, а иногда и вообще не ведется без дополнительного телефонного звонка. Некоторые публикации не обработаны администраторами уже в течение пяти месяцев.

Все вышесказанное имеет отношение к обслуживанию публикаций и иных документов. Однако, ведение отчетных данных предполагает также учет конференций, проектов, грантов и иной научной деятельности. Для этих целей не очень годятся стандартные библиографические описания. Нужна доработка входных форм и форматов вывода.

Отметим в итоге, что ведение и обслуживание репозитория перестало быть обслуживанием только библиографических данных. Записи пополняются наукометрическими данными, остро стоит вопрос пополнения базы данных иными отчетными данными – сведениями о конференциях, грантах и проектах и т.п. База данных превращается в средство оценки научной деятельности организации и извлечения отчетных данных. Для описания таких данных не существует стандартов, и каждая организация ведет их по-своему. Если проблемы, связанные с вводом публикаций, известны и имеют решения, пусть и не всегда окончательные, ввод иных отчетных данных ждет своего решения.

### Литература

1. Мазов, Н.А. Web of Science и Scopus: новые технологические решения для наукометрических исследований персоны (организации) [Электронный ресурс] / Н.А. Мазов // Система обеспечения российских организаций научно-технической информацией в электронном виде. Итоги и перспективы проекта МОН: материалы конференции. – СПб., 2012. - URL: <http://conf.neicon.ru/index.php/science/mon2012/paper/view/29/26>
2. Мазов, Н.А. Идентификация библиографических метаданных научных публикаций в различных базах данных: проблемы и решения [Электронный ресурс] / Н.А. Мазов // Научное издание международного уровня: проблемы и решения при подготовке и включении в индексы цитирования и реферативные базы данных: материалы международной научно-практической конференции. – М.: ВИНТИ РАН, 2012. - URL: <http://www.viniti.ru/download/russian/conf/MAT/16-10.ppt>
3. Ковязина, Е.В. Использование баз данных научного цитирования в формировании корпоративных информационных ресурсов // 1-я международная конференция НЭИКОН о российских и международных научных ресурсах: материалы конференции. – Будва, 2013. – URL: <http://conf.neicon.ru/index.php/science/overseas2013/paper/view/93/147>