

**Интеграция базы данных публикаций организации  
с индексами научного цитирования: реализация средствами САБ ИРБИС64**

**Integrating database of organization publications  
with scientific citation indices with IRBIS64 ALIS functionality**

*С. Р. Баженов*

*Государственная публичная научно-техническая библиотека СО РАН,  
Новосибирск, Россия*

*М. В. Данилин, О. А. Rogoznikova*

*Красноярский научный центр СО РАН, Центральная научная библиотека,  
Красноярск, Россия*

*Sergey Bazhenov*

*State Public Library for Science and Technology,  
Russian Academy of Sciences Siberian Branch,  
Novosibirsk, Russia*

*Mikhail Danilin and Olga Rogoznikova*

*Central Scientific Library, Krasnoyarsk Research Center,  
Russian Academy of Sciences Siberian Branch,  
Krasnoyarsk, Russia*

В работе описывается разработка и использование системы автоматического получения количества ссылок цитирований на статьи сотрудников организации с использованием зарубежных и отечественных ресурсов Web of Science, Scopus и РИНЦ.

The issues of the development and use of system automatic count of citation links to staff publication using national and foreign resources: Web of Science, Scopus, Russian Science Citation Index, are discussed.

База данных публикаций сотрудников всегда была неременным атрибутом научной организации. В последние годы, несмотря на появление масштабных государственных проектов в области учета и качественной оценки публикаций российских авторов, работа по учету публикаций на уровне отдельной организации не утратила своей актуальности по следующим причинам:

1. Любые наукометрические измерения начинаются на уровне конкретных организаций, именно эти данные лежат в основе более общих, – по ведомственной принадлежности, географическим признакам, предметным областям. Именно конкретные организации на основе информации о публикациях своих сотрудников могут упорядочить информацию о своих авторах, научных группах и организации в целом во внешних реферирующих базах данных [1].
2. В научных организациях принята система материального стимулирования сотрудников, основанная на оценке результативности их научной деятельности. Количество публикаций разного вида, их цитируемость входят в группу параметров, по которым происходит оценка, а правильно организованная база данных может служить инструментом для расчета этих параметров.

Однако, к базе данных, которая могла бы служить основой для наукометрических измерений, предъявляются серьезные требования: полнота учета публикаций, содержательность записей и актуальность изменяющихся данных. И если решить проблему полноты учета публикаций и качества записей можно на уровне отдельной организации, то поддерживать актуальность информации о количестве цитирований публикации в Web of Science без интеграции с этой базой данных невозможно.

Тема интеграции базы данных публикаций сотрудников с Web of Science, Scopus и РИНЦ отражена в работах Н. А. Мазова (ИНГГ СО РАН) и В. Н. Гуреева (ГНЦ ВБ «Вектор») [2,3]. Для интеграции базы данных публикаций сотрудников ИНГГ СО РАН, ведущейся в среде CDS/ISIS, разработан программно-технологический комплекс, позволяющий снабжать метаданные соответствующими ссылками, связь устанавливается посредством идентификаторов записей в индексах цитирования, отраженных также в записях базы данных публикаций.

Собственная программная система, позволяющая получать данные из индексов цитирования и других баз данных, а также формирующая статистические отчеты по библиометрическим показателям, представлена сотрудниками ИПИ РАН [4].

В КНЦ СО РАН была поставлена цель обеспечить интеграцию с Web of Science, Scopus и РИНЦ базы данных публикаций, ведущейся в САБ ИРБИС64. Для связи записей базы данных с записями индексов цитирования предполагалось, подобно ИНГГ СО РАН, использовать идентификаторы записей в индексах цитирования (UT, Scopus-ID, ID-РИНЦ). В стандартных рабочих листах САБ ИРБИС64 отсутствовало поле, пригодное для их фиксации, было принято решение о добавлении поля *321 Примечания об отдельно изданных указателях / рефератах / ссылках, отражающих каталогизируемый документ* из списка полей RUSMARC [5].

Выбор индексов цитирования был обусловлен тем, что в соответствии с Приказом Министерства образования и науки Российской Федерации от 5 марта 2014 года №162 [6] обязательными элементами сведений о результатах деятельности научных организаций являются число и совокупная цитируемость публикаций, индексируемых Web of Science. Показатели двух оставшихся чаще всего фигурируют в отчетах научных организаций КНЦ в качестве факультативных.

Были выделены две задачи:

1. Обеспечение полноты и оперативности отражения в базе данных публикаций, прореферированных перечисленными выше индексами цитирования, а также качества записей.
2. Поддержание информации о цитированиях в Web of Science, Scopus и РИНЦ публикаций, учтенных в базе данных, в актуальном состоянии.

Решением первой задачи является возможность загрузки в базу данных САБ ИРБИС64 готовых записей из перечисленных выше индексов цитирования. Возможность выгрузки записей есть во всех трех индексах цитирования, в случае Web of Science и Scopus требуется наличие подписки на эти ресурсы, а в случае РИНЦ необходима подписка на информационно-аналитическую систему SCIENCE INDEX. Из Web of Science и Scopus можно выгружать любые записи, существует ограничение только на разовую порцию выгрузки. SCIENCE INDEX допускает скачивание только тех записей, которые приписаны к организации, имеющей подписку. Из Web of Science и Scopus возможен экспорт записей в текстовом формате, есть возможность определения набора элементов выгружаемых записей. В SCIENCE INDEX выгрузка в формате XML полного набора всех элементов записи. Web of Science и SCIENCE INDEX предоставляют при выгрузке идентификаторы своих записей, в записях, экспортированных из Scopus, идентификаторов нет.

Для загрузки в базу данных САБ ИРБИС64 данных, выгруженных в текстовом формате, используется конвертор, созданный сотрудниками ГПНТБ СО РАН.

Конвертор из XML был создан в ЦНБ КНЦ СО РАН.

После первичного отбора и загрузки записей за весь доступный период для получения информации о появлении новых публикаций в Web of Science и Scopus были созданы постоянно действующие запросы (Search Alert), информирование о новых поступлениях осуществляется еженедельно, с этой же периодичностью происходит пополнение базы данных публикаций САБ ИРБИС64.

Экспорт информации из индексов цитирования, конвертирование и загрузка в базу данных выполняются библиотекарем.

Предпосылкой решения второй задачи для Web of Science и Scopus явилось существование большого спектра Web-сервисов для подписчиков этих ресурсов.

Thomson Reuters Article Match Retrieval Service [7] позволяет организациям, имеющим лицензионный доступ, получать по запросу актуальные значения следующих показателей: количество цитирований, список цитирующих публикаций, список связанных публикаций, импакт-фактор

журнала. Правила составления запроса строго регламентированы, определен список элементов записи, которые могут быть использованы в запросе для идентификации публикации и получения ответа. Наиболее целесообразным представляется использование в запросе UT- идентификатора записи в индексе цитирования, который ранее был помещен в поле 321 соответствующей записи в базе данных публикаций. Сервис предоставляется подписчикам Web of Science бесплатно. Есть ограничение на количество запросов в единицу времени, однако, на практике мы ни разу с ним не сталкивались.

Abstract Citations Count API (Elsevier Content APIs) [8] позволяет подписчикам Scopus получать актуальные сведения о количестве цитирований публикации. Поиск на основе идентификаторов, в том числе Scopus-ID. Ответ возможен в виде графического изображения или метаданных (JSON, XML). Есть ограничение на количество запросов в период времени.

Получение актуальных данных о количестве цитирований публикаций в Web of Science и Scopus полностью автоматизировано.

Для поддержания информации о цитировании в актуальном состоянии используются пакетные задания, которые выполняются при помощи службы «Планировщик задач» Microsoft Windows.

Каждое пакетное задание последовательно выполняет три задачи:

1. Выгрузка идентификаторов из базы ИРБИС64
2. Получение данных о количестве цитирований из базы Scopus или Web of Science
3. Обновление данных в базе ИРБИС64.

Обмен данными между задачами осуществляется посредством текстовых файлов специального формата.

Так в текстовый файл для каждой записи базы данных публикаций сотрудников выгружается только поле 321, содержащее только идентификатор системы (Scopus или Web of Science) и идентификатор записи в соответствующей базе данных. Далее для получения данных о количестве цитировании по базам Scopus и Web of Science используется специально разработанное отдельное приложение Windows. Основными задачами приложения является чтение входного файла с идентификаторами, выполнение запросов к API, формирование выходного файла, содержащего также поле 321, дополненное информацией о количестве цитирований данной статьи. Затем средствами импорта со слиянием системы ИРБИС64 эта информация загружается в БД публикаций сотрудников.

Достоинством такой системы является то, что обновление количества ссылок на статью происходит автоматически и с любым заданным временным интервалом. В данный момент этот интервал у нас равен суткам, т.е. задание запускается каждую ночь. Время работы этого пакетного задания составляет 1.5-2 часа на базе из 6500-7000 записей. В дальнейшем планируется увеличить скорость выполнения запросов за счет использования многопоточного приложения.

В разработанной системе автоматически получаем только общее количество ссылок на статью. Однако следующим шагом является доработка системы для получения данных и о количестве цитирований по годам.

К сожалению, с базой данных РИНЦ все сложнее, т.к. использование API РИНЦ затруднено, из-за ограничений на количество выполняемых запросов. Однако в настоящий момент проходит этап тестирования по получению информации и из РИНЦ.

В результате разработки этой системы получена возможность автоматического получения актуальной информации о количестве ссылок на публикации всех сотрудников большой научной организации.

### Литература:

1. Мазов Н. А. Новые методы формирования публикационного профиля научной организации в сети науки / Н. А. Мазов, В. Н. Гуреев // Научные и технические библиотеки. - 2013. - № 12. - С. 42-48.
2. Мазов Н. А. Влияние библиометрических методов на формирование рейтинга научной организации / Н. А. Мазов, В. Н. Гуреев // Электронные библиотеки: перспективные методы и технологии, электронные коллекции: труды XV Всероссийской научной конференции RCDL'2013. 2013. -С. 118-121.
3. Мазов Н. А. Использование библиометрических методов для решения библиотечных задач и задач оценки результативности научной деятельности организации / Н. А. Мазов, В. Н. Гуреев // Новые электронные технологии в информационном обслуживании ученых и специалистов Сибирского отделения Российской академии наук: материалы научно-практического семинара. 2013. -С. 30-33.
4. Заикин М. Ю. Библиографическая информационно-аналитическая система ИПИ РАН / М. Ю. Заикин, О. Л. Обухова, И. В. Соловьев // Системы и средства информатики. - 2014. - Т. 24, № 1. - С. 244-259.
5. RUSMARC [Электронный ресурс].- Режим доступа: <http://www.rusmarc.ru/rusmarc/format.html> (дата обращения 30.04.2015)
6. Об утверждении порядка предоставления научными организациями, выполняющими научно-исследовательские, опытно-конструкторские и технологические работы гражданского назначения, сведений о результатах их деятельности и порядка подтверждения указанных сведений федеральными органами исполнительной власти в целях мониторинга, порядка предоставления научными организациями, выполняющими научно-исследовательские, опытно-конструкторские и технологические работы гражданского назначения, сведений о результатах их деятельности в целях оценки, а также состава сведений о результатах деятельности научных организаций, выполняющих научно-исследовательские, опытно-конструкторские и технологические работы гражданского назначения, предоставляемых в целях мониторинга и оценки: Приказ Министерства образования и науки Российской Федерации от 5 марта 2014 года № 162 [Электронный ресурс] // Российская газета. – 2014. – 14 мая. – Режим доступа : <http://www.rg.ru/2014/05/14/minobrnauki2-dok.html> (дата обращения 30.04.2015)
7. Article Match Retrieval (AMR) [Электронный ресурс] – Режим доступа: <http://wokinfo.com/directlinks/amrfaq/> (дата обращения 30.04.2015)
8. Abstract Citations Count API (Elsevier Content APIs) [Электронный ресурс]. – Режим доступа: <http://api.elsevier.com/documentation/AbstractCitationCountAPI.wadl> (дата обращения 30.04.2015)