

УДК 001.102:011/016–057.212:004.63(470-25)БЕН

**А. А. Ивановский,**канд. биол. наук,  
зав. отделом ЦБС БЕН РАН**ТЕХНОЛОГИИ ОПЕРАТИВНОГО СИГНАЛЬНОГО  
ИНФОРМИРОВАНИЯ: НОВЫЕ РАЗРАБОТКИ БЕН РАН**

Автор освещает итоги исследования по разработке модели сетевой распределенной системы пертинентного информирования (2015), в результате которого были реализованы две технологии оперативного сигнального информирования: первая предусматривает использование специальной программы – менеджера библиографии – как основной среды работы библиотекаря; вторая построена на использовании собственной базы данных библиотеки.

*Ключевые слова:* оперативное сигнальное информирование, библиографическая база данных, менеджер библиографии.

Одна из главных задач научной библиотеки – помочь пользователю ориентироваться в мировом информационном пространстве [1]. И в первую очередь эта задача актуальна в случае с сетевыми версиями иностранных научных журналов. Целью работы системы оперативного сигнального информирования мы видим адресное предоставление пользователям пертинентной библиографической информации с возможностью получения полнотекстовых материалов для экономии рабочего времени исследователя. Сложность работы с сетевыми научными информационными ресурсами, как для читателей, так и для библиотекарей, заключается в чрезвычайном разнообразии интерфейсов сайтов разных журналов, разных издательств, разных научных обществ. Соответственно, встаёт вопрос об унифицированных способах обработки этой информации и о снижении затрат рабочего времени.

Подавляющее большинство иностранных издателей научных журналов предоставляют пользователям возможность выгрузки со своих сайтов развёрнутых библиографических описаний статей в различных коммуникационных форматах, в частности, формате RIS – Research Information Systems Citation File. RIS – простой теговый формат, разработанный фирмой Research Information Systems, в настоящее время

принадлежащей Thomson Reuters. Набор тегов у формата довольно велик. Для наших целей важно, что среди них есть теги, идентифицирующие аннотацию статьи, её DOI и адрес в Интернете.

В рамках системы оперативного сигнального информирования [2] выгруженные в этом формате библиографические данные сразу поступают в базу данных специальной программы обработки библиографической информации – *менеджера библиографии*, и уже в его оболочке осуществляется дальнейшая работа. При этом выбор конкретной программы мало влияет на суть работы.

Дальнейшая работа заключается в фильтрации – отборе по определённым критериям – информации, предназначенной для конкретного читателя. Такие фильтры, работающие как постоянные запросы, создаются в программе один раз, при подключении нового читателя к такому обслуживанию.

Следующий, заключительный, этап – выгрузка персональных подборок в конечном виде. Это могут быть библиографические списки в виде html-страниц или pdf-файлов, в зависимости от предпочтений пользователя.

Описанная технология, по мере её внедрения в опытное обслуживание пользователей некоторых библиотек БЕН РАН, показала ряд недостатков с организационной точки зрения.

Прежде всего, в системе ОСИ задействовано большое число наименований журналов, которые оказываются востребованными реальными пользователями. При обслуживании даже десятка пользователей перечень журналов составляет несколько десятков наименований.

В исходной схеме библиотекарь должен был работать с сайтом каждого журнала в отдельности – это несколько десятков сайтов. Это приводило к существенным затратам рабочего времени.

Второй момент – настройки менеджеров библиографии определяются их создателями. В нашем случае это приводит к тому, что в информации, получаемой пользователем, содержатся лишние сведения: номера ISSN, даты получения статей редакцией, принятия к опубликованию, факта опубликования, варианты названий журналов и т. п.

Чтобы решить первую проблему, мы перенесли работу с журналами с сайтов самих журналов в библиографические базы данных. Автоматические оповещения и, соответственно, готовые списки статей из нужных журналов формируются в аккаунтах библиотеки в Web of Science или Scopus.

С избыточностью информации можно было бы и смириться. Но оказалось, что в RIS-файлах, формируемых библиографическими базами данных, часть информации передаётся неполно.

Поэтому мы были вынуждены пойти по пути создания собственной базы данных.

Во-первых, это решает проблему с избыточностью информации: в собственной базе данных мы сами управляем параметрами запросов на выдачу результата.

Во-вторых, выгрузка информации из внешних библиографических баз данных – Web of Science и Scopus – осуществляется в формате CSV: текст с разделителями. Это позволяет обрабатывать исходную информацию из такого файла более гибко, извлекая из неё нужные фрагменты библиографических записей в нужном виде.

Остановимся подробнее на некоторых технологических аспектах, связанных с использованием баз данных Web of Science и Scopus.

Первый аспект – это выбор конкретной базы данных. Учитывая профиль Библиотеки по естественным наукам, с точки зрения перечней индексируемых журналов, востребованных пользователями БЕН РАН, две базы данных существенно не различаются. Отметим, что переводные версии отечественных научных журналов (издательства «Наука») по имеющейся у нас выборке читательских предпочтений полнее представлены в БД Scopus. Также Scopus полнее индексирует содержание журнала Nature, довольно популярного у пользователей разных по тематике библиотек БЕН РАН. Однако эта полнота, в отличие от полноты перечня индексируемых отечественных переводных журналов, имеет негативную сторону. Пользователям, которые пожелали получать оглавления свежих выпусков журналов, из журнала Nature интересны только полноценные научные статьи. В то время как в БД Scopus отражаются все типы публикаций этого журнала, включая короткие новостные заметки. Поэтому при использовании для получения первичной информации для БД оперативного сигнального информирования БД Scopus по этому изданию библиотеке необходимо осуществлять дополнительную фильтрацию материала.

Добавление новых записей в БД Web of Science происходит с известной и чётко определённой периодичностью: один раз в неделю. Библиографическая информация в БД Scopus обновляется постоянно, без фиксированной периодичности; во всяком случае, нам при работе с

несколькими обширными запросами систематичности в обновлении информации обнаружить не удалось. При этом оказывается, что информация по статьям из одного номера конкретного журнала попадает в БД Scopus за несколько обновлений, при этом предсказать, насколько растянется появление всех новых статей из конкретного выпуска, невозможно. Другими словами, в каждый момент времени, основываясь только на БД Scopus, нельзя быть уверенными в том, что мы располагаем информацией обо *всех* статьях из последнего номера конкретного журнала – чтобы убедиться в этом, необходимо обратиться к собственному сайту журнала. Контроль полноты перечня статей из конкретного выпуска в БД Scopus при этом тоже оказывается затруднённым: на настоящий момент, сортировка результатов по дате в Scopus работает некорректно.

Появление новых записей в БД Web of Science для некоторых востребованных нашими пользователями журналов может отставать от выхода в свет номеров на срок до месяца (по нашим наблюдениям на нескольких десятках наименований журналов). В этом аспекте информация в БД Scopus оказывается более оперативной. Более того, в БД Scopus индексируются и т. н. Articles in Press, что отдельно подчёркивает фирма-создатель. Индексирование такого типа публикаций ещё больше увеличивает оперативность БД Scopus по сравнению с БД Web of Science.

Наконец, обратимся к вопросу импорта информации из внешних библиографических баз данных в собственную БД библиотеки – базу данных оперативного сигнального информирования.

Наиболее оптимальным форматом для импорта информации представляется формат CSV. В общем виде подразумевается, что в файле такого формата отдельные поля библиографической записи разделены специальным символом (символом-разделителем), и каждая новая библиографическая запись представляет собой новую строку файла.

При экспорте информации из БД Scopus в файл CSV, по сути, используются два разделителя. Для полей библиографической записи, имеющих непустое значение, разделителем является символ кавычки. Для пустых полей библиографической записи (например, если журнал выходит по томам без деления на выпуски, то поле «выпуск» будет пустым) используется разделитель «запятая». Для файлов, содержащих записи, однотипные по набору непустых полей, такой подход не представляет неудобства. Однако в случае работы с результатами обширных запросов, содержащих десятки наименований журналов, записи в экспортируемом

файле оказываются разнотипными по набору непустых полей, что делает результаты последующего импорта непредсказуемыми.

Создатели БД Web of Science решили проблему экспорта данных в файл CSV более удачно. Символом-разделителем в этом файле является знак табуляции. Это снимает и ещё одну проблему при работе с файлами CSV, имеющуюся в БД Scopus: в БД Web of Science кавычки, являющиеся частью поля библиографической записи (например, кавычки в тексте аннотации статьи), не нарушают структуру файла, т. к. не воспринимаются в качестве разделителя полей.

Неудачный выбор символа-разделителя в БД Scopus вынудил нас осуществлять экспорт информации из этой БД в формате RIS. Эта информация импортируется в менеджер библиографии, средствами которого готовится новый CSV-файл. И уже этот новый CSV-файл импортируется в базу данных библиотеки.

В заключение следует обратить внимание, что в двух обсуждаемых базах данных одни и те же поля библиографической записи, например ISSN, экспортируются в разных форматах. В БД Web of Science ISSN представлен в формате «четыре знака–дефис–четыре знака». В БД Scopus ISSN экспортируется в виде последовательности из восьми знаков (без дефиса). Для нас это различие оказалось критически важным, т. к. в качестве уникальных идентификаторов журналов в базе данных оперативного сигнального информирования используются именно ISSN, в частности из-за того, что названия одного и того же журнала различаются в двух базах данных (как, впрочем, и в каталогах разных библиотек). Поэтому при импорте библиографической информации в собственную БД библиотеки приходится приводить записи ISSN к одному формату.

При условии наличия наименования журнала в обеих реферативных базах данных мы отдаём предпочтение БД Web of Science. Во-первых, Web of Science позволяет импортировать библиографическую информацию в собственную БД библиотеки за одну технологическую операцию, используя файл с разделителями, генерируемый непосредственно базой данных Web of Science. Во-вторых, БД Web of Science в силу предсказуемости её обновлений представляет более полную и достоверную информацию для системы оперативного сигнального информирования. Одновременно для тех же наименований журналов, информация по свежим выпускам которых импортируется нами из БД

Web of Science, мы используем БД Scopus для включения в пользовательские оповещения информации о публикациях типа Article in Press, отсутствующей в Web of Science.

(Работа выполнена при поддержке гранта РФФИ № 16–07–00450.)

#### Список использованных источников

1. *Ивановский А. А.* Присутствие научных изданий в Интернете и значение научной библиотеки / А. А. Ивановский, Е. В. Ткачева // Теория и практика общественно-научной информации : сб. науч. тр. – Москва, 2014. – С. 157–161.
2. *Ивановский А. А.* Современные программные средства оперативного сигнального информирования в практике библиотек ЦБС БЕН РАН / А. А. Ивановский // Информационное обеспечение науки: новые технологии : сб. науч. тр. / ред.: Н. Е. Каленов, В. А. Цветкова. – Москва : БЕН РАН, 2015. – С. 275–278.

Статья поступила 24.06.2016

#### Aleksandr Ivanovskii

#### TECHNOLOGIES OF OPERATIVE SIGNAL INFORMING: THE LAST INNOVATIONS OF BEN RAN

The author gives a review of the research aimed to model of net distribute system of pertinent informing (2015), which resulted in realization of two technologies of operative signal informing. The first supposes using of special program – manager of bibliography as main space of librarian work, and the second uses own database of library.

*Keywords:* operative signal informing, bibliographical database, manager of bibliography.