

ОТ ЭЛЕКТРОННЫХ КАТАЛОГОВ К ЭЛЕКТРОННОЙ БИБЛИОТЕКЕ

С.А.Власова, Н.Е.Каленов, Библиотека по естественным наукам РАН,
119890 Москва, Знаменка, 11
E-mail: osiat@ben.irex.ru

FROM ELECTRONIC CATALOGUES TO ELECTRONIC LIBRARY

S.Vlasova, N.Kalenov, Library for Natural Sciences of Russian Academy of
Sciences, Znamenka, 11, 119890 Moscow, Russia
E-mail: osiat@ben.irex.ru

Library for Natural Sciences of RAS (LNS RAS) heads a centralized library system (CLS) that includes more than 150 RAS institutions libraries. LNS RAS supports on Internet server (<http://www.benran.ru>) a lot of different resources (electronic catalogues, new acquisitions etc.)

One of resources is being created by LNS and CLS libraries is database “Science of Russia” that contains information about Russian scientists and their publications. Technology for creating, supporting this database and its connection with RAS Integrated information System are described in the paper.

Библиотека по естественным наукам (БЕН) РАН, являющаяся одним из федеральных информационных центров, возглавляет централизованную библиотечную систему (ЦБС), включающую более 100 библиотек академических учреждений. Одной из основных функций БЕН РАН как головной библиотеки ЦБС является ведение сводных каталогов изданий, поступающих в фонды библиотек системы. Начиная с 1990 года, БЕН РАН поддерживает сводный электронный каталог журналов, включающий более 6500 наименований, а с 1993 года - сводный электронный каталог книг и продолжающихся изданий. В каталогах отражаются все материалы, поступающие в библиотеки ЦБС, как централизованно через БЕН, так и непосредственно в библиотеки НИИ РАН.

С 1996 года каталоги ЦБС БЕН РАН представлены в Интернет по адресу <http://ben.irex.ru>.

Развитие сетевых технологий и широкое распространение электронных публикаций, предоставляемых через Интернет, потребовали от БЕН РАН разработки новых технологий, обеспечивающих достаточно высокий уровень информационной поддержки научных исследований. Работа ведется в двух взаимосвязанных направлениях. Первое из них обеспечивает отражение в Интернет ресурсов, формируемых в академических организациях, второе направлено на предоставление пользователям мировых ин-

формационных Интернет-ресурсов. Работы поддерживаются РФФИ (проекты 00-07-9097 и 02-07-90088).

В рамках первого направления, наряду с развитием сводных электронных каталогов, БЕН РАН реализует проект «Наука России», целью которого является создание системы децентрализованного ввода данных о публикациях сотрудников академических организаций с последующей централизованной обработкой этих данных [1].

Информационную основу системы составляют традиционные «картотеки трудов сотрудников». Ведение таких картотек, содержащих данные об изданных статьях, монографиях и т.п., подготовленных сотрудниками данного института, является одной из обязательных функций библиотеки каждого НИИ РАН.

В процессе реализации проекта разработаны два программных комплекса – локальный (обеспечивающий подготовку информации в библиотеках НИИ РАН) и центральный, позволяющий объединять локальные базы данных и обеспечивать к ним доступ через Интернет.

Локальный комплекс системы обеспечивает:

- инсталляцию системы с назначением перечня полей данных в каждой из 4 взаимосвязанных подбаз – «персоналии», «организации», «публикации» (сведения о статьях в журналах, сборниках и т.п.), «источники» (сведения о журналах, сборниках и т.п. в целом);

- ввод данных – возможен ввод данных автономно в каждую из подбаз «организации» или «источники»; в две подбазы - «персоналии» и «организации»; во все 4 подбазы, начиная с подбазы «публикации», с возможностью ввода пристатейной библиографии;

- корректировку содержимого полей данных в любой подбазе;
- ввод новых полей данных в уже существующую базу;
- развитые возможности поиска данных с использованием булевских операторов;

- возможность автоматического получения списка публикаций, цитируемых и цитирующих работ любого автора (группы авторов), найденного по тем или иным запросам в подбазе «персоналии»;

вывод результатов поиска на экран и в файлы различной структуры, в том числе во внутреннем формате базы данных, что создает возможность в произвольный момент времени разделить накопленную базу данных на фрагменты (например, публикации сотрудников отдельных лабораторий) и в дальнейшем поддерживать их независимо.

Центральный комплекс системы «Наука России», поддерживаемый в БЕН РАН, обеспечивает:

- слияние баз данных (имеющих «непротиворечивые» структуры, установленные при инсталляции системы), сформированных с помощью локального комплекса в различных библиотеках;

- исключение дублированных записей, введенных в различных библиотеках – ситуация достаточно часто встречающаяся, когда авторами публикации являются сотрудники разных институтов;
- поддержка Интернет-версии системы, обеспечивающей развитый поиск данных через стандартные Интернет-браузеры.

Алгоритм исключения дублированных записей является одной из отличительных особенностей системы «Наука России». Он основан на специальном анализе содержимого записей (, позволяющем предполагать, что две записи описывают одну и ту же публикацию, даже если содержаемое их полей несколько различается), работает в диалоговом режиме и является «советующим» - предлагает пользователю выбрать одну из пары «подозрительных» записей или оставить в базе данных обе.

Еще одной особенностью системы «Наука России», отличающей ее, в частности, от известной системы Science Citation Index, является возможность различать записи персоналий, имеющих одинаковые фамилии, имена и отчества.

Локальный комплекс системы «Наука России» находится в промышленной эксплуатации с 1998 года, он установлен в библиотеках многих институтов РАН, где накоплены, в общей сложности, десятки тысяч описаний публикаций.

Базы данных трудов сотрудников ряда НИИ РАН, поддерживаются на сервере БЕН РАН, поиск информации в них возможен при обращении по адресу http://195.178.196.159/cgi-bin/sr1_n.exe?inaros/inaros.ini.

К сожалению, сотрудники библиотек институтов РАН, используют возможности системы в минимальной степени, вводя только те данные, которые они привыкли отражать в традиционных картотеках, что существенно обедняет информационные возможности системы. Однако это - проблемы организационные (связанные, в первую очередь, с уровнем оплаты библиотечных работников), поэтому, считая, что рано или поздно они будут решены, мы стараемся развивать систему в направлении преобразования ее в современную электронную библиотеку.

Если на начальных этапах развития системы предполагалось отражать в ней только библиографические сведения о публикациях, то сейчас она также позволяет подключать их полные тексты, а также файлы изображений, в частности, фотографии сотрудников. Возможности работы с этими ресурсами демонстрируются на примере базы данных публикаций сотрудников БЕН РАН, представленной в Интернете.

В настоящее время принята программа создания Единой информационной системы (ЕИС) РАН. Очевидно, что сведения, формируемые в рамках системы «Наука России» в библиотеках институтов РАН, являются существенной составляющей ресурсов ЕИС. В связи с этим специалистами БЕН РАН, ВЦ РАН и ЦНТК РАН начаты работы по включению информации, формируемой в рамках системы «Наука России», в «Интегрирован-

ную систему информационных ресурсов (ИСИР) РАН [2, 3], являющуюся подсистемой ЕИС. При этом необходимо было решить ряд задач, первая из которых – разработка обменного формата данных, обеспечивающего загрузку информации, формируемой в системе «Наука России», в ИСИР РАН. Такой формат, основанный на схеме DUBLIN CORE METADATA и языке XML в стандарте RDF, был разработан для представления данных о ресурсах типа «Публикация», «Персона» и «Организация». Формат достаточно подробно описан в [3], в дальнейшем, в процессе отладки технологии загрузки реальных данных в ИСИР, формат был несколько модифицирован.

В частности, статья из журнала представляется тремя унифицированными метаописаниями ресурсов типа «публикация», отличающихся полем “Type”, которое принимает значения: “S” – сводный уровень (наименование журнала, общие сведения о нем), “M” – монографический уровень (спецификация конкретного выпуска журнала, в котором опубликована данная статья, со ссылкой на соответствующий ресурс типа “S”), “A” – аналитический уровень (название статьи, ссылки на авторов, страницы, ссылка на соответствующий ресурс типа “M”). Аналогично, но двумя уровнями, описывается статья из сборника. Такой подход позволяет однократно описывать каждый ресурс (при повторном упоминании журнала или его выпуска используется только ссылка на его идентификатор) и достаточно легко реализовать навигацию по сформированной электронной библиотеке.

Специалистами БЕН РАН была разработана программа-конвертер, формирующая файлы согласованной структуры из баз данных, подготовленных с использованием локального комплекса системы «Наука России». В настоящее время на примере данных о сотрудниках БЕН и их публикациях отлаживается практическая технология взаимодействия ИСИР РАН с системой «Наука России». На сервер ИСИР РАН (<http://isir.ras.ru>) загружены данные, примерно, о 700-х публикациях со ссылками на полные тексты ряда статей с обеспечением навигации по различным элементам библиографических описаний изданий и сведениям об авторах.

Таким образом технология, разрабатываемая, ориентированная первоначально на создание традиционных для библиотек ресурсов (хотя и в электронной форме), становится составной частью процесса формирования принципиально новой среды – электронной библиотеки, предоставляющей пользователям Интернет развитый информационный сервис.

Элементы распределенной электронной библиотеки создаются в БЕН РАН и в рамках второго направления, связанного с предоставлением пользователям внешних информационных ресурсов. БЕН РАН разработала систему отсылок от журналов, отраженных в ее сводном Интернет-каталоге, к соответствующим разделам сайтов издательств и ресурсам «Научной электронной библиотеки» (НЭБ) РФФИ.

Журнальный каталог БЕН РАН включает данные о более чем 600 000 выпусков около 6500 наименований журналов (из которых 80% - зарубежные), поступающих в библиотеки ее ЦБС. Однако из-за финансовых проблем за последние годы количество выписываемых зарубежных журналов неуклонно снижается – например, в 2001 году БЕН РАН смогла выписать немногим более 200 наименований. Благодаря поддерживаемой системе отсылок пользователь, просматривающий каталог БЕН, может ознакомиться с аннотациями (а авторизованный пользователь – с полными тестами) статей из многих журналов, даже если они отсутствуют в фондах ЦБС БЕН РАН.

Литература

1. Васильев А.В., Власова С.А., Глушановский А.В., Каленов Н.Е. Типовая система для распределенной подготовки библиографических баз данных // Научные и технические библиотеки, 1999. – № 6. – с.36-38.
2. Каленов Н.Е. Власова С.А., Глушановский А.В. Развитие системы «Наука России» как части Интегрированной системы информационных ресурсов РАН // Современные технологии в информационно-библиотечном обеспечении научных исследований: материалы 8-го научно – практического семинара «Информационное обеспечение науки: новые технологии». – М., 1999. – с.65-68.
3. Бездушный А.Н., Меденников А.М., Серебряков В.А., Власова С.А., Каленов Н.Е. Подход к интеграции информационных коллекций В ИСИР РАН // Электронные библиотеки: перспективные методы и технологии, электронные коллекции. 2-я Всероссийская научная конференция. Сб. докладов. Сентябрь, 26-28, 2000. – С. 129-135.
4. Власова С.А., Каленов Н.Е., Куриков П.М., Серебряков В.А. Библиотечная составляющая ИСИР РАН // Новые технологии в информационном обеспечении науки. Сб. статей. / Ред.: Каленов Н.Е. – М.: Биоинформ-сервис, 2001. С. 24-32.