

ОПЫТ СОЗДАНИЯ БАЗ ДАННЫХ В ОТДЕЛЕНИИ ГПНТБ СО РАН

Н.В.Васильева, О.А.Федотова

Государственная публичная научно-техническая библиотека

Сибирского отделения Российской академии наук

e-mail: natavale@mail.ru, o4f8@mail.ru

Аннотация

В докладе обсуждается опыт создания каталогов и баз данных Отделения ГПНТБ. Основное внимание уделяется разработке каталога, содержащего информацию о научных журналах, доступных в полнотекстовом виде с компьютеров ГПНТБ СО РАН.

Введение

Информационное обеспечение науки и образования всегда являлось главной целью деятельности Отделения ГПНТБ СО РАН. Развитие технологий в области передачи и обработки информации, появление электронных ресурсов, создание современных телекоммуникационных систем, привели к появлению принципиально новых возможностей организации практически всех этапов научно-образовательного процесса, что, в свою очередь, обусловило качественный рост информационных потребностей научных работников [1]. Очевидно, что на первом месте перед библиотекой встала задача создания электронных каталогов как средства навигации в огромных по объему информационных библиотечных ресурсах и предоставление этой информации ученым, преподавателям и студентам.

Система JISIS

Сайт Отделения ГПНТБ СО РАН открылся в 1997 году. Практически с самого начала его функционирования были доступны каталоги онлайн. Для этого программистами Отделения была создана web-ориентированная информационно-поисковая система JISIS, которая отражала и давала возможности поиска в библиографических базах и каталогах в СУБД ISIS, IBIS и IRBIS32.

На начало 2011 года функционировали восемь баз данных в ISIS, две в IBIS, и в IRBIS32 – тридцать три. Вне поисковой системы JISIS существовали еще две базы данных: WODA – зарубежные журналы, полученные Отделением ГПНТБ СО РАН, и «Научный потенциал НСО», содержащая 25 тысяч публикаций докторов наук по 71 научным учреждениям.

Система IRBIS64

В 2010 году ГПНТБ переходит на современную систему доступа к удалённым ресурсам – WEB-ИРБИС. В то же время в Отделении при обновлении программного обеспечения на сервере стало понятно, что информационно-поисковая система JISIS, работавшая около 15 лет, устарела и более не соответствует настоящей реальности современных программных продуктов. В одно из возможных решений – перенос каталогов и баз данных на одну платформу – прекрасно вписывался IRBIS64 со своим web-шлюзом. Однако возникла проблема в том, что сервер Отделения находится под управлением системы

из семейства ОС UNIX, а IRBIS64 работает только под операционной системы MS Windows. У Отделения ГПНТБ СО РАН нет возможности самостоятельно развернуть и поддерживать Windows Server. Сервера библиотеки, работающие под ОС FreeBSD, не имеют необходимого дискового пространства. Выходом из этой ситуации стало размещения каталогов и баз данных на виртуальном Windows Server'е в «Корпоративном облаке СО РАН»¹.

Виртуальная платформа была предоставлена бесплатно, так же был организован удаленный рабочий стол, с помощью которого ведущими программистами Отделения осуществляется работа по созданию и ведению каталогов и баз данных. Поддержку всего программного обеспечения, в том числе и виртуального сервера, осуществляют администраторы корпоративного облака [2].

Работа по всем ресурсам, размещенным в корпоративном облаке СО РАН, в настоящее время активно продолжается. В текущем году были полностью перенесены существовавшие базы данных и созданы новые. В рамках сотрудничества с КТИ ВТ СО РАН были размещены на web-шлюзе базы данных публикаций сотрудников Института (только за 2012 год) и каталог библиотеки института. Продолжается развитие ресурса «Научные школы ННЦ»: за 2014 год были добавлены пять персон и соответствующих баз данных.

Благодаря возможностям АБИС IRBIS была изменена технология создания информационных ресурсов. Теперь основополагающим моментом является ведение базы данных, из которой с помощью созданных форм вывода формируются библиографические списки. Большое число созданных электронных ресурсов тематической библиографии на нашем сайте за 15 лет своего существования послужила толчком к созданию БД «Библиографические ресурсы Отделения ГПНТБ СО РАН».

Путем объединения баз данных в определённые группы, меню доступа к ним было модернизировано (по аналогии с ГПНТБ). Это позволило сэкономить пространство на экране и более удобно ориентироваться в базах данных. Продолжается работа по улучшению форм вывода для разных баз данных, расширению поисковых возможностей по полям, размещению каталогов отечественных и зарубежных периодических изданий.

В 2015 году большое внимание будет уделено создающимся патентным базам данных.

ИС «Журналы, полные тексты которых доступны с компьютеров ГПНТБ СО РАН»

Большие, постоянно увеличивающиеся, объемы ресурсов на сайте Отделения ГПНТБ СО РАН вынуждают думать об изменении созданной изначально структуры. Для пользователей нашей библиотеки интересен не только физический фонд, но и отражение на сайте доступных полнотекстовых ресурсов, в частности, журналов. Такой электронный ресурс был создан². Он представляет списки сайтов 8013 научных журналов, предоставляющих доступ к полным текстам с компьютеров Отделения ГПНТБ СО РАН. Так как сайты издательств предоставляют различные дополнительные услуги и иные возможности доступа к информации, то наряду с алфавитным указателем представлены списки тех же журналов и по издательствам.

¹ http://irbis.prometeus.sbras.ru/cgi-bin/irbis64r_11/cgiirbis_64.exe?LNG=&C21COM=F&I21DBN=CATNO

² <http://www.prometeus.nsc.ru/resource/web/journal/2009/>

Однако, такой ресурс в виде связанных HTML страниц, состоящий из нескольких тысяч названий журналов, предоставленных десятками издательств и агрегаторов, трудно поддерживать в актуальном виде. Также для такого ресурса необходимо присутствие поиска журналов по названию и ISSN.

При проектировании каталога было принято решение отказаться от WEB-ИРБИСа, так как необходим более дружелюбный интерфейс, как для пользователей, так и для сотрудников, поддерживающих этот ресурс. Информационная система (ИС) «Журналы, доступные в полнотекстовом виде с компьютеров ГПНТБ СО РАН» создается с использованием СУБД MySQL.

Профиль ИС

Основным требованием, предъявляемым к информационным системам, является интероперабельность. Особенно важно это при создании каталогов.

Под интероперабельностью любой информационной системы, в том числе и каталогов, понимается степень ее способности взаимодействовать с другими информационными системами, в том числе и с человеком [3]. Но если при взаимодействии с последним (как с информационной системой) основная нагрузка на обеспечение взаимопонимания ложится на человека, который в состоянии обработать даже очень плохо организованную информацию, то для обеспечения эффективного взаимодействия между информационными системами требуются специальные технологические подходы и общие соглашения. Обеспечение интероперабельности систем невозможно без строгого выполнения соответствующих международных стандартов и рекомендаций [4]. При этом стандартам должны соответствовать:

- протоколы и интерфейсы доступа к данным;
- поисковые языки и интерфейсы;
- схемы и форматы представления данных;
- правила контроля доступа к данным.

Для формирования метаданных системы применяются несколько стандартов, являющихся расширениями рекомендаций Dublin Core³ и Qualified Dublin Core (QDC).

В нашем случае в первую очередь необходимо обеспечить взаимодействие с каталогами ИРБИС. Для обмена данными, содержащими библиографические описания документов, применяется формат MARC (Machine Readable Cataloguing), который является основой для каталогизации и библиографических баз данных.

Для того чтобы выполнять обмен метаданными между ИРБИСом планируется создать сервис, выполняющий преобразование схем метаданных из схемы ИРБИС в схему Dublin Core с использованием квалификаторов.

Базовый уровень описания ресурсов ИС

Первым шагом при создании информационной системы является разработка ее структуры – определение таблиц в базе данных, полей в каждой таблице, связей между элементами (рис. 1). Необходимо включить средства управления по наличию/отсутствию доступа к полнотекстовым документам.

³ Dublin Core Metadata Initiative (DCMI) (<http://www.dublincore.org/>).

ИС содержит следующие коллекции метаданных:

1. Список журналов (таблица *journals*).
2. Список издательств (таблица *publishers*).
3. Список агрегаторов (таблица *agents*).
4. Словари-классификаторы:

4.1. ББК – ссылка на рубрику Библиотечно-Библиографической Классификации, либо вложенное описание рубрики с указанием кода и, возможно, словесной расшифровки.

4.2. Основной код УДК – тематическая классификация с помощью ссылки на рубрику основной таблицы УДК (Универсального Десятичного Классификатора).

4.3. ГРНТИ – тематическая классификация с помощью ссылки на рубрику основной таблицы Государственного рубрикатора научно-технической информации.

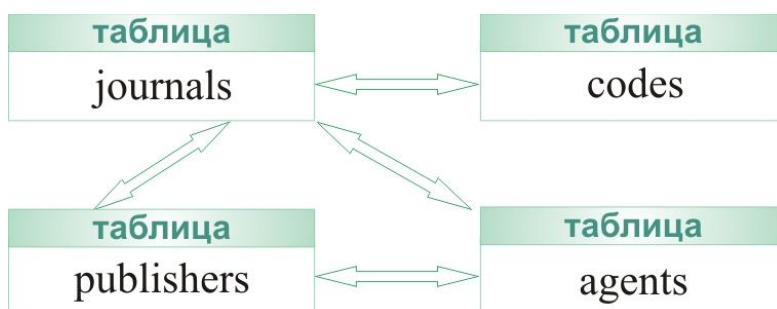


Рис. 1. Логические связи между основными коллекциями

За основу схем данных коллекций приняты рекомендации DCMI. Базовое метаописание журнала включает следующие элементы и связи:

- *id* – локальный идентификатор журнала;
- *dc.title* – название журнала;
- *dc.title.alternative* – английское название журнала;
- *dc.description* – описание журнала;
- *dc.identifier.uri* – основной URL адрес в сети;
- *dc.publisher* – ссылка на издательство журнала из таблицы *publishers* (идентификатор журнала);
- *dc.identifier.issn* – ISSN;
- *dc.identifier.eissn* – ISSN электронного издания;
- *access* – доступ (есть/нет/тестовый) (выбор из 3 вариантов);
- *dc.date.available* – за какие годы доступен журнал;
- *dc.subject.code* – код классификатора (*code = udc, code = bbk* и т.д.);
- *rus_or_foreign* – отечественный/зарубежный (выбор из 2 вариантов);
- *agent* – ссылка на агрегатор из таблицы *agents* (идентификатор агрегатора);
- *inst* – есть ли доступ в Институтах ННЦ (выбор из списка) – это поле включено в структуру в надежде на сотрудничество с институтами для более полного информирования научных сотрудников.

Коллекция *publishers* содержит список издательств и имеет следующий набор элементов:

- *id* – уникальный идентификатор организации (параметр для ссылок);
- *dc.title* – название издательства;

- *dc.title.alternative* – сокращенное название издательства;
- *dc.identifier.uri* – основной URL адрес в сети.

Коллекция *agents* содержит список агрегаторов и имеет следующий набор элементов:

- *id* – уникальный идентификатор агрегатора (параметр для ссылок);
- *dc.title* – название агрегатора;
- *list_journals* – ссылки на журналы из таблицы *journals*;
- *dc.identifier.uri* – основной URL адрес в сети.

В системе организовано два типа поиска документов: контекстный и атрибутивный. Контекстный поиск организован средствами СУБД. Атрибутивный поиск идет только по полям метаданных.

На рис. 2 можно увидеть варианты представление записей в базе данных для пользователей.

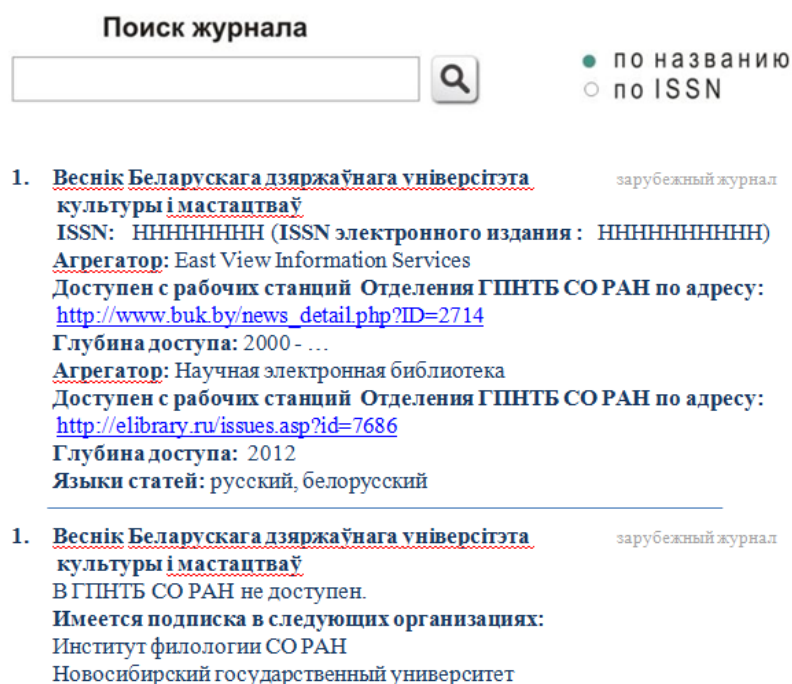


Рис. 2. Варианты выдачи результата поискового запроса

Планируется вывести просмотр записей в БД по алфавиту, издательствам и темам.

ЛИТЕРАТУРА

1. Федотов А.М., Барахнин В.Б., Жижимов О.Л., Федотова О.А. Информационная модель электронной библиотеки по научному наследию // Сборники Президентской библиотеки им. Б.Н. Ельцина / Вып. 5: Научное и организационно-технологическое формирование цифрового библиотечного, музейного и архивного контента: сборник научных трудов. – 2014. – XXX с.: ил. – (Серия «Электронная библиотека» / науч. ред. Е.Д. Жабко). – 2014. – Санкт-Петербург: ФГБУ «Президентская библиотека имени Б.Н. Ельцина». – С.175-202. - ISBN 978-5-905273-51-3.
2. Васильева Н.В. «Облачные технологии»: возможности использования при обслуживании ученых Сибирского отделения Российской академии наук // Новые

электронные технологии в информационном обслуживании ученых и специалистов Сибирского отделения Российской академии наук: материалы науч.-практ. семинара (г. Красноярск, 1-5 июля 2013 г.) / ГПНТБ СО РАН, Красноярский науч. центр СО РАН; Ред. кол.: Е.Б.Артемяева (отв. ред.), Е.А.Базылева (отв. ред.) и др. – Новосибирск: ГПНТБ СО РАН, 2014. – С.50-53.

3. Жижимов О.Л., Федотов А.М. Обеспечение интероперабельности электронных библиотек // Информационные технологии и математическое моделирование в науке, технике и образовании (Бишкек, Кыргызстан, 05.10 – 09.10.2011): Известия Кыргызского Государственного Технического Университета им. И.Раззакова. – №24. Материалы международной конференции. – Бишкек, 2011. – С. 331-335.
4. Жижимов О.Л., Федотов А.М., Федотова О.А. Построение типовой модели информационной системы для работы с документами по научному наследию // Вестник НГУ. Сер.: Информационные технологии. – 2012. – Т. 10. – Вып: 3. – С. 5-14.