

Электронная коллекция «Электронное государство»: технологические аспекты *

И.А. Мбого, А.В. Чугунов

Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики

Irina.Mbogo@gmail.com, chugunov@egov-center.ru

Аннотация

Работа описывает комплекс задач построения электронных коллекций, отображения, распространения и интеграции библиографической информации. В статье рассматривается принцип построения электронной библиотеки на примере коллекции «Электронное государство», методы формирования библиографического описания, интеграция электронных коллекций посредством OAI-PMH и совместимость данных с пользовательскими программами библиографических менеджеров типа Zotero и Mendeley. Электронная коллекция может использовать многоуровневые классификаторы. Формирование записей осуществляется на основании заполнения полей форм с привязкой к разделу классификатора. В качестве описываемых ресурсов могут выступать различные печатные и электронные издания, др. В качестве отдельной опции реализована возможность формирования библиографического описания в соответствии с ГОСТом в полуавтоматическом режиме.

1. Публикация электронных коллекций

Задачей современной электронной коллекции является создание многофункционального комплекса на основе масштабируемого программного обеспечения, выполняющего набор функций, необходимых для визуализации исходных материалов, для интеграции в международное информационное научное пространство, для обеспечения пользователей необходимым инструментарием и реализующими современные методы библиографического поиска.

При создании электронной коллекции «Электронное государство», используется система управления контентом Drupal и подключаемый модуль biblio. Drupal является не только системой управления контентом, но и каркасом (framework), содержащим большое количество разнообразных библиотек (рис. 1).

В ходе проекта создается электронная библиотека, предоставляющая комплекс услуг, наиболее

полно удовлетворяющая потенциальные потребности пользователя. Библиотека предоставляет списки публикаций в виде библиографических описаний в соответствии с ГОСТ 7.0.5 – 2008, в качестве OAI-PMH репозитория и для работы с библиографическими данными с использованием библиографических менеджеров.

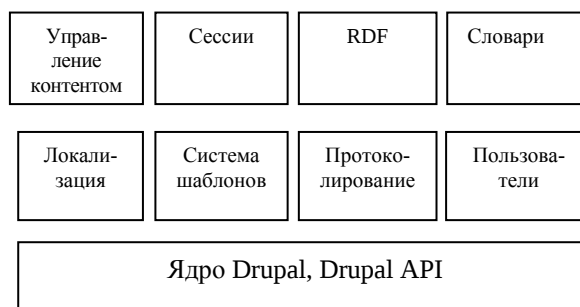


Рис. 1. Ядро и основные модули Drupal

Модуль biblio предоставляет дополнительный тип материала «запись в библиотеку» с расширенным набором полей. Информационная система позволяет создавать описания, хранить материалы как в виде текстов так и в виде прикрепленных файлов. Могут формироваться следующие типы записей: книга, глава книги, журнал, статья в периодическом издании, статья в сборнике трудов конференций, web ресурс, автореферат, отчет. Записи формируются на основе заполненных полей: название, автор, год издания, место издания, издатель, том, выпуск, страницы и т.д. На рис. 2 представлена форма ввода информации в электронную библиотеку.

Архитектура модуля строится на базе Drupal API и включает в себя набор функций, определяющих основную функциональность:

- biblio_admin,
- biblio_contributors,
- biblio_feeds,
- biblio_fields,
- biblio_import,
- biblio_keywords,
- biblio_pages,
- biblio_search,
- biblio_util,
- biblio_theme,
- biblio_xml.

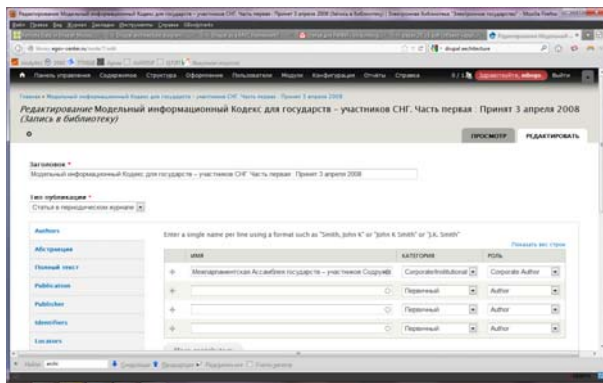


Рис. 2. Форма ввода данных в электронную библиотеку

Отдельной задачей является представление данных. Варианты представления библиографической информации в электронных коллекциях могут быть различны: неформатированная композиция из набора данных, в виде сформированных библиографических описаний по какому-либо стандарту – для отображения человеку или в виде XML для дальнейшей компьютерной обработки.

Модуль biblio имеет возможность представлять списки библиографических описаний в соответствии с различными стандартами, в основном, американскими. Одной из задач, решаемых в рамках проекта, было формирование библиографического описания в соответствии с ГОСТ 7.0.5 — 2008. На данный момент создан стиль оформления списков записей из набора полей в соответствии с ГОСТ. Сформированный список представлен на рис. 3.

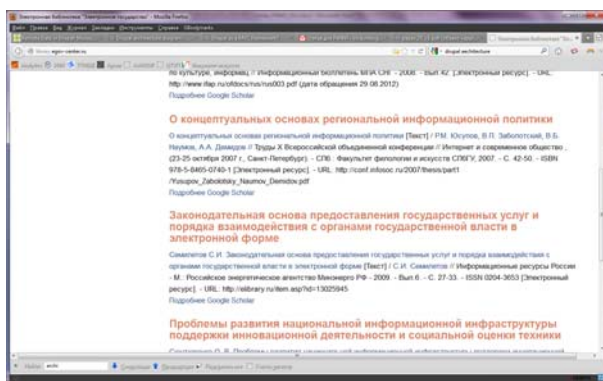


Рис. 3. Списки автоматически сформированных библиографических описаний

2. Интеграция ресурсов

На сегодняшний день общепризнанным механизмом интеграции электронных коллекций и библиотек является инициатива открытых архивов и протокол OAI-PMH, который определяет механизм сбора записей, содержащих метаданные из различных хранилищ. Протокол предоставляет возможность хранилищам сделать метаданные доступными для сервисов, основанных на открытых стандартах HTTP и XML. Использование разработчиками рос-

сийских электронных коллекций технологий и методологии OAI позволяет обеспечить информационную интеграцию как с мировыми ресурсами (для тех публикаций, которые имеют набор метаинформации на английском языке: библиографическое описание, реферат или аннотацию, набор ключевых слов), так и с российскими научными электронными коллекциями, поддерживающими шлюзы на базе протоколов OAI.

Модуль OAI-PMH динамически формирует XML на все типы запросов протокола в нотации Dublin Core и представляет библиотеку в виде OAI хранилища.

3. Стили представления информации

Существует довольно много форматов описания метаданных, таких как MARC21, RUSMARC, USMARC. Приведенные форматы являются широко распространенными, но весьма сложными.

Большое количество международных представлений библиографической информации предполагает наличие различных конвертеров для преобразования из одного представления в другое. Такой подход является достаточно традиционным, но не самым эффективным. На сегодняшний день разработан новый открытый XML формат – язык CSL (Citation Style Language) [3]. Язык описывает, в том числе, и представление библиографий. К настоящему моменту насчитывается около 2500 стилей CSL, поддерживаемых как национальными ассоциациями так и частными лицами. Для представления библиографических описаний в соответствии с российским стандартом в репозитории существует 3 стиля. Апробация этих стилей показала, что они требуют доработки.

Необходимо отметить, что использование стилей, основанных на языке CSL позволяет их использовать не только на стороне сервера при создании электронной коллекции, но и на стороне клиента с использованием библиографических менеджеров для построения списков библиографий при подготовке научных работ.

4. Библиографические менеджеры

Библиографические менеджеры предназначены для сбора и хранения библиографических данных. К наиболее удачным можно отнести два программных продукта Zotero и Mendeley [4] [5]. Эти программы позволяют сохранять информацию о требуемых источниках и предоставлять ее в том виде, который необходим для конкретной публикации. Для генерации библиографий используется язык CSL. Zotero является дополнением к браузеру Firefox и распространяется как продукт с открытым исходным кодом. Mendeley представляет собой обычное приложение, распространяется бесплатно, но доступа к исходным кодам у пользователей нет [5]. Главным преимуществом персональных менеджеров является возможность импорта-экспорта библиографической

структурированной информации, а также их встраиваемость в ведущие текстовые редакторы, что позволяет вести учёт библиографическим ссылкам в рамках одного файла и часто не задумываться о правилах оформления библиографических описаний, потому что программа их может оформить на лету.

Проект [6] предполагает работу с тремя группами пользователей:

- авторы научных публикаций — инициаторы (модераторы) тем научных дискуссий, члены редколлегии, рецензенты и редакционные работники (зарегистрированные пользователи со специфическим набором прав доступа к материалам и разделам системы);
- научные работники и преподаватели, заявившие о заинтересованности участия в сетевых обсуждениях и публикациях своих материалов, потенциальные авторы получающие доступ к неопубликованным рабочим материалам системы (зарегистрированные пользователи со стандартными правами);
- незарегистрированные пользователи, имеющие возможность ознакомиться с материалами, находящимися в открытом доступе.

Заключение

Оторванность российских исследователей, работающих в социально-гуманитарных отраслях науки, от мирового научного информационного пространства существенным образом препятствует вхождению россиян в международные исследовательские сети. Одним из путей решения этой проблемы является создание электронных библиотек, использующих методы и технологий интеграции информационных ресурсов и сервисов, которые развиваются в рамках инициативы открытых архивов ОАИ (Open Archives Initiative). [7]

Использование разработчиками российских электронных коллекций технологий и методологии ОАИ позволяет обеспечить информационную интеграцию как с мировыми ресурсами (для тех публикаций, которые имеют набор метаинформации на английском языке: библиографическое описание, реферат или аннотацию, набор ключевых слов), так и с российскими научными электронными коллекциями, поддерживающими шлюзы на базе протоколов ОАИ. Использование данных технологий обеспечивает постепенное появление в мировом научном информационном пространстве описаний российских научных публикаций (со ссылками на их полные тексты). При этом важным фактором является распространение современной культуры электронных научных публикаций, ставших стандартами де-факто в мировом научном коммуникационном пространстве.

Литература

- [1] Шварцман М. Дополнительные критерии оценки ЭБС. URL: <http://www.aselibrary.ru/blogs/archives/840/>.
- [2] Борисовский В.Ф. Организация открытого архива научных публикаций сотрудников ОИЯИ. URL: http://rcdl2009.krc.karelia.ru/doc/full_text/451_458_Section15-1.pdf.
- [3] The Citation Style Language - open and free citation styles <http://citationstyles.org>.
- [4] <http://zotero.org>.
- [5] Четвериков А. FAQ по Mendeley и Zotero <http://chetvericov.ru/faq-po-mendeley-i-zotero/#.UEj03pZGwa8>.
- [6] Мбого И.А., Чугунов А.В. Электронная библиотека «электронное государство»: система управления контентом и задачи интеграции ресурсов // Интернет и современное общество: Труды XIII Всероссийской объединенной конференции. Санкт-Петербург, 19 – 22 октября 2010 г. — СПб., 2010. С. 45 – 48.
- [7] Мбого И.А. Инициатива открытых архивов: информационные, методические и технологические аспекты / И.А. Мбого, И.Н. Карнуп, А.В. Чугунов // Интернет и современное общество: Труды XII Всероссийской объединенной конференции. Санкт-Петербург, 27 – 29 октября 2009 г. — СПб.: Факультет филологии и искусств СПбГУ, 2009. С. 56 – 59.

«E-Government» Electronic Collection: Technology Aspects

Irina A. Mbogo, Andrei V. Chugunov

The electronic collection can use multilevel qualifiers. Formation of records is carried out on the basis of filling the fields of forms bounded a binding to qualifier section. Various printing editions (books, articles, author's abstracts), electronic editions, artifacts of the real world can act as described resources. The form fields support all types of data Dublin Core and include additional possibilities for the expanded description.

* Работа выполняется в поддержке Российского фонда фундаментальных исследований (грант 11-07-00555-а).