

Формирование тематических электронных коллекций вуза культуры и искусств

© Халиков Л. И.

Казанский государственный университет культуры и искусств
kafedra@kazguki.ru

Аннотация

Одним из важных направлений в разработке систем электронных библиотек, активно развивающихся в последние годы, является создание для них коллекций информационных ресурсов. Изложены возможные методы и средства формирования тематических электронных коллекций электронной библиотеки вуза культуры и искусств.

1 Цели и задачи

В научных и образовательных учреждениях существуют значительные объемы информации в виде множества разнообразных коллекций и отдельных информационных ресурсов (ИР). Становится очевидным путь повышения эффективности использования в вузе накопленной и вновь создаваемой информации, основанный на создании таких информационных систем (ИС), которые способны надежно сохранять информацию и обеспечивать ее целенаправленное использование. Новым и интенсивно развивающимся классом подобных систем являются электронные библиотеки (ЭБ).

В настоящее время в Казанском государственном университете культуры и искусств (КГУКИ) по решению ученого совета принята концепция создания ЭБ вуза и начата разработка ее модели [2]. Целями формирования этой ЭБ являются создание тематических коллекций электронных документов и обеспечение их доступности всем категориям пользователей (локальным и удаленным). Для достижения названных целей были поставлены следующие задачи: изучение принципов организации, технологии формирования электронных коллекций; форматов метаданных как инструмента описания и идентификации электронных документов и принципов создания лингвистического обеспечения коллекций. Методологической основой разработки электронных коллекций для вуза культуры и искусств послужила работа [1].

В ЭБ вуза культуры и искусств предполагается формирование следующих полнотекстовых коллекций:

- авторефератов диссертаций, защищенных в КГУКИ;
- трудов сотрудников университета;
- материалов конференций, проводимых в КГУКИ;
- учебно-методических материалов;
- электронная версия научного журнала «Вестник КГУКИ».

2 Модель создания документов

Создание электронного документа коллекции предполагает отделение содержательных данных документа (контента) от информации, описывающей формат его представления. Для этого с помощью XML-технологий можно задать логическую разметку документа в соответствии с некоторым шаблоном, называемым моделью документа, и описать его представление и правила трансформации.

Алгоритм создания электронного документа предполагает:

- разработку соответствующих DTD или XML Schema описания структуры XML-документов коллекций;
- формирование соответствующих XML-документов на основе разработанных DTD или XML Schema;
- разработку XSLT-преобразований XML-документов в различные форматы (HTML для браузера, PDF для печати).

Применение XML и сопутствующих технологий позволяет формировать следующие электронные документы: краткое библиографическое описание, полное описание, полнотекстовый документ.

3 Методика разработки и развития схем

Одной из наиболее перспективных моделей метаданных на сегодняшний день является модель RDF (Resource Description Framework). Она позволяет представлять семантическую структуру XML-документов и выражать смысл этих и иных ресурсов WWW.

В настоящее время заметна тенденция стандартизации RDF-словарей свойств метаданных для конкретных предметных областей - так называемых «обменных схем» или «профилей метаданных». Использование терминов (свойств, словарей значений и пр.), зафиксированных в стандартах, позволяет приложениям легко интегрироваться между собой,

обмениваться информацией, понятной им всем. Например, Dublin Core Metadata Initiative (DCMI) определил минимальный набор свойств для описания цифровых ресурсов Web, а также их детализацию в рамках «общего профиля». Возможность развития схем метаданных (наращивание уровней, глубины описания) в рамках стандарта DC позволяет описывать характеристики отдельных видов электронных документов в ЭБ.

Разбиение схемы метаданных на последовательно наращиваемые подсхемы становится возможным благодаря свойственной RDF «децентрализации» данных: каждая схема рассматривается как набор утверждений, а расширенная схема - как набор дополнительных утверждений, вдобавок к утверждениям базовой схемы.

Таким образом, в качестве верхнего уровня описания ИР вуза можно использовать систему метаданных DC, а в качестве нижнего уровня - уточняющие квалификаторы, учитывающие специфику конкретного ИР коллекции.

4 Описание предметной области

Помимо метаописаний, содержащих стандартизованную информацию о ресурсах, большое значение для эффективной работы с метаданными представляют поля метаописаний, из которых выбираются элементы, характеризующие данный ресурс в рамках конкретной предметной области научного знания. Тематическим полем формирования электронных коллекций в создаваемой ЭБ является область культуры и искусств.

По степени «продвинутости» можно определить следующую градацию описаний предметной области (ПО):

- неформальное описание; разработчик метаданных принимает самостоятельные решения о том, какие значения вводить в элементы метаописаний;
- значения для метаданных выбираются из фиксированного словаря; словарь - неупорядоченное множество слов (символьных строк), не обладающее внутренней структурой;
- таксономия - значения элемента метаданных классифицируются; это формирует ПО как некоторую иерархическую структуру; значения разбиваются на классы объектов, классы взаимодействуют друг с другом на основе правил наследования;
- онтология - это таксономия плюс описание свойств предметной области на некотором формальном языке, имеющем логическую семантику.

В онтологии объекты через иерархию классов и описание свойств связаны различными отношениями с другими объектами. Сами отношения могут

обладать свойствами, полезными для получения информации (например, транзитивностью или симметричностью). Поэтому, как только некоторое значение из онтологии присваивается элементу метаданных ресурса, сам ресурс вовлекается в целую систему взаимосвязей, определяемую этой онтологией.

В настоящее время для создания и поддержки онтологии существует целый ряд инструментов, позволяющих экспертам по предметной области легко и свободно вводить свои знания в ПО. Например, это Protege - свободно распространяемое интегрированное инструментальное программное средство.

Таким образом, с одной стороны у экспертов в области культуры и искусств появляется возможность накопления знаний в виде онтологии. С другой стороны, формализованное представление предметной области позволяет осуществлять семантический поиск в ЭБ в рамках заданной предметной области и выдавать соответствующие ИР, релевантные содержанию запроса.

Использование международных стандартов представления метаданных ИР, метаязыка XML, онтологического подхода, позволяющего описывать термины и отношения между ними, модели RDF, устанавливающей способ представления значений, определенных в онтологии, дает возможность организовать функционирование ЭБ вуза культуры и искусств на базе современных технологий Semantic Web.

Литература

- [1] Коголовский М.Р. Систематика коллекций информационных ресурсов в электронных библиотеках. Программирование. МАИК "Наука"/Интерпериодика. - 2000. - No 3.
- [2] Федоров А.О., Халиков Л.И. Модель электронной библиотеки как база формирования информационного пространства ИД-ПО//Модернизация и развитие дополнительного профессионального образования в едином культурно-образовательном пространстве региона: материалы Межвузов, науч.-практ. конф., Казань, 28 янв. 2005 г./Сост. М.С. Галявиева. - Казань: Изд-во Казанск. ун-та, 2005. - С. 162-165

Development of subject-matter digital collection of culture and arts university

Khalikov L. I

The work presents possible methods and conditions of development of subject-matter digital library of culture and arts university.