

СЕРВИСЫ ЭЛЕКТРОННЫХ ЕСТЕСТВЕННОНАУЧНЫХ КОЛЛЕКЦИЙ, ПОСТРОЕННЫЕ НА ОСНОВЕ ТЕХНОЛОГИИ MATHML

А.М. Елизаров, Е.К. Липачев, М.А. Малахальцев

Организация специализированных электронных научных ресурсов и управление ими как научная проблема исследовалась нами в рамках нескольких проектов Российского фонда фундаментальных исследований (РФФИ), см., например, [1]. В качестве технологического решения при формировании и поддержке физико-математических электронных научных изданий были выбраны инструменты семантического веба (<http://www.w3c.org>). Поскольку основным компонентом (хотя бы с точки зрения создания полного цикла обработки) в этих изданиях являются математические формулы, ключевой технологией созданных систем стал язык математической разметки MathML (см., например, [2]).

Процесс замещения книг их электронными аналогами ставит целый ряд задач по организации системы хранения, учитывающей сетевой характер как получения информации, так и всех последующих операций с ней. Любую электронную коллекцию с точки зрения сервисной парадигмы можно рассматривать как набор сетевых сервисов, существенно расширяющих возможности работы с информацией, как правило, недоступные для привычных бумажных источников (мобильность – доступ из зоны сети, «живая» связь с источником – гипертекст, расширенный поиск, в том числе по формулам и т. д.). Ряд таких сервисов создан при реализации проекта автоматизации работы электронного журнала «Lobachevskii Journal of Mathematics» (<http://ljm.ksu.ru>). Один из них можно рассматривать как прототип сервиса электронной математической коллекции.

Ставилась задача создать «переносимую» коллекцию электронных математических документов, которая может быть размещена на внешнем носителе (компакт-диске или флэш-накопителе), жестком диске персонального компьютера, сетевом диске или в интернете. Результатом реализации такого сервиса коллекции стал компакт-диск, содержащий статьи журнала «Lobachevskii Journal of Mathematics» [3].

Структура коллекции состоит из двух частей – хранилища статей и поисковой системы. Хранилище статей представляет собой множество каталогов, в каждом из которых помещены HTML-файлы, содержащие информацию о статье (автор, название, выходные данные) и ссылки на файлы текста статьи в форматах .dvi, .ps, .pdf, XHTML (в последнем хранится текст статьи в формате MathML), а также сами файлы текстов статей в указанных форматах.

Поисковые функции реализованы на Javascript. Связка Firefox + Javascript оптимально реализует сервис со следующими возможностями:

- вывод списка ссылок на статьи, входящие в коллекцию; пройдя по ссылке, пользователь попадет на страницу статьи, содержащую реферат и ссылки на тексты статьи в форматах .dvi, .ps, .pdf и MathML;
- вывод списка авторов статей, входящих в коллекцию; пройдя по ссылке, пользователь попадет на страницу, содержащую все статьи данного автора, включенные в коллекцию;
- поиск по авторам, заглавиям, ключевым словам, рефератам, тексту статей, математическим формулам; при поиске по тексту статей и формулам выводится список ссылок на тексты статей в формате MathML, а в открывающихся MathML-документах – текст, входящий в поисковый запрос; в качестве достоинств такого подхода отметим, что сервис функционирует одинаково, независимо от размещения коллекции, способа доступа к ней и операционной системы, установленной на компьютере пользователя;
- перенос коллекции с одного носителя на другой осуществлялся простым копированием файлов и не требует перенастройки;
- в коллекции реализован не только стандартный поиск (по авторам, названиям, ключевым словам и т. д.), но и по фрагментам формул.

Поскольку создано целое семейство специализированных языков разметки, предложенный подход может использоваться для управления междисциплинарным контентом (см., например, [2,4]).

В докладе будут подробно представлены описанные выше результаты, а также выходящая из печати монография [2].

Работа выполнена при финансовой поддержке РФФИ (проект 09-07-12059офи_м).

ЛИТЕРАТУРА:

1. Веселаго В.Г., Елизаров А.М., Липачёв Е.К., Малахальцев М.А. Формирование и поддержка физико-математических электронных научных изданий: переход на технологии семантического веба// В кн. «Научно-исследовательский институт математики и механики им. Н.Г. Чеботарева Казанского государственного университета. 2003 – 2007 гг.». Кол. монография под ред. А.М. Елизарова. – Казань: Изд-во Казан. ун-та, 2008. – С. 456-476.
2. Елизаров А. М., Липачев Е. К., Малахальцев М. А. Веб-технологии для математика: Основы MathML. Практическое руководство. – М.: Физматлит, 2010. – 200 с.

3. Lobachevskii Journal of Mathematics (1998 - 2007). LJM disk is created by M.A. Malakhaltsev, 2009.
4. Елизаров А. М., Липачев Е. К., Малахальцев М. А. Организация взаимодействия языков разметки в системе автоматизации электронных научных хранилищ: семантический подход // Научный сервис в сети Интернет: масштабируемость, параллельность, эффективность: Труды Всероссийской суперкомпьютерной конференции (21-26 сентября 2009 г., г. Новороссийск). – М.: Изд-во МГУ, 2009. – С.456-457.