

Электронные научные издания: переход на технологии семантического Веба

В.А. Глухов, А.М. Елизаров, Е.К. Липачёв, М.А. Малахальцев

Аннотация

Рассмотрено применение технологий семантического Веба в электронных научных коллекциях. Изложены подходы к поддержке электронных коллекций на основе XML, RDF и других технологий семантического Веба. В частности, решены вопросы генерации метаданных, организации хранения и поиска данных. Эти подходы реализованы в проекте Научная электронная библиотека (НЭБ) eLibrary.ru и электронном журнале «Lobachevskii Journal of Mathematics» (<http://ljm.ksu.ru>).

1 Введение

В конце 20-го века развитие информационно-телекоммуникационных технологий и средств представления информации привело к тому, что появились электронные версии научных журналов, чтение которых сначала было возможно лишь на отдельных компьютерах в локальных сетях, а вскоре и через Интернет. Западные издатели стали предлагать своим подписчикам электронные версии журналов как дополнительную услугу. Например, только на электронные издания издательства Elsevier (www.elsevier.ru) зарегистрировано 6390 подписчиков, из которых более двухсот из России и СНГ (по данным доклада менеджера по работе с подписчиками издательства Elsevier Г. Якшонов – <http://isl.ksu.ru/images/konf2006/yakshonok2.pdf>). Впоследствии электронные версии стали реализовываться независимо от их печатных аналогов.

Семантический Веб разрабатывается консорциумом W3C (<http://www.w3c.org>) как перспективная «машиноориентированная» технология, которая предназначена для замены традиционных веб-технологий, требующих непосредственного участия человека в большинстве операций по обработке данных (см., например, [1]). Текущие проблемы, связанные с развитием семантического Веба, постоянно обсуждаются специалистами (см., например, материалы конференции WWW2002 Workshop on Real World RDF and Semantic Web Applications – <http://www.cs.rutgers.edu/~shklar/www11/>). Поддержка технологий семантического Веба реализуется в ряде крупных исследовательских проектов (например, http://www.ilrt.bris.ac.uk/projects/semantic_web, <http://www.ilrt.bris.ac.uk/people/cmdjb/>), направленных на обеспечение семантической интероперабельности в Web.

В докладе (<http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html>; см. также [2]) Бернерс-Ли предложил диаграмму «SemanticWeb layer cake», которая дает представление о многослойной архитектуре семантического Веба, включающей, в частности, Extensible Markup Language (XML), Resource Description Framework (RDF), RDF Schema и Web Ontology Language (OWL). Структурирование информации производится на основе XML, а RDF используется для создания расширенной системы метаданных на основе иерархических моделей.

Наиболее злободневные технологические проблемы развития электронных библиотек приведены в работе [3, 4]. Решение этих проблем необходимо и для развития сети электронных журналов. В первую очередь, это относится к совершенствованию

методов представления электронных ресурсов и определения состава метаданных. Методы представления математических текстов в электронной форме с использованием семантического Веба описаны в [5].

В 1996 году в мире существовало всего около 250 электронных журналов, однако уже спустя два года их количество увеличилось до 2 тысяч. Сегодня издается, по разным оценкам, от 20 до 30 тысяч электронных научных журналов. Все крупнейшие западные издательства публикуют свои журналы в электронном виде. Появилось значительное количество электронных журналов открытого доступа (см., например, <http://www.doaj.org/>). Подавляющее большинство электронных журналов доступно через Интернет и объединено в крупные базы данных. По данным издательства Elsevier электронный архив Science Direct включает 2347 полнотекстовых электронных журналов, содержащих более 7,8 млн. статей.

Создание и широкое распространение электронных журналов стало возможным благодаря развитию технологий электронного книгоиздания, специализированных форматов (в основном на основе SGML и XML), средств телекоммуникаций и Интернет, программных средств обработки данных. Появление электронных журналов обуславливалось еще рядом факторов. Например, высокой стоимостью подписки на печатные версии журналов и ограничением распространения традиционных изданий. Из-за высокой стоимости печатных изданий многие журналы и книги не приобретаются российскими библиотечными учреждениями или доступны в регионах в одной – двух библиотеках; 75% объема рынка российских печатных СМИ, по оценкам Гильдии издателей периодической печати, приходится на Москву. Преобладающий объем столичного рынка обусловлен также тем, что издания не могут пробиться в регионы из-за отсутствия системы распространения печатной периодики.

В области электронного книгоиздания Россия не является лидером. В 2005 году нами была проведена экспертиза почти тысячи журналов, входящих в список Высшей аттестационной комиссии (ВАК) РФ. Из них в Интернете так или иначе присутствовало около 300 журналов. Детальное изучение сайтов показало, что большинство журналов представлено только в виде оглавлений с аннотациями. Часть журналов представлена 1 – 2 годами изданий, публикация которых в электронном виде была осуществлена несколько лет назад. Количество «действующих» электронных журналов из списка ВАК (т. е. тех, которые были опубликованы в 2004 – 2005 гг.) не превышало 100 наименований. Практически на всех сайтах отсутствует возможность поиска по авторам, названиям статей, ключевым словам, аннотациям и особенно – по полным текстам статей. Тексты представлены в разных форматах – html, PDF, DjVu, Word и т. д. Недостаточное развитие российских электронных научных журналов в целом снижает рейтинг российских изданий и приводит к тому, что ученые стали отдавать предпочтение публикации своих работ в западных журналах. Отметим несколько причин недостаточного развития электронного книгоиздания.

Прежде всего, не разработаны форматы представления электронных изданий, обеспечивающие структурирование текстов статей и учет этой структуры для загрузки журналов в базы данных. Форматы записей, давно и успешно применяемые в библиотечном деле, такие, как US Marc, Unimarc или Rusmarc, предназначены лишь для описания печатных источников (книг, журналов в целом, статей) на библиографическом уровне. Для описания полных текстов журналов или отдельных публикаций эти форматы не могут быть использованы из-за отсутствия необходимых спецификаций. Поскольку библиотечная каталогизация направлена только на описание печатных источников, то указанные форматы (при всей их детализации) зачастую не включают описания таких

элементов, как, например, подробные сведения об индивидуальных авторах (ученое звание, место работы, почтовый и электронный адрес). Ни один из библиотечных форматов не поддерживает описаний полного текста статьи из журнала, главы из книги или текста книги целиком. Полностью отсутствует описание (тем более детализированное) библиографических списков, что весьма важно при построении индексов научного цитирования. Точно так же не могут применяться описания, построенные на языках метаописания, подобных Dublin Core (DC) (см. <http://purl.oclc.org/dc/>), используемых в основном для библиографического описания Интернет-страниц. Поэтому требуется разработка форматов, предназначенных специально для описания полнотекстовых электронных изданий.

Кроме того, отсутствует специализированное программное обеспечение, позволяющее производить углубленную обработку электронных версий журналов с целью их последующей загрузки в базы данных. Применяется достаточно много издательских пакетов программ (PageMaker, Word и др.), однако почти все они предназначены лишь для набора текста, оформления и создания макетов изданий, с которых производится типографское тиражирование. Использование таких макетов для загрузки электронных изданий в базы данных не предусматривается и не может быть выполнено. Необходима разработка программно-технологических комплексов, предназначенных для структурирования (разделения на поля) электронных версий печатных изданий. При разработке следует учитывать особенности представления специализированных текстов, например, по математике.

Издательства и редакции не всегда идут на публикацию электронных изданий. Возможно, это связано с опасением утратить подписку на печатный вариант издания. Кроме того, юридическая база электронных публикаций недостаточно проработана и, как следствие, традиционные (печатные) варианты публикаций пока считаются более привлекательными. Необходимым условием функционирования электронного журнала является признание научным сообществом равноценности электронной и традиционной (в «бумажном» научном журнале) публикаций. Это накладывает на редакции электронных журналов ту же ответственность, что и в любом традиционном научном журнале, в частности, по организации независимого научного рецензирования.

Необходимо предложить владельцам журналов разумные схемы взаимодействия с электронными библиотеками, включая подготовку пакетов юридических документов по прямому лицензированию. Кроме того, необходимо подготовить юридическую документацию, регулирующую проблемы авторского права применительно к электронным публикациям – на уровне «автор – редакция» или «автор – издательство».

По проекту Министерства образования и науки «Создание электронных версий российских научных журналов» (шифр РИ-13.0/006) в 2005 году Казанским государственным университетом (КГУ) и Научной электронной библиотекой были разработаны программы и технологии обработки электронных версий печатных журналов, созданы форматы данных, подготовлен пакет юридических документов, регулирующих отношения авторов, издателей и электронных библиотек. При этом решались задачи обработки как печатных изданий, не имеющих «представительства» в Интернет, так и электронных журналов.

Научная электронная библиотека (НЭБ) eLibrary.ru была создана в 1998 году в результате выполнения одного из проектов РФФИ и впоследствии приобрела характер национального проекта. История формирования и развития НЭБ подробно отражена, например, в [6, 7, 8].

Одной из целей исследований, проведенных в рамках проектов РФФИ 02-07-90230 и 03-07-90252 и проекта 03-03-12007 Российского гуманитарного научного фонда в 2003 – 2005 гг., была разработка полнофункциональной программной среды электронного научного журнала по математике, позволяющей автоматизировать ряд процессов, стандартных для научного издания.

Такая среда была создана для электронного журнала LJM (Lobachevskii Journal of Mathematics). LJM – одного из первых отечественных научных электронных журналов – издаваемого с 1998 года (дата регистрации в Министерстве по делам печати и информации – 2 августа 1996 года). За это время вышло более двадцати томов журнала. Статьи российских авторов занимают примерно половину объема журнала. В журнале публикуются также работы авторов из Армении, Индии, Испании, Канады, Китая, Марокко, Норвегии, Сербии, США, Финляндии, Японии. Редколлегия журнала возглавляет крупнейший российский математик, академик РАН С. П. Новиков. В составе редколлегии – ведущие специалисты в большинстве областей современной математики. LJM представлен на сервере Европейского математического общества, статьи реферируются в журналах Mathematical Reviews, Zentralblatt fur Mathematik, информация о журнале регулярно появляется в Notices of American Mathematical Society. В настоящее время журнал включен в базы данных Science Direct издательства Elsevier (<http://www.elsevier.com>) и Научной электронной библиотеки eLibrary.ru (<http://elibrary.ru>).

Особенностью журнала является разнообразие его тематики: в журнале представлены работы по алгебре, геометрии и топологии, комплексному анализу, функциональному анализу, теории вероятностей и математической статистике, оптимальному управлению, теории алгоритмов. Это создает дополнительные сложности в процессе автоматизации документооборота редколлегии электронного журнала.

В силу указанных обстоятельств руководство журнала особое внимание уделяет внедрению современных информационных технологий. Разработан и внедряется автоматизированный комплекс редакционной обработки статей на основе технологий семантического Веба. Он позволяет производить регистрацию статьи и ее предварительную обработку, в том числе преобразование в форматы представления. Наряду со стандартными наборами форматов представления математических статей (dvi, pdf, ps) журнал представляет статьи в формате MathML (впервые в мире среди электронных журналов). Проведена адаптация имеющихся программных средств конвертации TeX-файлов в MathML-файлы и встраивания их в автоматизированную систему публикации статей.

LJM функционирует в системе международных электронных научных журналов, издаваемых Казанским университетом («Journal of Formal, Computational and Cognitive Linguistics», LJM, «Magnetic Resonance in Solids», «Информационные технологии и телерадиокоммуникации»). Переход на новые технологии позволил вывести работу этих журналов на совершенно иной уровень, приближающийся к современным информационным стандартам, и существенно повысил эффективность их использования. Отметим, что в работе журнала LJM применены подходы, которые в определенной степени отличаются от принятых. Разработанные нами технологии апробированы в «Lobachevskii Journal of Mathematics» и предлагаются в качестве возможного варианта автоматизации работы редколлегий электронных журналов, прежде всего, математических (см. [9]).

2 Технологии семантического Веба в электронных изданиях

В настоящее время технологии семантического Веба используется в ряде зарубежных библиотек для долговременного хранения данных. В то же время, отсутствуют российские форматы представления электронных изданий, предполагающие детальное разбиение элементов изданий, которые могут быть использованы для загрузки полнотекстовых журналов и книг в электронные хранилища. Издатели электронных журналов сейчас применяют, как правило, html-разметку выпусков своих изданий, не следуя при этом никаким общепринятым правилам, поскольку их просто не существует. Кроме того, такая разметка не может быть использована для загрузки в электронные хранилища.

Для структурирования изданий в рамках проекта «Научная электронная библиотека eLibrary.ru» были разработаны описание структуры XML-документа на языке DTD (Sarticle.dtd) и программа Sarticle, использующая это описание

Перечислим основные составляющие DTD-описания Sarticle.dtd:

- раздел описания журнала в целом, куда входят сведения о названии журнала, издатель, ISSN, обобщенной структуре издания (том – номер – часть – спецвыпуск), а также поля, позволяющие описать отдельный выпуск журнала;
- сведения о статье из выпуска журнала, куда входят описание индивидуальных и/или коллективных авторов статьи с подробной информацией о них, название статьи, ключевые слова, реферат (аннотация), полный текст статьи без списка литературы, коды наиболее распространенных классификаторов (УДК, ББК, ГРНТИ, DOI для электронных изданий и др.), а также подраздел, описывающий пристатейные списки литературы; при этом каждая позиция в списке литературы (или сноске) разбита на отдельные элементы – например, автор(ы) работы, название, источник, год издания и т. д.;
- раздел тематических рубрик журнала, куда входит описание подразделов выпуска журнала.

Отметим, что технически возможно в одном XML-документе описать любое количество журналов, но с точки зрения удобства хранения и заполнения эффективнее использовать для каждого выпуска журнала отдельный XML документ.

Используемые в современном книгоиздании программы, как правило, предназначены лишь для набора текста, оформления и создания макетов изданий, с которых производится типографское тиражирование. Для создания файлов, предназначенных для загрузки в электронные хранилища, необходимо было разработать специальное программное обеспечение., которое обеспечивает структурирование (разметку) текстов электронных версий изданий. Отметим, что попытки создания специальных программ обработки макетов изданий предпринимались неоднократно, однако ни одна из них не увенчалась успехом. Проблема в том, что корректная работа подобных программ с макетами изданий возможна лишь в том случае, если сам макет подготовлен с тщательным соблюдением правил, что практически неосуществимо.

Для решения проблемы были разработаны три версии программного обеспечения для разметки электронных изданий, в основу которого был положен принцип выделения фрагментов текста в соответствии с описанием Sarticle.dtd.

Кратко опишем взаимодействие пользователя с программой Sarticle. Входными данными для программы являются электронные документы в форматах MS Word (rtf и doc), Adobe Acrobat^[1], текстовые файлы (html и txt), содержащие полные версии журналов, сборники статей или иную библиографическую информацию. Кроме того, возможна обработка Интернет-страниц журналов. Разметка документа производится пользователем последовательным выделением в тексте этого документа элементов библиографического описания. Каждый выделенный элемент связывается с соответствующим элементом, входящим в DTD-описание. На этапе такого выделения и связывания элементов формируется визуальный образ будущего XML-документа, представляющий собой пакет древовидных иерархических структур.

В программе используется два типа таких структур: для отображения библиографических данных журнала и для отображения библиографических данных статьи из журнала. Каждая такая структура формируется как самостоятельный визуальный объект и размещается на отдельной странице. Программа производит верификацию и конвертацию данных с формированием выходного XML-документа отдельного выпуска (номера) журнала.

Был разработан специальный модуль программного обеспечения, позволяющий производить автоматическое структурирование библиографических списков и сносок. Разработка такого модуля необходима для подготовки библиографических материалов, включаемых в различные индексы научного цитирования.

Модуль поддерживает работу с библиографическими списками, которые подготовлены в соответствии с ГОСТ 7.1-84 «Библиографическое описание документа». Работа модуля заключается в автоматическом выделении из библиографического описания книги, статьи из журнала или сборника, патента, авторского свидетельства, на которые ссылаются авторы оригинальных статей. Производится выделение следующих фрагментов текста: автор, название, источник (название журнала или сборника), место издания и издательство, год издания, том, номер, часть, страницы, а также URL-адрес электронной версии источника.

Разработанное программное обеспечение прошло успешное тестирование в ряде редакций научных журналов. Следующим шагом было создание новой версии программы, позволяющей работать с большим количеством исходных файловых форматов и максимально автоматизирующей процесс структуризации текста. Модифицированная версия программы, созданная в рамках выполнения проекта, позволила расширить перечень обрабатываемых файловых форматов. Кроме того, добавлены два новых специальных модуля для программного распознавания группы авторов с их описаниями (место работы, почтовый и электронный адреса и т. д.), ключевых слов, а также библиографических списков. Отметим, что программа обрабатывает диакритические, математические и другие специализированные символы и позволяет сохранять XML-файлы как в формате ASCII, так и в формате Unicode (UTF-16). В результате разработки и использования программы время обработки усредненного выпуска журнала (115 страниц, 3 автора каждой статьи, 17 – 18 статей и 15 – 20 пристатейных ссылок) составляет 2 – 3 часа рабочего времени.

Загрузка данных в базу данных реализована с использованием технологии Windows Script Host (WSH). Создан универсальный загрузчик на VBScript, осуществляющий парсинг данных и записывающий результат в базу данных. При этом исходные тексты статей в форматах pdf и/или html записываются на файловый сервер. Апробация программ загрузки данных в базу электронной библиотеки выполнялась на 60 выпусках журналов

издательства Института научной информации по общественным наукам РАН. Пользователи электронной библиотеки могут проводить поиск статей из журналов по следующим параметрам: авторы, названия статей, аннотации (рефераты), ключевые слова, слова из полных текстов статей, библиографические описания источников из пристатейных списков литературы.

По разработанной технологии, описанной выше, в НЭБ были размещены более 200 научных журналов. Среди них такие авторитетные издания, как «Успехи физических наук», «Успехи химии», «Биология моря», «Ученые записки Казанского университета», «Известия вузов. Авиационная техника», «Известия вузов. Математика», «Известия вузов. Радиофизика», «Казанский медицинский журнал», «Вестник Казанского государственного технического университета», «Вестник Дальневосточного отделения РАН», «Дальневосточный математический журнал», «Тихоокеанская геология» и ряд других. Большинство из этих журналов до 2005 года не было представлено в Интернете.

Некоторые разработки в рамках совместных проектов НЭБ и КГУ были положены в основу создаваемого Научной электронной библиотекой Российского индекса научного цитирования (см. <http://www.elibrary.ru/projects/citation/proposal.doc>).

^[1] Требуется предварительная обработка исходных файлов в этом формате

3 Технологии семантического Веба в электронном математическом журнале

Выбор технологий семантического Веба для организации работы математического электронного журнала объясняется наличием в них инструментов, позволяющих учитывать структурную и семантическую составляющие информации. Наиболее полно современное представление о методологии применения электронных технологий к формированию и хранению научной информации отражено в рекомендациях консорциума W3C (www.w3.org), а также в разработках корпорации Wolfram Research [10].

Программные решения, применяемые в настоящее время при подготовке научных изданий, основаны на технологиях HTML и, в значительной степени, предполагают участие человека в обработке информации, поскольку теговая разметка языка HTML позволяет структурировать текст только в части отображения документа. Подготовка математических текстов производится в TeX-формате (обязательное требование большинства математических журналов). Это текстовый формат с теговой разметкой, обеспечивающей форматирование документа и включение математических формул (<http://ctan.org>). TeX-формат также является слабоструктурированным. Автоматизация процедур обработки слабоструктурированной информации затруднительна и не всегда эффективна. В частности, поиск информации нельзя полностью автоматизировать из-за сложности процедуры извлечения данных.

Отдельной задачей, стоящей перед научными изданиями, является обеспечение возможности извлечения метаданных («информации об информации») поисковыми системами, большая часть которых для индексирования применяет программы-роботы, использующие метаописание ресурсов сети. Метаданные содержат обобщенную информацию о структуре и содержании информационного источника (автор, дата, источник, ключевые слова, предметная область и т. д.). В формате HTML описание метаданных возможно только через метатеги. Для наиболее унифицированного описания

и каталогизации ресурсов в сети создаются специальные метаязыки, наиболее распространенным из них является Dublin Core (DC). Автоматизация генерации метаданных затруднена в силу слабой структурированности представления информации. Более того, для составления блока метаданных каждого документа требуется участие квалифицированного специалиста.

Основная проблема при создании, хранении и отображении электронных публикаций по математике касается представления математических формул. Наиболее распространенное на данный момент решение – представление формул в виде графических файлов – неудовлетворительно с точки зрения структурной обработки математических текстов [11].

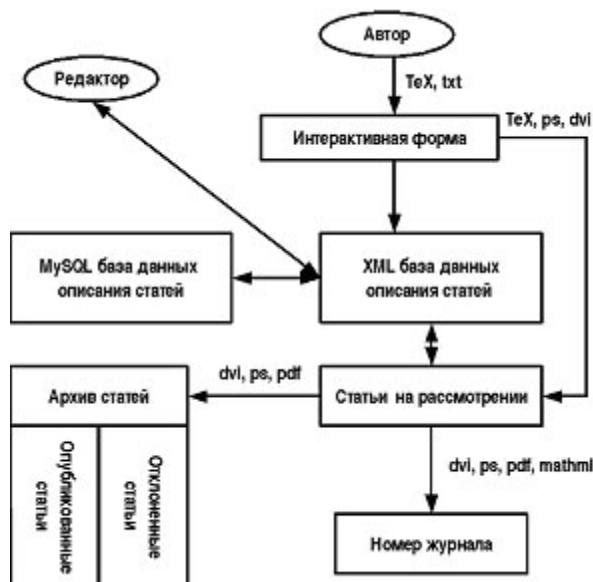
Для решения указанных проблем можно применить стандарты и рекомендации, разработанные консорциумом W3C в рамках проекта Semantic Web. В частности, консорциумом W3C (<http://www.w3.org>) разработан язык математической разметки – Mathematical Markup Language (MathML). ([12], [13]). MathML – это официальная рекомендация консорциума W3C, одобренная математическим сообществом, представляющая из себя определение типа документа (DTD) для языка XML. Применение XML-технологий позволяет решать задачи программной обработки документов (в частности, задачу поиска) на новом технологическом уровне и, поэтому, изменяет принципы организации и управления электронными публикациями по математике.

В настоящее время язык MathML становится стандартом представления математической информации в электронной форме в силу следующих причин: технология обработки данных на основе языка MathML реализует одну из основных тенденций современной информатики – разделение разметки и данных, поэтому она представляет широкие возможности многоуровневого структурирования данных и расширенного поиска; имеется возможность создания программного обеспечения, использующего технологию MathML; созданы и продолжают совершенствоваться программные средства, позволяющие конвертировать в MathML документы, подготовленные с помощью имеющихся стандартных технологий (LaTeX, Mathematica, Maple, MS Word). Из наиболее распространенных инструментов отметим редактор WebEq (www.dessci.com/en/products/webeq), конверторы TeX в MathML: TtM (<http://hutchinson.belmont.ma.us/tth/mml/>), TeX4ht (входит в стандартный пакет MiKTeX), а также конвертор XML в TeX (MikTeX). MathML поддерживается основными просмотрщиками: Internet Explorer (при установке соответствующего модуля [14, 15]), Mozilla, Amaya; технология MathML поддерживается системами Maple[®] и MathCAD 2001, а компания Wolfram Research предложила собственную концепцию использования технологии MathML [16], которая реализована в пакете Mathematica[®], в частности, в этом пакете предусмотрено сохранение документов в формате MathML.

Работа по автоматизации электронного журнала велась нами на основе CASE-моделирования. Была создана модель электронного математического хранилища, состоящая из комплекса UML-диаграмм.

Разработаны методы автоматической обработки электронной математической информации на основе архитектуры семантического Веба. Семантическая разметка веб-страниц выполняется на основе стандарта XML, при этом для разметки математической части текста использован MathML. Разработана иерархическая модель метаданных для ресурсов электронного математического журнала в соответствии со стандартом RDF. Разработаны методы использования технологии XSLT для преобразования математических текстов и служебной информации, представленной в XML-формате, в частности, MathML. Спроектирована архитектура таблиц MySQL базы данных

электронного математического журнала. Согласно разработанной схеме, в XML-файле накапливается информация, поступающая в результате пользовательского запроса, далее применяется XSLT-преобразование и результат перенаправляется в MySQL-базу. Это технологическое решение позволило существенно снизить нагрузку (количество запросов) на MySQL-сервер, поскольку каждый запрос работает с отдельным XML-файлом, без постоянного подключения к системам баз данных.



На диаграмме представлена принципиальная схема процесса редакционной обработки статьи. Каждому прямоугольнику отвечает программный модуль, выполняющий соответствующую процедуру. На стрелках указаны форматы, участвующие в операциях обмена данными

Указанные подходы были реализованы в виде программного комплекса для обработки и управления потоками данных в электронном журнале на основе XML/XSLT и MathML-технологий. Одной из составляющих этого комплекса является система PHP-скриптов, которая включает в себя, в частности, сервисы автоматизированного представления рукописей статей в журнал и автоматизированного прохождения рукописи. Ряд задач был решен с помощью скриптов на языках AWK и Perl. Для генерации и обработки запросов используется язык RDQL [17]. Создано программное обеспечение, осуществляющее автоматический перевод в RDF-формат метаданных, поступающих из интерактивной формы, предоставленной пользователю. Преобразование XML/RDF-информации в HTML-формат производится с помощью скриптов (Java, PHP). Осуществлена программная реализация управления потоками данных с веб-сайта в MySQL-базу с использованием XML-файлов в роли буфера.

Сервис автоматизированного представления рукописей включает в себя регистрацию авторов и внесение их данных в MySQL-базу данных; сбор метаданных из формы, заполняемой авторами публикаций; компиляцию TeX-файлов в форматы dvi, ps, pdf для редакторской работы; внесение основных данных о рукописи (название, AMS-классификация ((AMS – American Mathematical Society), ключевые слова, дата подачи рукописи) в MySQL-базу; извещение секретаря журнала о поступившей рукописи.

Сервис автоматизированного прохождения рукописей позволяет получить информацию о статусе статьи (наличие отзывов с датами представления), дает возможность технического

редактирования рукописи, генерирует файлы в форматах dvi, ps, pdf, mathml, а также генерирует html-файлы с информацией о статьях и о томе журнала. Также организовано автоматическое генерирование метаданных, представляемых средствами Dublin Core и RDF.

Разработанное программное обеспечение позволило осуществить перевод в формат MathML полных текстов ранее опубликованных статей. В настоящее время статьи хранятся как в формате MathML, так и в стандартных форматах (dvi, ps, pdf). Кроме того, на сайте журнала LJM размещены аннотации статей в формате MathML.

Работа с сайтом журнала предполагает, что клиент выполнит необходимые настройки своего браузера. Прежде всего, требуется обеспечить поддержку MathML. На сайте журнала (<http://ljm.ksu.ru>) имеются инструкция по настройке наиболее распространенных браузеров и список узлов с необходимым программным обеспечением.

В связи с тенденцией перехода на открытые Unix-подобные операционные системы разработанное программное обеспечение адаптировано к работе под управлением ОС Linux и использует средства Unix-ориентированных систем.

В заключение отметим, что разработанные программные средства внедрены в работу электронного журнала Lobachevskii Journal of Mathematics. В настоящее время все публикуемые в журнале статьи снабжаются комплексом метаданных, включающим смешанный набор метатегов. Этот набор состоит из традиционного набора метаданных и метаданных в формате DC. Это позволяет индексировать страницы журнала как поисковым системам, роботы которых распознают только традиционный формат метаданных, так и поисковым системам, рассчитанным на метаданные в формате DC.

Авторы благодарят М. Р. Когаловского за внимание к работе и полезные замечания.

Литература

1. Berners-Lee T. Semantic Web Road Map – <http://www.w3.org/DesignIssues/Semantic.html>
2. Passin T. M. Explorer's Guide to the Semantic Web //Manning Publication Company, 2004. – 280 с.
3. Когаловский М.Р. Стандарты XML и электронные библиотеки //Электронные библиотеки. – 2003. – Т. 6. – Выпуск 2.
4. Когаловский М.Р. Тенденции развития технологий управления информационными ресурсами //Труды 8^{ой} Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» – RCDL'2006, Суздаль, Россия, 2006. – С. 46-55.
5. Елизаров А. М., Липачев Е. К., Малахальцев М. А. Основы MathML. Представление математических текстов в Internet. Практическое руководство. – Казань: Изд-во Казан. матем. об-ва, 2004. – 60 с.
6. Глухов В.А., Новиков В.Д., Петров А.Н. Научная электронная библиотека: итоги и перспективы //Вестник Национального комитета «Интеллектуальные ресурсы России». – Москва, 2004. – № 2. – С. 9-14.
<http://www.elibrary.ru/item.asp?id=8826534>
7. Глухов В.А. Проект «Научная электронная библиотека eLibrary.Ru» и перспективы развития электронного книгоиздания в России //Educational Technology & Society. 2005. – Т. 8, № 1. – С. 191-197. <http://www.elibrary.ru/item.asp?id=8370428>

8. Глухов В. А., Елизаров А. М. Проект «Научная электронная библиотека eLibrary.ru» и российские электронные журналы: новый этап развития //Труды 8^{ой} Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» – RCDL'2006, Суздаль, Россия, 2006. – С. 203-207.
9. Елизаров А. М., Липачев Е. К., Малахальцев М. А. Технологии Semantic Web в практике работы электронного журнала по математике //Труды 8^{ой} Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» – RCDL'2006, Суздаль, Россия, 2006. – С.215-218.
10. Wolfram Research Contributes Central Ideas to Web Math Standard. – <http://www.wolfram.com/news/archive/mathml.html>
11. Митюнин В. А. Обзор средств публикации и просмотра математических документов в сети Интернет. – <http://mathmag.spbu.ru/article/4/>
12. Mathematical Markup Language (MathML) Version 2.0 (Second Edition) W3C Recommendation 21 October 2003 – <http://www.w3.org/TR/MathML2/>
13. Sandhu P. The MathML Handbook. – Charles River Media, 2003. <http://www.tphenwolfram.com/ews/>
14. MathPlayer Display MathML. – <http://www.dessi.com/n/product/mathplayer>
15. TechExplorer. – <http://www-3.ibm.com/software/network/techexplorer/>
16. Wolfram Stephen. Mathematical Notation: Past and Future. – <http://www.stephenwolfram.com/publications/talks/mathml/>
17. RDQL – A Query Language for RDF. – <http://www.w3.org/Submission/2004/SUBM-RDQL-20040109/>

Об авторах

Глухов Виктор Алексеевич - ИНИОН РАН, зам. директора
E-mail: vglukhov@inion.ru

Елизаров Александр Михайлович - НИИ математики и механики им. Н. Г. Чеботарева
Казанского государственного университета, директор
E-mail: elizarov@ksu.ru

Липачев Евгений Константинович - Казанский государственный университет, доцент
E-mail: lipachev@ksu.ru

Малахальцев Михаил Арменович - Казанский государственный университет, доцент
E-mail: Mikhail.Malakhaltsev@ksu.ru

© В.А. Глухов, А.М. Елизаров, Е.К. Липачёв, М.А. Малахальцев, 2007

Статья написана по материалам докладов авторов на 8ой Всерос. научн. конф. «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» – RCDL'2006, Суздаль, 17 – 19 октября 2006 г.