

ТЕХНОЛОГИИ ФОРМИРОВАНИЯ И ПОДДЕРЖКИ ЭЛЕКТРОННЫХ НАУЧНЫХ МАТЕМАТИЧЕСКИХ КОЛЛЕКЦИЙ: ОПЫТ КАЗАНСКОГО УНИВЕРСИТЕТА

А.М. Елизаров, Е.К. Липачёв

Казанский (Приволжский) федеральный университет, г. Казань, Россия

amelizarov@gmail.com, elipachev@gmail.com

Введение. Как известно (см., например, материалы электронного журнала «Электронные библиотеки», URL: <http://www.iis.ru/el-bib>), для современного информационного общества характерно стремительное развитие и активное использование таких информационно-коммуникационных технологий (ИКТ), которые обеспечивают не только сетевой обмен информацией, но и возможность интеграции локальных информационных ресурсов в единое информационное пространство. Эти ресурсы существенно влияют на интенсивность процессов обучения и научных исследований, поэтому обеспечение публичного (в том числе удаленного) доступа к ним стало одной из первоочередных задач оптимизации информационных процессов в образовании, науке и культуре. Сегодня общепризнано, что наиболее эффективный путь решения этой задачи связан с созданием электронных библиотек (ЭБ) – «распределенных информационных систем, позволяющих надежно сохранять и эффективно использовать разнообразные коллекции электронных документов (текст, графика, аудио, видео и др.), доступные в удобном для конечного пользователя виде через глобальные сети передачи данных» [1]. Ключевую роль в ЭБ играют сформировавшиеся в 1990-е годы технологии управления информационными ресурсами и коммуникационные средства, на которых базируются национальные программы развития электронных библиотек, принятые и реализуемые сегодня во многих странах. Основными составными элементами ЭБ являются коллекции электронных информационных ресурсов.

2. Электронные научные журналы и отечественная система электронного книгоиздания. Наиболее распространенными коллекциями научных ЭБ, обеспечивающими формирование новых видов информационных ресурсов и обмен научной информацией на базе современных ИКТ, являются сегодня электронные научные журналы.

В конце 20-го века развитие ИКТ и средств представления информации привело к тому, что появились электронные версии традиционных научных журналов. Сначала издательства стали предлагать своим подписчикам электронные версии журналов как дополнительную услугу, впоследствии электронные версии стали реализовываться независимо от их печатных аналогов.

В 1996 году в мире существовало всего 250 электронных журналов, однако уже спустя два года их количество увеличилось до 2 тысяч. По данным [2], в 2008 году 11 стран выпускали 76 % всех изданий (США – 34 %, Великобритания – 10 %, Германия – 7 %, Канада – 4 %, Франция, Италия, Австралия, Нидерланды, Япония, Китай, Испания – остальное), в том числе 75 % – научных, 84% – всех находящихся в интернете и 86% – научных, доступных в он-лайн. Издания России в этом ряду составляли всего 1,3 %.

Сегодня издается около 30 тысяч электронных научных журналов, и их число растет. Все крупные издательства публикуют свои журналы в электронном виде. Например, только в базы данных крупнейшего европейского издательства Elsevier включено более 14 тысяч рецензируемых электронных журналов. Появилось значительное количество электронных журналов открытого доступа (например, <http://www.doaj.org/>). В настоящее время практически все бумажные периодические издания имеют электронные версии, доступные через интернет и объединенные в крупные базы данных. На качественно новом технологическом уровне развиваются и электронные журналы, не имеющие печатных аналогов. Тем самым в мире формируется альтернативная система научных публикаций.

Создание и широкое распространение электронных журналов стало возможным благодаря развитию технологий электронного книгоиздания, специализированных форматов (как правило, на основе технологий Семантического Веба), средств телекоммуникаций и интернета, а также программных средств обработки данных. Однако появление электронных журналов обусловлено не только этими факторами – стоит упомянуть высокую стоимость подписки на печатные версии журналов и отсутствие системы распространения печатной периодики. Для электронных журналов нет таких ограничений, хотя система подписки на электронные издания в нашей стране начала формироваться только в последние годы. Как следствие, отсутствует единая система электронных научных публикаций, практически нет методик и пакета юридических документов, обеспечивающих процесс электронной публикации научных изданий, до сих пор большое число бумажных журналов, издающихся в России, не имеет общедоступных электронных версий. Все это затрудняет распространение результатов научных исследований и обмен последними достижениями между членами научного сообщества и представителями промышленности. Вместе с тем, издание электронных научных журналов в России является скорее инициативой отдельных научных или образовательных организаций, не связанной с предпринимательской деятельностью, чем развитием системы электронного книгоиздания в стране. По-видимому, одна из причин такой ситуации – в отсутствии в России научных издательств, имеющих большие пакеты научных журналов. Исключение составляет только академическое издательство «Наука», выпускающее около 200 научных журналов и сравнимое лишь с западным издательством среднего уровня. При этом многие академические научные и образовательные организации (научно-исследовательские институты и университеты), традиционно издающие бумажные научные журналы, нашедшие своего читателя и пользующиеся заслуженным авторитетом, создают электронные

версии своих изданий, выставляя их в открытый доступ сразу после выпуска бумажного экземпляра или с некоторой задержкой, а иногда распространяя электронную версию также по подписке.

Отдельную группу электронных научных журналов составляют чисто электронные издания, не имеющие бумажных версий. Преимущество этих изданий – в оперативности публикаций, отсутствии ограничений по объему, возможности оперативного ознакомления с публикуемыми научными материалами (сразу после принятия этих материалов в печать) самой широкой аудитории пользователей глобальной Сети. Кроме того, в чисто электронных научных журналах может быть резко ускорен и упрощен весь цикл подготовки, пересылки и рецензирования статей, а их издание существенно дешевле издания бумажных журналов, так как оно исключает все типографские проблемы.

С развитием интернета возрастают количество и роль электронных публикаций, при этом наличие чисто электронных журналов никак не может изменить баланса публикаций в бумажных журналах. В то же время надо понимать, что существующие ныне бумажные журналы смогут сохранить себя и по-прежнему играть заметную роль в процессе научных коммуникаций только в том случае, если начнут энергично развивать свои электронные версии. Уже общепризнано, что будущее – за эффективным сочетанием бумажных и электронных изданий.

Таким образом, задача включения российского научного сообщества в уже сложившуюся общемировую систему электронных научных публикаций является чрезвычайно актуальной. На этом пути уже имеется ряд ярких достижений. Это, прежде всего, реализованные проекты «Научная электронная библиотека (НЭБ) eLibrary.ru» и «Общероссийский математический портал MathNet.ru».

НЭБ eLibrary.ru была создана в 1998 году в результате выполнения одного из проектов Российского фонда фундаментальных исследований (РФФИ) и впоследствии приобрела характер национального проекта. История формирования и развития этой электронной библиотеки отражена, например, в [3].

Код	Название рубрики	Журналов	
27.00.00	Математика	962	
ISSN	Название журнала	Издательство	Выпусков
1619-4500	4OR: Quarterly Journal of the Belgian, French and Italian Operations Research Societies	Springer Germany, Heidelberg	13
1085-3375	Abstract and Applied Analysis		19
0360-0300	ACM Computing Surveys	Proquest ABI/INFORM	22
1549-6325	ACM Transactions on Algorithms	ACM	7
1544-3558	ACM Transactions on Applied Perception	ACM	2
1529-3785	ACM Transactions on Computational Logic		9
0098-3500	ACM Transactions on Mathematical Software		15
1049-3301	ACM Transactions on Modeling and Computer Simulation		7
0167-8019	Acta Applicandae Mathematicae: An International Survey Journal on Applying Mathematics and Mathematical Applications	Springer Science+Business Media B.V.	190
0065-1036	Acta Arithmetica		52
0324-721X	Acta Cybernetica		2

Математическая коллекция НЭБ eLibrary.ru

Информационная система MathNet.Ru (см. <http://www.mathnet.ru>) – это общероссийский математический портал, созданный и развиваемый Математическим институтом им. В.А. Стеклова РАН совместно с Отделением математических наук РАН, предоставляющий российским и зарубежным математикам различные возможности в поиске математической информации. Ключевой компонент портала – «Журналы» – связывает российские периодические издания в области математических наук в единую информационную систему. В базы данных проекта MathNet включен целый ряд математических изданий и Казанского университета, в том числе первый российский электронный математический журнал «Lobachevskii Journal of Mathematics» (LJM; ljm.ksu.ru).

The screenshot shows the Math-Net.Ru website interface. On the left, there is a list of institutions and their associated journals. On the right, there are thumbnail images of the journal covers.

Московское математическое общество
 Труды Московского математического общества

Евразийский национальный университет им. Л. Н. Гумилёва, г. Астана, Казахстан
 Eurasian Mathematical Journal

Казанский (Приволжский) федеральный университет
 Учёные записки Казанского государственного университета
 Известия высших учебных заведений. Математика
 Исследования по подземной гидромеханике
 Исследования по прикладной математике и информатике
 Исследования по теории пластин и оболочек
 Конструктивная теория функций и функциональный анализ
 Труды геометрического семинара
 Труды семинара по краевым задачам
 In memoriam N. I. Lobachevskii
 Lobachevskii Journal of Mathematics

Независимый Московский университет – МЦНМО
 Математическое просвещение
 Moscow Mathematical Journal

Новосибирский государственный университет – Сибирский фонд алгебры и логики
 Алгебра и логика

Российский университет дружбы народов
 Современная математика. Фундаментальные направления

Самарский государственный технический университет
 Вестник Самарского государственного технического университета. Серия «Физико-математические науки»
 Математическое моделирование и краевые задачи

The right side of the screenshot displays a grid of journal covers, including titles like 'Eurasian Mathematical Journal', 'Moscow Mathematical Journal', 'Алгебра и логика', and 'Вестник Самарского государственного технического университета'.

Математические коллекции университетов в MathNet

3. Технологии формирования и поддержки электронных научных коллекций. Выделим некоторые результаты в области электронного книгоиздания, полученные при активном участии Казанского университета.

В 2005 году Казанским университетом и Научной электронной библиотекой в рамках проекта РИ-13.0/006 Министерства образования и науки «Создание электронных версий российских научных журналов» 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 интернете, так и электронных журналов.

В результате выполнения проекта был разработан формат Sarcicle, отличительными особенностями которого являются: вложенность полей, возможности описания любого количества информации одним файлом, проверки правильности составления файлов описаний на стороне издательства, использования файлов описаний для наполнения собственных сайтов изда-

тельств и совместимости с другими форматами обмена метаданными, основанными на XML. Основные блоки формата – информация о журнале, выпуске, статье (основная информация файла). Большинство полей может дублироваться на нескольких языках с целью более удобного предоставления разным пользователям конечной информации в электронной библиотеке.

Формат исполнен в двух видах – в DTD и в MS Schema. Набор тегов формата не зависит от выбора между видами описания XML. Технически возможно в одном файле описать любое количество журналов, но с точки зрения удобства хранения и заполнения предпочтительна ситуация «один файл XML – один выпуск журнала».

Возможные способы создания XML-документов в формате Sarcticle могут включать использование специализированных программных средств создания документов XML, конформных формату Sarcticle, а также любого XML-ориентированного текстового редактора. Имеются дополнительные описания элементов формата (или «справочники»): «arcticle types» – список кодов типов статей для атрибута `arttype`; «language codes» – список кодов языков для атрибута `fieldlang`; «country codes» – список кодов стран для атрибута `jcountr`; «symbols.html» (в HTML) – список всех сущностей, заменяющих специальные символы; «dateUni format.txt» – описание формата поля `dateUni`.

Все современные научные журналы сейчас готовятся к изданию с использованием компьютерной техники, применяется большой перечень издательских пакетов программ. К сожалению, за редким исключением эти программы предназначены лишь для набора текста, оформления и создания макетов изданий, с которых производится типографское тиражирование, а использование макетов для загрузки электронных изданий в базы данных не может быть выполнено. Поэтому было необходимо разработать специальные программы структурирования (разметки) текстов электронных версий журналов для создания файлов, предназначенных для загрузки в базы дан-

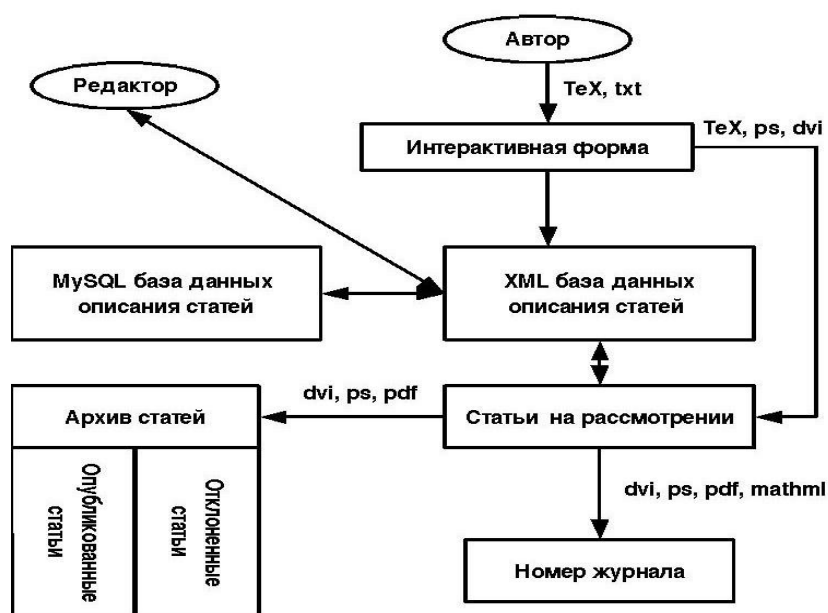
ных. Попытки создания специальных программ обработки макетов изданий предпринимались неоднократно, однако ни одна из них не увенчалась успехом. Проблема в том, что корректная работа подобных программ с макетами изданий возможна лишь в том случае, если сам макет сделан при тщательном соблюдении правил, что практически неосуществимо. Отчасти этим условиям соответствуют издания, подготовленные с использованием издательского пакета TeX, однако и в этих случаях требуется создание специализированных конверторов. Но, самое главное, что и в этом случае многие значимые элементы издания не могут быть программно выделены из текста и структурированы. Для решения названной проблемы были разработаны две версии программного обеспечения для разметки макетов электронных изданий, в основу которого был положен принцип выделения элементов текста макетов изданий и присвоения им меток полей XML-формата Sarcticle.

В результате выполнения описываемого проекта было обработано и загружено в базу данных Научной электронной библиотеки 47 российских научных журналов, открыт доступ пользователям.

Далее, в течение последнего пятилетия в Казанском университете в рамках ряда проектов РФФИ был выполнен цикл исследований, нацеленный на развитие системы электронного книгоиздания. В частности, разработана полнофункциональная программная среда электронного научного журнала по математике, позволяющая автоматизировать ряд процессов, стандартных для научного издания [4 – 7].

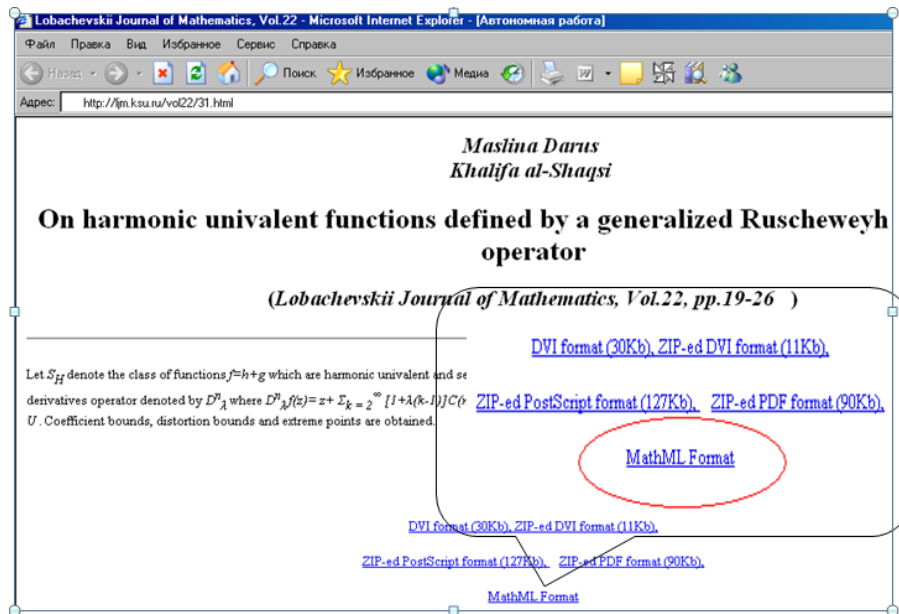
Современные математические электронные журналы в своей работе придерживаются тех же высоких стандартов качества публикаций, что и классические математические журналы. В частности, статьи проходят научное рецензирование, что отличает электронный научный журнал от электронных коллекций, например, всемирно известного архива физических и математических препринтов xxx.lanl.gov. Это обстоятельство влияет на одно

из главных преимуществ электронных журналов – скорость выхода публикации. Поэтому возникает задача такой организации информационных потоков в процессе функционирования электронного журнала, которая минимизирует временные задержки. Сокращение времени выхода публикации в электронном журнале по сравнению с обычным достигается не только за счет оперативного обмена информацией на шаге рецензирования, но, главным образом, в переходе на новые стандарты электронных публикаций.



Пример организации редакционной обработки статьи в электронном научном журнале ЛМ

С точки зрения обработки информации процесс электронного издания представляет собой XML/MathML-поток (<http://www.dessci.com/en/products/mathflow/>). На входе этого потока находятся материалы, представленные авторами в LaTeX-формате с использованием стандартных для журнала стилевых правил, на выходе – очередной том (выпуск) журнала с возможностью для читателя выбора подходящего формата представления данных (pdf, djvu, html и т. д). Данная схема реализована в электронном журнале ЛМ. Журнал представляет читателем статьи в форматах pdf, ps, dvi, а также в формате MathML.



Пример выбора формата отображения статьи в LJM

В связи с основной тенденцией развития современной науки – переходом на электронную форму представления информации – актуальной является проблема представления и обработки математических текстов в электронной форме. В основном эта проблема касается представления математических формул. Наиболее распространенное на данный момент решение – представление формул в виде графических файлов – неудовлетворительно с точки зрения структурной обработки математических текстов. Поэтому в 1999 году консорциумом W3C (<http://www.w3.org>) начата разработка языка математической разметки – Mathematical Markup Language (MathML). Язык MathML представляет собой подмножество языка разметки XML (см. библиографию в [7]). Поскольку XML является новым способом организации веб-информации, позволяющим на более высоком уровне решать задачи программной обработки документов (в частности, задачу поиска), MathML способен изменить принципы организации и управления электронными публикациями по математике.

- Семантический Веб
- В частности, MathML



1 www.w3.org/Math/



MathML – стандарт представления математической информации

В настоящее время язык MathML можно рассматривать как стандарт представления математической информации в электронной форме. Технология обработки данных на основе языка MathML реализует одну из основных тенденций современной информатики – разделение разметки и данных – и поэтому открывает новые возможности многоуровневого структурирования данных и расширенного поиска. Созданы и совершенствуются программные средства конвертации в MathML документов, подготовленных с помощью имеющихся стандартных технологий (TeX, Mathematica, Maple, Word). Технология MathML поддерживается широко известными системами Maple, MathCAD, Mathematica и рядом других.

В журнале ЦИМ с использованием ряда технологических приемов, примененных разработчиками проекта НЭБ eLibrary.ru, организован XML/MathML-поток получения, преобразования и хранения данных. Один из разрезов потока представляет собой преобразование материалов (LateX-документов и данных используемых форм), представленных авторами, в XML-файлы со вставками MathML. Формы для авторов содержат сведения об авторе и ключевые слова, отражающие содержание публикации. XSL-модули

отвечают за преобразование информации, содержащейся в XML-хранилище, в представление, выбранное читателем (например, предназначенное для печати). Кроме того, рассматривались вопросы конвертации в MathML документов, подготовленных с помощью имеющихся стандартных технологий (LaTeX, Mathematica, Maple, Word). Система метаданных DublinCore (<http://purl.oclc.org/dc/>), а также Resource Description Framework (<http://www.w3.org/TR/2002/WD-rdf-syntax-grammar-20021108>) были использованы для включения метаданных в xml-файлы. Это позволило оптимизировать процессы поиска по библиографическим данным опубликованных статей.

В настоящее время описанное направление исследований развивается в рамках проектов РФФИ 12-07-00667 и 12-07-97018-р_поволжье.

4. Литература

1. Ершова Т. В., Хохлов Ю. Е. Межведомственная программа «Российские электронные библиотеки»: подходы и перспективы// Электронные библиотеки. – 1999. – Т. 2. – Вып. 2. – URL: <http://www.iis.ru/el-bib>.

2. Кириллова О.В. Место информационных продуктов винити в среде электронных ресурсов по науке и технике// Информационные ресурсы России. – 2008. – № 2. – С. 9-11. – URL:http://www.aselibrary.ru/datadocs/doc_771ga.pdf.

3. Глухов В.А. Проект «Научная электронная библиотека eLibrary.ru» и перспективы развития электронного книгоиздания в России// Educational Technology & Society. – 2005. – V. 8. – No 1. – P. 191-197. – URL: http://ifets.ieee.org/russian/periodical/V_81_2005EE.html.

4. Елизаров А.М., Липачев Е.К., Малахальцев М.А. Технологии Semantic Web в практике работы электронного журнала по математике //Тр. 8-й Всерос. науч. конф. «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» – RCDL'2006, Суздаль, Россия, 2006. – С. 215-218.

5. Глухов В.А., Елизаров А.М., Липачев Е.К., Малахальцев М.А. Электронные научные издания: переход на технологии семантического Веба // Электронные библиотеки. – 2007. – Т. 10, Вып. 1. – URL: <http://www.elbib.ru>.

6. Елизаров А.М., Липачев Е.К., Малахальцев М.А. Технологии управления разнородным естественнонаучным контентом на основе семантического Веба // Электронные библиотеки: перспективные методы и технологии, электронные коллекции: Труды XI Всероссийской научной конференции RCDL-2009. – Петрозаводск: КарНЦ РАН, 2009. – С. 325-328.

7. Елизаров А.М., Липачев Е.К., Малахальцев М.А. Сервисы электронных естественнонаучных коллекций, построенные на основе технологии MathML // Труды Всероссийской суперкомпьютерной конф. «Научный сервис в сети Интернет: суперкомпьютерные центры и задачи». – М.: Изд-во Моск. ун-та, 2010. – С. 533-534.

8. Елизаров А.М., Липачев Е.К., Малахальцев М.А. Веб-технологии для математика: Основы MathML. Практическое руководство. – М.: Физматлит, 2010. – 192 с.