

Уважаемые читатели!

Перед вами очередной, 18-й том (№№ 1, 2) журнала «Электронные библиотеки». С 2015 года соучредителем журнала (наряду с Институтом развития информационного общества, г. Москва) стал Казанский (Приволжский) федеральный университет. Обновлен состав редакционной коллегии журнала.

РЕДАКЦИОННАЯ КОЛЛЕГИЯ ЖУРНАЛА «ЭЛЕКТРОННЫЕ БИБЛИОТЕКИ»

Главный редактор

А.М. Елизаров, Институт математики и механики им. Н.И. Лобачевского Казанского (Приволжского) федерального университета (КФУ), г. Казань, Россия;

Заместители главного редактора

Е.К. Липачёв, Институт математики и механики им. Н.И. Лобачевского КФУ, Казань, Россия;

Е.Н. Струков, Научная библиотека им. Н.И. Лобачевского КФУ, г. Казань, Россия;

А.Ф. Хасьянов, Высшая школа информационных технологий и информационных систем (ВШ ИТИС) КФУ, г. Казань, Россия;

Ответственный секретарь

Д.С. Зуев, Институт математики и механики им. Н.И. Лобачевского КФУ, Казань, Россия;

Члены редакционной коллегии

И.С. Антипин, Институт химии КФУ, г. Казань, Россия;

В.Б. Барахнин, Институт вычислительных технологий Сибирского отделения российской академии наук (СО РАН), Новосибирский государственный университет, г. Новосибирск, Россия;

А.В. Богданов, Санкт-Петербургский государственный университет, г. Санкт-Петербург, Россия;

М.Р. Биктимиров, Всероссийский институт научной и технической информации (ВИНИТИ) РАН, г. Москва, Россия;

А.И. Вислый, Российская государственная библиотека, г. Москва, Россия;

В.Э. Вольфенгаген, Московский физико-технический институт, г. Москва, Россия;

М.М. Горбунов-Посадов, Институт прикладной математики им. М.В. Келдыша РАН, г. Москва, Россия;

А.О. Грачев, Технопарк в сфере высоких технологий «ИТ-парк», г. Казань, Россия;
Стивен Гриффин (Stephen Griffin), University of Pittsburgh, Mellon Cyberscholar, Pittsburgh, USA;

Милена Добрева (Milena Dobрева), University of Malta, Faculty of Media and Knowledge Sciences, Department of Library Information and Archive Sciences, Malta;

Б.В. Добров, Лаборатория анализа информационных ресурсов Научно-исследовательского вычислительного центра (НИВЦ) Московского государственного университета (МГУ) им. М.В. Ломоносова, г. Москва, Россия;

О.Л. Жижимов, Институт вычислительных технологий СО РАН, г. Новосибирск, Россия;

С.В. Знаменский, Институт программных систем им. А.К. Айламазяна РАН, г. Переславль-Залесский, Россия;

А.Д. Изаак, Математический институт им. В.А. Стеклова РАН, г. Москва, Россия;

Н.Е. Каленов, Библиотека по естественным наукам РАН, г. Москва, Россия;

М.Р. Когаловский, Институт проблем рынка РАН, г. Москва, Россия;

С.Д. Кузнецов, Институт системного программирования РАН, г. Москва, Россия;

Д.В. Ландэ, Институт проблем регистрации информации НАН Украины, г. Киев, Украина;

Н.В. Лукашевич, НИВЦ МГУ им. М.В. Ломоносова, г. Москва, Россия;

Н.А. Мазов, Институт нефтегазовой геологии и геофизики им. академика А.А. Трофимука СО РАН, г. Новосибирск, Россия;

А.Г. Марчук, Институт вычислительных систем СО РАН, г. Новосибирск, Россия;

М.Е. Мошков, НИИ системных исследований РАН, г. Москва, Россия;

Н.А. Никифоров, Министерство связи и массовых коммуникаций РФ, г. Москва, Россия;

М.Н. Овчинников, Институт физики КФУ, г. Казань, Россия;

С.И. Паринов, Центральный экономико-математический институт РАН, г. Москва, Россия;

Ю.Н. Прошин, Институт физики КФУ, г. Казань, Россия;

А.Н. Райков, Институт проблем управления РАН; ООО «Агентство Новых Стратегий», г. Москва, Россия;

А.А. Рогов, Петрозаводский государственный университет, г. Петрозаводск, Россия;

В.А. Серебряков, Федеральный исследовательский центр «Информатика и управление» РАН, г. Москва, Россия;

В.Д. Соловьев, Институт филологии и межкультурной коммуникации, ВШ ИТИС КФУ, г. Казань, Россия;

С.А. Ступников, Федеральный исследовательский центр «Информатика и управление» РАН, г. Москва, Россия;

Д.Ш. Сулейманов, Институт вычислительной математики и информационных технологий КФУ, Институт семиотики АН РТ, г. Казань, Россия;

Аарон Трегуб (Aaron Trehub), Auburn University, Alabama, USA;

А.З. Фазлиев, Институт оптики атмосферы имени В.Е. Зуева СО РАН, г. Томск, Россия;

А.О. Федоров, Чувашский государственный институт культуры и искусств, г. Чебоксары, Россия;

А.М. Федотов, Институт вычислительных технологий СО РАН, Новосибирский государственный университет, г. Новосибирск, Россия;

Кевин Хокинс (Kevin S. Hawkins), University of North Texas Libraries, USA;

Ю.Е. Хохлов, Институт развития информационного общества, г. Москва, Россия;

М.Г. Храмченков, Институт математики и механики им. Н.И. Лобачевского КФУ, г. Казань, Россия;

Я.Л. Шрайберг, Государственная публичная научно-техническая библиотека России, г. Москва, Россия;

Л.Н. Щур, Научный центр РАН в Черноголовке, г. Черноголовка Московской обл., Россия.

ОТ СОСТАВИТЕЛЯ

Настоящий выпуск журнала «Электронные библиотеки», включающий два его первых номера за 2015 год, представляет собой тематический сборник статей, посвященный вопросам поддержки жизненного цикла как отдельных научных статей, так и научных журналов в целом с использованием программных платформ управления электронными научными изданиями.

Сегодня электронные издания являются основным источником научной информации, а печатные издания имеют электронные варианты, доступные по подписке или в свободном доступе. Переход к электронной форме научных изданий обусловлен постоянно возрастающими объемами информации, обработка которой уже невозможна без применения информационно-коммуникационных технологий. Несмотря на их использование на всех этапах подготовки научных публикаций, организация большинства рабочих процессов остается традиционной, ориентированной на печатные издания. Причина такого положения – в сложившейся системе рассмотрения материалов в редакциях журналов, которые по-разному регламентируют правила их оформления и систему научного рецензирования.

В области хранения информации широко применяются технологии электронных библиотек, созданы соответствующие информационные системы, этой проблематике посвящено большое количество исследований. Однако вопросы создания самого контента, размещаемого в электронных коллекциях, менее исследованы.

На рубеже веков для обработки постоянно растущего объема научной информации стали создаваться специализированные информационно-издательские системы – ведущие мировые научные издательства внедрили системы автоматического управления рабочими процессами, в числе которых – наиболее сложные и длительные по времени редакционные процессы независимого научного рецензирования. Таким образом, к настоящему времени в мире внедрены десятки таких систем, построенных на различных программных платформах и позволяющих управлять редакционно-издательским процессом. Одновременно стал более оперативным доступ к результатам научных исследований. Информационные системы управления научными журналами и публикациями являются

подклассом систем управления электронными библиотеками. Следовательно, при их создании могут быть использованы развитые и широко применяемые технологии электронных библиотек с учетом специфики бизнес-процессов, характерных для научного издания.

В соответствии с международными стандартами публикация научного журнала предполагает дальнейшую обработку публикуемых материалов информационно-аналитическими системами. К последним относятся международные библиографические и реферативные базы данных Scopus, Web of Science и Российский индекс научного цитирования. В течение последнего десятилетия они активно используются для оценки научного уровня как самих журналов, так и публикуемых ими статей с помощью ряда показателей (импакт-факторы журналов, индексы цитирования по различным базам данных и др.). Метаданные являются важным элементом научных публикаций, а их наличие и полнота важны для ориентации в научном пространстве и учета в аналитических системах. Программные платформы управления научными журналами позволяют автоматизировать процесс загрузки метаданных научного контента в базы научного цитирования.

В настоящем тематическом выпуске представлены статьи, посвященные созданию и развитию систем управления электронными научными журналами. В частности, проанализированы свободно распространяемые системы автоматизации редакционных бизнес-процессов и возможности их применения в работе электронных научных журналов; обсуждены дополнительные функции, востребованные редакторами таких журналов, предложены варианты их реализации с использованием дополнительных плагинов для самой распространенной в мире свободной программной платформы Open Journal Systems (OJS). Предложен алгоритм автоматического извлечения библиографических данных и метаданных из публикаций для экспорта в международные информационно-аналитические системы. Развита методика интеграции платформы OJS и международных баз научного цитирования. Рассмотрен целый ряд близких вопросов и проблем.

Главный редактор журнала «Электронные библиотеки»,
составитель тематического выпуска

А.М. Елизаров

УДК 004.82+004.9

ФОРМИРОВАНИЕ МЕТАДААННЫХ ДЛЯ МЕЖДУНАРОДНЫХ БАЗ ЦИТИРОВАНИЯ В СИСТЕМЕ УПРАВЛЕНИЯ ЭЛЕКТРОННЫМИ НАУЧНЫМИ ЖУРНАЛАМИ

А.Н. Герасимов¹, А.М. Елизаров², Е.К. Липачёв³

Институт математики и механики им. Н.И. Лобачевского Казанского (Приволжского) федерального университета

¹sav241@mail.ru, ²amelizarov@gmail.com, ³elipachev@gmail.com

Аннотация

Предложен алгоритм автоматического извлечения библиографических данных из однородного массива публикаций (в частности, выпусков научного журнала) и формирования блоков метаданных для экспорта в международные информационно-аналитические системы. Развита платформа интеграции платформы управления электронными научными журналами Open Journal Systems и международных баз научного цитирования.

Ключевые слова: *издательские системы, электронный научный журнал, интеграция электронных ресурсов, базы данных научного цитирования, экстракция метаданных*

ВВЕДЕНИЕ

В соответствии с современными международными стандартами публикация (в том числе, в электронной форме) научного журнала (текущего номера или выпуска с соответствующим набором статей) предполагает дальнейшую обработку публикуемых материалов информационно-аналитическими системами. К последним, например, относятся международные библиографические и реферативные базы данных Scopus (<http://www.scopus.com/>) и Web of Science (<http://apps.webofknowledge.com/>), а также российская национальная информационно-аналитическая система Российский индекс научного цитирования (РИНЦ, http://elibrary.ru/project_risc.asp). В течение последнего десятилетия они активно

используются для оценки научного уровня, как самих журналов, так и публикуемых ими статей с помощью целого набора показателей (импакт-факторы журналов, индексы цитирования по различным базам данных и др.) (см., например, [1–7]). В частности, индекс цитирования принят в научном мире как показатель значимости трудов ученого, представляет собой число ссылок на публикации этого ученого в реферируемых научных периодических изданиях и используется для оценки качества научной деятельности. Другой оценкой качества статей служит полнота библиографических ссылок. Как отмечено в [4], англоязычные статьи содержат в среднем около 30, а русскоязычные – около 10 ссылок. Такое количественное различие объясняется, в частности, сложившейся отечественной практикой издания преимущественно бумажных периодических изданий, предусматривающей ограничения на объем как самой публикации, так и количество пристатейных ссылок.

Блок метаданных любой научной публикации обязательно включает ее библиографическое описание (авторы, название, источник (например, журнал), год издания, том, номер, начальная и конечная страницы), авторское резюме (аннотация, реферат) и ключевые слова, а также различную дополнительную информацию. Таковой могут быть названия и места расположения организаций, от имени которых авторы представили свои материалы (аффилиация); фамилии ученых (членов редколлегии), представивших статью к публикации (как, например, в таких журналах, как Доклады академии наук и *Lobachevskii Journal of Mathematics*), указание текущей версии публикации (как в ставших сегодня модными «живых публикациях» [8, 9]) и другая информация. Отметим, что для размещения в международных информационно-аналитических системах неанглоязычных научных материалов необходимо дополнительно включать в блок метаданных список авторов, название работы, аннотацию, аффилиацию и ключевые слова на английском языке, а список литературы привести в транслитерации.

Для автоматизированного формирования блока метаданных сегодня с успехом могут быть использованы современные информационные технологии, особенно в тех случаях, когда они применяются для подготовки и издания самого научного журнала. В этом отношении передовые позиции занимают те издания, которые базируются на современных издательских системах (например, Open

Journal Systems – OJS, см., например, <https://pkp.sfu.ca/ojs/>, https://ru.wikipedia.org/wiki/Open_Journal_Systems) и специализированных программных пакетах, служащих для набора текстов статей [8]. К таким изданиям относятся, прежде всего, многие физико-математические и инженерно-технические журналы.

ОСОБЕННОСТИ ПОДГОТОВКИ БИБЛИОГРАФИЧЕСКОГО ОПИСАНИЯ НАУЧНЫХ ПУБЛИКАЦИЙ

Библиографическим описанием является совокупность библиографических сведений о документе, его составной части или группе документов, составленных по определенным правилам и необходимых для общей характеристики этого документа. Как правило, оформление библиографического описания публикаций выполняется в соответствии с существующими библиографическими правилами, установленными Государственными стандартами. Для России эти стандарты таковы (см., например, <http://www.gosthelp.ru/gost/gost1560.html>):

- ГОСТ 7.1-2003. «Библиографическая запись. Библиографическое описание»;
- ГОСТ 7.80-2001. «Библиографическая запись. Заголовок. Общие требования и правила составления»;
- ГОСТ 7.82-2001. «Библиографическая запись. Библиографическое описание электронных ресурсов»;
- ГОСТ 7.12-93. «Библиографическая запись. Сокращение слов на русском языке. Общие требования и правила».

Основные положения и правила стандартов основаны на принципах Международного стандарта библиографического описания (ISBD – International Standard Bibliographical Description) [10], принятые в системе международного библиографического учета мировых информационных документов и широко применяемые в национальных библиографических изданиях и библиотеках многих стран мира. Благодаря следованию единым стандартам облегчаются использование библиографи-

ческих материалов в мировой системе коммуникации, обмен информацией в машиночитаемой и электронной форме, преодолеваются языковые барьеры: записи, составленные в одной стране или на другом языке, могут быть легко поняты в другой стране или на другом языке.

Следование единым стандартам оформления значительно облегчает задачу обработки больших объемов разнородных электронных научных коллекций, что позволяет, как проводить аналитические исследования (выполнять запросы) по различным тематическим срезам, так и получать индексы цитирования [11]. Однако не все журналы в научном пространстве следуют единым стандартам, многие из них предлагают свои стилевые файлы, задающие общие правила оформления научной публикации и списка литературы, что затрудняет процесс автоматической обработки.

РЕГУЛЯРНЫЕ ВЫРАЖЕНИЯ В АЛГОРИТМАХ ИЗВЛЕЧЕНИЯ МЕТАДААННЫХ ИЗ КОЛЛЕКЦИЙ НАУЧНЫХ СТАТЕЙ

Как известно, большинство статей набирается с помощью офисных систем (например, MS Word, OpenOffice и другие) или систем, основанных на T_EX-нотации. Как правило, научные журналы принимают к рассмотрению и дальнейшей публикации только те статьи, которые подготовлены в соответствии с требованиями, четко сформулированными в каждом издании, в одном формате .rtf, .doc, .docx, .odt и .tex. Выдвинутые требования регламентируют стилевые особенности оформления статей для журнала, а именно, порядок следования информационных блоков, оформление списка литературы и другие. Как правило, в статьях используются простые (с семантической точки зрения) средства структурирования. Это усложняет автоматическую обработку публикаций. Тем не менее, на основе анализа шрифтового выделения текста и порядка следования текстовых единиц (например, название, автор, аннотация и т. д.) и стандартных заголовков («Аннотация», «Ключевые слова», «Литература» и другие) можно определить структуру таких документов, что дает возможность проанализировать текст и автоматически

извлечь метаданные [12]. В рамках описанного подхода может быть предложен следующий алгоритм разбора библиографических списков:

- из текста статьи выделяется блок, содержащий список литературы: в качестве признака используются ключевые слова «References», «Список литературы», «Литература», «thebibliography» (для материалов в T_EX-нотации);

- извлекаются отдельные библиографические записи, признаками которых служат принятые в журнале правила оформления библиографических источников: квадратные скобки, нумерация, тег «bibitem» (для материалов в T_EX-нотации);

- каждая библиографическая запись разделяется на элементы описания: список авторов, название и т. д.; для этого используется техника регулярных выражений (см., например, [14, 15]);

- для каждой коллекции подбирается система паттернов регулярных выражений, охватывающая возможные варианты написания библиографических источников;

- на заключительном этапе формируется XML-файл, содержащий метаописание коллекции статей, который используется для последующей конвертации в форматы баз научного цитирования.

Описанный алгоритм был апробирован на электронных коллекциях математических изданий: «Труды Математического центра им. Н.И. Лобачевского», «Ученые записки Казанского университета» и «Известия вузов. Математика» [16]. Общий объем коллекций – более 3 тыс. документов в форматах .pdf и .tex. Отметим, что эти документы имеют различную структуру, что потребовало разработки нескольких вариантов алгоритмов извлечения метаданных. Для работы с коллекцией документов каждого из указанных журналов подготовлена система паттернов, позволяющих выполнить извлечение и разбор библиографии.

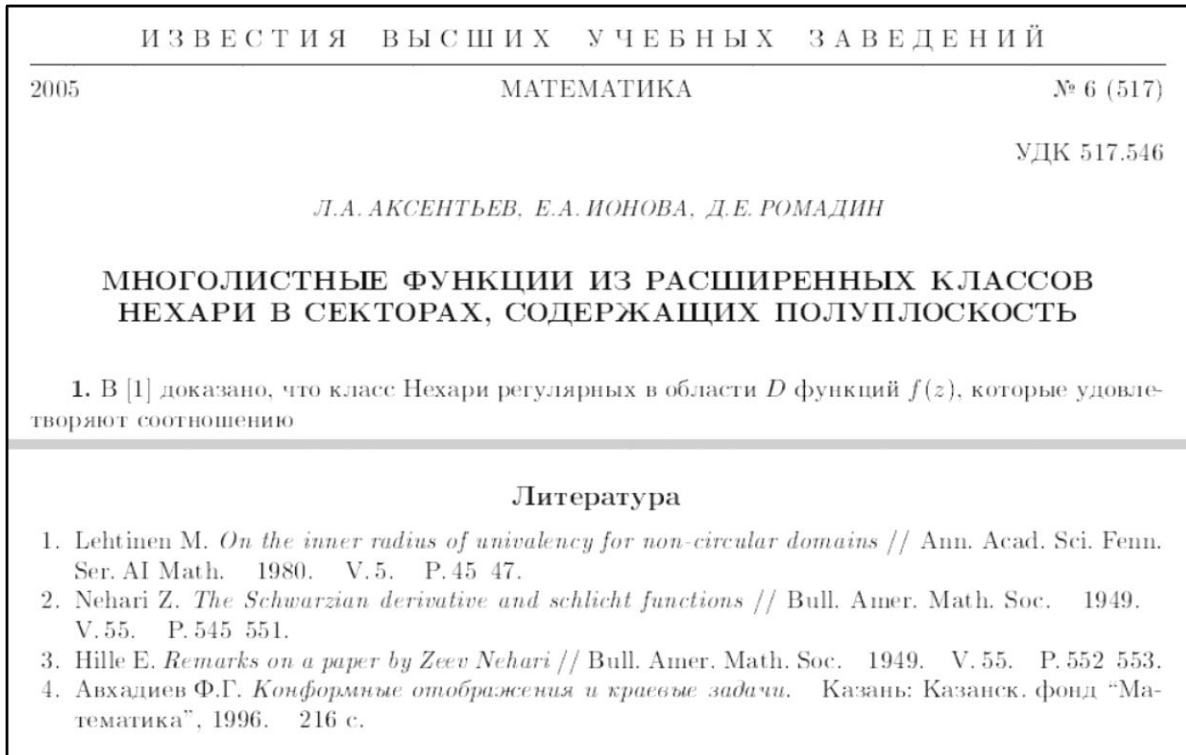


Рис. 1. Фрагмент статьи из журнала «Известия вузов. Математика», содержащий список авторов, название статьи и часть списка литературы

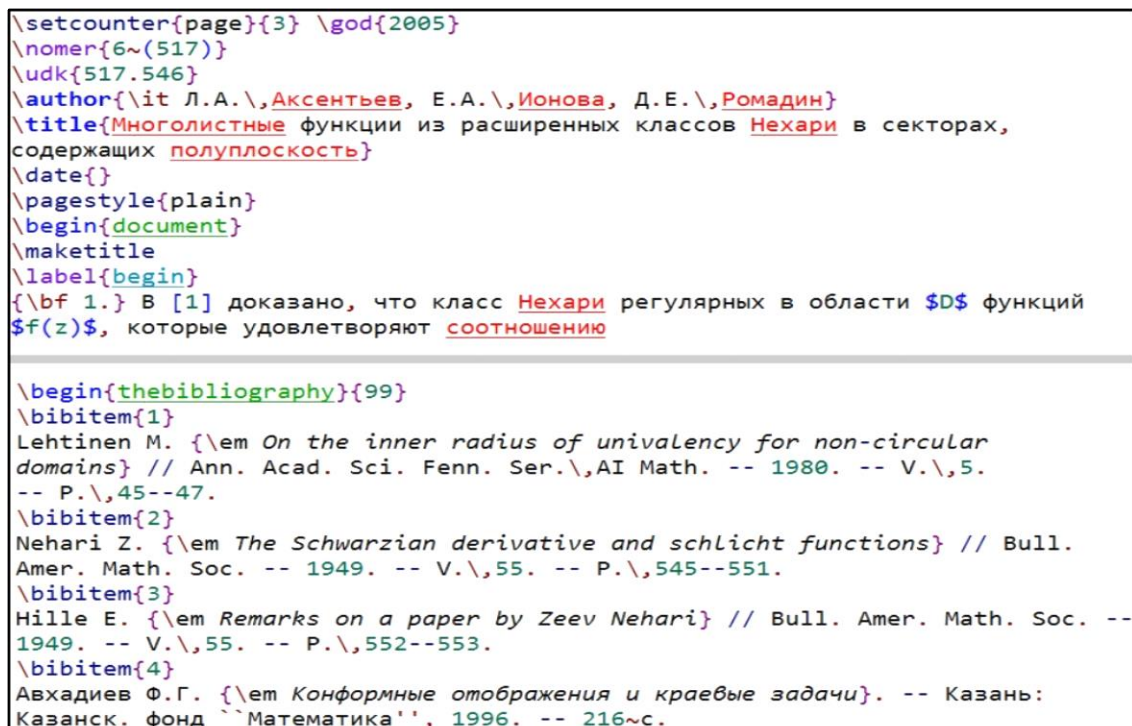


Рис. 2. Фрагмент TeX-кода статьи из журнала «Известия вузов. Математика»

LINEAR SPACES WITH PROBABILITY MEASURES,
ULTRAPRODUCTS AND CONTIGUITY

D.H.MUSHTARI AND S.G.HALIULLIN

Keywords: Ultraproducts, linear probability spaces, Gaussian measures

ABSTRACT. The ultraproducts of measurable linear spaces with probability

REFERENCES

- [1] S. Heinrich, J.fur die reine und angewandte Math. **313**, 72-104 (1980).
- [2] X. Férnique, Lecture Notes in Math. **480**. 1-96 (1975).
- [3] Y. Okazaki, Ann. Inst. H.Poincare, **21**, 393-400 (1985).
- [4] E. Szpilrajn, Comptes Rendus, **207**, 768-770 (1938).
- [5] N.N. Vakhaniya, V.I. Tarieladze and S.A. Chobanyan, *Probabilities distributions in Banach spaces* (M: Nauka, 1985)

Рис. 3. Фрагмент статьи из журнала «Lobachevskii Journal of Mathematics»

```
\begin{document}
\author{D.H.Mushtari and S.G.Haliullin}
\address{Kazan Federal University, 420008, Kremlevskaya, 18, Kazan, Russia}
\email{Samig.Haliullin@kpfu.ru}
\title{Linear spaces with probability measures, ultraproducts and contiguity}
\maketitle
\keywords{Keywords: Ultraproducts, linear probability spaces, Gaussian measures}

\begin{thebibliography}{99}
\bibitem{1} S. Heinrich, J.fur die reine und angewandte Math. \textbf{313}, 72
--104 (1980).
\bibitem{2} X. F'ernique, Lecture Notes in Math. \textbf{480}. 1--96 (1975).
\bibitem{3} Y. Okazaki, Ann. Inst. H.Poincare, \textbf{21}, 393--400 (1985).
\bibitem{4} E. Szpilrajn, Comptes Rendus, \textbf{207}, 768--770 (1938).
\bibitem{5} N.N. Vakhaniya, V.I. Tarieladze and S.A. Chobanyan, \textit
{Probabilities distributions in
Banach spaces} (M: Nauka, 1985)
```

Рис. 4. Запись приведенного ранее фрагмента статьи в T_EX-нотации

Статьи журнала «Известия вузов. Математика» и «Lobachevskii Journal of Mathematics» размечены по правилам языка T_EX с использованием стилевых инструкций, разработанных для данных журналов, в которых для библиографических списков используется окружение \thebibliography (см. рис. 1).

Для разбора библиографической записи использовалось регулярное выражение, где отличительным признаком было наличие окружения \bibitem для каждой записи:

```
@"item{.*?}{.*?}{.em(.*?)}.*?[-|/|](.*?)(\d\d\d\d).*?--(.*?)[bib|end"];
```

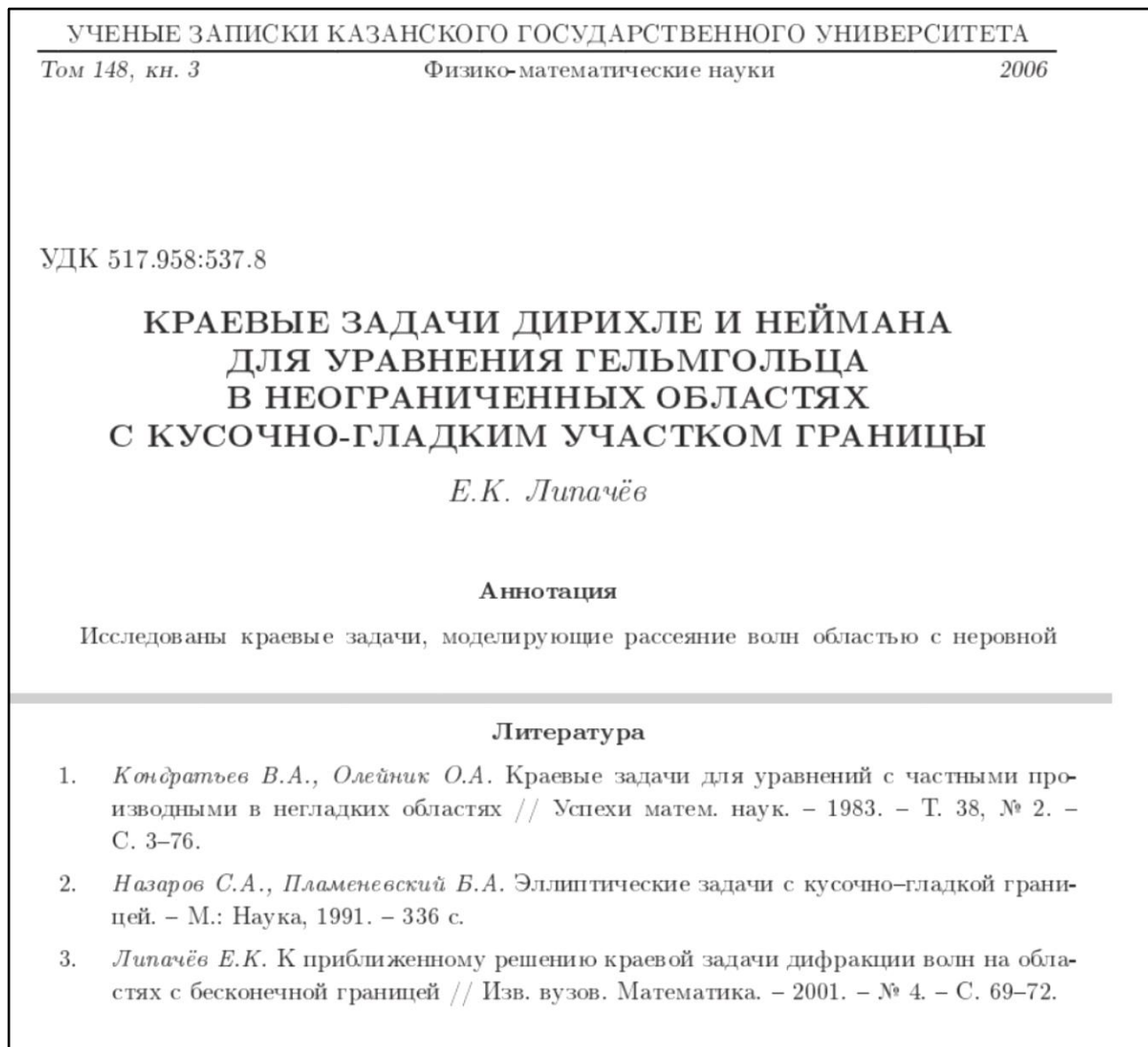


Рис. 5. Фрагмент статьи из журнала «Ученые записки Казанского университета. Серия Физико-математические науки»


```

\UDK{517.958:537.8}

\ArticleNAME{КРАЕВЫЕ ЗАДАЧИ ДИРИХЛЕ И НЕЙМАНА \\
ДЛЯ УРАВНЕНИЯ ГЕЛЬМГОЛЬЦА В ОБЛАСТЯХ \\
С БЕСКОНЕЧНОЙ КУСОЧНО--ГЛАДКОЙ ГРАНИЦЕЙ }
\ArticleAUTHOR{Е.К.~Липачёв}
\ArticleHEAD{Краевые задачи Дирихле и Неймана}
\ArticleAUTHORHEAD{Е.К.~Липачёв}
\makeabstitle

\begin{abstract}
Исследованы краевые задачи, моделирующие рассеяние волн областью с
неровной границей.

\References{
\bibitem{kon}
{\it Кондратьев~В.А., Олейник~О.А.} Краевые задачи для уравнений
с частными производными в негладких областях
// Успехи матем. наук. -- 1983. -- Т.~38, N~2. -- С.~3 -- 76.
\bibitem{nazplam} {\it Назаров~С.А., Пламеневский~Б.А.}
Эллиптические задачи с кусочно--гладкой границей. -- М.: Наука,
1991. -- 336~с.
\bibitem{lip2001} {\it Липачёв~Е.К.}
К приближенному решению краевой задачи дифракции волн на областях
с бесконечной границей // Изв. Вузов. Математика. -- 2001. -- N~4
(467). -- С.~69 -- 72.

```

Рис. 6. Запись приведенного ранее фрагмента статьи в Т_EX-нотации

В журнале «Ученые записки Казанского университета. Серия Физико-математические науки» каждая запись списка выделялась окружением \bibitem, фамилии и имена авторов были выделены в окружение \it или \em, а весь блок библиографии – в \References. Было построено и применено регулярное выражение

@"item.*?{.[em|it](.*?)}(.*?)[--|//](.*?)(\d\d\d\d).*?--(.*?)[bib|end]"

В. И. Жегалов

*Казанский (Приволжский) федеральный университет,
Институт математики и механики им. Н. И. Лобачевского,
vzhegalov@yandex.ru*

**ОБ ОДНОМ НАПРАВЛЕНИИ В ТЕОРИИ
УРАВНЕНИЙ С ЧАСТНЫМИ ПРОИЗВОДНЫМИ**

В течение ряда последних десятилетий ведется интенсивное исследование краевых задач, граничные условия в которых представляют собой соотношения, связывающие значения

ЛИТЕРАТУРА

1. Стеклов В. А. *Задача об охлаждении неоднородного твердого тела* // Сообщение Харьковского матем. об-ва, 1896. – Сер. 2. – Т. 5. – № 3–4. – С. 136–181.
2. Стеклов В. А. *Основные задачи математической физики*. – М.: Наука, 1983. – 432 с.
3. Франкль Ф. И. *Обтекание профилей газом с местной сверхзвуковой зоной, оканчивающейся прямым скачком уплотнения* // ПММ. – 1956. – Т. 20. – № 2. – С. 196–202.

Рис. 7. Фрагмент статьи из сборника «Труды Математического центра им. Н.И. Лобачевского»

```

\begin{document}
\ArticleAUTHOR{В.\,И.\,Жегалов \label{Zhegalov}}
\ArticleADDRESS{Казанский (Приволжский) федеральный университет,
Институт математики и механики им. Н.И.~\u041b\u043e\u0431\u0430\u0447\u0435\u0432\u0441\u043a\u043e\u0433\u043e, \\
vzhegalov@yandex.ru}
\ArticleNAME{ОБ ОДНОМ НАПРАВЛЕНИИ В ТЕОРИИ УРАВНЕНИЙ С ЧАСТНЫМИ \u041f\u041e\u041e\u0410\u0414\u0415\u0412\u0410\u041d\u041d\u0418\u041c\u0418}
\makeabstitle
В течение ряда последних десятилетий ведется \u0438\u043d\u0442\u0435\u043d\u0441\u0438\u0432\u043d\u043e\u0435 исследование \u043a\u0440\u0430\u0435\u0432\u044b\u0445 задач,
граничные условия в которых представляют собой соотношения, связывающие значения
искомых функций в различных переменных точках, а также на линиях, расположенных
внутри рассматриваемой области.

\centerline{Л И Т Е Р А Т У Р А}
\smallskip
1.\; \u0421\u0442\u0435\u043a\u043b\u043e\u0432\, \u0412.\, \u0410. {\it Задача об охлаждении \u043d\u0435\u043e\u0434\u043d\u043e\u0440\u043e\u0434\u043d\u043e\u0433\u043e твердого тела}
// \u0421\u043e\u043e\u0431\u0449\u0435\u043d\u0438\u0435 \u0425\u0430\u0440\u044c\u043a\u043e\u0432\u0441\u043a\u043e\u0433\u043e \u043c\u0430\u0442\u0435\u043c. \u043e\u0431-\u0432\u0430, 1896.~-- \u0421\u0435\u0440.~2.~-- \u0422.~5.~--
\u2013 \u041d\u043e.~3--4.~-- \u0421.~136--181.

2.\; \u0421\u0442\u0435\u043a\u043b\u043e\u0432\, \u0412.\, \u0410. {\it Основные задачи математической физики.}~--
\u041c.: \u041d\u0430\u0443\u043a\u0430, 1983.~-- 432\u0441.

3.\; \u0424\u0440\u0430\u043d\u043a\u043b\u044c\, \u0424.\, \u0418. {\it \u041e\u0431\u0442\u0435\u043a\u0430\u043d\u0438\u0435 \u043f\u0440\u043e\u0444\u0438\u043b\u0435\u0439 \u0433\u0430\u0437\u043e\u043c \u0441 \u043c\u0435\u0441\u0442\u043d\u043e\u0439 \u0441\u0432\u0435\u0440\u0445\u0437\u0432\u0443\u043a\u043e\u0432\u043e\u0439 \u0437\u043e\u043d\u043e\u0439,
\u043e\u043a\u0430\u043d\u0447\u0438\u0432\u0430\u044e\u0449\u0435\u044f\u0441\u044f \u043f\u0440\u044f\u043c\u044b\u043c \u0441\u043a\u0430\u0447\u043a\u043e\u043c \u0443\u043f\u043b\u043e\u0442\u043d\u0435\u043d\u0438\u044f}
// \u041f\u041c\u041c.~-- 1956.~-- \u0422.~20.~-- \u2013 \u041d\u043e.~2.~-- \u0421.~196--202.

```

Рис. 8. Запись приведенного ранее фрагмента статьи в Т_ЕX-нотации

В сборнике «Труды Математического центра им. Н.И. Лобачевского» фамилии и имена авторов были выделены в окружение `\it`, а весь блок библиографии – по ключевому слова «Литература». Каждая запись выделялась с помощью регулярного выражения

`@"[0-9](.*?){\it(.*?)}(.*?)(\d{4}).*?([CP].~\d+--\d+|\d+~[cp])."`



The screenshot shows the header of the Russian Digital Libraries Journal. On the left is a logo with the Cyrillic letters 'ЭБ' (EB) inside a shield. To the right of the logo, the text 'Russian Digital Libraries' is written in a serif font, with 'JOURNAL' underneath in a smaller, spaced-out font. In the top right corner, the ISSN number 'ISSN 1562-5419' is displayed. Below the header, there is a navigation bar with the text 'RDLР / Russian Digital Libraries Journal / Year 2014 / Issue 2'. The main content area has a light yellow background. At the top right of this area, it says '2014 | Volume 17 | Issue 2'. The article title is 'A Model for Integrating the Publication and Preservation of Journal Articles' in a bold, black font. Below the title is the author's name, 'Kevin S. Hawkins'. The 'Abstract' section follows, containing a paragraph of text. Below the abstract is the 'Key words' section, listing terms like 'magazines published online', 'digital repository HathiTrust', etc. The 'References' section is at the bottom, listing six numbered references with their respective URLs and publication details.

Рис. 9. Страница журнала «Электронные библиотеки»

На сегодняшний день для загрузки статей в базу данных Российского индекса научного цитирования (РИНЦ) используется онлайн-программа разметки Articuluss (<http://elibrary.ru/projects/contracts/publisher/messages/messages.asp>), предназначенная для подготовки выпусков журналов и книг в формате XML [17, 18]. Используя систему Articuluss, разметчик вынужден с помощью ручного ввода заполнять поля метаданными, что требует от него значительных ресурсов (человеко-часов). При подготовке проведения крупных конференций с большим количеством материалов и сжатыми сроками размещения на сайтах соответствующих

электронных ресурсов, когда о ручной обработке не может идти речи, задача становится особенно актуальной. Естественно, что традиционными методами оперативно выполнить эту работу невозможно.

АЛГОРИТМ ФОРМИРОВАНИЯ БЛОКА МЕТАДАННЫХ

Автоматизация процесса формирования блока метаданных выполнялась в несколько этапов на массивах естественно-научных журналов, перечисленных выше, и гуманитарного журнала «Вестник КазГУКИ». Статьи предварительно проверялись на соответствие стилевым требованиям журналов, далее проводились обработка библиографических данных и их проверка на соответствие единому стандарту.

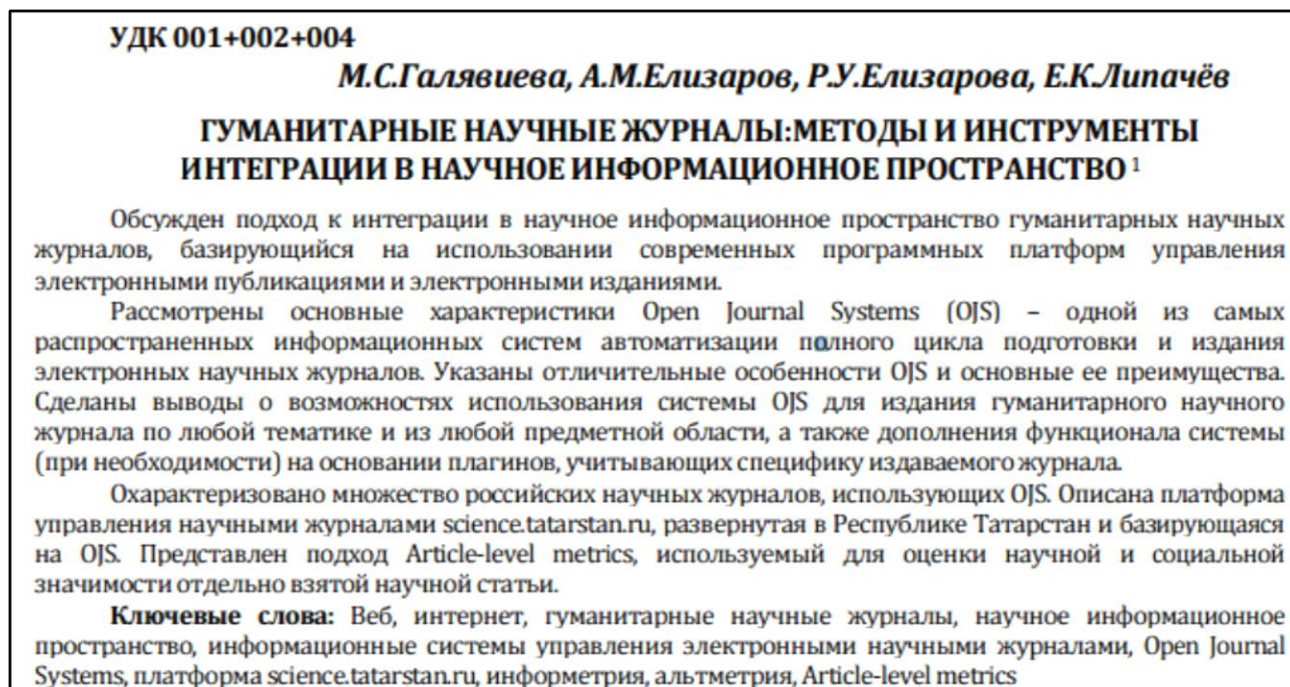


Рис. 10. Страница из журнала «Вестник КазГУКИ» содержит основные метаданные

Для извлечения метаданных автоматически производился разбор файла статьи: файл преобразовывался в формат .docx, который основан на стандарте [19], построенном на базе стандартов XML, ZIP и XML-Schema. С использованием структуры .docx-файла извлекался основной компонент документа – файл *document.xml*. На рис. 11 приведен фрагмент PHP-скрипта, предназначенного для выделения библиографического списка из текста статьи.

```
//извлекаем document.xml из каждого docx файла в коллекции.
foreach($files as $file){
    $zip = new ZipArchive;
    $res = $zip->open($file);
    $zip->extractTo('./tmp', array('word/document.xml'));
    $zip->close(); $file=iconv("cp1251","UTF-8", $file);
    //echo $file.'  
';
    $file=str_replace("docx", "pdf",$file);
    $xml = new DOMDocument; $xml->load('tmp/word/document.xml')
    $w_ps=$xml->getElementsByTagName(
    'http://schemas.openxmlformats.org/wordprocessingml/2006/main','p');
    //цикл по абзацам документа
    while($k<$w_ps->length){
        ...
        //Поиск библиографии
        if(preg_match("[0-9] (.*)//(.*) (\d{4}) (.*) (P.|C.) (.*) (\.)" ,
        $w_ps->item($k)->nodeValue )){
        }
    }
    ...
}
```

Рис. 11. Фрагмент скрипта выделения библиографического списка

Отличительными признаками блока библиографии журнала является наличие библиографических записей на русском и английском языках, заголовка «Литература» и следование инициалов автором после фамилии. В силу перечисленных особенностей блок библиографии выделялся с помощью регулярного выражения, учитывающего особенности оформления (см. рис. 12).

Литература	
1. D'Arcy P. CIO strategies for consumerization: the future of enterprise mobility. – Dell Power Solutions Special Issue, Dell Inc., 2012. – P. 22-25. – URL: http://www.dell.com/Learn/us/en/555/power-solutions-magazine-2012-special-issue .	1. D'Arcy P. CIO strategies for consumerization: the future of enterprise mobility. – Dell Power Solutions Special Issue, Dell Inc., 2012. – P. 22-25. – URL: http://www.dell.com/Learn/us/en/555/power-solutions-magazine-2012-special-issue .
2. Gantz J., Reinsel D. The Digital Universe in 2020: big data, bigger digital shadows, and biggest growth in the Far East. – IDC Digital Universe Study, December 2012. – URL: http://www.emc.com/leadership/digitaluniverse/iview/index.htm .	2. Gantz J., Reinsel D. The Digital Universe in 2020: big data, bigger digital shadows, and biggest growth in the Far East. – IDC Digital Universe Study, December 2012. – URL: http://www.emc.com/leadership/digitaluniverse/iview/index.htm .
3. Елизаров А.М., Зуев Д.С., Липачёв Е.К. Свободно распространяемые системы управления электронными научными журналами и технологии электронных библиотек// Тр. 15-й Всерос. науч. конф. «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» – RCDL-2013, Ярославль: ЯрГУ, 2013. – С. 227-236.	3. Elizarov A.M., Zuev D.S., Lipachjov E.K. Svobodno rasprostranjaemye sistemy upravlenija jelektronnymi nauchnymi zhurnalami i tehnologii jelektronnyh bibliotek// Tr. 15-j Vseros. nauch. konf. «Jelektronnye biblioteki: perspektivnye metody i tehnologii, jelektronnye kolekcii» – RCDL-2013, Jaroslavl': JarGU, 2013. – S. 227-236.
4. Елизаров А.М., Липачёв Е.К. Системы интеграции электронной научно-образовательной информации и повышение поисковой	4. Elizarov A.M., Lipachjov E.K. Sistemy integracii jelektronnoj nauchno-obrazovatel'noj informacii i

Рис. 12. Оформление списка литературы в журнале «Вестник КазГУКИ»

Для автоматического разбора библиографии было использовано несколько шаблонов регулярных выражений. Отличительным признаком статьи считалось наличие знаков “//” в библиографическом описании. В этом случае применялось регулярное выражение, в котором отдельные блоки метаданных выделялись в группы:

```
@"[0-9](.*?)/(.*?)\{4}(.*?)(P.|C.)(.*?)\.
```

Приведенное регулярное выражение позволило выделить такие группы, как список авторов, номер тома, страницы и другие блоки библиографического описания. Далее проводился разбор каждой группы, в частности, осуществлялось выделение отдельного автора из списка авторов статьи.

Для разбора библиографического описания книг использовался шаблон

```
@"([0-9]\.)(.*?)\.(.*?)\.(.*?)\{4}(.*?)(c.|p.)
```

В качестве следующего этапа улучшения алгоритма рассматривается возможность программного построения регулярных выражений на основе шаблона представленных материалов. В случае неудовлетворительного результата и указания ошибочных блоков предусмотрена возможность повторного построения выражений.

ИМПОРТ МЕТАДАННЫХ В БАЗЫ ДАННЫХ НАУЧНОГО ЦИТИРОВАНИЯ

Необходимым условием загрузки метаданных в международные информационно-аналитические базы является оформление пристатейной библиографии на латинице [5]. Соответствующий алгоритм реализован как веб-приложение, позволяющее загружать файлы статей в форматах .tex, .doc и в результате получить XML-файл, содержащий метаданные, структурированные согласно заданному формату выбранной базы цитирования. Приложение позволяет выбрать группу файлов указанием пути к папке с документами и автоматизировать процесс выделения метаданных статей целых сборников или коллекций документов, имеющих схожее форматирование, например, выпусков журнала или статей конференций. С помощью разработанного веб-приложения, реализованного на языке PHP, были получены XML-файлы, содержащие блоки метаданных, записанных в соответствии с правилами РИНЦ.

На рис. 13 приведен фрагмент XML-файла, содержащий метаданные статьи (см. рис. 10):

```
<article>
  <pages>86-96</pages>
  <artType>PRC</artType>
  <authors>
    ...
    <author num="002" id="2521">
      <individInfo lang="RUS">
        <surname>Елизаров</surname>
        <initials>А.М.</initials>
        <orgName>Казанский (Приволжский) федеральный университет</orgName>
      </individInfo>
    </author>
    ...
    <author num="004" id="10719">
      <individInfo lang="RUS">
        <surname>Липачев</surname>
        <initials>Е.К.</initials>
        <orgName>Казанский (Приволжский) федеральный университет</orgName>
      </individInfo>
    </author>
  </authors>
  <artTitles>
    <artTitle lang="RUS">ГУМАНИТАРНЫЕ НАУЧНЫЕ ЖУРНАЛЫ: МЕТОДЫ И ИНСТРУМЕНТЫ ИНТЕГРАЦИИ В НАУЧНОЕ ИНФОРМАЦИОННОЕ ПРОСТРАНСТВО
    </artTitle>
    <artTitle lang="ENG">HUMANITARIAN SCIENTIFIC JOURNALS: METHODS AND TOOLS OF INTEGRATION IN SCIENTIFIC INFORMATION SPACE
    </artTitle>
  </artTitles>
```

Рис. 13. XML-файл метаданных в формате РИНЦ

В документе, приведенном выше, атрибут id тега author есть уникальный регистрационный номер автора, присвоенный при регистрации. Заполнение данного атрибута является обязательным для автоматической привязки статьи к ее автору. Для получения уникального номера, а также дополнительной информации (представление на английском языке) был реализован отдельный модуль поиска в системе РИНЦ, который имитирует работу пользователя в веб-браузере.

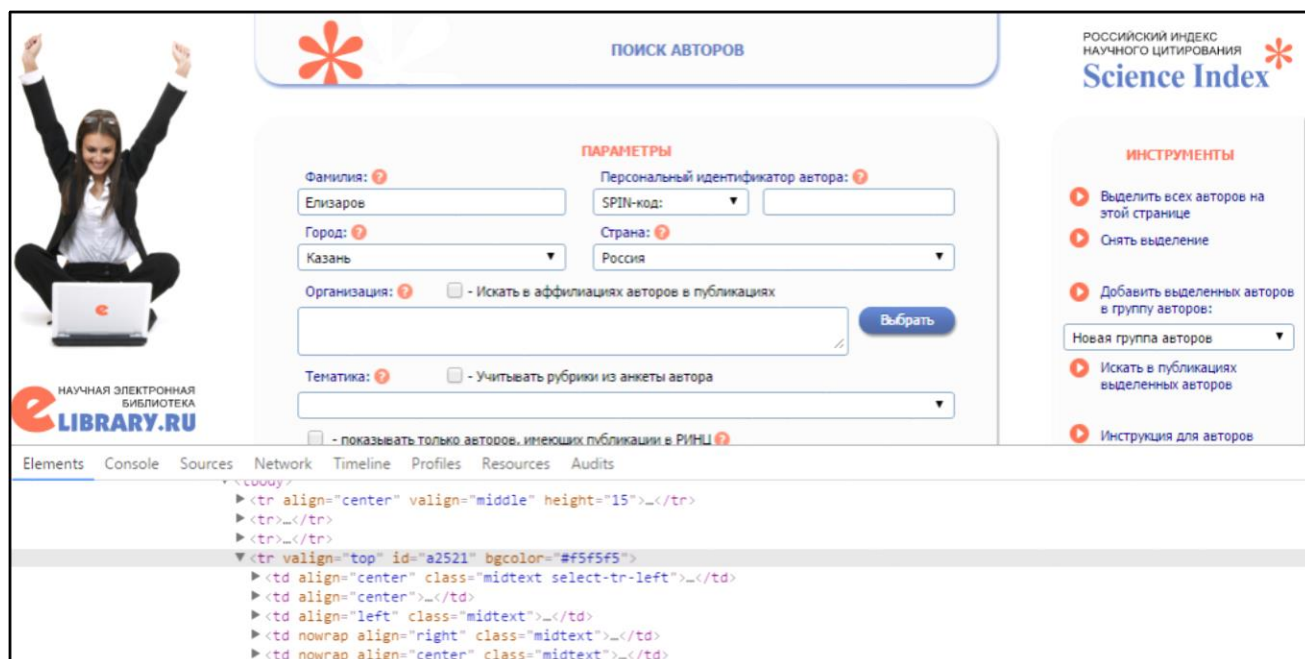


Рис. 14. Поиск автора в системе РИНЦ.

Следующий фрагмент (рис. 15) содержит часть списка литературы статьи (см. рис. 12), размеченную в автоматическом режиме:

```
<references>
<reference>D'Arcy P. CIO strategies for consumerization: the future of enter-prise
mobility. – Dell Power Solutions Special Issue, Dell Inc., 2012. – P. 22-25. – URL:
http://www.dell.com/Learn/us/en/555/powersolutions-magazine-2012-special-issue.
</reference>
...
<reference>Ахметов Д.Ю., Елизаров А.М., Липачев Е.К. Система автома-тизации редакционных
процессов на платформе электронных научных жур-налов // Ученые записки ИСГЭ. – 2014. –
№ 1-2 (12). – С. 228-233. </reference>
...
</references>
```

Рис. 15. Результат автоматической разметки списка литературы

Фрагмент XML-документа, построенного с разбиением списка литературы (см. рис. 12) на библиографические элементы, представлен на рис. 16.

```
<references>
  <reference id=1>
    <authors>
      <author>D'Arcy P.</author>
    </authors>
    <title-paper> CIO strategies for consumerization: the future of enterprise mobility.</title-paper>
    <year>2012</year>
    <pages>22-25</pages>
    <inf-bib>CIO strategies for consumerization: the future of enterprise mobility. - Dell
    Power Solutions Special Issue, Dell Inc., 2012. - P. 22-25. - URL:
    http://www.dell.com/Learn/us/en/555/powersolutions-magazine-2012-special-issue.</inf-bib>
  </reference>
  ...
  <reference id=20>
    <authors>
      <author>Ахметов Д.Ю.</author>
      <author>Елизаров А.М.</author>
      <author>Липачёв Е.К.</author>
    </authors>
    <title-paper>Свободно распространяемые системы управления электронными научными журналами
    и технологии электронных библиотек</title-paper>
    <year>2014</year>
    <pages>228-233</pages>
    <inf-bib>Система автоматизации редакционных процессов на платформе электронных научных
    журналов // Ученые записки ИСГЗ. - 2014. - № 1-2 (12). - С. 228-233.</inf-bib>
  </reference>
</references>
```

Рис. 16. Результат автоматической разметки списка литературы с разбиением на библиографические элементы

СЕРВИСЫ ИНТЕГРАЦИИ ИНФОРМАЦИОННЫХ СИСТЕМ УПРАВЛЕНИЯ ЭЛЕКТРОННЫМИ ЖУРНАЛАМИ И БАЗ НАУЧНОГО ЦИТИРОВАНИЯ

Все больше журналов начинают использовать различные системы управления электронными научными журналами, такие, как Open Journal System (OJS), ePublishing Toolkit, Ambra Publishing System и другие. Наибольшую популярность получила система OJS, являющаяся онлайн издательской системой полного цикла, что связано с наличием достаточного функционала и доступностью к его расширению [18, 12].

Внедрение платформ управления бизнес-процессами научного журнала позволяет автоматизировать наиболее трудоемкие рабочие процессы, в том числе дает возможность интеграции журнала с различными информационными системами. Платформа OJS обладает модулями экспорта метаданных в форматах:

- XML, по шаблону native.dtd;
- Erudit, определенном в виде DTD;
- CrossRef XML;
- PubMed XML для индексирования MEDLINE;

- XML для архивации в DOAJ.

В качестве расширения базовой функциональности платформы OJS была выполнена реализация алгоритмов экстракции метаданных в виде отдельного модуля, что позволило выгружать XML-файлы, сформированные в соответствии с заранее заданным форматами различных внешних информационных систем. Данный модуль представлен в виде отдельного веб-сервиса, записанного на языке WSDL (Web Services Description Language) [21] и предоставляющего доступ к журналам по протоколу SOAP (Simple Object Access Protocol) [22]. Для поддержки мультиязычности при выгрузке метаданных был реализован отдельный модуль, позволяющий выполнять транслитерацию в автоматическом режиме.

Предложенный алгоритм активно внедряется в виде веб-сервиса в информационную систему управления научными публикациями Science.Tatarstan.ru [21], при этом используются программные инструменты интеграции с международными базами цитирования системы Open Journal System [24, 25].

ЗАКЛЮЧЕНИЕ

Метаданные являются важным элементом научных публикаций, а их наличие и полнота очень важны для ориентации в научном пространстве и учета в аналитических системах. Подход, представленный выше, является шагом в развитии взаимодействия информационных систем научного характера, а внедрение платформ управления научными журналами позволит автоматизировать процесс загрузки метаданных научного контента в базы научного цитирования.

Благодарности

Работа выполнена при финансовой поддержке Российского гуманитарного научного фонда, проект 14-03-12004, а также Российского фонда фундаментальных исследований и Правительства Республики Татарстан в рамках научного проекта № 15-47-02472.

СПИСОК ЛИТЕРАТУРЫ

1. *Кириллова О.В.* Подготовка научных журналов по международным стандартам и требованиям индексов цитирования. URL: <http://lib.ss-au.ru/uploaded/Publ/IC%20and%20science%20journals.pdf>.

2. *Кириллова О.В.* О системе включения журналов в БД Scopus: основные требования и порядок представления. URL: <http://www.webci-tation.org/68vOlqztg>.

3. *Кириллова О.В.* Редакционная подготовка журналов по международным стандартам требованиям глобальных индексов цитирования. URL: http://www.mgsu.ru/science/Nssled_i_innovac_deyat/UNP/naukometriya/redakcionnaya_podgotovka_zhurnalov.pdf.

4. *Гусев А.Л.* Индексы цитирования и аналитический аппарат современной издательской платформы // *Альтернативная энергетика и экология*. 2013. №4. С. 87-91.

5. *Кириллова О.В.* Критерии отбора и рекомендации по подготовке журнала в индекс цитирования Scopus. URL: http://fano.gov.ru/common/upload/library/2014/12/main/kriterii_journals.pdf.

6. *Когаловский М.Р.* Онтологическое аннотирование библиографических ссылок в научных публикациях и его использование в наукометрии// *Информационные ресурсы России*. 2013. № 5. С. 5-13.

7. *Когаловский М.Р., Паринов С.И.* Новый источник данных для наукометрических исследований// *Труды 15-й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» – RCDL-2013*, Ярославль, Россия, 14–17 октября 2013 г. С. 107-117.

8. *Паринов С.И., Когаловский М.Р.* Технология поддержки электронных научных публикаций как «живых» документов // *Труды XI Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции»*. — Петрозаводск: Карельский научный центр РАН, 17–21 сентября 2009. С. 53-58.

9. *Горбунов-Посадов М.М.* Живая публикация // *Открытые системы*. СУБД. 2011. № 4. С. 48-49.

10. Оформление библиографического списка. URL: http://lib.samgtu.ru/making_list.

11. *Афонин С.А., Бахтин А.В., Бухонов В.Ю., Васенин В.А., Ганкин Г.М., Гаспарянц А.Э., Голомазов Д.Д., Иткес А.А., Козицын А.С., Тумайкин И.Н., Шапченко К.А.*

Интеллектуальная система тематического исследования научно-технической информации (ИСТИНА). Под ред. академика В.А. Садовниченко. М.: Издательство Московского университета, 2014. 262 с.

12. *Елизаров А.М., Зуев Д.С., Липачёв Е.К.* Информационные системы управления электронными научными журналами // Научно-техническая информация. Серия 1: Организация и методика информационной работы. 2014. № 3. С. 31-38.

13. *Елизаров А.М., Липачёв Е.К., Хохлов Ю.Е.* Семантические методы структурирования математического контента, обеспечивающие расширенную поисковую функциональность // Информационное общество. 2013. № 1–2. С. 83-92.

14. *Елизаров А.М., Липачёв Е.К., Невзорова О.А., Соловьев В.Д.* Методы и средства семантического структурирования электронных математических документов // Доклады Академии наук. 2014. Т. 457, № 6. С. 642-645.

15. *Гойвертс Я., Левитан С.* Регулярные выражения. Сборник рецептов. СПб: Символ-Плюс, 2013. 608 с.

16. *Герасимов А.Н., Елизаров А.М., Липачёв Е.К.* Паттерны регулярных выражений в алгоритмах семантической обработки коллекций математических документов // Труды Казанской школы по компьютерной и когнитивной лингвистике. Казань, 2014. Вып. 16. С. 43-45.

17. *Глухов В.А.* Книжная коллекция. О размещении книг, сборников и материалов конференций в Российском индексе научного цитирования: Научная электронная библиотека. URL: http://www.library.spbu.ru/blog/wp-content/uploads/2014/12/books_Glukhov.pdf.

18. *Мбого И.А., Прокудин Д.Е., Чугунов А.В.* Комплексная интеграция цифровых коллекций в информационное пространство научных исследований // Технологии информационного общества в науке, образовании и культуре: сборник научных статей. Материалы XVII Всероссийской объединенной конференции «Интернет и современное общество» IMS-2014, Санкт-Петербург, 19 – 20 ноября 2014 г. С. 48-53.

19. Standard ECMA-376: Office Open XML File Formats. URL: <http://www.ecmainternational.org/publications/standards/Ecma-376.htm>.

20. *Willinsky J., Stranack K., Smecher A., MacGregor J.* Open Journal Systems: a complete guide to online publishing. Simon Fraser University Library, 2010. URL: <https://pkp.sfu.ca/ojs/docs/userguide/2.3.3/index.html>.

21. Web Services Description Language (WSDL) Version 2.0. W3C Recommendation. – <http://www.w3.org/TR/2007/REC-wsdl20-20070626>.

22. *Snell J., Tidwell D., Kulchenko P.* Programming Web Services with SOAP. O'Reilly Media, 2001. 262 p.

23. *Ахметов Д.Ю., Грачев А.О., Герасимов А.Н., Елизаров А.М., Липачёв Е.К.* Облачная платформа поддержки электронных научных изданий // Учёные записки Института социальных и гуманитарных знаний. 2014. № 1 (12), ч. 1. С. 13-19.

24. *Stranack K.* Getting found, staying found, increasing impact. Enhancing readership and preserving content for OJS journals // Public Knowledge Project. 2006. 40 p.

25. Importing and Exporting Data with OJS // URL: <http://www.informatica.si/manual/ojs-import-export.pdf>.

SUBSYSTEM OF FORMATION METADATA FOR SCIENCE INDEX DATABASES ON MANAGEMENT PLATFORM ELECTRONIC SCIENTIFIC JOURNALS

A.N. Gerasimov¹, A.M. Elizarov², E.K. Lipachev³

N.I. Lobachevskii Institute of Mathematics and Mechanics. Kazan Federal University

¹sav241@mail.ru, ²amelizarov@gmail.com, ³elipachev@gmail.com

Abstract

An algorithm for automatic extraction of bibliographic data from a one-dimensional array of publications (in particular, the issues of the scientific journal) and the formation of metadata blocks for export to international information-analytical system. The methods of integration management platform electronic scientific journals (Open Journal Systems) and Science Citation Index.

Keywords: *publishing systems, digital scientific journal, the integration of electronic resources, databases, scientific citation, metadata extraction*

REFERENCES

1. *Kirillova O.V.* Podgotovka nauchnykh zhurnalov po mezhdunarodnym standartam i trebovaniyam indeksov tsitirovaniya. URL: <http://lib.ss-au.ru/uploaded/Publ/IC%20and%20science%20journals.pdf>.
2. *Kirillova O.V.* O sisteme vklyucheniya zhurnalov v BD Scopus: osnovnye trebovaniya i poryadok predstavleniya. URL: <http://www.webci-tation.org/68vOlqztg>.
3. *Kirillova O.V.* Redaktsionnaya podgotovka zhurnalov po mezhdunarodnym standartami trebovaniyam global'nykh indeksov tsitirovaniya. URL: http://www.mgsu.ru/science/Nssled_i_innovac_deyat/UNP/naukometriya/redakcionnaya_podgotovka_zhurnalov.pdf.
4. *Gusev A.L.* Indeksy tsitirovaniya i analiticheskij apparat sovremennoj izdatel'skoj platformy // *Al'ternativnaya ehnergetika i ehkologiya*. 2013. №4. S. 87-91.
5. *Kirillova O.V.* Kriterii otbora i rekomendatsii po podgotovke zhurnala v indeks tsitirovaniya Scopus. URL: http://fano.gov.ru/common/upload/library/2014/12/main/kriterii_journ-als.pdf.
6. *Kogalovskij M.R.* Ontologicheskoe annotirovanie bibliograficheskikh ssylok v nauchnykh publikacijah i ego ispol'zovanie v naukometrii// *Informacionnye Resursy Rossii*. 2013. № 5. S. 5-13.
7. *Kogalovskij M.R., Parinov S.I.* Novyj istochnik dannyh dlja nauko-metricheskikh issledovanij// *Trudy 15-j Vserossijskoj nauchnoj konferencii «Jelektronnye biblioteki: perspektivnye metody i tehnologii, jelektronnye kollekcii» — RCDL-2013, Jaroslavl', Rossija, 14–17 oktjabrja 2013 g.* S. 107-117.
8. *Parinov S.I., Kogalovskij M.R.* Tekhnologiya podderzhki ehlektronnykh nauchnykh publikatsij kak «zhivykh» dokumentov // *Trudy XI Vserossijskoj nauchnoj konferentsii «EHlektronnye biblioteki: perspektivnye metody i tehnologii, ehlektronnye kollekcii»*. Petrozavodsk: Karel'skij nauchnyj tsentr RAN, 17–21 sentyabrya 2009. S. 53-58.
9. *Gorbunov-Posadov M.M.* Zhivaya publikatsiya // *Otkrytye sistemy. SUBD*. 2011. № 4. S. 48-049.

10. Oformlenie bibliograficheskogo spiska. URL: http://lib.samgtu.ru/making_list.

11. *Afonin S.A., Bahtin A.V., Buhonov V.Ju., Vasenin V.A., Gankin G.M., Gasparjanc A.Je., Golomazov D.D., Itkes A.A., Kozicyan A.S., Tumajkin I.N., Shapchenko K.A.* Intel'ktual'naja sistema tematicheskogo issledovanija nauchno-tehnicheskoy informacii (ISTINA). Pod red. akademika V.A. Sadovnichego. M.: Izdatel'stvo Moskovskogo universiteta, 2014. 262 s.

12. *Elizarov A.M., Zuev D.S., Lipachev E.K.* Informatsionnye sistemy upravleniya ehlektronnyimi nauchnymi zhurnalami // Nauchno-tehnicheskaya informatsiya. Seriya 1: Organizatsiya i metodika informatsionnoj raboty. 2014. № 3. S. 31-38.

13. *Elizarov A.M., Lipachev E.K., Khokhlov Yu.E.* Semanticheskie metody strukturovaniya matematicheskogo kontenta, obespechivayushhie rasshirennuyu poiskovuyu funktsional'nost' // Informatsionnoe obshhestvo. 2013. № 1-2. S. 83-92.

14. *Elizarov A.M., Lipachev E.K., Nevzorova O.A., Solov'ev V.D.* Metody i sredstva semanticheskogo strukturovaniya ehlektronnykh matematicheskikh dokumentov // Doklady Akademii nauk. 2014. T. 457. № 6. S. 642-645.

15. *Gojverts J., Levithan C.* Reguljarnye vyrazhenija. Sbornik receptov. SPB: Simvol-Pljus, 2013. 608 p.

16. *Gerasimov A.N., Elizarov A.M., Lipachev E.K.* Patterny reguljarnykh vyrazhenij v algoritmakh semanticheskoy obrabotki kolleksij matematicheskikh dokumentov // Trudy Kazanskoj shkoly po komp'yuternoj i kognitivnoj lingvistike. Kazan', 2014, Vyp. 16. S. 43-45.

17. *Glukhov V.A.* Knizhnaja kolleksija. O razmeshhenii knig, sbornikov i materialov konferencij v Rossijskom indekse nauchnogo citirovanija: Nauchnaja jelektronnaja biblioteka. URL: http://www.library.spbu.ru/blog/wp-content/uploads/2014/12/books_Glukhov.pdf.

18. *Mbogo I.A., Prokudin D.E., Chugunov A.V.* Kompleksnaya integratsiya tsifrovyykh kolleksij v informatsionnoe prostranstvo nauchnykh issledovanij // Tekhnologii informatsionnogo obshhestva v nauke, obrazovanii i kul'ture: sbornik nauchnykh statej. Materialy XVII Vserossijskoj ob"edinennoj konferentsii «Internet i sovremennoe obshhestvo» IMS-2014, Sankt-Peterburg, 19 – 20 noyabrya 2014 g. S. 48-53.

19. Standard ECMA-376: Office Open XML File Formats. URL: <http://www.ecmainternational.org/publications/standards/Ecma-376.htm>.

20. *Willinsky J., Stranack K., Smecher A., MacGregor J.* Open Journal Systems: A complete guide to online publishing. Simon Fraser University Library, 2010. URL: <https://pkp.sfu.ca/ojs/docs/userguide/2.3.3/index.html>.

21. Web Services Description Language (WSDL) Version 2.0. W3C Recommendation. URL: <http://www.w3.org/TR/2007/REC-wsdl20-20070626>.

22. *Snell J., Tidwell D., Kulchenko P.* Programming Web Services with SOAP. O'Reilly Media, 2001. 262 p.

23. *Akhmetov D.Yu., Grachev A.O., Gerasimov A.N., Elizarov A.M., Lipachev E.K.* Oblachnaya platforma podderzhki ehlektronnykh nauchnykh izdanij // Uchyonye zapiski Instituta sotsial'nykh i gumanitarnykh znaniy. 2014. № 1 (12). ch. 1. S. 13-19.

24. *Stranack K.* Getting found, staying found, increasing impact. Enhancing Readership and Preserving Content for OJS Journals // Public Knowledge Project. 2006. 40 p.

25. *Importing and Exporting Data with OJS.* URL: <http://www.informatica.si/manual/ojs-import-export.pdf>. URL: <http://www.informatica.si/manual/ojs-import-export.pdf>.

СВЕДЕНИЯ ОБ АВТОРАХ



ГЕРАСИМОВ Алексей Николаевич – аспирант Института математики и механики им. Н.И. Лобачевского Казанского (Приволжского) федерального университета.

Alex Nikolaevich GERASIMOV, received MS degree in Mathematics from Kazan Federal University, (2012). Currently is a graduate student at the N.I. Lobachevskii Institute of Mathematics and Mechanics of Kazan Federal University. Current scientific interests: data mining, knowledge extraction technologies, integration of scientific resources.

email: gerasimov.mailstore@gmail.com



ЕЛИЗАРОВ Александр Михайлович – доктор физико-математических наук, профессор, заслуженный деятель науки Республики Татарстан, зам. директора по научной деятельности Института математики и механики им. Н.И. Лобачевского Казанского (Приволжского) федерального университета.

Alexander Mikhailovich ELIZAROV – Doctor of Physics and Mathematics, Professor, Honoured Worker of Science of the Republic of Tatarstan, Deputy Director of the Lobachevskii Institute of Mathematics and Mechanics of Kazan Federal University. Current scientific interests: data mining, recommender systems, cloud computing, knowledge extraction technologies.

email: amelizarov@gmail.com



ЛИПАЧЁВ Евгений Константинович – кандидат физико-математических наук, доцент кафедры теории функций и приближений Института математики и механики им. Н.И. Лобачевского Казанского (Приволжского) федерального университета.

Evgeny Konstantinovich LIPACHEV – Candidate of Physics and Mathematics, Associate Professor, Lobachevskii Institute of Mathematics and Mechanics of Kazan Federal University. Current scientific interests: data mining, recommender systems, cloud computing, knowledge extraction technologies.

email: elipachev@gmail.com

Материал поступил в редакцию 15 января 2015 года

УДК 004.82

АВТОМАТИЗАЦИЯ РЕДАКЦИОННЫХ ПРОЦЕССОВ В ИНФОРМАЦИОННОЙ СИСТЕМЕ УПРАВЛЕНИЯ ЭЛЕКТРОННЫМИ НАУЧНЫМИ ЖУРНАЛАМИ

Д.Ю. Ахметов¹, А.М. Елизаров², Е.К. Липачёв³

*Институт математики и механики им. Н.И. Лобачевского Казанского
(Приволжского) федерального университета*

¹akhmetov.dy@gmail.com, ²amelizarov@gmail.com, ³elipachev@gmail.com

Аннотация

Исследованы особенности использования информационных систем в процессе издания электронных научных журналов и проведено их сравнение с точки зрения автоматизации редакционных процессов. Описаны программные модули, созданные для расширения функционала платформы Open Journal Systems в целях автоматизации ряда редакционных процессов электронного научного журнала. Приведены алгоритмы автоматической стилевой валидации текстов на этапе регистрации автором статьи в информационной системе электронного научного журнала, автоматического подбора рецензентов, рассылки уведомлений и контроля сроков рецензирования.

Ключевые слова: *издательские системы, электронный научный журнал, интеграция электронных ресурсов, данные научного цитирования, экстракция метаданных, Open Journal Systems*

ВВЕДЕНИЕ

Сегодня электронные издания являются основным источником научной информации, а подавляющее число современных печатных изданий имеет электронные варианты, доступные по подписке или в свободном доступе.

Переход к электронной форме научных изданий обусловлен многими причинами, прежде всего, постоянно возрастающими объемами создаваемой информации, обработка которой уже невозможна без компьютера. Несмотря на повсеместное использование информационно-коммуникационных технологий

практически на всех этапах подготовки и рассмотрения научной работы в редакции научного журнала, организация редакционных процессов по-прежнему остается традиционной – сложившейся в научных издательствах, ориентированных на печатные издания. При этом редакции научных журналов регламентируют правила оформления поступающих материалов и самостоятельно организуют научное рецензирование. Одновременно для обработки постоянно растущего объема информации стали создаваться и внедряться специализированные информационно-издательские системы – к настоящему времени в мире внедрены десятки таких систем, построенных на различных программных платформах и позволяющих управлять редакционно-издательскими процессами (см., например, [1]). Развитие мобильных устройств и облачных технологий и поддержка инициативы использования персональных устройств позволяют ученому работать в любом месте, где есть доступ в интернет. Вместе с тем, согласно исследованиям [2], уже к 2020 году объем информации в Сети достигнет величины 40 зеттабайт, из которой более 40% будет сгенерировано самими инструментами Cloud Computing для обеспечения их оптимальной работы. Все это значительно усложняет поиск научной информации.

Проблема оперативного доступа к результатам научных исследований начала успешно разрешаться с внедрением информационно-издательских систем (см., например, [1, 3]), часть из которых разработана научными сообществами. К настоящему времени сформировано несколько подходов к построению таких систем. Отметим семантическую сеть систем, реализованную в HyperJournal [4], – несколько журналов на этой платформе можно объединить в единое информационное пространство. Полный цикл управления редакционно-издательскими процессами поддерживается платформой Open Journal Systems (OJS) (<https://pkp.sfu.ca/ojs/>).

В настоящей работе исследованы особенности использования информационно-издательских систем в процессе подготовки выпусков электронных научных журналов с точки зрения автоматизации редакционных процессов. Предложен алгоритм автоматизации процесса экспертной оценки, включая подбор рецензентов и сопровождение основных этапов рецензирования. На основе технологии

расширения функционала открытой системы OJS созданы программные модули, обеспечивающие автоматизацию ряда редакционных процессов.

РЕДАКЦИОННО-ИЗДАТЕЛЬСКИЕ ПРОЦЕССЫ И СЕРВИСЫ ИХ АВТОМАТИЗАЦИИ

К редакционно-издательским процессам, связанным с деятельностью авторов и редколлегии электронного научного журнала, как правило, относят: представление материалов, их рецензирование и доработку, учитывающую замечания редакции и рецензентов. Далее следуют классификация, аннотирование, метаописание, объединение в коллекцию и публикация. Специфическими редакционно-издательскими процессами, характерными для электронных научных журналов, являются: конвертирование в различные форматы представления, распространение, одновременное распространение информации на различные страницы или веб-сайты (синдикация), рассылка уведомлений, сбор и архивирование сетевых ресурсов (веб-харвестинг), сбор статистики использования ресурсов, взаимодействие с институциональными репозиториями, контроль доступа, подписка, долгосрочное хранение (подробнее см. [5, 6]).

В функционале программных платформ управления электронными научными журналами можно выделить основные и дополнительные (опциональные) сервисы, которые, в свою очередь, можно разделить на сервисы подготовки научной публикации – сервисы для автора (например, инструменты подготовки и оформления материалов в соответствии с редакционными требованиями), сервисы стилевой валидации и поддержки научного рецензирования (например, сервисы подбора рецензентов и подготовки текстов рецензий по шаблонам редакции), сервисы поддержки редакционных процессов и сервисы для читателя (например, поиск статей, «близких» к рассматриваемой публикации).

Сервис стилевой валидации предназначен для автоматической проверки загруженного материала (рукописи статьи) на соответствие установленным в журнале требованиям оформления. В качестве примера назовем сервис проверки

физико-математических статей, набранных в Т_EX-нотации, на правильное использование стилевых конструкций из стилевого файла, разработанного редакцией журнала [7].

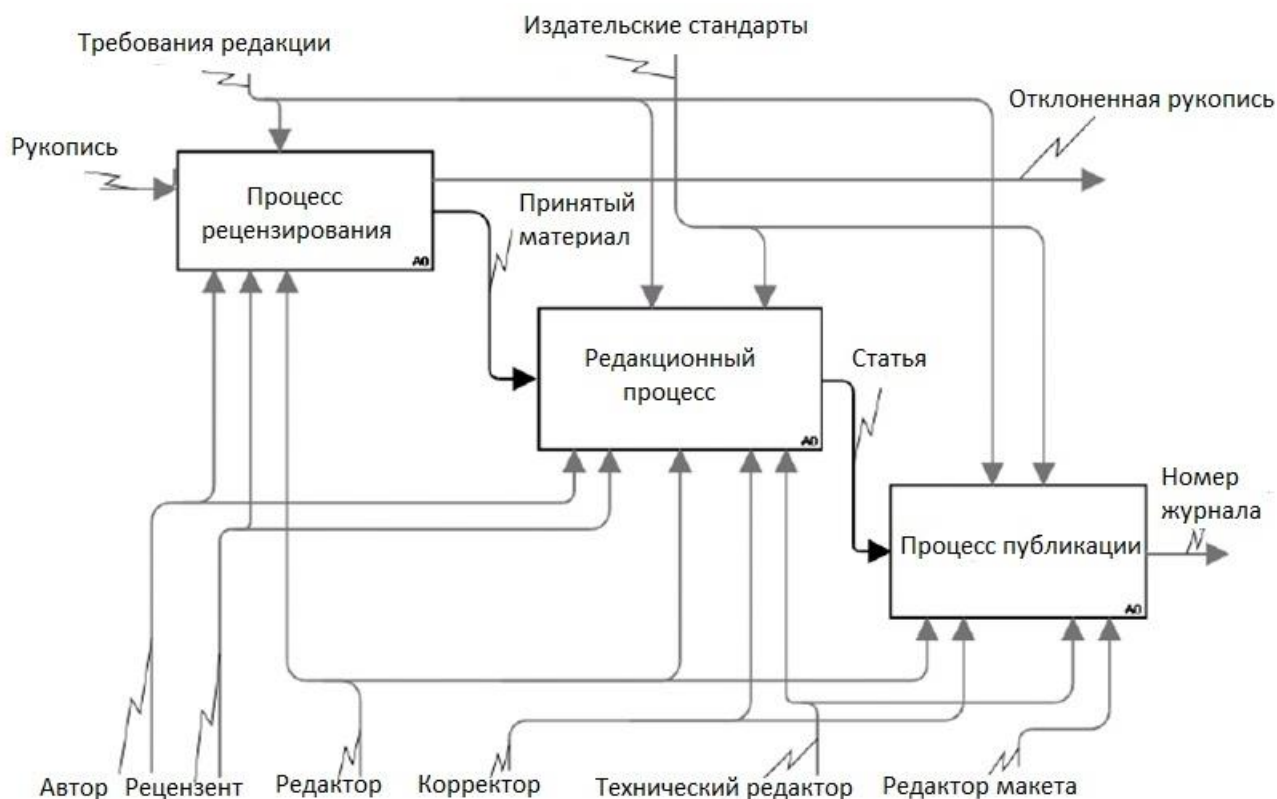


Рис. 1. Модель редакционно-издательских процессов в нотации IDEF0 [8]

Сервис поддержки научного рецензирования позволяет подобрать рецензентов для каждой работы и в автоматическом режиме управлять процессом рецензирования, включая рассылку уведомлений и напоминаний. Один из вариантов такой автоматизации реализован в журнале Lobachevskii Journal of Mathematics [9]. Этот подход получил дальнейшее развитие на платформе science.tatarstan.ru [10, 11].

Информационная система управления электронным научным журналом должна обеспечивать автоматизацию редакционных процессов, включая классификацию, аннотирование, выделение метаданных, публикацию, объединение в коллекцию, организацию поиска и навигации, долгосрочное хранение, конвертирование в различные форматы и распространение, контроль доступа, подписку,

рассылку уведомлений, анализ статистики использования, формирование наукометрических данных. Также в функционале систем должны присутствовать дополнительные сервисы, обеспечивающие проверку загружаемых текстов на заимствования; подготовку OAI-метаданных (например, на основе Open Harvester Systems); информетрический анализ (например, подключением сервиса Article-Level Metrics в системе OJS); поддержку научных блогов и новостных лент научных журналов.

СЕРВИС ПОДДЕРЖКИ НАУЧНОГО РЕЦЕНЗИРОВАНИЯ

Важной отличительной особенностью работы редколлегии научного журнала является организация независимого рецензирования с привлечением специалистов в предметных областях, определенных тематикой журнала. Процесс научного рецензирования – как правило, наиболее трудоемкий при рассмотрении научных материалов в редакции журнала. Критическими по времени при этом являются подбор рецензентов для квалифицированной оценки работы, поступившей в журнал, а также само рецензирование. Автоматизация этого процесса способна сократить затраты времени – это подтверждает и практический опыт (см., например, [9]). Такая автоматизация предполагает выбор рецензента из базы экспертов, сформированной в журнале, с учётом их профессиональных компетенций, объёма рецензирования и текущей загруженности.

Для автоматического подбора рецензентов предложен следующий алгоритм. С помощью ключевых слов, указанных в рукописи и введенных автором при регистрации научной работы в информационной системе журнала, а также классификационных признаков (УДК, ББК, Mathematics Subject Classification), в автоматическом режиме определяется область исследований. Далее на основе информации из базы данных информационной системы журнала создается список возможных рецензентов. При выборе конкретных рецензентов учитываются их загруженность и история рецензирования. Кроме того, исключается возможность саморецензирования (автор не может рецензировать свою работу). Процесс экспертной оценки организован по стандартному принципу «принять в печать» – «вернуть с замечаниями на доработку» – «отклонить». В автоматическом режиме

осуществляются контроль сроков рецензирования, а также рассылка уведомлений авторам и рецензентам. Опишем подробнее основные шаги алгоритма автоматического подбора рецензентов.

На первом шаге алгоритма формируются входные параметры, идентифицирующие материал, загруженный автором. На следующем шаге из базы данных экспертов составляется список специалистов по тематике представленной работы, из которого на основании указанных ранее правил отбираются возможные рецензенты. Далее по критерию наилучшего соответствия (число совпадений по ключевым словам, количество рецензируемых работ и др.) в зависимости от настроек системы производится отбор необходимого количества рецензентов (как правило, не менее двух). Затем осуществляется рассылка уведомлений рецензентам и редакторам. Отказ кого-либо из экспертов от рецензирования приводит к уменьшению необходимого количества экспертов – в этом случае система формирует новый список.

СЕРВИС АВТОМАТИЧЕСКОГО НАЗНАЧЕНИЯ РЕЦЕНЗЕНТОВ В OPEN JOURNAL SYSTEMS

Назначение рецензентов в системе OJS производится пользователем, имеющим роль «Редактор» или «Редактор раздела». Выбор рецензентов производится из списка пользователей системы, имеющих роль «Рецензент». Выбор можно выполнить только в ручном режиме – на основании профессионального опыта редактора. Автоматизация процедуры подбора рецензента требует изменения программного кода системы.

Отметим, что система OJS предоставляет возможности изменения своего функционала путем добавления модулей (плагинов), составленных по определенным правилам (см., например, [12–16]). Поскольку система написана на языке PHP и является открытой, имеется возможность внести изменения в функционал системы без специального инструментария OJS.

Все модули в системе разбиты на категории, поэтому разработка нового модуля начинается с выбора соответствующей категории. Системные модули, расширяющие возможности OJS и не попадающие ни в одну из категорий, определяются как «Остальные модули» и располагаются в директории `\plugins\generic`.

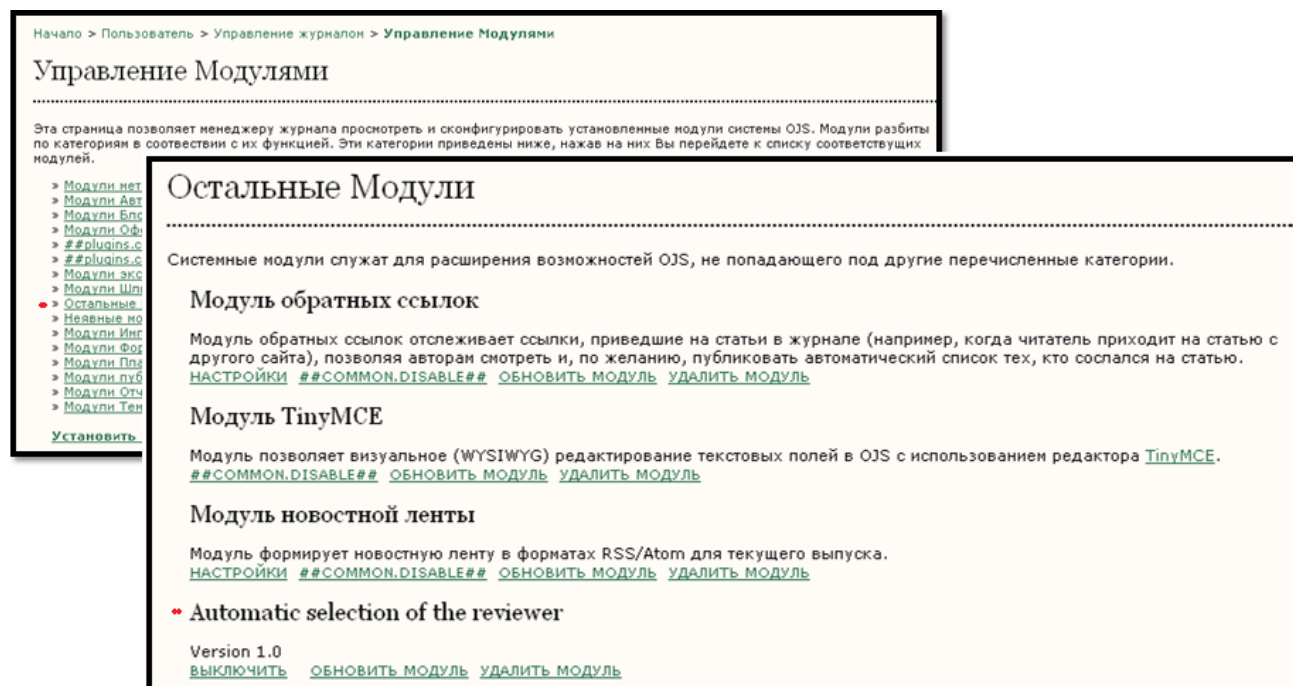


Рис. 2. Управление модулями в OJS

При разработке системных модулей OJS все создаваемые файлы размещаются в отдельной папке; служебная информация сохраняется в файле version.xml; код модуля записывается на языке PHP в виде файла myPlugin.inc.php, где myPlugin – название модуля; для вызова модуля создается файл index.php my.

Реализация алгоритма автоматического выбора рецензентов научной статьи потребовала провести ряд изменений как в программном коде OJS, так и в используемой базе данных. Прежде всего, в файле peerReview.tpl к пункту «Выбрать рецензента» добавлен новый пункт меню «Automatic selection of the reviewer», передающий управление по внешней ссылке. Идентификатор рецензируемой работы записывается в адресную строку с указанием абсолютного пути к файлу скрипта и передается методом GET (см. рис. 3).

```

<form>
  {assign var="idij" value=$submission->getId()}
  <a href='http://.../plugins/generic/autoReviewer/myPlugin.inc.php
  ?iu={$idij}'>Automatic selection of the reviewer</a>
</form>
    
```

Рис. 3. Фрагмент кода

После активации модуля считывается значение параметров из адресной строки и формируется список первичных ключей таблицы «roles». Далее создаются массивы, содержащие идентификаторы рецензентов, производятся сортировка и отбор рецензентов. Формируются переменные для вычисления текущего времени и времени, к которому истекает срок подачи рецензии (по умолчанию OJS для составления рецензии выделяет 4 недели):

```
$newdate=date("Y-m-d H:i:s");
```

```
$newdate2=date ('Y-m-d 00:00:00', strtotime ('+4 weeks')).
```

Затем в базе данных устанавливается связь между статьями и возможными рецензентами. После окончания всех операций с данными организуется возврат в систему.

The screenshot shows the '#2 Рецензии' (Reviews) page in OJS. It includes a navigation bar with 'РЕЗЮМЕ', 'РЕЦЕНЗИИ', 'РЕДАКЦИЯ', 'ИСТОРИЯ', and 'ЛИТЕРАТУРА'. The 'Статья' (Article) section displays metadata: 'Авторы: Единственный Показатель Автор', 'Название: Пример для автоматизации', 'Раздел: Статьи', 'Редактор: Иван Иванов', and 'Версия рецензии: 2-4-1-RV.TXT 2013-05-07'. There is a file upload field for 'Загрузить отредактированный файл' and buttons for 'Обзор...' and 'Закрыть'. Below this, the 'Рецензия' (Review) section shows 'Раунд 1' and a link for 'Автоматический выбор рецензента' (Automatic selection of the reviewer), which is circled in red. A code snippet is provided below:

```
Фрагмент кода:  
<form>  
  {assign var="idij" value=$submission->getId();  
  <a href='http://system2.ru/plugins/generic/autoReviewer/myPlugin.inc.php?iu=  
    {$idij}'>Automatic selection of the reviewer</a>  
</form>
```

Рис. 4. Страница выбора рецензента

ЗАКЛЮЧЕНИЕ

Предложены модель редакционно-издательского процесса и алгоритм автоматизации выбора экспертов для проведения квалифицированного рецензирования. Практическая реализация алгоритма осуществлена для системы Open Journal Systems.

Благодарности

Работа выполнена при финансовой поддержке Российского гуманитарного научного фонда, проект 14-03-12004, а также Российского фонда фундаментальных исследований и Правительства Республики Татарстан в рамках научного проекта № 15-47-02472.

СПИСОК ЛИТЕРАТУРЫ

1. *Елизаров А.М., Зуев Д.С., Липачёв Е.К.* Информационные системы управления электронными научными журналами // Научно-техническая информация. Серия 1: Организация и методика информационной работы. 2014. № 3. С. 31-38.

2. *Gantz J., Reinsel D.* The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East. – IDC Digital Universe Study, December 2012. URL: <http://www.emc.com/leadership/digitaluniverse/iview/index.htm>.

3. *Елизаров А.М., Зуев Д.С., Липачёв Е.К.* Свободно распространяемые системы управления электронными научными журналами и технологии электронных библиотек // Тр. XV Всерос. науч. конф. «Электронные библиотеки: перспективные методы и технологии, электронные коллекции, RCDL'2013», г. Ярославль, 14 – 17 октября 2013 года. Ярославль: Изд-во ЯрГУ им. П.И. Демидова, 2013. С. 227-236. URL: http://rcdl2013.uni-yar.ac.ru/doc/full_text/s3_1.pdf.

4. *Barbera M., Donato F., Morbidoni C., Tummarello G.* HyperJournal software, PHP scripting and Semantic Web technologies for the Open Access// In ESWC: European Semantic Web Conference, Heraklion (Greece), 2005. URL: <http://eprints.rclis.org/8295/>.

5. *Елизаров А.М., Липачёв Е.К., Хохлов Ю.Е.* Технологии облачных вычислений для поддержки функционирования электронного научного журнала // Материалы Межд. науч.-практ. конф. ИТОН-2012. Казань: Казанский ун-т, 2012. С. 82-85. URL: <http://vuz.exponenta.ru/PDF/NAUKA/Sbornik12ito.pdf>.

6. *Elizarov A., Lipachev E., Zuev D.* Mathematical content semantic markup methods and open scientific e-journals management systems // Proceedings of 5th Int. Conference «Knowledge engineering and Symantec Web» – KESW 2014, Kazan, Russia, 29 September 2014 – 1 October 2014, CCIS No 468. P. 242-251.

7. Ахметов Д.Ю., Елизаров А.М., Липачёв Е.К. Автоматизация процесса первичной обработки математической статьи в информационной системе электронного научного журнала // Тр. Математического центра имени Н.И. Лобачевского. Материалы Двенадцатой молодежной науч. шк.-конф. «Лобачевские чтения – 2013». Казань: Изд-во Казан. матем. об-ва, 2013. Т. 47. С. 6-10.

8. *Методология функционального моделирования IDEF0*. Руководящий документ, Госстандарт России. М.: ИПК Издательство стандартов, 2000. 75 с.

9. Елизаров А.М., Липачёв Е.К., Малахальцев М.А. Веб-технологии в работе электронного математического журнала Lobachevskii Journal of Mathematics // Научный сервис в сети Интернет: многоядерный компьютерный мир. 15 лет РФФИ. Труды Всерос. науч. конф. М.: Изд-во МГУ, 2007. С. 355-356.

10. Ахметов Д.Ю., Елизаров А.М., Липачёв Е.К. Система автоматизации редакционных процессов на платформе электронных научных журналов // Учёные записки Института социальных и гуманитарных знаний. 2014. № 1 (12), Ч. 2. С. 228-233.

11. Ахметов Д.Ю., Герасимов А.Н., Грачев А.О., Елизаров А.М., Липачёв Е.К. Облачная платформа поддержки электронных научных изданий // Учёные записки Института социальных и гуманитарных знаний. 2014. № 1 (12), Ч. 1. С. 13-19. URL: http://www.isgz.ru/images/Alexey/Chirko/ek%202014_i.pdf.

12. *Open Journal Systems. Technical Reference*. – Public Knowledge. Simon Fraser University Library. URL: <https://pkp.sfu.ca/ojs/OJSTechnicalReference.pdf>.

13. *Willinsky J., Stranack K., Smecher A., MacGregor J.* Open Journal Systems: a complete guide to online publishing. – Simon Fraser University Library, 2010. 273 p. URL: <https://pkp.sfu.ca/ojs/docs/userguide/2.3.3/index.html>.

14. *Open Journal Systems. OJS hosting, installation, upgrade, theme and plugin, customization, training and support*. – URL: <http://openjournalssystems.com>.

15. *Plugin Gallery – PKP Support*. URL: <https://pkp.sfu.ca/support/forum/viewforum.php?f=28>.

16. *Chen S.* LATEXGalley Plugin for Open Journal Systems. – Simon Fraser University and Zhejiang University, 2010. 30 p.

INFORMATION SYSTEMS OF ELECTRONIC SCIENTIFIC JOURNALS AND EDITORIAL PROCESS AUTOMATION

D.Yu. Akhmetov¹, A.M. Elizarov², E.K. Lipachev³

N.I. Lobachevskii Institute of Mathematics and Mechanics. Kazan Federal University

¹akhmetov.dy@gmail.com, ²amelizarov@gmail.com, ³elipachev@gmail.com

Abstract

Features of use of information systems in the process of electronic scientific publications issues investigated. From the standpoint of automation editorial processes, we compared these systems. Algorithms for automatic validation of mathematical texts style introduced into the information system of the electronic scientific journal. The method of automating the process of peer review, including the selection of reviewers, e-mail notifications and control the timing of reviewing proposed. Automation technology a number of editorial processes of electronic scientific journal, operates on a platform Open Journal Systems is presented.

Keywords: *publishing systems, digital scientific journal, the integration of electronic resources, databases, scientific citation, OJS*

REFERENCES

1. *Elizarov A., Zuev D., Lipachev E.* Electronic scientific journal management systems // Scientific and Technical Information Processing. 2014. V. 41, No 1. P. 66-72.
2. *Gantz J., Reinsel D.* The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East. – IDC Digital Universe Study, December, 2012. URL: <http://www.emc.com/leadership/digitaluniverse/iview/index.htm>.
3. *Elizarov A.M., Zuev D.S., Lipachev E.K.* Open scientific e-journals management systems and digital libraries technology // CEUR Workshop Proc. Selected Papers of the 15th All-Russian Scientific Conference "Digital libraries: Advanced Methods and Technologies, Digital Collections", Yaroslavl, Russia, October 14 – 17, 2013. V. 1108. P. 102-111. URL: <http://ceur-ws.org/Vol-1108/paper13.pdf>.

4. *Barbera M., Donato F., Morbidoni C., Tummarello G.* HyperJournal software, PHP scripting and Semantic Web technologies for the Open Access // In ESWC: European Semantic Web Conference, Heraklion (Greece), 2005. URL: <http://eprints.rclis.org/8295/>.

5. *Elizarov A.M., Lipachev E.K., Khokhlov Yu.E.* Tekhnologii oblachnykh vychislenij dlya podderzhki funktsionirovaniya ehlektronnogo nauchnogo zhurnala // Materialy Mezhd. nauch.-prakt. konf. ITON-2012. – Kazan: Kazanskij un-t, 2012. S. 82-85. URL: <http://vuz.exponenta.ru/PDF/NAUKA/Sbornik12ito.pdf>.

6. *Elizarov A., Lipachev E., Zuev D.* Mathematical content semantic markup methods and open scientific e-journals management systems // Proceedings of 5th Int. Conference «Knowledge engineering and Symantec Web» – KESW 2014, Kazan, Russia, 29 September 2014 – 1 October 2014, CCIS No 468. P. 242-251.

7. *Akhmetov D.Yu., Elizarov A.M., Lipachev E.K.* Avtomatizatsiya protsessa pervichnoj obrabotki matematicheskoy stat'i v informatsionnoj sisteme ehlektronnogo nauchnogo zhurnala // Tr. Matematicheskogo tsentra imeni N.I. Lobachevskogo. Materialy Dvenadtsatoj molodezhnoj nauch. shk.-konf. «Lobachevskie chteniya – 2013». – Kazan: Izd-vo Kazan. matem. ob-va, 2013. T. 47. S. 6-10.

8. *Metodologiya funktsional'nogo modelirovaniya IDEF0.* Rukovodyashhij dokument, Gosstandart Rossii. – M.: IPK Izdatel'stvo standartov, 2000. 75 s.

9. *Elizarov A.M., Lipachev E.K., Malakhaltsev M.A.* Web-tekhnologii v rabote ehlektronnogo matematicheskogo zhurnala Lobachevskii Journal of Mathematics // Nauchnyj servis v seti Internet: mnogoyadernyj komp'yuternyj mir. 15 let RFFI. Trudy Vseros. nauch. konf. – M.: Izd-vo MGU, 2007. S. 355-356.

10. *Akhmetov D.Yu., Elizarov A.M., Lipachev E.K.* Sistema avtomatizatsii redaktsionnykh protsessov na platforme ehlektronnykh nauchnykh zhurnalov // Uchyonye zapiski Instituta sotsial'nykh i gumanitarnykh znaniy. 2014. № 1 (12), Ch. 2. S. 228-233.

11. *Akhmetov D.Yu., Gerasimov A.N., Grachev A.O., Elizarov A.M., Lipachev E.K.* Oblachnaya platforma podderzhki ehlektronnykh nauchnykh izdaniy // Uchyonye zapiski Instituta sotsial'nykh i gumanitarnykh znaniy. 2014. № 1 (12), Ch. 1. S. 13-19. URL: http://www.isgz.ru/images/Alexey/Chirko/ek%202014_i.pdf.

12. *Open Journal Systems. Technical Reference.* – Public Knowledge. Simon Fraser University Library. URL: <https://pkp.sfu.ca/ojs/OJSTechnicalReference.pdf>.

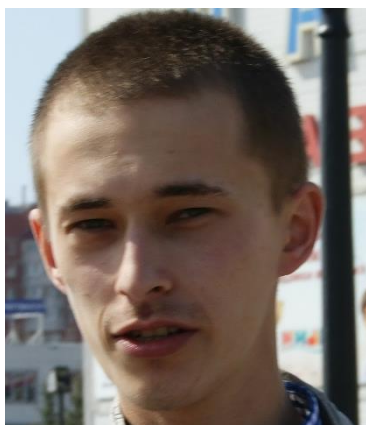
13. *Willinsky J., Stranack K., Smecher A., MacGregor J.* Open Journal Systems: a complete guide to online publishing. – Simon Fraser University Library, 2010. 273 p. URL: <https://pkp.sfu.ca/ojs/docs/userguide/2.3.3/index.html>.

14. *Open Journal Systems. OJS hosting, installation, upgrade, theme and plugin, customization, training and support.* URL: <http://openjournalssystem.com/>.

15. *Plugin Gallery – PKP Support.* URL: <https://pkp.sfu.ca/support/forum/viewforum.php?f=28>.

16. *Chen S.* LATEXGalley Plugin for Open Journal Systems. – Simon Fraser University and Zhejiang University, 2010. 30 p.

СВЕДЕНИЯ ОБ АВТОРАХ



АХМЕТОВ Дмитрий Юрьевич – аспирант Института математики и механики им. Н.И. Лобачевского Казанского (Приволжского) федерального университета.

Dmitriy Yurevich AKHMETOV, received MS degree in Mathematics from Kazan Federal University (2012). Currently is a graduate student at the N.I. Lobachevskii Institute of Mathematics and Mechanics of Kazan Federal University. Current scientific interests: knowledge extraction technologies, integration of scientific resources.

email: akhmetov.dy@gmail.com



ЕЛИЗАРОВ Александр Михайлович – доктор физико-математических наук, профессор, заслуженный деятель науки Республики Татарстан, зам. директора по научной деятельности Института математики и механики им. Н.И. Лобачевского Казанского (Приволжского) федерального университета.

Alexander Mikhailovich ELIZAROV – Doctor of Physics and Mathematics, Professor, Honoured Worker of Science of the Republic of Tatarstan, Deputy Director of the Lobachevskii Institute of Mathematics and Mechanics of Kazan Federal University. Current scientific interests: data mining, recommender systems, cloud computing, knowledge extraction technologies.

email: amelizarov@gmail.com



ЛИПАЧЁВ Евгений Константинович – кандидат физико-математических наук, доцент кафедры теории функций и приближений Института математики и механики им. Н.И. Лобачевского Казанского (Приволжского) федерального университета.

Evgeny Konstantinovich LIPACHEV – Candidate of Physics and Mathematics, Associate Professor, Lobachevskii Institute of Mathematics and Mechanics of Kazan Federal University. Current scientific interests: data mining, recommender systems, cloud computing, knowledge extraction technologies.

email: elipachev@gmail.com

Материал поступил в редакцию 14 декабря 2014 года

УДК 004.82

ТЕХНОЛОГИИ CLOUD COMPUTING В ЭЛЕКТРОННЫХ НАУЧНЫХ ЖУРНАЛАХ

Д.С. Зуев

*Отдел информационных технологий Казанского государственного
медицинского университета*

dzuev11@gmail.com

Аннотация

Рассмотрены свободно распространяемые системы автоматизации редакционных бизнес-процессов и их применения в работе электронных научных журналов. Обсуждены дополнительные функции, востребованные редакциями таких журналов, предложены варианты их реализации с использованием дополнительных плагинов для Open Journal System.

Ключевые слова: *облачные сервисы, Open Journal System, издательская система, электронные библиотеки, информационная система, метаданные, семантический анализ*

ВВЕДЕНИЕ

В настоящее время использование информационно-коммуникационных технологий (ИКТ) в научно-образовательном процессе поставило новые задачи в издательской деятельности. В частности, для включения в мировые информационные потоки необходимо решить задачу интеграции в едином информационном пространстве электронных документов научного и образовательного содержания. Один из подходов к решению этой задачи основан на создании специализированных информационных систем. Интеграция электронных документов дает возможность: обращения к данным независимо от места их размещения; анализа информации, включающего определение характеристик, взаимосвязей и источников; преобразования информации для ее обогащения и приведения в соответствие заданным целям, а также объединения информации, чтобы сделать ее доступной людям, процессам и приложениям [1].

Распространение мобильных устройств и развитие облачных технологий выдвигают на первый план задачу создания сервисов, позволяющих обеспечить взаимодействие ученых с системами поддержки научных изданий, прежде всего, с издательствами и журналами. Такое взаимодействие должно обеспечиваться в любом месте, где есть доступ в интернет. Одним из путей повышения поисковых возможностей информационных систем для поддержки научных изданий является применение в них технологий Семантического Веба (см., например, [1]), а также использование журналами и издательствами единой промышленной информационной системы с необходимым функционалом.

Серьезными проблемами отечественного научного книгоиздания являются разобщенность издательских инфраструктур, недостаточная системная работа по инновационному развитию этой отрасли, а также решению вопросов интеграции отечественных информационных ресурсов в мировые базы данных.

Изменения в издательской практике, связанные с переходом к электронной форме изданий и использованию сервисов, ориентированных на пользователей, авторов и редакторов, предполагают системную работу, возможную только на основе промышленной издательской платформы.

Сравнительный анализ существующих информационных систем управления электронными научными журналами проведен в [2, 3]. Этот анализ основан на использовании критериев оценки, заложенных в модели DELOS DLRM [4–6].

СИСТЕМЫ АВТОМАТИЗАЦИИ РЕДАКЦИОННЫХ БИЗНЕС-ПРОЦЕССОВ

Практически все ведущие научные издательства внедряют системы автоматического управления рабочими процессами, в числе которых – наиболее сложные и длительные по времени редакционные процессы, обеспечивающие независимое научное рецензирование.

Для поддержки жизненного цикла как отдельных научных статей, так и научных журналов в целом целесообразно использовать в качестве ядра системы управления электронными научными журналами программную платформу Open Journal Systems – OJS (<https://pkp.sfu.ca/ojs/>). Выбор OJS обоснован в [2]. Там же предложена архитектура универсальной платформы управления электронными

научными журналами, которая содержит три уровня – физический, базовый и уровень сервисов.

Физический уровень характеризует аппаратную составляющую системы, обеспечивающую функционирование верхних уровней, и содержит системное и прикладное программное обеспечение. Эти компоненты предполагают техническую поддержку с использованием технологий виртуализации и облачных вычислений.

Базовый уровень реализует основные сервисы управления электронными научными журналами, в том числе, регистрацию авторов и пользователей, прием и первичную обработку статей, включая автоматизацию проверки соблюдения правил редакции и рецензирования, контроль соблюдения сроков рассмотрения статей, назначение рецензентов и рассылку уведомлений. Базовый уровень включает также сервисы удаленного взаимодействия и совместной работы, поиска в электронном хранилище и автоматического извлечения метаданных, структурирования входящей информации, управления пользователями и ролями, платного доступа к контенту.

На уровне сервисов размещены дополнительные надстройки и функции, учитывающие специфику предметной области научного журнала. Например, для математических журналов востребованы сервисы конвертации в специализированные форматы (TeX, MathML и др.). Здесь реализуется front-end системы и происходит взаимодействие с конечным пользователем.

Взаимодействие с системой управления электронными научными журналами может быть организовано либо через собственный веб-портал, либо через специальные программные интеграционные адаптеры с сайта конкретного журнала, размещающего свой контент в хранилище системы.

При первом способе взаимодействия зарегистрированный пользователь получает доступ ко всем журналам, размещенным в системе, а веб-портал служит единой точкой входа. Такой способ наиболее удобен для новых журналов, не имевших собственных сайтов в Сети.

Для журналов, уже имеющих историю и поддерживающих собственные сайты, более приемлемым, на наш взгляд, является второй способ взаимодействия. В частности, это позволяет сохранить привычный адрес сайта журнала и его

«историю» в интернете, при этом максимально автоматизировав редакционные процессы.

Внедрение информационной системы управления бизнес-процессами научного журнала позволяет, прежде всего, автоматизировать наиболее трудоемкие рабочие процессы, а порталное решение дает возможность интегрировать журнал в мировое информационное научное пространство.

Вместе с тем, для развития электронного научного журнала необходимо расширение функциональности его базовой информационной системы, что связано как с особенностями предметной области этого журнала, так и со сложившимися традициями работы его редакции и редколлегии. Подобная проблема решается разработкой специализированных модулей с последующей интеграцией их в базовую информационную систему.

Функциональные возможности такой информационной системы, в частности, должны включать сервисы, регулирующие процесс рецензирования и обеспечивающие коллективное редактирование электронных документов, – должна быть обеспечена поддержка редакционных сервисов. Важной составляющей информационной системы управления научным контентом являются сервисы сбора наукометрических данных, учет которых в последнее время стал повсеместной практикой в научных и образовательных организациях. Для журналов физико-математического направления необходима программная поддержка процесса обработки электронных документов, созданных в Т_EX-нотации.

Однако, несмотря на весьма гибкий набор функций, OJS обладает также типичными недостатками любого свободно распространяемого программного обеспечения (ПО): ее внедрение требует глубокой первоначальной настройки; необходимо содержать штат технических специалистов, отвечающих за функционирование программной системы, что может быть проблематичным, в частности, для журналов социо-гуманитарной сферы и небольших редакций. Известно, что любое ПО, свободно распространяемое по принципу «as-is», часто может содержать ошибки и неточности в работе и поэтому требует для использования дополнительных настроек.

В силу сказанного выше, все более востребованными становятся облачные сервисы автоматизации редакционных процессов, построенные на принципах

аутсорсинга и предоставляющие весь набор ПО «под ключ». Одним из примеров такого подхода может служить проект платформы science.tatarstan.ru [7, 8], который разрабатывается в Республике Татарстан при поддержке ГАУ «Технопарк в сфере высоких технологий «ИТ-парк»». В основе платформы также лежит система OJS [9], на ней размещено несколько электронных научных журналов. Проект некоммерческий, на текущий момент времени фактически работает в тестовом режиме и используется в основном для отладки различных дополнительных сервисов, потенциально интересных редакциям электронных научных журналов.

Одним из журналов, размещенных на этой платформе, является Lobachevskii Journal of Mathematics, что потребовало развивать сервисы управления контентом, содержащим формулы и другие элементы математической нотации.

Для развития электронного научного журнала необходимо расширение функциональности его базовой информационной системы, что связано как с особенностями предметной области этого журнала, так и со сложившимися традициями работы его редакции и редколлегии, что достигается разработкой специализированных модулей, например, для математических журналов – сервисов конвертации в специализированные форматы (TeX, MathML и др.). В настоящее время для системы TeX созданы расширения, включающие семантическую разметку [10, 11]. Использование MathML позволяет создавать системы хранения с более эффективным управлением, в частности, предоставить поиск по формулам ([12–14]), а использование онтологий предметных областей – формировать списки «близких» публикаций [15, 16].

Важным аспектом работы редколлегии любого научного журнала является рецензирование поступающих в редакцию научных работ, при этом предполагается, что рецензент или рецензенты компетентны в вопросах, обсуждаемых в рассматриваемой работе, что позволяет им высказывать замечания по существу работы.

С ростом аудитории журнала редакция неизбежно сталкивается с проблемой корректного назначения рецензента. Другая проблема связана с корректным отслеживанием процесса рецензирования – необходимо, чтобы статья дошла до

рецензента вовремя, а рецензия была сделана качественно и в срок. Это становится особенно важным, когда редколлегия журнала распределенная, а сам журнал – международный. Нами реализован сервис, позволяющий автоматизировать назначение рецензента из базы данных рецензентов журнала и сопровождающий процесс от выбора и назначения рецензента до принятия решения по рассматриваемой работе. Сервис основан на анализе упоминаний фамилий и ссылок в текстах статей журнала, сфере научных интересов, ключевых словах и иных атрибутах, позволяющих сделать вывод о компетентности того или иного ученого в области исследований, которой посвящена поступившая работа. Сервис реализован в качестве плагина к системе OJS [7, 13]. Основой этого сервиса послужил алгоритм автоматического выбора рецензента статьи, внедренный в 2006 году в работу редакции журнала *Lobachevskii Journal of Mathematics* [17].

Для облегчения работы редакции предложен метод автоматизации процесса первичной обработки научной статьи, использующей $\text{T}_\text{E}\text{X}$ -нотацию, в журнальную информационную систему. Первичная обработка предполагает валидацию электронного документа, включая анализ электронного документа на соответствие требованиям редакции по стилевому оформлению публикации, а также $\text{T}_\text{E}\text{X}$ -компиляцию документа. Система позволяет на этапе загрузки статьи исключить возможные отклонения от редакционных правил и сообщить автору об обнаруженных ошибках при подготовке электронной публикации. Автоматизированная обработка электронных публикаций на платформе OJS реализована по схеме, первоначально созданной для электронного математического журнала *Lobachevskii Journal of Mathematics* (см. [7, 13]).

Отдельно необходимо коснуться проблемы отображения формул в математических текстах, обрабатываемых с помощью OJS. Проблема касается, в основном, отображения формул и спецсимволов в заглавиях и аннотациях статей. Математические формулы, в частности, сложные, могут плохо восприниматься в исходном виде, поэтому для решения проблемы необходима компиляция формул непосредственно в рамках системы. Данная проблема также была решена в рамках платформы [8] путем расширения функциональности OJS с помощью ряда пла-

гинов и их последующей тонкой настройкой. Подобная проблема также решалась, например, на портале MathNet [11–13, 18, 19] подключением JavaScript-платформы MathJax (<https://www.mathjax.org/>).

Помимо вышесказанного логичным представляется предоставление следующих дополнительных функций:

- предоставление доступа к информационным ресурсам, размещаемым в системе на платной основе, и сервис оплаты услуг, например, OJS имеет возможность работы с электронным кошельком PayPal (www.paypal.com), также разумно рассматривать варианты с использованием любых других сервисов приема и процессинга платежей;
- информетрический анализ, например, в системе OJS реализована поддержка сервиса Article-Level Metrics (<http://article-level-metrics.plos.org>);
- сервисы организационной поддержки научных конференций, например, система www.easychair.org или система автоматизации конференций Open Conference System, созданная в рамках Public Knowledge Project;
- онлайн-общение (вебинары и видеоконференции, в том числе для распределенных редколлегий);
- поиск и сбор OAI-метаданных, например, система индексирования метаданных Open Harvester Systems;
- проверка загружаемых ресурсов на плагиат.

ТИПОВАЯ АРХИТЕКТУРА СЕРВИСА

В работах [20, 21] рассмотрены возможные наборы сервисов, расширяющие функциональность электронных журнальных систем, и предложены варианты их интеграции в программную платформу. Учитывая собственный опыт реализации, а также результаты анализа имеющихся решений, можно сделать выводы о типовой организации подобных сервисов. Все сервисы располагаются в облаке и предоставляют примерный типовой набор услуг – хостинг и размещение портала журнала с сохранением доменного имени, предоставление типовых дизайнов и шаблонов документов, загрузка и разметка первоначального архива статей, организация редакционного процесса, администрирование и техническая поддержка системы. Также могут предлагаться услуги по технической экспертизе портала журналов с целью оптимизации их работы.

Фактически каждым агрегатором формируется система, которая де-факто содержит внутри себя ядро – OJS или ей подобную систему и набор различных дополнительных услуг. Такой набор, вообще говоря, различен у каждого из существующих порталов, но возможно выделить следующие особенности.

Как правило, все изменения в дизайне сайта электронного научного журнала сводятся к применению одного из существующих стилей оформления, поддерживаемого ядром системы. Такой подход может накладывать некоторые ограничения при построении портала журнала, особенно для изданий с уже существующими сайтами, хотя он безусловно проще для дальнейшей технической поддержки. Основная тонкость связана с необходимостью адаптации имеющегося портала журнала с учетом особенностей редакционной системы. В этой ситуации логично предложить несколько иной подход к построению такого облачного сервиса.

Условно в организации сервиса можно выделить два уровня – уровень портала журнала и системный уровень, взаимосвязанных между собой. В общем случае эти уровни могут быть физически разнесены между собой на разные домены и/или сервера.

На портале журнала размещаются вся общая информация о научном журнале, новости, сведения о редколлегии, тематике и правилах оформления статей, контактная информация о редакции или любая другая необходимая информация, не требующие сложного процесса рецензирования и публикации. При этом технически портал журнала может быть вообще не совместим с одной из электронных журнальных систем. Это также позволяет при переходе на облачный сервис автоматизации редакции не переносить полностью существующий сайт «внутри» системы (см. рис.).

Все процессы, связанные с подачей статей, рецензированием и формированием новых выпусков журналов, выносятся отдельно и образуют уровень системы. Как правило, это не публичный раздел портала – на этом уровне расположена сама база данных всех ресурсов, реализованы бизнес-процессы рецензирования и публикации, организован обмен метаданными со сторонними системами по различным протоколам. Межуровневое взаимодействие может быть организовано несколькими способами, начиная от примитивной прямой гиперссылки и

заканчивая бесшовной интеграцией (фреймы, веб-сервисы, и т. д.). Современные веб-технологии обеспечивают полный набор инструментов для организации такого взаимодействия, и это не представляет большой технической сложности.

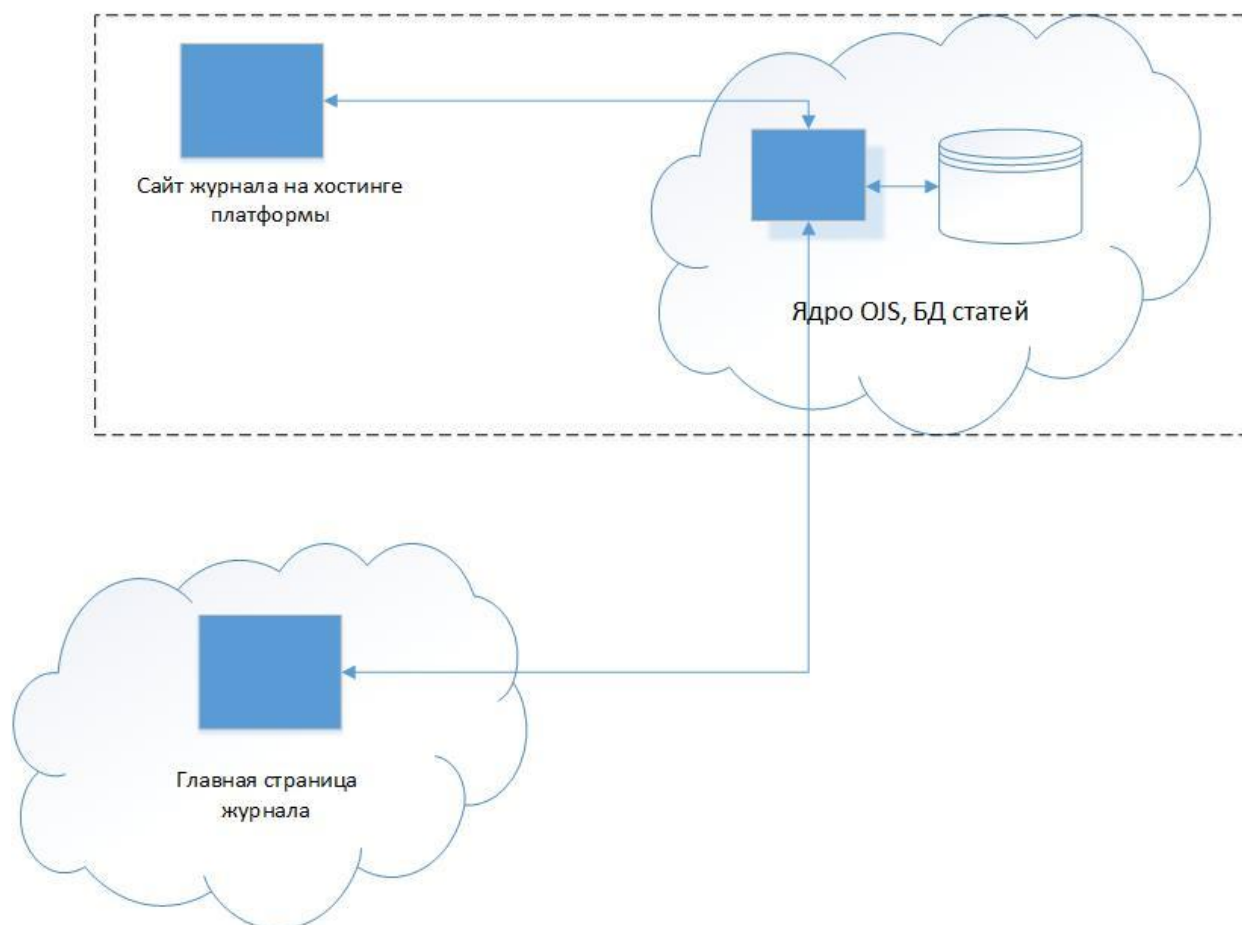


Рис. Принципиальная схема редакционного сервиса

Важно отметить, что при таком подходе сохраняются история и место портала журнала в Сети во всех рейтингах, а системный уровень может быть единым для большого пула редакций.

Благодарности

Работа выполнена при финансовой поддержке Российского гуманитарного научного фонда, проект 14-03-12004, а также Российского фонда фундаментальных исследований и Правительства Республики Татарстан в рамках научного проекта № 15-47-02472.

СПИСОК ЛИТЕРАТУРЫ

1. *Елизаров А.М., Липачёв Е.К.* Системы интеграции электронной научно-образовательной информации и повышение поисковой функциональности с помощью семантических технологий // Учёные записки Института социальных и гуманитарных знаний. 2013. №1 (11). С. 113-119.
2. *Елизаров А.М., Зуев Д.С., Липачёв Е.К.* Информационные системы управления электронными научными журналами // Научно-техническая информация. Серия 1. Организация и методика информационной работы. 2014. № 3. С. 31-38.
3. *Елизаров А.М., Зуев Д.С., Липачёв Е.К.* Свободно распространяемые системы управления электронными научными журналами и технологии электронных библиотек // Тр. XV Всерос. науч. конф. «Электронные библиотеки: перспективные методы и технологии, электронные коллекции, RCDL'2013», г. Ярославль, 14 – 17 октября 2013 года. – Ярославль: Изд-во ЯрГУ им. П.И. Демидова, 2013. С. 227-236. URL: http://rcdl2013.uni-yar.ac.ru/doc/full_text/s3_1.pdf.
4. *Candela L., Castelli D., Fuhr N., Ioannidis Y., Klas C.-P., Pagano P., Ross S., Saidis C., Schek H.-J., Schuldt H., Springmann M.* DELOS Workpackage 1. D1.4.1 – Current digital library systems: user requirements vs provided functionality. – 2005.
5. *Candela L., Castelli D., Fuhr N., Ioannidis Y., Klas C.-P., Pagano P., Ross S., Saidis C., Schek H.-J., Schuldt H., Springmann M.* Current digital library systems: user requirements vs provided functionality. IST-2002-2.3.1.12. Technology-enhanced Learning and Access to Cultural Heritage. March 2006.
6. *Candela L., Castelli D., Dobрева M., Ferro N., Ioannidis Y., Katifori H., Koutrika G., Meghini C., Pagano P., Ross S., Agosti M., Schuldt H., Soergel D.* The DELOS Digital Library Reference Model Foundations for Digital Libraries. IST-2002-2.3.1.12. Technology-enhanced Learning and Access to Cultural Heritage. Version 0.98, December 2007. URL: http://www.delos.info/files/pdf/ReferenceModel/DELOS_DLReferenceModel_0.98.pdf.
7. *Ахметов Д.Ю., Герасимов А.Н., Грачев А.О., Елизаров А.М., Липачёв Е.К.* Облачная платформа поддержки электронных научных изданий // Учёные записки Института социальных и гуманитарных знаний. 2014. № 1 (12), Ч. 1. С. 13-19. URL: http://www.isgz.ru/images/Alexey/Chirko/ek%202014_i.pdf.

8. *Science.Tatarstan*. URL: www.science.tatarstan.ru.
9. *Public knowledge project*. URL: <https://pkp.sfu.ca/ojs/>.
10. *Kohlhase M.* STEX: Semantic markup in TEX/LATEX. Jacobs University, Bremen, 2008. URL: <http://kwarc.info/kohlhase>.
11. *Kohlhase M.* Using LATEX as a semantic markup format // *Math. Comput. Sci.* 2008. V. 2. P. 279-304.
12. *Malakhaltsev M.A.* Lobachevskii Journal of Mathematics (1998–2007). Portable electronic collection of mathematical documents. Kazan University, Kazan (2009).
13. *Елизаров А.М., Липачев Е.К., Малахальцев М.А.* Сервисы электронных естественнонаучных коллекций, построенные на основе технологии MathML // В сборнике: Научный сервис в сети Интернет: суперкомпьютерные центры и задачи Труды Международной суперкомпьютерной конференции. Российская академия наук, Суперкомпьютерный консорциум университетов России. г. Москва, 2010. С. 533-534.
14. *Елизаров А.М., Липачёв Е.К., Малахальцев М.А.* Веб-технологии для математика. Основы MathML. М.: Физматлит, 2010. 194 с.
15. *Гусев А.Л.* Анализ рынка услуг издательских платформ по управлению деятельностью распределенных коллегий электронных изданий // *Int. Scientific Journal for Alternative Energy and Ecology*. 2014. No 04/1 (123). С. 82-86.
16. *Общероссийский математический портал Math-Net.Ru*. URL: <http://www.mathnet.ru/>.
17. *Елизаров А.М., Липачев Е.К., Малахальцев М.А.* Веб-технологии в работе электронного математического журнала Lobachevskii Journal of Mathematics // В сборнике: Научный сервис в сети Интернет: многоядерный компьютерный мир. 15 лет РФФИ. Труды Всероссийской научной конференции. Московский государственный университет им. М.В. Ломоносова, Южный федеральный университет, Институт вычислительной математики РАН. г. Москва, 2007. С. 355-356.
18. *Zhizhchenko A.B., Izaak A.D.* The information system MathNet.Ru. Application of contemporary technologies in the scientific work of mathematicians // *Russian Math. Surveys*. 2007. V. 62, No 5. P. 943-966.

19. *Chebukov D., Izaak A., Misyurina O., Pupyrev Yu., Zhizhchenko A.* Math-Net.Ru as a digital archive of the Russian mathematical knowledge from the XIX century to today // *Lecture Notes in Computer Science*. ed. J. Carette et al. 2013, No 7961. P. 344-348, arXiv: 1305.5655.

20. *Елизаров А.М., Зуев Д.С., Липачев Е.К.* Управление жизненным циклом электронных публикаций в информационной системе научного журнала // *Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии*. 2014. № 4. С. 81-88.

21. *Elizarov A., Lipachev E., Zuev D.* Mathematical content semantic markup methods and open scientific e-journals management systems // *Proceedings of 5th int. conference «Knowledge engineering and Symantec Web» – KESW 2014, Kazan, Russia, 29 September 2014 – 1 October 2014, CCIS No 468*. P 242-251.

CLOUD PUBLISHING SERVICE FOR SCIENTIFIC JOURNALS

D.S. Zuev

Kazan State Medical University

dzuev11@gmail.com

Abstract

The paper summarizes the experience of the author in investigation of open source publishing information systems that was done for several years in Kazan Federal University. The emphasis placed on practical application of results for scientific journals of mathematics. We discuss additional functions that are needed for editorial office of electronic journals. We also suggest implementation of these functions by additional plug-ins for the Open Journal System.

Keywords: cloud services, Open Journal System, publishing platform, digital library, information system, metadata, semantic analysis

REFERENCES

1. *Elizarov A.M., Lipachev E.K.* Sistemi integratsii elektronnoy nauchno-obrazovatelnoy informatsii i povyishenie poiskovoy funktsionalnosti s pomoschyu semanticheskikh tehnologiy // Uchyonyie zapiski Instituta sotsialnyih i gumanitarnyih znaniy. 2013. № 1 (11). S. 113-119.
2. *Elizarov A., Zuev D., Lipachev E.* Electronic scientific journal management systems // Scientific and Technical Information Processing. 2014. V. 41, No 1. P. 66-72.
3. *Elizarov A.M., Zuev D.S., Lipachev E.K.* Open scientific e-journals management systems and digital libraries technology// CEUR Workshop Proc. Selected Papers of the 15th All-Russian Scientific Conference "Digital libraries: Advanced Methods and Technologies, Digital Collections", Yaroslavl, Russia, October 14 – 17, 2013. V. 1108. P. 102-111. URL: <http://ceur-ws.org/Vol-1108/paper13.pdf>.
4. *Candela L., Castelli D., Fuhr N., Ioannidis Y., Klas C.-P., Pagano P., Ross S., Saidis C., Schek H.-J., Schuldt H., Springmann M.* DELOS Workpackage 1. D1.4.1 – Current digital library systems: user requirements vs provided functionality. 2005.
5. *Candela L., Castelli D., Fuhr N., Ioannidis Y., Klas C.-P., Pagano P., Ross S., Saidis C., Schek H.-J., Schuldt H., Springmann M.* Current digital library systems: user requirements vs provided functionality. IST-2002-2.3.1.12. Technology-enhanced Learning and Access to Cultural Heritage. March 2006.
6. *Candela L., Castelli D., Dobрева M., Ferro N., Ioannidis Y., Katifori H., Koutrika G., Meghini C., Pagano P., Ross S., Agosti M., Schuldt H., Soergel D.* The DELOS Digital Library Reference Model Foundations for Digital Libraries. IST-2002-2.3.1.12. Technology-enhanced Learning and Access to Cultural Heritage. Version 0.98, December 2007. URL: http://www.delos.info/files/pdf/ReferenceModel/DELOS_DLReferenceModel_0.98.pdf.
7. *Ahmetov D.Yu., Gerasimov A.N., Grachev A.O., Elizarov A.M., Lipachev E.K.* Oblachnaya platforma podderzhki elektronnyih nauchnyih izdaniy // Uchyonyie zapiski Instituta sotsialnyih i gumanitarnyih znaniy. 2014. № 1 (12), Ch. 1. S. 13-19. URL: http://www.isgz.ru/images/Alexey/Chirko/ek_2014_i.pdf.
8. *Public knowledge project.* URL: <https://pkp.sfu.ca/ojs/>.
9. *Science.Tatarstan.* URL: www.science.tatarstan.ru.

10. *Kohlhase M.* STEX: Semantic markup in TEX/LATEX. Jacobs University, Bremen, 2008. URL: <http://kwarc.info/kohlhase>.
11. *Kohlhase M.* Using LATEX as a semantic markup format // *Math. Comput. Sci.* 2008. V. 2. P. 279-304.
12. *Malakhaltsev M.A.* Lobachevskii Journal of Mathematics (1998–2007). Portable electronic collection of mathematical documents. Kazan University, Kazan (2009).
13. *Elizarov A.M., Lipachev E.K., Malakhaltsev M.A.* Servisyi elektronnykh estestvennonauchnykh kollektiy, postroennyye na osnove tehnologii MathML // V sbornike: Nauchnyy servis v seti Internet: superkompyuternyye tsenyri i zadachi. Trudy Mezhdunarodnoy superkompyuternoy konferentsii. Rossiyskaya akademiya nauk, Superkompyuternyy konsortsiy universitetov Rossii. g. Moskva, 2010. S. 533-534.
14. *Elizarov A.M., Lipachev E.K., Malakhaltsev M.A.* Veb-tehnologii dlja matematika. Osnovy MathML. Moskva: Fizmatlit, 2010. 194 s.
15. *Gusev A.L.* Analiz ryinka uslug izdatelskikh platform po upravleniyu deyatelnostyu raspredelennykh kollegiy elektronnykh izdaniy // *Int. Scientific Journal for Alternative Energy and Ecology.* 2014. No 04/1 (123). S. 82-86.
16. *Obscherossiyskiy matematicheskiy portal Math-Net.Ru.* URL: <http://www.mathnet.ru/>.
17. *Elizarov A.M., Lipachev E.K., Malakhaltsev M.A.* Veb-tehnologii v rabote elektronnoy matematicheskogo zhurnala Lobachevskii Journal of Mathematics // V sbornike: Nauchnyy servis v seti Internet: mnogoyadernyy kompyuternyy mir. 15 let RFFI. Trudy Vserossiyskoy nauchnoy konferentsii. Moskovskiy gosudarstvennyy universitet im. M.V. Lomonosova, Yuzhnyy federalnyy universitet, Institut vyichislitelnoy matematiki RAN. g. Moskva, 2007. S. 355-356.
18. *Zhizhchenko A.B., Izaak A.D.* The information system MathNet.Ru. Application of contemporary technologies in the scientific work of mathematicians// *Russian Math. Surveys.* 2007. V. 62, No 5. P. 943-966.
19. *Chebukov D., Izaak A., Misyurina O., Pupyrev Yu., Zhizhchenko A.* Math-Net.Ru as a digital archive of the Russian mathematical knowledge from the XIX cen-

ture to today // Lecture Notes in Computer Science. ed. J. Carette et al. 2013. No 7961. P. 344-348, arXiv: 1305.5655.

20. *Elizarov A.M., Zuev D.S., Lipachev E.K.* Upravlenie zhiznennym tsiklom elektronnykh publikatsiy v informatsionnoy sisteme nauchnogo zhurnala // Vestnik Voronezhskogo gosudarstvennogo universiteta. Seriya: sistemnyy analiz i informatsionnyye tehnologii. 2014. № 4. S. 81-88.

21. *Elizarov A., Lipachev E., Zuev D.* Mathematical content semantic markup methods and open scientific e-journals management systems // Proceedings of 5th int. conference "Knowledge engineering and Symantec Web" – KESW 2014, Kazan, Russia, 29 September 2014 – 1 October 2014. CCIS No 468. P. 242-251.

СВЕДЕНИЯ ОБ АВТОРЕ



ЗУЕВ Денис Сергеевич – кандидат технических наук, Казанский государственный медицинский университет.

Denis Sergeevich ZUEV – PhD of technical sciences, Kazan State Medical University.

email: dzuev11@gmail.com

Материал поступил в редакцию 15 декабря 2014 года

УДК 004.91

СЕМАНТИЧЕСКИЙ АНАЛИЗ ДОКУМЕНТОВ В СИСТЕМЕ УПРАВЛЕНИЯ ЦИФРОВЫМИ НАУЧНЫМИ КОЛЛЕКЦИЯМИ

Ш.М. Хайдаров

*Институт математики и механики им. Н.И. Лобачевского
Казанского (Приволжского) федерального университета
15jkeee@gmail.com*

Аннотация

Предложены методы семантического анализа документов в системе управления цифровыми научными коллекциями, в том числе электронными научными журналами. Рассмотрены методы обработки документов, содержащих математические формулы, а также способы конвертации этих документов из формата OpenXML в формат TeX. Разработан алгоритм поиска по формулам в коллекциях математических документов, хранящихся в формате OpenXML. Алгоритм реализован в виде онлайн-сервиса на платформе science.tatarstan.

Ключевые слова: семантический анализ, издательские системы.

ВВЕДЕНИЕ

В настоящее время объемы электронных данных возрастают с колоссальной скоростью, а научная деятельность неразрывно связана с использованием информационно-коммуникационных технологий. Электронное представление научных документов вошло в каноны научной деятельности; уже многие журналы перешли на электронный способ приема материалов, практически все журналы выставляют опубликованные работы в открытый доступ в интернет (с временным эмбарго либо непосредственно после опубликования соответствующего номера или выпуска) (см., например, [1]).

В «Издательской Вселенной» – множество «жителей»: авторы, рецензенты, редакторы, издатели, библиотеки, читатели, университеты и научное сообщество. Для упорядочения жизненного цикла создания публикаций, а также последующего их распространения и хранения уже существуют различные систе-

мы, распределяющие как роли взаимодействующих пользователей, так и позволяющие группировать различные издания и хранить публикации по определенным нормам и правилам (см., например, [2, 3]). Подобные системы значительно упрощают публикационный процесс, перестраивают классические этапы получения и передачи статей рецензентам и дальнейшей их обработки. В результате появляется возможность автоматизировать многие из этих процессов. При этом сами такие системы нуждаются в дополнительных сервисах, позволяющих оптимизировать процессы, поддерживаемые этими. Одна из имеющихся проблем заключается в различиях форматов представления (оформления) публикаций, используемых в тех или иных журналах. Отличаются также методы набора текстов и форматы их последующего хранения, не говоря уже о способе воспроизведения математических формул. Мало того, что формулы можно воспроизвести различными средствами, с помощью этих средств одни и те же формулы могут быть представлены разнообразными способами. Таким образом, в большинстве случаев ошибки при наборе формул в материалах, представляемых к публикации, однотипны и связаны, прежде всего, с оформлением математических выражений по шаблону, применяемому в конкретном журнале или издании.

Настоящая статья посвящена решению одной из таких проблем, когда различия в форматах представления и хранения документов создают значительные неудобства как авторам научных статей, так и редколлегиям журналов, принимающих эти материалы к публикации и использующих определенные форматы их представления (см., например, [4]). Таким образом, речь идет о системах, позволяющих извлекать данные из одних форматов и преобразовывать их в другой. В частности, это касается документов, созданных с использованием офисных пакетов.

ОБЩЕПРИНЯТЫЕ ФОРМАТЫ ХРАНЕНИЯ ДАННЫХ

Нотация TeX. На момент создания в 1978 году Дональдом Кнудом цифровой системы TeX для представления математических документов она не была ни первой, ни единственной подобной системой. Быстро завоевав признание, сегодня она стала основным инструментом набора математических текстов и важнейшим средством коммуникации научного сообщества. В системе TeX пользо-

ватель задает текст и его структуру, а система на основе выбранного пользователем шаблона самостоятельно форматирует документ, заменяя при этом дизайнера и верстальщика. На данный момент TeX является основным инструментом представления математических формул. Однако, в виду сложности представления документов, созданных в TeX, был разработан другой метод представления математической информации в Вебе. Этим форматом стал **язык разметки MathML**, созданный в 1999 году консорциумом W3C как способ представления математических формул в Вебе. В результате MathML значительно поменял принципы организации и управления электронными публикациями по математике.

Офисные пакеты. При подготовке научных документов все чаще используют офисные пакеты. Их широкое распространение обусловлено простотой ввода и наглядностью конечного (печатного) варианта документов (статей, докладов, книг и пр.). Наиболее распространённый формат хранения офисных документов *.doc устарел, и на его замену пришел формат OpenXML, основанный на открытых семантических технологиях и имеющий расширение *.docx. Ниже описаны методы машинного извлечения данных из документов в формате OpenXML.

ИЗВЛЕЧЕНИЕ ДАННЫХ ИЗ ДОКУМЕНТОВ В ФОРМАТЕ OpenXML

Формат OpenXML основан на открытых технологиях семантического Веба и представляет собой сжатый ZIP-контейнер, содержащий как разметку документа, так и вспомогательные его части, такие, как шрифты и стили (см., например, [5, 6]). Основной скелет документа в структурированном виде хранится в файле «word/document.xml». Здесь содержится текст документа с его форматированием (см. фрагмент кода на рис. 1, 2) и всеми его объектами или ссылками на них, например, изображения хранятся в отдельном каталоге, однако их позиция и свойства описаны в данном файле.

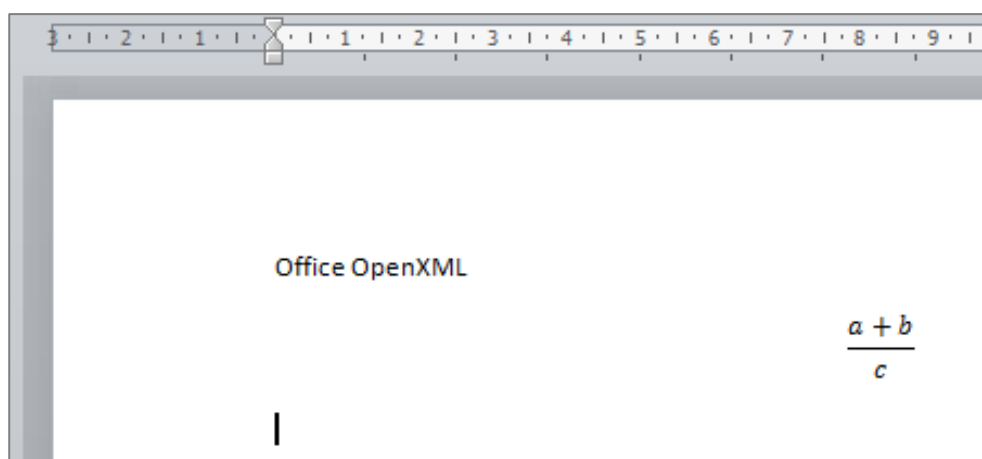


Рис. 1. Представление кода

```
1 <?xml version="1.0" encoding="UTF-8" standalone="yes"?>
2 <w:document xmlns:wpc="
  http://schemas.microsoft.com/office/word/2010/wordprocessingCanvas"
3 <w:body> <!--начало документа-->
4 <w:p w:rsidR="00AF6381" w:rsidRPr="00DA088A" w:rsidRDefault="
  00B300E9"> <!--Начало абзаца-->
5 <w:pPr> <!--Свойства абзаца-->
6 <w:rPr>
7 <w:rFonts w:ascii="Calibri" w:eastAsia="Calibri"/>
8 </w:rPr>
9 </w:pPr>
10 <w:r w:rsidRPr="00DA088A">
11 <w:t>Office OpenXML</w:t> <!--Текст.-->
12 </w:r>
13 </w:p>
14 <m:oMath><!--Тег начала формул, пространство имён OpenMathML-->
15 <m:f> <!--Тег дроби-->
16 <m:fPr> <!--Свойства дроби и стили текста-->
17 </m:fPr>
18 <m:num> <!--Тег числителя-->
19 <m:t>a+b</m:t>
20 </m:num>
21 <m:den> <!--Тег знаменателя-->
22 <m:t>c</m:t>
23 </m:den>
24 </m:f>
25 </m:oMath>
26 </w:body>
27 </w:document>
```

Рис. 2. Фрагмент кода в OpenXML

Извлечение элементов документа. При обработке документа сначала необходимо извлечь из архива XML-файлы с метаданными. В случае с текстом достаточно из архива получить файл «word/document.xml». Если стоит задача лишь выделения обычного текста без всякого форматирования, то достаточно вернуть значения тегов «w:t». Однако, чтобы применить выбранные стили, необходимо обработать атрибуты тегов [7, 8]. На рис. 3 показан фрагмент шаблона XSLT, обрабатывающего текст с основными стилями, на примере перевода

его в TeX-нотацию.

```
22 <xsl:template match="w:p">
23   <xsl:apply-templates/>
24   <xsl:if test="position() !=last()"><xsl:text>
25     .
26   </xsl:text>
27   </xsl:if>
28 </xsl:template>
29
30 <xsl:template match="w:r">
31   <xsl:if test="w:footnoteReference">
32     <xsl:text>\footnote{</xsl:text>
33     <xsl:call-template name="footnote">
34       <xsl:with-param name="fid">
35         <xsl:value-of select="//@w:id"/>
36       </xsl:with-param>
37     </xsl:call-template>
38     <xsl:text>}</xsl:text>
39   </xsl:if>
40   <xsl:if test="w:rPr/w:b">
41     <xsl:text>\textbf{</xsl:text>
42   </xsl:if>
43     <xsl:call-template name="pastb"/>
44   <xsl:if test="w:rPr/w:i">
45     <xsl:text>}</xsl:text>
46   </xsl:if>
47 </xsl:template>
48
49 <xsl:template name="pastb">
50   <xsl:if test="w:rPr/w:i">
51     <xsl:text>\textit{</xsl:text>
52   </xsl:if>
53   <xsl:call-template name="pasti"/>
54   <xsl:if test="w:rPr/w:i">
55     <xsl:text>}</xsl:text>
56   </xsl:if>
57 </xsl:template>
58
```

Рис. 3. Фрагмент шаблона, обрабатывающего текст в OpenXML

Что касается изображений, графиков и элементов из других приложений, то они хранятся в отдельной папке в бинарном виде, а в приведенном xml-файле содержится лишь ссылка, характеризующая положение этой папки, примененные фильтры и др. (см. рис. 4).

```

78 <a:graphic xmlns:a="http://schemas.openxmlformats.org/drawingml/2006/main">
79 <a:graphicData uri="http://schemas.openxmlformats.org/drawingml/2006/picture">
80 <pic:pic xmlns:pic="http://schemas.openxmlformats.org/drawingml/2006/picture">
81 <pic:nvPicPr>
82 <pic:cNvPr id="0" name="P1030639.JPG"/>
83 <pic:cNvPicPr/>
84 </pic:nvPicPr>
85 <pic:blipFill>
86 <a:blip r:embed="rId5" cstate="print">
87 <a:extLst>
88 <a:ext uri="{28A0092B-C50C-407E-A947-70E740481C1C}">
89 <a14:useLocalDpi xmlns:a14=
90 "http://schemas.microsoft.com/office/drawing/2010/main" val="0"/>
91 </a:ext>
92 </a:extLst>
93 </a:blip>
94 <a:stretch>
95 <a:fillRect/>
96 </a:stretch>
97 </pic:blipFill>
98 <pic:spPr>
99 <a:xfrm>
100 <a:off x="0" y="0"/>
101 <a:ext cx="5940425" cy="4455160"/>
102 </a:xfrm>
103 <a:prstGeom prst="rect">
104 <a:avLst/>
105 </a:prstGeom>
106 </pic:spPr>
107 </pic:pic>
108 </a:graphicData>
109 </a:graphic>

```

Рис. 4. Фрагмент OpenXML со ссылкой на изображение

Извлечение математических формул. Одним из подмножеств формата OpenXML является Office Math Markup Language (OMML) – язык математической разметки, который встроен в WordprocessingML (кроме того, он может быть использован в SpreadsheetML и PresentationML). Во фрагменте, приведенном на рис. 2, показано, как выглядит простая дробь в формате OMML, это представление хранится в окружении тега «*m:oMath*».

Важно отметить следующее обстоятельство: вследствие того, что OMML является подмножеством основного языка разметки WordprocessingML, как у любого элемента во всем документе, у каждого элемента имеется свое форматирование, что позволяет применить конкретные стили для любой части математического текста.

Хотя языки разметки MathML и OMML весьма схожи между собой, между ними имеется существенное различие. Организация математических элементов в MathML происходит по положению, а в OMML – по точному названию (см. рис.

5 для сравнения, как выглядит уже знакомая формула).

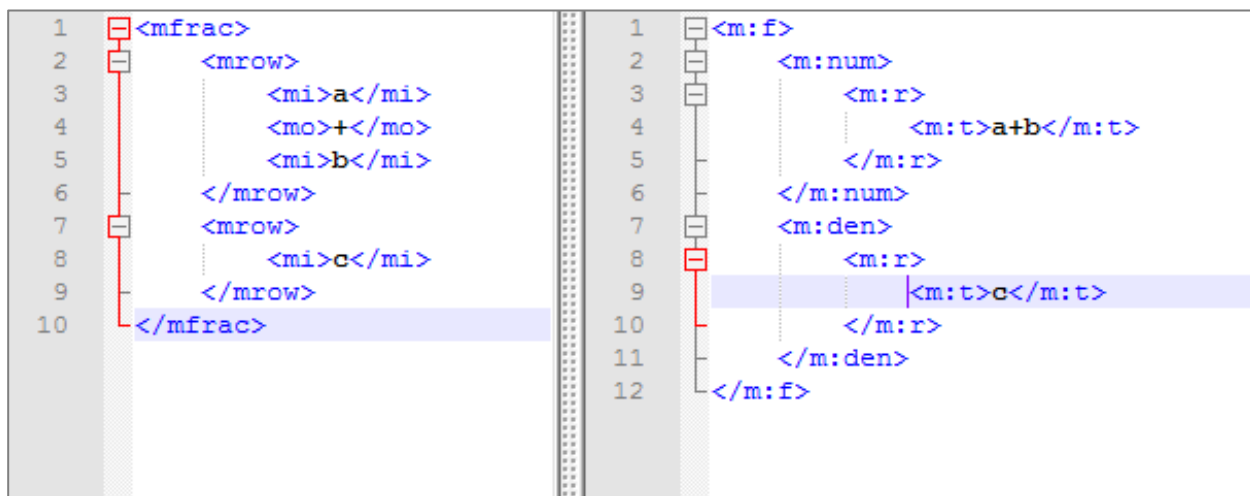


Рис. 5. Различие представления в MathML и OMML

Таким образом, в формате MathML каждый элемент (число, переменная или функция) является самостоятельным, а в OMML все они группируются по схожим характеристикам (например, все функции записываются в виде текста). По этой причине возникают неточности в переводе при использовании данных шаблонов. Например, при переводе OMML в MathML все элементарные функции, которые в MathML имеют свои теги, записываются в виде математического текста (сравните: `<cos/>` в «чистом» MathML и `<mml:mi mathvariant="normal">cos</mml:mi>` — в виде, переведенном из OMML). В MathML отсутствует поэлементное форматирование математического текста, что необходимо учитывать при реализации поиска по формульным фрагментам.

В пакете Microsoft Office содержатся средства для работы с OMML, реализованные в виде XSLT-шаблонов. Эти шаблоны позволяют получать из документа формата OMML документ формата MathML и обратно. В Microsoft Office версии 2010 эти функции выполняют шаблоны «MML2OMML.xsl» и «OMML2MML.xsl», расположенные в директории «%ProgramFiles%\Microsoft Office\Office14». Данные шаблоны предполагают, что используются пространства имен

«xmlns:mml="http://www.w3.org/1998/Math/MathML"»

и

«xmlns:m="http://schemas.openxmlformats.org/officeDocument/2006/math"».

Таким образом, теги из других пространств имен будут просто пропущены.

Чтобы их не пропустить, достаточно в таблицы стилей добавить ссылки на соответствующие пространства. Например, для текстовой разметки это «xmlns:w="http://schemas.openxmlformats.org/wordprocessingml/2006/main"» и соответственно шаблон, обрабатывающий эти теги. Например, в случае перевода в TeX-нотацию этот шаблон будет следующим (см. рис. 6):

```
23 <xsl:template match="w:body">
24 <xsl:text>\begin{document}
25 </xsl:text>
26 <xsl:apply-templates/>
27 <xsl:text>
28 \end{document}</xsl:text>
29 </xsl:template>
30
31 <xsl:template match="w:p">
32 <xsl:apply-templates/>
33 <xsl:if test="position() !=last()">
34 <xsl:text>
35
36 </xsl:text></xsl:if>
37 </xsl:template>
```

Рис. 6. Фрагмент шаблона, обрабатывающего текст

Для конвертации в TeX-формат разработана таблица стилей «d2t.xsl», работа которой сводится к извлечению файла с разметкой основного каркаса docx-документа с последующей обработкой этого файла XSLT-процессором.

На рис. 7 приведен фрагмент таблицы стилей «d2t.xsl», которая преобразует дроби. Код записи дроби в формате OpenXML приведен на рис. 2. Таким образом, в исходном файле обрабатываются только те теги, которые описаны в таблицах стилей.

```

1 <xsl:template match="m:f"> <!--если исходный файл содержит данный тег,
2   то будет выполнен этот шаблон-->
3   <xsl:variable name="sLowerCaseType" select=
4     "translate(m:fPr[last()]/m:type/@m:val,
5       'ABCDEFGHIJKLMNOPQRSTUVWXYZ', 'abcdefghijklmnopqrstuvwxyz')" />
6   <!--Создает переменную в которой написано свойство дроби, и
7     переводит в нижний регистр-->
8   <xsl:choose <!------>
9     <xsl:when test="$sLowerCaseType=''> <!--если свойство пустое,
10      то по-умолчанию выбирается обычная дробь-->
11      <xsl:text>\frac{</xsl:text> <!--Добавляется текст-->
12      <xsl:apply-templates select="m:num[1]" /> <!--числитель-->
13      <xsl:text>}</xsl:text>
14      <xsl:apply-templates select="m:den[1]" /> <!--знаменатель-->
15      <xsl:text></xsl:text>
16    </xsl:when>
17    <xsl:when test="$sLowerCaseType='lin'">
18      <xsl:apply-templates select="m:num[1]" />
19      <xsl:text>\!/</xsl:text>
20      <xsl:apply-templates select="m:den[1]" />
21    </xsl:when>
22  </xsl:choose>
23 </xsl:template>

```

Рис. 7. Шаблон, преобразующий дроби из формата OMMML в формат TeX

Конвертер поддерживает основное форматирование текста, формулы: все математические функции, поддерживаемые Word (дроби, радикалы, тригонометрические функции, пределы и пр.), операторы (интегралы, ряды и пр.), матрицы и системы уравнений, диакритические знаки, греческий алфавит и др.

Отметим, что нотация TeX не содержит тегов для всех символов кодировки Unicode. Так как OpenXML для записи символов использует Unicode, ряд символов в TeX отсутствует. Поэтому при конвертации необходимо использовать дополнительные программные средства.

На следующих рисунках приведен пример конвертации из OpenXML в TeX.

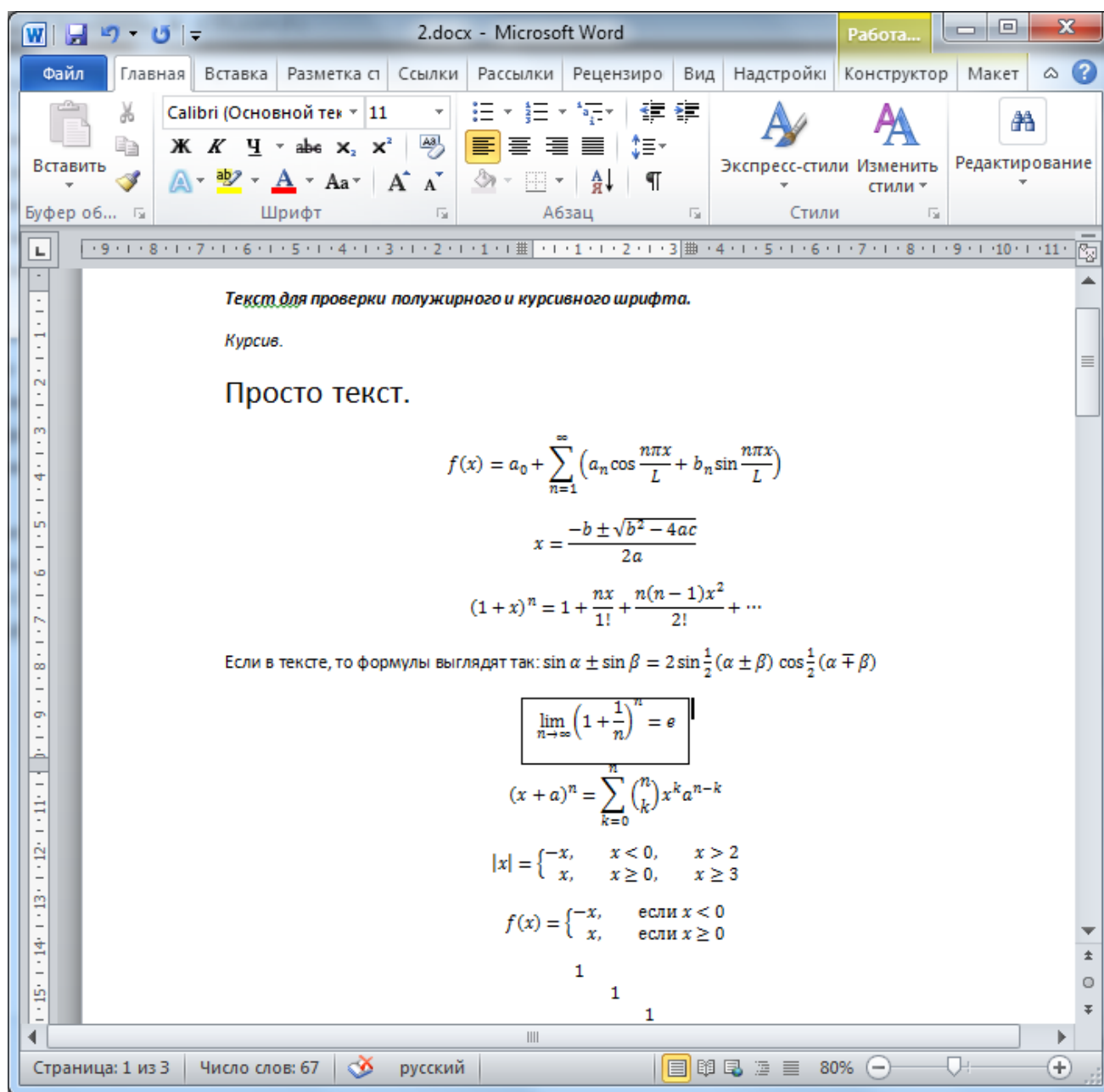


Рис. 8. Исходный документ в формате OpenXML

```

WinEdt 6.0 - [D:\Desktop\DOCX2TEX2\2.tex]
File Edit Search Insert Document Project View Tools Macros Accessories TeX Options Window Help Shashkov's
2.tex
\documentclass{article}
\usepackage{amsmath}
\usepackage{cmap}
\usepackage[utf8]{inputenc}
\usepackage[english, russian]{babel}
\begin{document}
\textbf{\textit{Текст для проверки полужирного и курсивного шрифта}}
\textit{Курсив.}
Просто текст.

$$f(x) = a_0 + \sum_{n=1}^{\infty} \left( a_n \cos\left(\frac{n\pi x}{L}\right) + b_n \sin\left(\frac{n\pi x}{L}\right) \right)$$


$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$


$$\left(1+x\right)^n = 1 + \frac{nx}{1!} + \frac{n(n-1)}{2!}x^2 + \dots$$

Если в тексте, то формулы выглядят так:  $\sin(\alpha) \pm \sin(\beta) = 2 \sin\left(\frac{1}{2}(\alpha \pm \beta)\right) \cos\left(\frac{1}{2}(\alpha \mp \beta)\right)$ 

$$\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n = e$$


$$\sum_{k=0}^n \binom{n}{k} x^k a^{n-k}$$


$$\left|x\right| = \begin{cases} -x, & x < 0, \\ x, & x \geq 0 \end{cases}$$


$$f(x) = \begin{cases} -x, & \text{если } x < 0 \\ x, & \text{если } x \geq 0 \end{cases}$$

\begin{matrix}
1 & & \\
& 1 & \\
& & 1
\end{matrix}
\end{document}
? A 5:37 30 Wrap Indent INS LINE Spell TeX:UTF-8:UNIX:UTF-8 --src WinEdt.prj

```

Рис. 9. TeX-код, полученный в результате конвертации

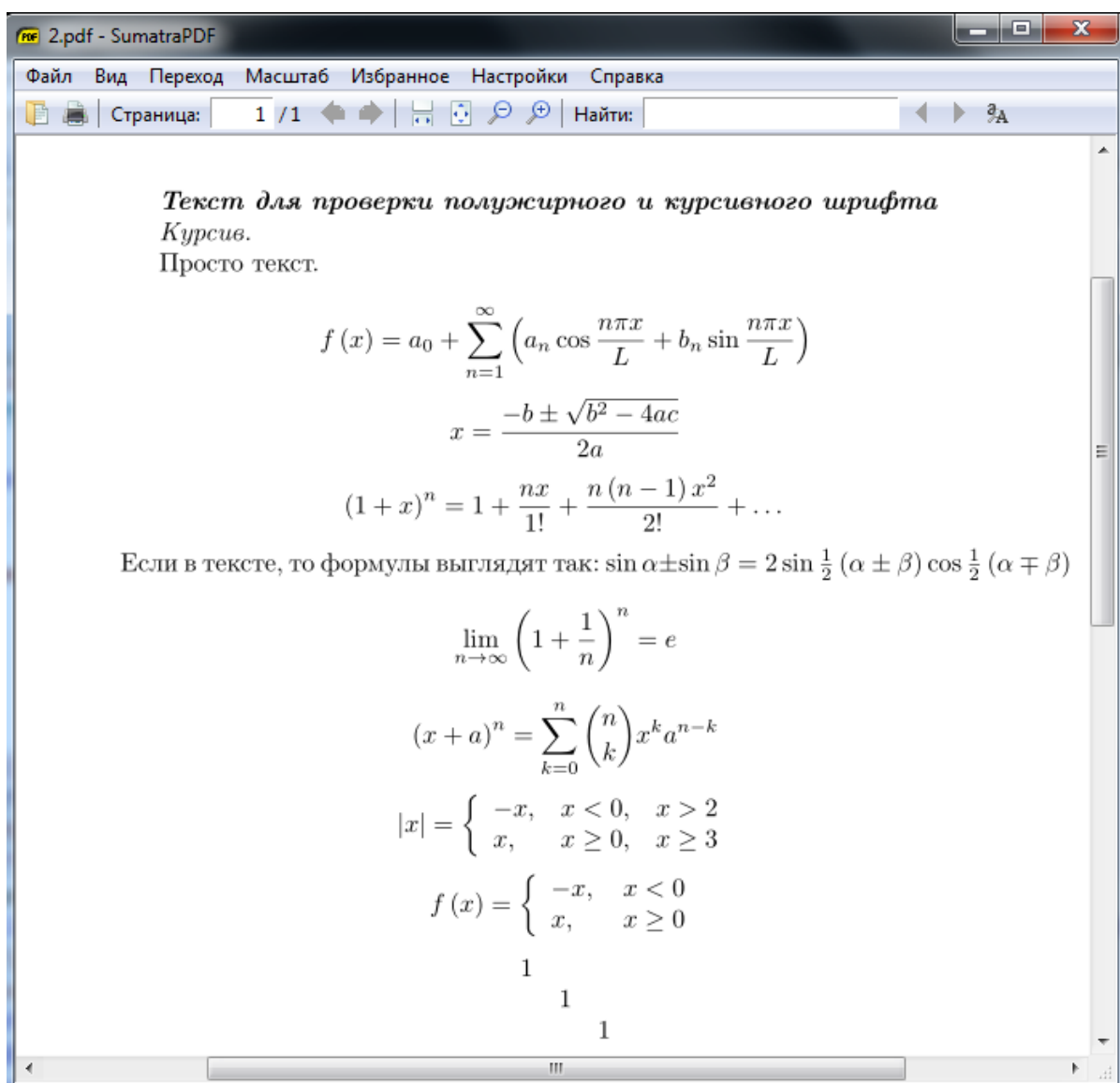


Рис. 10. Документ, полученный в результате работы конвертера

Алгоритм реализован в виде онлайн-сервиса на платформе science.tatarstan [9].

ПОИСК ПО МАТЕМАТИЧЕСКИМ ФОРМУЛАМ

С развитием веб-технологий в интернете размещается все больше результатов научных работ (это относится и к математическим исследованиям), одновременно все более важным становится развитие поисковых инструментов для анализа научного контента. Традиционный поиск, организованный по ключевым словам, как правило, не способен находить математические формулы, интере-

сующие пользователя. Правда, иногда поисковики выдают требуемый результат, но чаще всего это происходит из-за удачно подобранных ключевых слов и словосочетаний. Тем не менее, основная информация в математических текстах содержится именно в формулах. Поэтому актуальным является разработка методов, способных анализировать математический контент и осуществлять поиск по формулам (см., например, [8]).

Среди подходов к решению задачи поиска по математическим выражениям выделяют синтаксический поиск [2] и семантический поиск, основанный на онтологиях предметных областей [10–12]. Последний, в отличие от синтаксического поиска, разбивает документ на семантические области, выполняет семантическую разметку документа и создает онтологические связи [13]. Это позволяет вести поиск в различных фрагментах текста (определениях, теоремах, доказательствах и т. п.). После индексирования коллекции математических документов для каждого документа формируется семантическое описание (например, в форматах STeX и RDF (см. [12, 13])).

Канонизация представления. Важный этап обработки математических текстов – представление поисковых запросов в едином формате. Для этого необходима конвертация в этот формат документов, представленных в других форматах. При реализации названного этапа преимуществами обладают языки семантической разметки, такие, как MathML (например, [14, 15]). Эффективность использования MathML в качестве формата хранения документов подтверждена практикой его использования в журнале Lobachevskii Journal of Mathematics [2], [16]. Отметим, что MathML реализован в двух вариантах: Presentation MathML и Content MathML [17]. Использование Presentation MathML при обработке математического контента не эффективно в силу следующих причин [12, 18, 19]:

1. *Многообразие математических терминов и обозначений:* в разных разделах математики термины могут иметь различный смысл, то же касается и обозначений, например, формула $\frac{n!}{k!(n-k)!}$ может обозначаться как $\binom{n}{k}$, ${}_k C^n$, C_n^k , C_k^n .

2. *Выбор обозначений*: при переобозначении переменных смысл формулы не изменяется (например, результаты поиска формул $\int f(x)dx$ и $\int g(y)dy$ должны совпадать).

3. *Различные способы записи математических выражений*: например, степени можно записать словами, надстрочным знаком и т. п.

Часть указанных проблем решают языки разметки содержательного уровня, такие, как Content MathML и OpenMath.

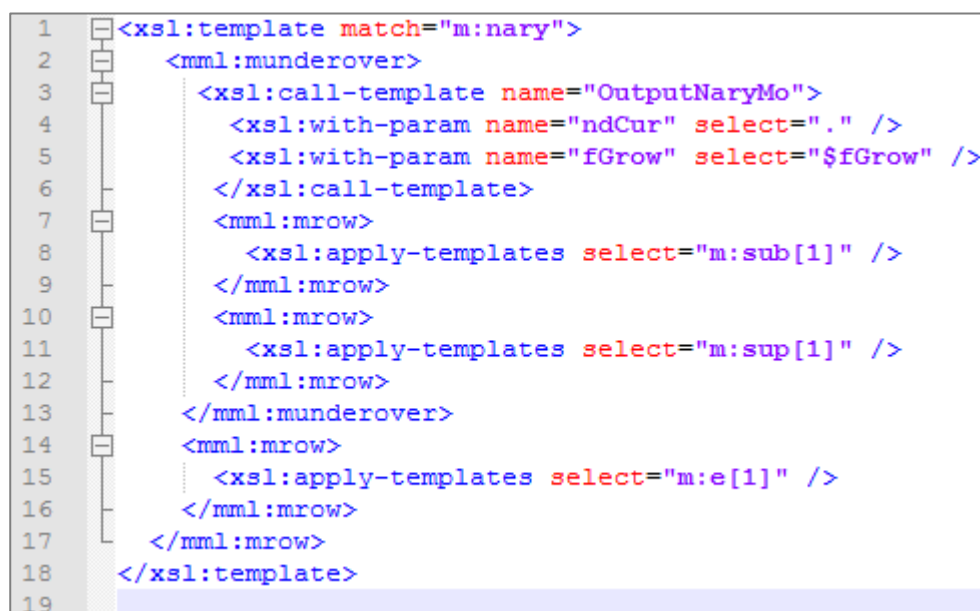
Для конвертации из Content MathML в Presentation MathML можно использовать XSLT-преобразование (см., например, [20]). Обратная же задача является одной из проблем искусственного интеллекта (см. [21]). Отметим в связи с этим Java-библиотеку SnuggleTeX [22], предназначенную для конвертации фрагментов математических формул в различных формах представления – из TeX в Presentation MathML, из Presentation MathML – в Content MathML, а также в формат Maxima. Наибольший интерес представляет преобразование в Content MathML: для этого используются как средства XSLT, так и средства Java, однако поддержка ограничивается только арифметическими операциями и функциями. Кроме того, Content MathML используется как формат хранения документов в пакете Wolfram Mathematica [23]. Обусловлено это тем, что этот пакет сохраняет структуру введенных данных, причем при вводе данные проходят валидацию.

Подход к решению проблемы поиска по формулам. В настоящей работе предложен метод структурного поиска по математическим выражениям, хранящимся в документах в форматах OpenXML и TeX. Использован язык программирования PHP для возможности интеграции с издательской платформой OJS. Алгоритм поиска по математическим выражениям аналогичен подходу, реализованному в переносимой коллекции электронных математических документов журнала Lobachevskii Journal of Mathematics (см. [2], [15]). Программа поиска включает два блока: первый индексирует математические выражения в коллекции, второй представляет пользовательский интерфейс поиска, обрабатывающий запрос и выдающий найденные результаты. Создание индексируемого

файла позволяет увеличить скорость поиска за счет ограничения времени обработки документов.

Индексирование документов в формате OpenXML проводится по следующей схеме:

- в выбранной директории создается список документов с расширением *.docx*;
- поскольку каждый файл *.docx* является архивом [5], [6], производится его распаковка и извлекается файл «*word/document.xml*»; на этом шаге производится анализ этого файла и исключаются документы без математических формул (по наличию тега «*o:Math*»);
- с использованием таблицы стилей «*OMML2MML.XSL*», входящей в состав Microsoft Office, документ конвертируется в формат MathML; дополнительно в таблицу стилей вносятся изменения для унификации записи (например, пределов интегрирования или суммирования, см. рис. 11);
- в полученном документе удаляются упоминания пространств имен, атрибуты и параметры форматирования;
- создается xml-файл (индекс-файл), в котором сгруппированы все формулы документов директории (рис. 12).



```
1 <xsl:template match="m:nary">
2   <mml:munderover>
3     <xsl:call-template name="OutputNaryMo">
4       <xsl:with-param name="ndCur" select="." />
5       <xsl:with-param name="fGrow" select="$fGrow" />
6     </xsl:call-template>
7     <mml:mrow>
8       <xsl:apply-templates select="m:sub[1]" />
9     </mml:mrow>
10    <mml:mrow>
11      <xsl:apply-templates select="m:sup[1]" />
12    </mml:mrow>
13  </mml:munderover>
14  <mml:mrow>
15    <xsl:apply-templates select="m:e[1]" />
16  </mml:mrow>
17 </mml:mrow>
18 </xsl:template>
19
```

Рис. 11. Изменения, вносимые в файл *OMML2MML.xml*, для унификации записи операторов

```
1      <?xml version="1.0"?>
2      <files>
3      <file name="files//2.docx">
4          <math>Формула 1..</math>
5          <math>Формула 2..</math>
6          <math>Формула 3..</math>
7      </file>
8      <file name="files//3.docx">
9          <math>Формула 1..</math>
10         <math>Формула 2..</math>
11     </file>
12 </files>
13
14
```

Рис. 12. Фрагмент xml-файла, содержащего сгруппированные формулы

Основные шаги алгоритма таковы (см. рис. 13):

- вводится поисковый запрос в TeX-нотации, для его отображения в браузере и дальнейшей обработки используются библиотеки MathJax [24, 25] и JQuery [26], средствами которых генерируется код запроса на языке MathML;
- в этом коде удаляются теги пространств имен, атрибутов и параметров форматирования;
- записывается регулярное выражение, позволяющее учесть различия в записях переменных в одной и той же формуле в различных документах коллекции, где осуществляется поиск (см. рис. 14, 15);
- выполняется поиск совпадений в индекс-файле, формируется набор гиперссылок на документы коллекции, содержащие искомую формулу (пример см. на рис. 16).

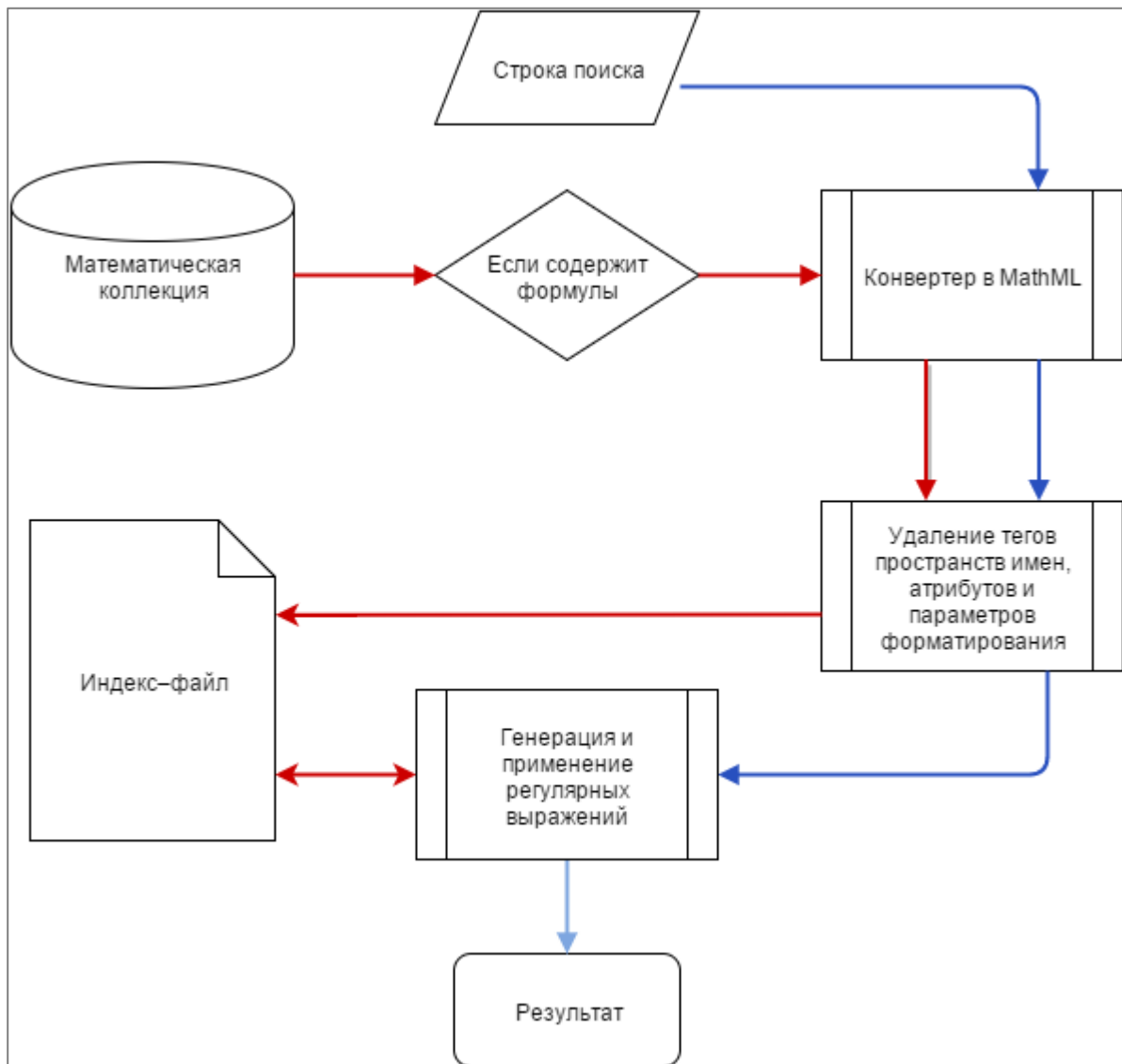


Рис. 13. Алгоритм работы программы поиска по формулам

```
21 //Эквивалентность.
22 preg_match_all("/(?<=<mi>)$excludedvars(=?<\mi>)/", $query, $vars);
23 $vars=array_count_values($vars[0]);
24 foreach($vars as $key=>$var){
25     if($var>1){
26         $query=preg_replace("/(?<=<mi>)$key(=?<\mi>)/", "(.)", $query, 1);
27         $query=str_replace("<mi>$key</mi>", "<mi>\\1</mi>", $query);
28     }else $query=preg_replace("/(?<=<mi>)$key(=?<\mi>)/", "$excludedvars", $query, 1);
29 }
30
```

Рис. 14. Фрагмент кода для учета различий в записях переменных в одной и той же формуле

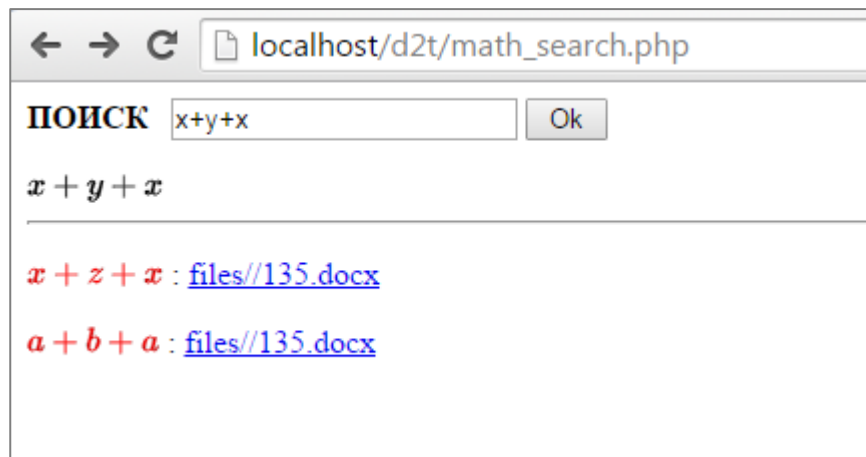


Рис. 15. Иллюстрация обработки свободных переменных

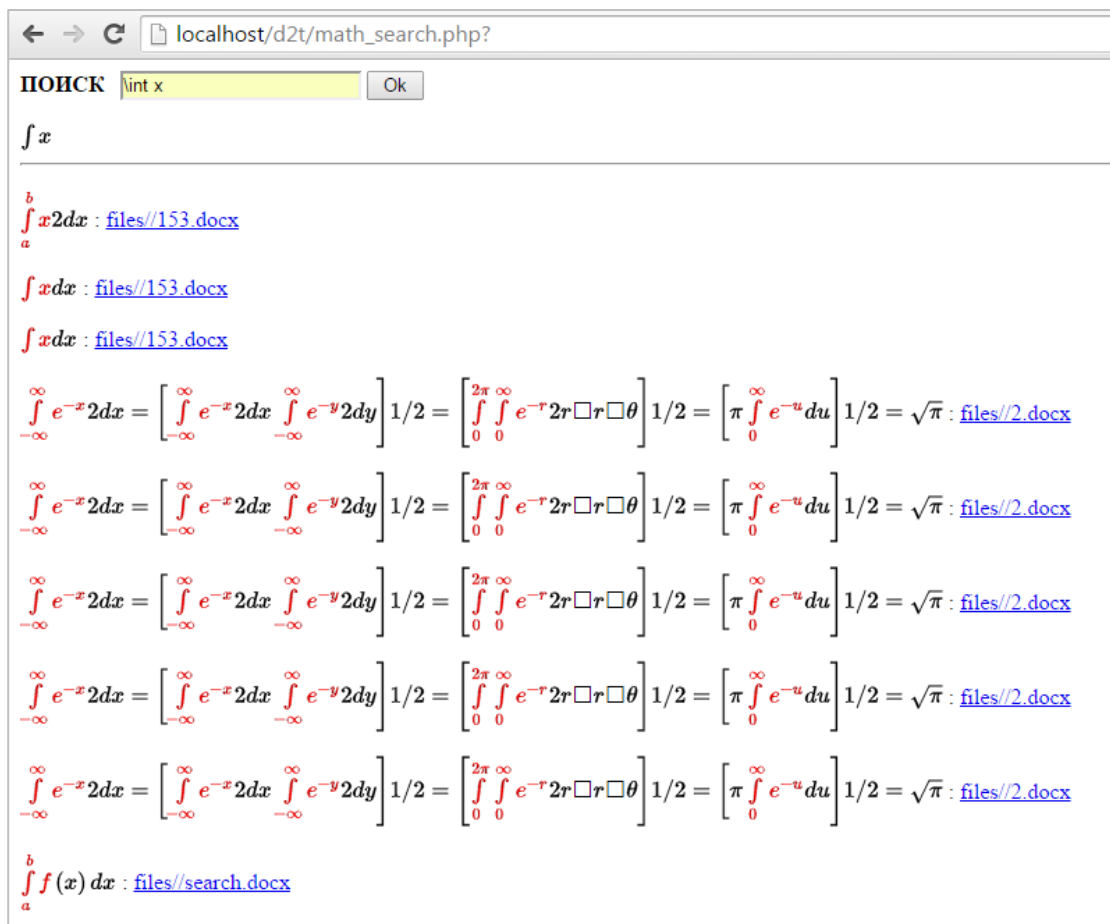


Рис. 16. Отображение результатов поиска по запросу $\int x$

ЗАКЛЮЧЕНИЕ

Наиболее эффективными в организации поиска представляются подходы, основанные на семантических языках содержательного уровня, таких, как Content MathML. Однако ввиду отсутствия программных средств, позволяющих качественно и полноценно извлекать код из форматов хранения презентационного уровня, возникают трудности в обработке таких выражений. Тем не менее, все форматы, описанные выше, можно привести к общему виду, используя такие технологии, как XSLT и JavaScript. Изложенный подход позволил реализовать поиск по документам в формате OpenXML. Кроме того, частично решена проблема поиска по математическим формулам.

БЛАГОДАРНОСТИ

Работа выполнена при финансовой поддержке Российского гуманитарного научного фонда, проект 14-03-12004, а также Российского фонда фундаментальных исследований и Правительства Республики Татарстан в рамках научного проекта № 15-47-02472.

СПИСОК ЛИТЕРАТУРЫ

1. *Елизаров А.М., Липачев Е.К., Хохлов Ю.Е.* Семантические методы структурирования математического контента, обеспечивающие расширенную поисковую функциональность // Информационное общество. 2013. № 1–2. С. 83-92.
2. *Елизаров А.М., Липачев Е.К., Малахальцев М.А.* Веб-технологии в работе электронного математического журнала Lobachevskii Journal of Mathematics // В сборнике: Научный сервис в сети Интернет: многоядерный компьютерный мир. 15 лет РФФИ. Труды Всероссийской научной конференции. Московский государственный университет им. М.В. Ломоносова; Южный федеральный университет; Институт вычислительной математики РАН, г. Москва. 2007. С. 355-356.
3. *Елизаров А.М., Зуев Д.С., Липачёв Е.К.* Информационные системы управления электронными научными журналами // Научно-техническая информация. Серия 1. Организация и методика информационной работы. 2014. № 3. С. 31-38.
4. *Хайдаров Ш.М.* Методы управления математическим контентом в информационных издательских системах // Тр. Математического центра им. Н.И.

Лобачев-ского. Материалы 14-й Всерос. Молодежной школы-конференции «Лобачевские чтения–2015» (Казань, 22–27 октября 2015 года). Казань. 2015. С. 162-165.

5. *Boyter B.B.* Open XML – Кратко и доступно. Open XML Technical Evangelist, Microsoft, 2007. 101 с.

6. *Standard ECMA-376: Office Open XML File Formats* [Электронный ресурс] URL: <http://www.ecmainternational.org/publications/standards/Еcma-376.htm>.

7. *Липачёв Е.К., Хайдаров Ш.М.* Система сервисов преобразования электронных математических документов на основе облачных технологий // Труды Математического центра им. Н.И. Лобачевского. Казань. 2013. Т. 47. С. 109-110.

8. *Елизаров А.М., Липачёв Е.К., Невзорова О.А., Соловьев В.Д.* Методы и средства семантического структурирования электронных математических документов // Доклады Академии наук. 2014. Т. 457. № 6. С. 642-645.

9. *Ахметов Д.Ю., Герасимов А.Н., Грачев А.О., Елизаров А.М., Липачёв Е.К.* Облачная платформа поддержки электронных научных изданий // Учёные записки Института социальных и гуманитарных знаний. 2014. № 1 (12), ч.1. С. 13-19.

10. *Elizarov A.M., Kirillovich A.V., Lipachev E.K., Nevzorova O.A., Solovyev V.D., Zhiltsov N.G.* Mathematical knowledge representation: semantic models and formalisms // Lobachevskii Journal of Mathematics. 2014. V. 35. No 4. P. 348-354.

11. *Биряльцев Е.В., Елизаров А.М., Жильцов Н.Г., Липачёв Е.К., Невзорова О.А., Соловьев В.Д.* Методы анализа семантических данных математических электронных коллекций // Научно-техническая информация. Серия 2: Информационные процессы и системы. 2014. № 4. С. 12-17.

12. *Биряльцев Е.В., Галимов М.Р., Жильцов Н.Г., Невзорова О.А.* Подход к семантическому поиску математических выражений // OSTIS-2012. 2012. С. 245-256.

13. *Nevzorova O., Zhiltsov N., Kirillovich A., Lipachev E.* OntoMathPRO Ontology: A Linked Data Hub for Mathematics // Knowledge Engineering and the Semantic Web Communications in Computer and Information Science. 2014. V. 468. P. 105-119.

14. *Веселаго В.Г., Елизаров А.М., Липачев Е.К., Малахальцев М.А.* Формирование и поддержка физико-математических электронных научных изданий: переход на технологии Семантического Веба // Научно-исследовательский институт математики и механики им. Н.Г. Чеботарева Казанского государственного университета. 2003–2007 гг. Коллективная монография под ред. А.М. Елизарова. Казань: Изд-во Казан. ун-та, 2008. С. 456-476.

15. *Елизаров А.М., Липачёв Е.К., Малахальцев М.А.* Веб-технологии для математика. Основы MathML. М.: Физматлит, 2010. 194 с.

16. *Елизаров А.М., Липачев Е.К., Малахальцев М.А.* Технологии Semantic Web в практике работы электронного журнала по математике // Труды 8 Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» – RCDL'2006, Суздаль. Ярославль: Ярославский госуниверситет, 2006. С. 215-218.

17. *Ausbrooks R., Buswell S., Carlisle D., Chavchanidze G. etc.* Mathematical Markup Language (MathML) Version 3.0 2nd Edition. W3C Recommendation 10 April 2014 // World Wide Web Consortium (W3C). 2014. URL: <http://www.w3.org/TR/MathML/mathml.pdf>.

18. *Kohlhase M., Şucan I.A.* A search engine for mathematical formulae // International Conference on Artificial Intelligence and Symbolic Computation. 2006.

19. *Muhammad Adeel, Hui Siu Cheung, Sikandar.* Math GO! prototype of a content based mathematical formula search engine // Journal of Theoretical and Applied Information Technology. 2008.

20. *Wiki Pages – web-xslt. Example XSLT code for transforming XML languages for the web.* URL: <https://code.google.com/p/web-xslt/>.

21. *Kohlhase M.* MathML presenting and capturing mathematics for the Web. URL: <http://www.w3.org/Math/Documents/mathml-tutorial.pdf>.

22. *SnuggleTeX – Overview & Features* [Электронный ресурс] // School of Physics and Astronomy. URL: <http://www2.ph.ed.ac.uk/snuggletex/documentation/overview-and-features.html>.

23. *Working with MathML & Wolfram Language Documentation.* URL: <http://reference.wolfram.com/language/XML/tutorial/MathML.html>.

24. *MathJax. Beautiful math in all browsers.* URL: <http://www.mathjax.org/>.

25. *Getting Started – MathJax 2.5 documentation*. URL: <http://docs.mathjax.org/en/latest/start.html>.

26. *jQuery – write less, do more* [Электронный ресурс] URL: <https://jquery.com>.

SEMANTIC ANALYSIS OF DOCUMENTS IN THE CONTROL SYSTEM OF DIGITAL SCIENTIFIC COLLECTIONS

S.M. Khaydarov

Kazan (Volga region) Federal University
15jkeee@gmail.com

Abstract

Methods of the semantic documents parsing in digital control system of scientific collections, including electronic journals, offered. The methods of processing documents containing mathematical formulas and methods for the conversion of documents from the OpenXML-format in TeX-format considered. The search algorithm for the mathematical formulas in the collections of documents stored in OpenXML-format designed. The algorithm is implemented as online-service on platform science.tatarstan.

Keywords: semantic analysis, publishing systems.

REFERENCES

1. *Elizarov A.M., Lipachev E.K., Khokhlov Yu.E.* Semanticheskie metody strukturirovaniya matematicheskogo kontenta, obespechivayushchie rasshirennuyu poiskovuyu funktsional'nost' // *Informatsionnoe obshchestvo*. 2013. № 1–2. S. 83-92.
2. *Elizarov A.M., Lipachev E.K., Malakhaltsev M.A.* Veb-tehnologii v rabote elektronnoho matematicheskogo zhurnala Lobachevskii Journal of Mathematics // *V sbornike: Nauchnyy servis v seti Internet: mnogoyadernyy kompyu-ternyy mir. 15 let RFFI. Trudyi Vserossiyskoy nauchnoy konferentsii. Moskovskiy gosudarstvennyy universitet im. M.V. Lomonosova, Yuzhnyy federalnyy universitet, Institut vychislitelnoy matematiki RAN*. g. Moskva, 2007. S. 355-356.

3. *Elizarov A.M., Zuev D.S., Lipachev E.K.* Informatsionnye sistemy upravleniya elektronnyimi nauchnymi zhurnalami // Nauchno-tekhnicheskaya informatsiya. Seriya 1. Organizatsiya i metodika informatsionnoy raboty. 2014. № 3. S. 31-38.

4. *Khaydarov Sh.M.* Metody upravleniya matematicheskim kontentom v informatsionnykh izdatel'skikh sistemakh // Tr. Matem. tsentra im. N.I. Lobachevskogo. Materialy 14-y Vseros. Molodezhnoy shkoly-konferentsii «Lobachevskie chteniya–2015» (Kazan', 22–27 oktyabrya 2015 goda). Kazan'. 2015. S. 162-165.

5. *Vouter V.V.* Open XML – Kratko i dostupno. Open XML Technical Evangelist, Microsoft, 2007. 101 s.

6. *Standard ECMA-376: Office Open XML File Formats.* URL: <http://www.ecmainternational.org/publications/standards/Ecma-376.htm>

7. *Lipachev E.K., Khaydarov Sh.M.* Sistema servisov preobrazovaniya elektronnykh matematicheskikh dokumentov na osnove oblachnykh tekhnologiy // Trudy Matematicheskogo tsentra im. N.I. Lobachevskogo. Kazan'. 2013. T. 47. S. 109-110.

8. *Elizarov A.M., Lipachev E.K., Nevzorova O.A., Solov'ev V.D.* Metody i sredstva semanticheskogo strukturirovaniya elektronnykh matematicheskikh dokumentov // Doklady Akademii nauk. 2014. T. 457. № 6. S. 642-645.

9. *Akhmetov D.Yu., Gerasimov A.N., Grachev A.O., Elizarov A.M., Lipachev E.K.* Oblachnaya platforma podderzhki elektronnykh nauchnykh izdaniy // Uchenye zapiski Instituta sotsial'nykh i gumanitarnykh znaniy. 2014. № 1 (12), ch.1. S. 13-19.

10. *Elizarov A.M., Kirillovich A.V., Lipachev E.K., Nevzorova O.A., Solovyev V.D., Zhiltsov N.G.* Mathematical knowledge representation: semantic models and formalisms // Lobachevskii Journal of Mathematics. 2014. V. 35. No 4. P. 348-354.

11. *Biryal'tsev E.V., Elizarov A.M., Zhil'tsov N.G., Lipachev E.K., Nevzorova O.A., Solov'ev V.D.* Methods for analyzing semantic data of electronic collections in mathematics // Automatic Documentation and Mathematical Linguistics. 2014. V. 48, No 2. P. 81-85.

12. *Biryal'tsev E.V., Galimov M.R., Zhil'tsov N.G., Nevzorova O.A.* Podkhod k semanticheskomu poisku matematicheskikh vyrazheniy // OSTIS-2012. 2012. S. 245–256.

13. *Nevzorova O., Zhiltsov N., Kirillovich A., Lipachev E.* OntoMathPRO Ontology: A Linked Data Hub for Mathematics // Knowledge Engineering and the

Semantic Web Communications in Computer and Information Science. 2014. V. 468. P. 105-119. URL: http://link.springer.com/chapter/10.1007/978-3-319-11716-4_9.

14. *Veselago V.G., Elizarov A.M., Lipachev E.K., Malakhal'tsev M.A.* Formirovanie i podderzhka fiziko-matematicheskikh elektronnykh nauchnykh izdaniy: perekhod na tekhnologii Semanticheskogo Veba // Nauchno-issledovatel'skiy institut matematiki i mekhaniki im. N.G. Chebotareva Kazanskogo gosudarstvennogo universiteta. 2003–2007 gg. Kollektivnaya monografiya pod red. A.M. Elizarova. 2008. S. 456-476.

15. *Elizarov A.M., Lipachev E.K., Malakhal'tsev M.A.* Veb-tekhnologii dlya matematika. Osnovy MathML. Moskva: Fizmatlit, 2010. 194 s.

16. *Elizarov A.M., Lipachev E.K., Malakhaltsev M.A.* Tekhnologii Semantic Web v praktike raboty ehlektronnogo zhurnala po matematike // Trudy 8 Vserossijskoj nauchnoj konferentsii «Elektronnye biblioteki: perspektivnye metody i tekhnologii, ehlektronnye kolleksii» – RCDL'2006, Suzdal'. Yaroslavl': Yaroslavskij gosuniversitet, 2006. S. 215-218.

17. *Ausbrooks R., Buswell S., Carlisle D., Chavchanidze G. etc.* Mathematical Markup Language (MathML) Version 3.0 2nd Edition. W3C Recommendation 10 April 2014. // World Wide Web Consortium (W3C). 2014. URL: <http://www.w3.org/TR/MathML/mathml.pdf>

18. *Kohlhase M., Şucan I.A.* A Search Engine for Mathematical Formulae // Inter-national Conference on Artificial Intelligence and Symbolic Computation. 2006.

19. *Muhammad Adeel, Hui Siu Cheung, Sikandar.* Math GO! prototype of a content based mathematical formula search engine // Journal of Theoretical and AP.ied Information Technology. 2008.

20. *Wiki Pages – web-xslt. Example XSLT code for transforming XML languages for the web.* URL: <https://code.google.com/p/web-xslt/>.

21. *Kohlhase M.* MathML Presenting and Capturing Mathematics for the Web // World Wide Web Consortium (W3C). URL: <http://www.w3.org/Math/Documents/mathml-tutorial.pdf>.

22. *SnuggleTeX – Overview & Features* // School of Physics and Astronomy. URL: <http://www2.ph.ed.ac.uk/snuggletex/documentation/overview-and-features.html>.

23. *Working with MathML& Wolfram Language Documentation* // Wolfram: Computation Meets Knowledge. URL: <http://reference.wolfram.com/language/XML/tutorial/MathML.html>.

24. *MathJax. Beautiful math in all browsers*. URL: <http://www.mathjax.org/>.

25. *Getting Started – MathJax 2.5 documentation*. URL: <http://docs.mathjax.org/en/latest/start.html>

26. *jQuery – write less, do more*. URL: <https://jquery.com>.

СВЕДЕНИЯ ОБ АВТОРЕ



ХАЙДАРОВ Шамиль Махматович – аспирант Института математики и механики им. Н.И. Лобачевского Казанского (Приволжского) федерального университета.

Shamil Mahmutovich KHAYDAROV, received MS degree in mathematics from Kazan Federal University (2015). Currently is a graduate student at the N.I. Lobachevskii Institute of Mathematics and Mechanics of Kazan Federal University. Current scientific interests: data mining, recommender systems, cloud computing, knowledge extraction technologies.

email: 15jkeee@gmail.com

Материал поступил в редакцию 12 февраля 2015 года