

ФОРМИРОВАНИЕ НАУЧНО-ИНФОРМАЦИОННОГО ПРОСТРАНСТВА РЕСПУБЛИКИ ТАТАРСТАН НА ОСНОВЕ СЕРВИСНО-ОРИЕНТИРОВАННОЙ ПЛАТФОРМЫ

А.М. Елизаров, Е.К. Липачёв, А.Ф. Хасьянов
Казанский (Приволжский) федеральный университет

Одной из черт происходящего сегодня перехода от индустриального общества к информационному является то, что информация и знание становятся одним из основных факторов развития, под воздействием которого значительно изменяются формы экономической деятельности и социальные взаимоотношения. Отличительной чертой сегодняшнего этапа развития человечества является представление информации и знаний не только в традиционной печатной, но и в электронной, цифровой форме, что позволяет принципиально иначе создавать, хранить, организовывать доступ и использовать информацию.

В информационном обществе одним из необходимых условий обеспечения конкурентоспособности является использование информационно-коммуникационных технологий (ИКТ). Применение ИКТ в научно-образовательном процессе поставило новые задачи в издательской деятельности, в частности, для полноценного включения в мировые информационные потоки необходимо решить задачу интеграции в едином информационном пространстве электронных документов научного и образовательного содержания. Это дает возможность обращения к данным независимо от места их размещения; анализа информации, включающего определение 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. Один из подходов к решению названной задачи и повышению поисковых возможностей систем поддержки научных изданий основан на создании специализированных информационных систем с применением технологий Семантического Веба (см., например, [1]). В программных документах консорциума W3C указано, что Семантический Веб – это расширение традиционного Веба в направлении существенно лучшего определения смысла информации, позволяющего компьютерам и людям эффективнее выполнять совместную работу. Основная цель состоит в создании среды, в которой программные агенты динамически обнаруживают и анализируют ресурсы, после чего взаимодействуют с ними, а также в решении 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. Функциональные возможности такой платформы, в частности, должны включать сервисы, регулирующие процесс рецензирования и обеспечивающие коллективное редактирование электронных документов. Также должна быть обеспечена поддержка редакционных сервисов, включая классификацию, аннотирование, выделение метаданных, публикацию, долгосрочное хранение, конвертирование, распространение, синдикацию, статистику использования, харвестинг, объединение в коллекции, взаимодействие с институциональными репозиториями, контроль доступа, подписку, рассылку уведомлений, новые поступления. 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. Исследования в описанном направлении проводились нашей научной группой в соответствии с ключевыми положениями консорциума W3C в рамках нескольких проектов РФФИ, в том числе совместных с АН Республики Татарстан (РТ).

В РТ с 2013 года на базе Технопарка в сфере высоких технологий «ИТ-парк» развивается программная платформа science.tatarstan.ru. Она построена в виде веб-портала и позволяет реализовать документооборот в научных журналах гуманитарного и естественнонаучного циклов. Кроме того, для научного журнала *Lobachevskii Journal of Mathematics* в пилотном варианте разработаны сервисы автоматизации ряда бизнес-процессов жизненного цикла электронных публикаций. Основная цель работы – создание на основе платформы science.tatarstan.ru информационной системы облачных сервисов, обеспечивающих многоуровневую поддержку процессов подго-

товки, публикации и интеграции научных знаний, образовательных и историко-культурных материалов в информационное пространство РТ. К настоящему времени получены следующие основные результаты (см. также [2–6]).

Проанализированы сервисы, обеспечивающие автоматизацию основных редакционных процессов издания электронных научных журналов. Описана типовая организация предлагаемых сервисов, предложен собственный подход к организации подобного сервиса. Развита методика автоматической обработки текстов, позволившие решать задачи извлечения знаний в терминах онтологий. Онтологические модели предметных областей использованы как технологическая основа построения рекомендательного сервиса, позволяющего выполнить персонализированный отбор научных документов в соответствии с семантическим профилем учёного.

Обсуждены сервисы поддержки жизненного цикла научной публикации, определяющие функциональность современных программных платформ управления электронными научными журналами. Особое внимание уделено сервису поддержки научного рецензирования – разработан алгоритм автоматизации экспертной оценки, указаны изменения как в программном коде открытой системы Open Journal Systems, являющейся ядром программной платформы, так и в используемой базе данных. Расширен функционал платформы science.tatarstan.ru включением сервисов автоматизации редакционных процессов, поисковой оптимизации, в том числе, автоматического формирования метаданных и использования микроданных.

Проанализированы методы хранения математических текстов в различных платформах и системах. Описан метод хранения математических выражений, созданный преимущественно для Веба и основанный на языке семантической разметки – MathML. Разработан новый формат документов офисного пакета Microsoft Office – OpenXML. Изучена проблема отображения математического контента в информационных системах. Приведены примеры отображения контента, хранящегося в нотации TEX и языке разметки MathML. Выделены особенности использования математических выражений в системе OJS. Предложен метод решения проблемы отображения TEX-файлов в издательской системе OJS на основе специализированного плагина, использующий библиотеку MathJax. Описаны методы конвертации данных между описанными форматами. Предложен метод извлечения данных из формата OfficeXML. Описаны методы обработки математических текстов, а именно: поиск по математическим документам. Предложен метод поиска по формулам. Описаны проблемы, возникающие при обработке математических выражений, и предложены варианты их решения.

Работа выполнена при финансовой поддержке РФФИ и АН РТ (проекты 15-07-08522, 15-47-02472).

Список литературы

1. *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.
2. *Ахметов Д.Ю., Елизаров А.М., Липачёв Е.К.* Автоматизация редакционных процессов в электронных научных журналах // Электронные библиотеки. 2015. Т. 17. № 1–2 (тематический выпуск). С. 3–14.
3. *Герасимов А.Н., Елизаров А.М., Липачёв Е.К.* Формирование метаданных для международных баз цитирования в системе управления электронными научными журналами // Электронные библиотеки. 2015. Т. 17. № 1–2 (тематический выпуск). С. 15–31.
4. *Елизаров А.М., Жижченко А.Б., Жильцов Н.Г., Кириллович А.В., Липачёв Е.К.* Онтологии математического знания и рекомендательная система для коллекций физико-математических документов // Докл. Академии наук. 2016. Т. 467. № 4. С. 392–395. DOI: 10.7868/S0869565216100042.
5. *Зуев Д.С.* Облачный редакционно-издательский сервис научного журнала // Электронные библиотеки. 2015. Т. 17. № 1–2 (тематический выпуск). С. 36–51.
6. *Хайдаров Ш.М.* Семантический анализ документов в системе управления цифровыми научными коллекциями // Электронные библиотеки. 2015. Т. 17. № 1–2 (тематический выпуск). С. 52–82.