ОБЛАЧНЫЕ СЕРВИСЫ В РАБОТЕ ЭЛЕКТРОННОГО НАУЧНОГО ЖУРНАЛА

А.М. Елизаров, Е.К. Липачёв, Ю.Е. Хохлов
Казанский (Приволжский) федеральный университет, Институт развития информационного общества (г. Москва)
amelizarov@gmail.com, lipachev@ksu.ru, yuri.hohlov@iis.ru

Аннотация. В докладе обсуждаются вопросы создания сервисов электронного научного журнала. Проведен сравнительный анализ технологий, используемых в электронных журналах в настоящее время. Отдельное внимание уделено управлению электронным контентом с высокой гранулированностью текста. Приведен опыт успешного внедрения в практику работы электронного научного журнала сервисов, функционирующих на основе технологий семантического веба.

Электронные научные журналы. Современная научная информация, представленная в электронной форме, по своим объемам значительно превышает выпуск научных книг, изданных полиграфическим способом. Сегодня издается более 30 тысяч электронных научных журналов. Все крупнейшие издательства сопровождают выпуск бумажных номеров журналов размещением в сети их электронных вариантов. Большинство электронных научных ресурсов представлено в сети Интернет и объединено в информационные системы с наборами сервисов по поиску и обработке документов (см., напр., [1]). Благодаря этому эффективность работы с научным материалом имеет более высокий, недоступный ранее, качественный уровень.

Распространение электронных научных журналов и электронных изданий, в целом, связано с прогрессом в области информационно-коммуникационных технологий. Увеличение числа и качества сетевых сервисов, расширение географии доступа в Интернет и «цифровое выравнивание» регионов (см. [2, 3]), способствует превращению электронной формы научного (как, впрочем, и другого) взаимодействия в основную.

Процесс развития электронного научного книгоиздания предполагает разработку специализированных форматов и правил, регулирующих электронный документооборот.

В Казанском университете и Научной электронной библиотеке в рамках нескольких проектов Министерства образования и науки были разработаны технологии обработки электронных версий печатных журналов, созданы форматы данных, формализованы и юридически оформлены правила, регулирующие отношения авторов, издателей и электронных библиотек (см. [1]).

Как часть этих исследований выполнена разработка шаблонной информационной системы управления электронным математическим журналом, способной в автоматическом режиме выполнять ряд процессов, стандартных для научного журнала (см. [4, 5]).

Пилотная реализация системы автоматизации управления научным журналом выполнена для электронного журнала LJM (Lobachevskii Journal of Mathematics) – одного из первых отечественных научных электронных журналов.

Сервисы электронного журнала. По своему составу список сервисов научного электронного журнала в значительной степени такой же, как и для другой информационной системы управления электронными документами. Это позволяет использовать методы и технологии, примененные при создании других информационных систем, например, государственных (см. [6, 7]).

Основными сервисами, ориентированными на работу авторов и редколлегии электронного журнала , можно назвать следующие: сервис представления статьи, сервис, регулирующий процесс рецензирования, сервис, обеспечивающий коллективное

редактирование электронного документа, поиск и навигация в электронном архиве. К редакционным сервисам можно отнести такие сервисы, как классификация, аннотирование, выделение метаданных, публикация, долгосрочное хранение, конвертирование, распространение, синдикация, статистика использования, харвестинг, объединение в коллекцию, взаимодействие с институтциональными репозиториями, контроль доступа, подписка, рассылка уведомлений, новые поступления.

Роль и место каждого из указанных сервисов в информационной системе электронного научного журнала отражена в докладе.

Использование методик, разработанных для государственных структур и бизнеса, позволяет выделить направление работ, но не решает проблем, специфических для каждой научной области.

Специфика научных электронных коллекций – в наличии специализированных для конкретной предметной области конструкций, например, формул для математических публикаций, высокой гранулированностью научного текста, жесткой структурой изложения. Дополнительным элементом является наличие расчетов в научных публикациях. Создание сервисов автоматической обработки электронных научных коллекций отдельная и многоплановая тема (см., напр., [1]). Как один из прототипов информационной системы электронной научной коллекции можно рассматривать программную среду управления электронным математическим хранилищем, реализованную в проекте автоматизации работы электронного журнала «Lobachevskii Journal of Mathematics» (см. [4]).

Семантический веб и семантические сервисы. Интернет сегодня является ключевой инфраструктурой современного информационного общества. Основополагающая роль в технологическом переоснащении глобальной сети принадлежит технологиям семантического веба, разработка которых координируется консорциумом W3C (http://www.w3c.org).

Выбор технологий семантического Веба для организации работы научного электронного журнала объясняется наличием в них инструментов, позволяющих учитывать структурную и семантическую составляющие информации (см., напр., [6, 8, 9]).

Технологии семантического веба обеспечивают стандартную процедуру создания языка разметки, адаптированного к определенной предметной области, что позволяет гибко адаптировать структуру электронного хранилища для включения контента новых предметных областей (см., напр., [8]).

На основе уже разработанных в последнее время языках разметки и предназначенных для математических, химических, биологических и других предметных областей можно проектировать структуру естественнонаучного электронного хранилища. Общая технологическая основа этих языков позволяет использовать практические рекомендации семантического веба для управления междисциплинарным контентом естественнонаучного электронного хранилища (см., напр., [8, 10]).

Одна из тем, обсуждаемых в докладе, – методы сведения языков разметки, созданных на основе XML для большинства научных областей в единую систему и построение информационной среды, обеспечивающей их взаимодействие. Одной из основных частей этой системы является интеллектуальный поиск, позволяющий пользователю отслеживать динамические связи между документами с учетом специфики предметной области.

Исследование задачи управления специализированными научными ресурсами на основе технологий семантического веба было проведено нами в рамках нескольких проектов РФФИ, в частности, на основе разработанных подходов была реализованы программная среда электронного научного журнала. Методы автоматизации основаны на

использовании XML-схем и DTD-описаний, а также XSLT-преобразований для управления информационными потоками.

Поисковая оптимизация. В докладе обсуждается также комплекс мероприятий по продвижению сайта электронного журнала и, в том числе использованию методов поисковой оптимизации на основе рекомендаций Search Engine Optimization (см., напр., [11]).

Эффективность поиска напрямую связана с организацией ключевых слов и метаданных. Метаданные содержат обобщенную информацию о структуре и содержании информационного источника (автор, дата, источник, ключевые слова, предметная область и т. д.). В научных электронных журналах комплексное использование метаданных, записанных по правилам Dublin Core (DC) и RDF позволяет повысить эффективность поиска (см. [5, 10]).

Значительная часть научного контента создана по технологиям, ориентированным на печать как заключительную фазу публикации. Оцифровка этих публикаций, как правило, включает в себя получение электронного варианта и создание сервисов отображения. Блок метаданных создается вручную и, в большинстве реализаций, содержат только информацию о публикации (автор, название, выходные данные, библиографические ссылки). Автоматизация генерации метаданных затруднена в силу слабой структурированности информации оцифрованного контента. В докладе представлены подходы к решению проблемы выделения метаданных.

Облачные технологии. Термин “Cloud Computing” часто переводят как “облачные вычисления”, что, на наш взгляд, отражает лишь одну из сторон этой технологии, а именно, “вычислительную”, когда вычислительные ресурсы облака используются как сервис (см., напр., [12]). Доклад посвящен другой сфере использования Cloud Computing – научные электронные публикации и распространение научных результатов с помощью сервисов Cloud Computing.

Cloud Computing определяется как эволюция IT-сферы, когда предоставление и использование сервисов осуществляется посредством самообслуживания через Интернет и оплаты (на договорной или иной основе) за фактическое потребление ресурсов. Отличительными особенностями облачных сред от традиционных, являются следующие: вычислительные функции и функции хранения логически разделены и предлагаются как сервисы, сервисы создаются с учетом инфраструктуры, допускающей масштабирование в очень широких пределах, сервисы предоставляются по требованию посредством выделения динамичных, гибко настраиваемых ресурсов, сервисы легко приобрести, и они оплачиваются в соответствии с фактическим потреблением, ресурсы используются совместно, сразу многими пользователями, доступ к сервисам осуществляется через Интернет или внутреннюю сеть с помощью любых устройств (см., напр., [12 – 14]).

Среда cloud computing обещает изменить способы создания, предоставления и использования IT-сервисов. Предполагается, что облачная среда сможет быть объединенной, автоматизированной и учитывающей особенности клиентских устройств, что означает гибкую адаптацию облачных сервисов к устройствам (и установленным на них системах), используемым пользователем для получения сервиса (см., напр., [15]).

Облачные технологии учитывают увеличение числа и многообразие мобильных устройств и ориентированы на пользователей, работающих с несколькими устройствами – типичной для нашего времени становится ситуация “1 пользователь = много устройств”. Появился термин мобильный сотрудник и изменилось само понятие рабочего места, – с закрепленного места присутствия в деятельность на основе сетевых технологий из любого места и в любое время. По прогнозам (см., напр. [15]) в 2015 году более 2,5 миллиардов

людей с более чем 10 миллиардами устройств будут иметь доступ в Интернет , что более чем в 2 раза превышает теперешний уровень.

Развитие электронных журналов и перевод сервисов в облако нужно рассматривать с учетом процессов глобализации образования (см., напр., [16]).

В докладе на обсуждение вынесены методы автоматизации, реализованные в электронном журнале LJM и реализуемый в настоящее время проект внедрения облачных сервисов в информационную среду электронного журнала «Электронные библиотеки».

Работа поддержана РФФИ (проекты № 12-07-00667 и 12-07-97018-р_поволжье)

Литература

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

[2] Анализ развития и использования информационно-коммуникационных технологий в субъектах Российской Федерации. Проблемы преодоления различий между регионами по уровню информационного развития / Под ред. Ю.Е.Хохлова, С.Б.Шапошника. – М.: Институт развития информационного общества, 2009. – 208 с.

[3] Индекс готовности регионов России к информационному обществу. 2008-2009 / Под ред. Ю.Е. Хохлова и С.Б. Шапошника.–М.: 2010.–296 с.

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

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

[6] Когаловский М.Р., Хохлов Ю.Е. Стандарты XML для электронного правительства. – М.: Институт развития информационного общества, 2008. – 416 с.

[7] Когаловский М.Р., Хохлов Ю.Е. Стандарты Всемирной паутины в разработках электронного правительства. – Информационное общество: научно-аналитический журнал. – 2009. – № 2. – С. 21-32.

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

[9] Yu L. Introduction to Semantic Web and Semantic Web services. – Taylor & Francis Group, 2007. – 331 p.

[10] Елизаров А.М., Липачев Е.К., Малахальцев М.А. Технологии управления разнородным естественнонаучным контентом на основе семантического веба // Электронные библиотеки: перспективные методы и технологии, электронные коллекции: Труды XI Всероссийской научной конференции RCDL-2009 (Петрозаводск, 17-21 сентября 2009 г.). – Петрозаводск: КарНЦ РАН, 2009. - С.325-328.

[11] Enge E., Spencer S., Stricchiola J., Fishkin R. The Art of SEO. – O'Reilly Media, 2012. – 714 p.

[12] Cloud Computing: Principles, Systems and Applications. Ed. Antonopoulos N., Gillam L. – Springer-Verlag, 2010. – 386 p.

[13] Handbook of Cloud Computing. Ed. Furht B., Escalante A.. – Springer Science+Business Media, 2010. – 655 p.

[14] Концепция Intel современного перехода к Cloud Computing. – Корпорация Intel, 2010 г. – www.intel.com/go/cloud.

[15] Преимущества среды cloud computing, учитывающей особенности клиентских устройств. – Корпорация Intel, 2010 г. – www.intel.com/go/cloud.
[16] Katz R.N. The Tower and the Cloud: Higher Education in the Age of Cloud Computing. – EDUCAUSE, 2008. – 296 p.
[17] Облачные сервисы: взгляд из России / Под ред. Гребнева. – Cnews, 2011. – 282 с.