

К проектированию технологической составляющей электронной библиотеки Сибирского федерального университета

Ковязина Елена Васильевна

Институт вычислительного моделирования СО РАН, г. Красноярск, Россия

elena@icm.krasn.ru

Аннотация. Доклад посвящен технологической составляющей проекта «Социально-образовательный контент электронной библиотеки СФУ для развития человеческого капитала Красноярского края: концепция и технологии реализации». Приведены результаты исследования потребностей пользователей, обоснование выбора технологической платформы и стандартов метаданных.

Сегодня ни у кого уже не вызывает сомнения, что для позиционирования российского образования, науки и культуры в глобальном Интернет–пространстве необходимо создавать, развивать и поддерживать собственные электронные информационные ресурсы, обеспечивая доступ к информации для потребителей, проживающих как на территории России, так и за ее пределами. Федеральные государственные учреждения образования, науки и культуры призваны занимать лидирующие позиции в создании таких электронных ресурсов, в частности, и электронных библиотек, развитие которых в перспективе может существенным образом отразиться на скорости и глубине социально–экономического развития региона и страны в целом.

Для решения поставленных задач такого плана и был инициирован проект «Социально-образовательный контент электронной библиотеки СФУ для развития человеческого капитала Красноярского края: концепция и технологии реализации», выполняемый при поддержке Красноярского краевого фонда поддержки научной и научно-технической деятельности. Объектом исследования в проекте являются электронные библиотеки, предметом – социально-образовательный контент электронной библиотеки Сибирского федерального университета (ЭБ СФУ). В процессе работы над проектом были организованы фокус-группы, проведены глубинные интервью и анкетный опрос по выявлению ключевых проблем, приоритетов, технологий и организационных форматов «школьного» контента электронной библиотеки СФУ. Участниками стали школьники, молодежь, учителя средних учебных заведений, представители муниципальных и региональных органов управления образованием.

В результате анализа социально-образовательного запроса преподавателей и школьников Красноярского края на первом этапе проекта были выявлены следующие особенности, позволяющие сделать выводы и определить приоритеты технологического воплощения социально-образовательной части контента электронной библиотеки СФУ:

1. Информационные потребности пользователей достаточно многообразны и охватывают весь спектр информационных ресурсов - от досуговых до научно-исследовательских, включая ресурсы на иностранных языках. Следовательно, электронная библиотека СФУ должна располагать полным набором технологий для предоставления пользователям всего спектра имеющихся у нее информационных ресурсов.

2. Все образовательные учреждения края подключены к сети Интернет, и большая часть опрошенных является активными его пользователями, хорошо знакома с социальной составляющей Интернет и умеет пользоваться поисковыми средствами Всемирной паутины. Следовательно, электронная составляющая библиотеки СФУ должна быть доступна пользователям через Web-интерфейс и индексироваться поисковыми системами Интернет.

3. Опрос выявил плохую информированность пользователей в области электронных библиотек, что позволяет сделать вывод об отсутствии или недостаточности навыков работы с

такими ресурсами. Этот пункт позволяет сделать выводы об особой важности доступа к электронной библиотеке СФУ с помощью поисковых систем Интернет.

4. Важным фактом исследования явилось наличие собственного Интернет-сайта в каждом образовательном учреждении. Это обязывает организовать информационный раздел сайта электронной библиотеки СФУ для образовательных учреждений, который возможно было бы использовать как точку входа пользователей в электронную библиотеку СФУ.

Среди характерных особенностей компьютерных технологий сегодняшнего дня в реализациях электронных библиотек можно выделить следующие:

1. Большая часть контента электронных библиотек состоит из электронных аналогов печатных текстов – электронных копий документов. Такой состав контента позволяет использовать для его описания стандарты традиционного библиографического описания документа. Несомненным плюсом такого подхода является обеспечение эффективного поиска по атрибутам, позволяющего практически однозначно идентифицировать нужный документ по набору полей библиографического описания и даже по одному полю. Такое представление электронных документов хорошо развито технологически и обеспечено программными продуктами и квалифицированными кадрами.

2. Напротив, пользовательские запросы в отношении контента электронных библиотек стремительно растут с расширением представлений о возможностях электронных документов и трансформациях самого определения этого понятия. Современные электронные документы, отнюдь, не представляют собой цифровой аналог документа традиционной библиотеки, даже если толковать его в расширенном смысле – как совокупность текстовых, графических, видео-, аудио- и даже мультимедийных файлов. С развитием XML-технологий электронные документы приобрели ряд новых, чрезвычайно важных для пользователей глобальных сетей качеств, трудно определяемых и даже невозможных в рамках традиционных библиотечных технологий. Эти качества привели к необходимости создания особых стандартов и форматов описания документов гипермедийной среды Интернет. Однако так как развитие Интернет идет эволюционным путем, многие из этих стандартов носят рекомендательный характер, продолжают развиваться и совершенствоваться по мере выявления новых качеств электронных документов.

Выход из этого противоречия видится в параллельном внедрении и развитии как традиционных, так и не устоявшихся еще технологий, их постепенном слиянии и взаимном использовании того лучшего, что имеется в каждой из них, предоставлении пользователям возможностей получить это лучшее уже сегодня. Таким образом, задачу предоставления доступа к электронной библиотеке необходимо решать в двух взаимодействующих направлениях:

1. Традиционным для библиотек способом, а именно, с приоритетом атрибутивного поиска необходимого документа в электронном каталоге библиотеки по элементам библиографического описания с последующим переходом к полному тексту по URL-ссылке. В традиционные технологии хорошо вписывается также организация распределенных баз данных из каталогов библиотек и библиографических баз данных, созданных различными программными средствами, на основе сетевых протоколов Z39.50.

2. Оформление электронных документов на платформе XML в технологиях Web 3.0, где метаданные документов существуют в XML/RDF-схемах, предлагаемых и продвигаемых сообществом Интернет. Электронные коллекции создаются в виде архивов открытого доступа, а поиск может быть обеспечен как средствами самого такого архива, так и с помощью поисковых систем Интернет.

Традиционный подход к построению электронных библиотек повсеместно применяется библиотеками и основан на использовании технологической среды систем автоматизации библиотек (САБ). В качестве поисковых метаданных используется библиографическое описание документа в виде записи в электронном каталоге. Полный текст документа представлен в библиографическом описании полем ссылки, которая может быть, в свою очередь, именем файла либо URL-адресом документа в сети Интернет. Такой подход предполагает равноправность всех документов в Интернет, от текстовых до мультимедийных. Поисковые возможности при таком подходе целиком и полностью определяются системой автоматизации библиотек (САБ), в которой

был создан электронный каталог. Наличие международных стандартов библиографического описания позволяет виртуально объединять электронные каталоги библиотек, сформированные в различных САБ, создавая распределенные информационные ресурсы. При построении распределенных поисковых систем для обеспечения совместимости каталогов используется специализированный Интернет-протокол Z39.50. Web-интерфейс электронной библиотеки предоставляется средствами САБ или клиентским местом и/или шлюзом Z39.50-HTTP z-сервера [1]. Указанные технологии работы налагают дополнительные ограничительные требования на хранилища данных цифровых документов. Очевидно, что в электронных библиотеках, в отличие от других документальных информационных систем, приоритетной является доступность данных, которую не всегда удастся полноценно сочетать с их сохранностью. Так, требование обеспечить доступ к электронному документу по URL-ссылке из Интернет часто противоречит требованиям безопасности в ряде готовых хорошо зарекомендовавших себя программно-аппаратных решений, вынуждая производить их доработку, либо использовать их только для архивного хранения. Организация связи между документами в хранилище обеспечивается полностью средствами САБ, а значит, хранилище поддерживает исключительно иерархические связи, определенные стандартом ISO2709. Традиционные технологии описания электронного документа не обеспечивают такие, достаточно уже устоявшиеся в области хранилищ данных свойства документа, как многоверсионность, а также не отражают большинства гипертекстовых связей.

Развитие Интернет и гипертекстовой среды, освоение компьютерных технологий производителями знаний, усложнение электронных документов и сама неоднозначность понятия электронный документ определяют наличие у таких документов некоторых новых качеств, которые отсутствуют у печатных документов. Новые качества документов открывают новые возможности, которые затруднительно или невозможно использовать в рамках традиционного подхода. Эти качества породили новые термины, понятия и определения, активно обсуждаемые в среде профессионалов. Например, одним из наиболее очевидных свойств электронного документа во Всемирной паутине является его гипертекстовость, а внедренные в объект связи должны быть гарантировано актуальными. Как следствие, актуально отличное от традиционных библиотек требование обеспечения *ссылочной целостности данных* [2]. Гипертекстовые ссылки, внедренные в электронный документ, достаточно широко используются сегодня в информационных системах, подобных Web of Science или Российскому индексу научного цитирования (РИНЦ), для наукометрических исследований и оценки публикационной активности ученых и организаций [3]. Еще один пример – географическая привязка документа. Ведь электронный документ, например, средствами GIS, может быть привязан к определенным географическим координатам, с которыми он может быть связан местом проживания автора, содержанием, а также и местом его нынешнего хранения. Следовательно, документ может быть найден по карте, с использованием заложенных в нем данных. Привязка к географическим координатам фиксируется в метаданных документа с определением принадлежности каждого координатного индекса [4]. И наконец, практика размещения самими учеными результатов исследований в форме научных статей и материалов в открытом доступе в сети Интернет постепенно получает организационную поддержку. Научные статьи и материалы, депонируемые их авторами в электронном репозитории организации, являются частью профессиональной информационной среды. Они цитируются наряду с "полноценными" публикациями в рецензируемых журналах. Как следствие, авторы научных статей и материалов могут вносить в них изменения в течение всей своей профессиональной жизни. Электронные научные статьи и материалы при этом получают статус "живого" документа (в зарубежной литературе "liquid publication"- «текущая (или неустойчивая) публикация»), и требуют специальных средств и методов работы с ними. [5]. Приведенные примеры позволяют уловить качественную новизну понятия электронного документа, заложенные в нем потенциальные возможности, а также невозможность их реализации в традиционных технологиях.

Учитывая переход Интернет на XML-платформу [6] и освоение новых технологий, при реализации проектов электронной библиотеки невозможно ограничиться только традиционным подходом к проектированию доступа к электронным ресурсам. Как бы эффективно ни была организована система доступа к документам с помощью САБ, с точки зрения Интернет любая из

предоставляемых в доступ коллекций документов являет собой при таком подходе закрытый архив, подобный данным какой-либо СУБД. Описание документа «не видит» ни одна глобальная поисковая система, а значит, подавляющее большинство пользователей Интернет не могут отыскать его. Выходом является оформление коллекций документов в соответствии со стандартами и рекомендациями Интернет-сообщества. Наиболее освоенной в настоящее время российскими университетами Интернет-технологией является технология архивов открытого доступа [7]. Основопологающей идеей таких архивов является принцип свободы обмена научной информацией, независимости ее распространения от платежеспособности потребителей. Архивы открытого доступа формируются из электронных документов, предоставляемых в такой архив правообладателями для свободного доступа. Такая идеологическая основа диктует использование для создания архивов свободно распространяемого программного обеспечения с открытым исходным кодом полностью совместимого с программным обеспечением, используемым другими архивами такого рода. Основой функционирования архивов, иначе называемых институциональными репозитариями, в сети Интернет являются протоколы сбора метаданных Инициативы открытых архивов – OAI PMH. Лидирующие позиции в этой области занимают два быстро развивающихся проекта – DSpace и Eprints [8]. Данные системы являются представителями одного класса и имеют много общего. Обе являются системами с открытыми исходными кодами, OAI-совместимыми, интероперабельными, эквивалентными по функциональности самоархивирования. Использование технологий архивов открытого доступа позволяет позиционировать контент электронной библиотеки в Интернет, а также решать некоторые задачи, которые при традиционном построении электронной библиотеки приходится решать дополнительно, как то:

- 1) для базовой организации данных зафиксирована определенная модель данных, отражающая структуру этой организации;
- 2) система хранит и индексирует данные в различных форматах;
- 3) система хранит информацию о пользователях системы и обладает функцией авторизации;
- 4) система способна «поглощать» приходящие материалы, обеспечивая функцию самоархивирования;
- 5) интеграция поиска документов обеспечивается протоколами OAI PMH;
- 6) поддержка данных различных форматов от текстовых документов до цифрового видео;
- 7) средства развития web-интерфейса;
- 8) поддержка преобразования данных между ссылками в описании и библиографическим описанием документа.

Технологии Web3.0 и архивов открытого доступа не могли бы использоваться в электронных библиотеках в отсутствие стандарта метаданных электронных документов, описанных в XML/RDF-схеме. С историей разработки и подробным описанием наиболее распространенных стандартов с обширным списком библиографии и ссылками на сайты стандартов можно ознакомиться в [9]. Наиболее известны в библиотечном мире MARC 21, DC (Dublin Core) и MODS (Metadata Object Description Schema). Для указанных стандартов имеется описание схем в XML/RDF-виде, а также таблицы соответствия полей в разделе стандартов Интернет-портала Библиотеки Конгресса США.

В процессе работы над проектом были сформированы следующие технологические рекомендации обработки электронных документов. Формирование описательных метаданных электронных документов, составляющих контент электронной библиотеки СФУ, логично производить средствами САБ. Это позволит составить максимально детальное описание документа без потерь информативности. Для обеспечения взаимодействия технологий потребуется конверсия библиографической записи из САБ в XML/RDF-описание документа в схемах MARC 21 и/или DC. Полученное описание должно быть дополнено полями, отражающими специфические качества электронного документа, а затем может быть инкапсулировано в электронный документ или храниться отдельно в директории доступной для индексирования роботами поисковых машин Интернет. Пример, информационной системы, включающей похожий конвертор форматов, представлен в [10]. Такая процедура обработки позволит не растерять

важную для атрибутивного поиска информацию о документе, и в то же время, использовать все возможности получения информации из документа в электронном виде. Автоматизация конверсии форматов обеспечит минимум трудозатрат на эту процедуру. Наличие же нескольких описаний позволит школьникам и преподавателям отыскать нужный им документ любым привычным способом, будь то электронный каталог библиотеки или поисковая машина Интернет.

Литература

1. Еремеев Л. Г. Обзор электронных библиотек России [Электронный ресурс] / Л. Г. Еремеев, А. В. Кузнецов // Библиотеки и информационные ресурсы в современном мире науки, культуры, образования и бизнеса: материалы конф. – Электрон. дан. – М.: ГПНТБ России, 2010. – 1 электрон. опт. диск (CD-ROM). – Систем. требования: IBM PC, Windows 2000 или выше. – Загл. с этикетки диска. – ISBN 978-5-85638-139-8. – № гос. регистрации 0321000673.
2. Жижимов О. Л. О некоторых отличиях электронных библиотек от хранилищ цифрового контента традиционных библиотек [Электронный ресурс] / О. Л. Жижимов, Н. А. Мазов, А. М. Федотов // Библиотеки и информационные ресурсы в современном мире науки, культуры, образования и бизнеса: материалы конф. – Электрон. дан. – М.: ГПНТБ России, 2010. – 1 электрон. опт. диск (CD-ROM). – Систем. требования: IBM PC, Windows 2000 или выше. – Загл. с этикетки диска. – ISBN 978-5-85638-139-8. – № гос. регистрации 0321000673. – URL: <http://www.gpntb.ru/win/inter-events/crimea2010/disk/131.pdf>.
3. Когаловский М. Р. Информационные ресурсы, наукометрические показатели и показатели качества метаданных системы Соционет [Электронный ресурс] / М. Р. Когаловский, С. И. Паринов // Труды 9^{ой} Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» (RCDL'2007) - Переславль Залесский, 2007. – URL: <http://www.cemi.rssi.ru/mei/articles/koga-pari07-1.pdf>.
4. Жижимов О. Л. Мазов Н.А. Об использовании географических координат при поиске библиографической информации / О. Л. Жижимов, Н. А. Мазов // Научные и технические библиотеки. - 2009. -№ 1. - С. 54-60.
5. Паринов С. И. «Живые» документы в электронных библиотеках [Электронный ресурс] / С. И. Паринов, М. Р. Когаловский // Прикладная информатика. – 2009. - № 6. – URL: <http://www.cemi.rssi.ru/mei/articles/koga-pari09-2.pdf>.
6. Елизаров А. М. Веб-технологии для математика: основы MathML / А. М. Елизаров, Е. К. Липачев, М. А. Малахальцев. – М.: Физматлит, 2010. – 192 с.
7. Новицкий А. В. Обзор некоторых направлений интеграции гетерогенных ресурсов в электронных библиотеках / А. В. Новицкий // Труды 11й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» (RCDL'2009). – Петрозаводск, 2009. – С. 350-356.
8. Кудим К. А.
Создание научных электронных библиотек с помощью системы DSpace [Электронный ресурс] / К. А. Кудим, Г. Ю. Прокудина, В. А. Резниченко // Проблемы программирования – 2007. - № 3. - С. 49-60.
9. Негуляев Е. А. Универсальные схемы метаданных и задача описания веб-ресурсов / Е. А. Негуляев // Библиосфера. – 2006. - № 1. – С. 43-54.
10. Борисовский В. Ф. Организация открытого архива научных публикаций сотрудников ОИЯИ / В. Ф. Борисовский [и др.] // Труды 11й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» (RCDL'2009). – Петрозаводск, 2009. – С. 451-458.