

УДК 02-029:004

О. Л. Жижимов¹, Н. А. Мазов², А. М. Федотов³

НЕКОТОРЫЕ ЗАМЕТКИ ОБ ЭВОЛЮЦИИ ЦИФРОВЫХ РЕПОЗИТАРИЕВ ТРАДИЦИОННЫХ БИБЛИОТЕК К ПОЛНОФУНКЦИОНАЛЬНЫМ ЭЛЕКТРОННЫМ БИБЛИОТЕКАМ

В настоящее время работы по созданию электронных библиотек (ЭБ), начатые за рубежом в конце 80-х годов прошлого века, ведутся во всех развитых странах и приобрели статус национальных программ и международных проектов. Об этом свидетельствует, например, достаточное количество разработанных электронных архивов, электронных библиотек и других, смежных информационных систем [1-7]. Электронные библиотеки предоставляют множество новых возможностей практически для всех пользователей, которые имеют с ними дело. Так, например, библиотечные работники видят в электронных библиотеках новые возможности для совершенствования автоматизированных библиотечных систем, превращения их в публичные электронные библиотеки нового поколения с развитыми средствами представления разнообразных цифровых информационных ресурсов и доступа к ним, создаваемые с учетом необходимости интеграции издательских и библиотечных технологий.

Отличительная черта сегодняшнего этапа развития информационного общества это представление информации и знаний не только в традиционной печатной, но и в электронной, цифровой форме, что позволяет принципиально по-иному создавать, хранить, организовывать доступ и использовать информацию в любой форме. Наряду с этим современные информационно-телекоммуникационные технологии привели к тому, что большое число современных

¹ © Олег Львович Жижимов, доктор технических наук, заведующий лабораторией информационных ресурсов ИВТ СО РАН, пр.Лаврентьева, 6, Новосибирск, 630090, E-mail: zhizhim@sbras.ru.

² © Николай Алексеевич Мазов, канд. тех. наук, заведующий информационно-библиотечным центром ИНГГ СО РАН, пр.Коптюга,3, г. Новосибирск, 630090, E-Mail: MazovNA@ipgg.nsc.ru.

³ © Анатолий Михайлович Федотов, доктор физико-математических наук, чл.-корр. РАН, профессор, проректор по информатизации НГУ, главный научный сотрудник ИВТ СО РАН, ул.Пирогова,26, г. Новосибирск, 630090, E-Mail: fedotov@ict.nsc.ru.

информационных ресурсов сразу создается в электронном виде, т. е. формируется новый класс информационных систем, предназначенных для управления электронными информационными ресурсами — электронные библиотеки.

Несмотря на то, что ЭБ — явление достаточно новое, тем не менее, сегодня их следует рассматривать как множество слабосвязанных сущностей, объединяемых, на первый взгляд, только общим названием [1-6]. При этом, под общим названием «Электронная библиотека» сегодня фигурируют следующие объекты:

1. Архивы цифрового контента — хранилища переведенной в цифровую форму информации, снабженные минимальными интерфейсами доступа к этой информации, при этом не всегда даже сетевыми интерфейсами. Электронной библиотекой может называться DVD-диск вместе с прилагаемым программным обеспечением для доступа к цифровому контенту, организованного в виде файловой системы на этом диске.

2. Набор программного обеспечения, реализующего основные функции управления цифровым контентом и организации интерфейсов доступа к этому контенту.

3. Системы сетевых сервисов, предоставляющих доступ к цифровому контенту, объединенных единой системой управления этим доступом [8-9].

4. Некоторые организации, которые берут на себя ответственность не только за исполнение функций управления цифровым контентом и предоставления к нему доступа всем заинтересованным лицам, но и ответственность за соблюдение в процессе исполнения этих функций соответствия последних текущему законодательству в части обеспечения авторского права, приватности персональной информации, фильтрации паразитного трафика, сетевой безопасности клиента и т.д. и т.п.

Здесь особенно следует подчеркнуть тот факт, что последнее определение электронной библиотеки полностью соответствует определению традиционной библиотеки как организации в системе, например, министерства культуры. Именно это определение электронной библиотеки, наиболее импонирует руководству наших ведущих традиционных библиотек (РГБ, РНБ, ГПНТБ, ГПНТБ СО РАН и др.), поскольку обеспечивает преемственность методов работы в эпоху перехода к цифровому способу обработки и хранения информации. Однако при этом, упускается из виду тот факт, что работа с цифровым контентом, для обеспечения наибольшей эффективности от его использования, требует совершенно новых подходов к процессу обработки информации, новых кадров, обладающих новыми знаниями и умениями, и наличия новых правил и стандартов, регламентирующих

технологические процессы. При этом, последнее должно, несомненно, регламентироваться не региональными, не федеральными, а международными правилами и стандартами, для обеспечения глобальной интероперабельности не только интерфейсов доступа, но и схем и форматов представления цифрового контента.

Мы рассматриваем электронную библиотеку, как *информационную систему*, включающую все вышеперечисленные аспекты организации обработки, хранения и доступа к цифровому контенту.

Ниже обозначены некоторые моменты, которые с нашей точки зрения характеризуют различия между традиционными библиотеками, предоставляющими доступ к цифровому контенту, и полнофункциональными электронными библиотеками (информационными системами), которые не связаны старыми представлениями о хранении и управлении информационными ресурсами.

Хранение информации.

Основной канал пополнения фондов электронных библиотек — перевод в цифровую форму (сканирование, оцифровка и др. способы перевода в цифровую форму) традиционной печатной продукции. Получаемые в результате этого процесса цифровые объекты, не содержат никакой новой информации по сравнению со своими традиционными оригиналами. Функциональность созданной на их основе электронной библиотеки будет недостаточна, поскольку при работе с цифровыми объектами общество уже выработало определенный набор стереотипов, отсутствие которых вызывает дискомфорт. Так, например, одним из таких элементов является требование наличия взаимных ссылок между цифровыми объектами, проявляющихся, например, в виде гиперсвязей в пользовательских графических интерфейсах просмотра информации. Реализация взаимных ссылок в цифровых документах не представляет большой сложности, однако при этом электронный документ приобретает новое качество. Во-первых, электронный объект с реализованными связями уже не совсем соответствует своему печатному оригиналу. Это уже другой объект! И этот факт должны учитывать все юридические нормы. Во-вторых, внедренные в объект связи должны быть гарантировано актуальными. Никого, например, не интересуют гиперссылки, ссылающиеся на несуществующие документы. Так появляется отличное от традиционных библиотек требование обеспечения *ссылочной целостности* данных. Это очень жесткое требование, которое тяжело обеспечить даже в хорошо формализованных системах управления базами данных. Результат — новый цифровой

объект как самосогласованное хранилище цифрового контента, или база данных цифровых объектов.

С другой стороны, в электронной библиотеке объекты хранения могут содержать информацию, которая не имеет к объектам хранения традиционных библиотек вообще никакого отношения. Речь может идти:

- об электронных копиях элементов хранения традиционных архивов;
- об изображениях элементов хранения традиционных музеев;
- о видео-, аудио- информации, полученной разными способами, например, видеозапись доклада, сделанного на конференции;
- и т.п.

Отметим, что существование перечисленных объектов регламентируется нормами и правилами, которые не применяются в традиционных библиотеках. Последний тезис будет проиллюстрирован в разделе о каталогизации.

Каталогизация и метаданные

В традиционных библиотеках каталогизация реализует основную парадигму упорядочивания информации и обеспечения ее поиска по заранее определенным критериям. При каталогизации порождается новый вторичный информационный ресурс как массив стандартизованных описаний основных информационных объектов — элементов учета и хранения в традиционной библиотеке. Создание вторичного информационного ресурса регламентируется некоторыми правилами каталогизации, которые фиксируются в специальных нормативных документах. Первоначально конечным результатом каталогизации первичного объекта было создание каталожной карточки, в которой прописывались основные свойства первичного объекта в соответствии с общими правилами. Появление технологий машинного учета и баз данных привело к переводу вторичных информационных массивов традиционных библиотек в так называемые электронные каталоги, которые упростили доступ к вторичным ресурсам, но сохранили родовой отпечаток ущербности как следствие существующих правил каталогизации. Всякие попытки скрещивания технологий электронных и бумажных каталогов, выражаемых в создании необъяснимых в терминах логики здравого смысла монстров (см. RUSMARC и т.п.), не могут принести новых технологических элементов, использование которых могло бы оказаться полезным и продуктивным при каталогизации объектов не традиционных, а цифровых.

Здесь следует еще раз обратить внимание на то, что, с одной стороны, в электронных библиотеках имеют право существовать цифровые объекты, не имеющие аналогов в традиционных библиотеках,

и, как следствие, не попадающие под действующие правила каталогизации. С другой стороны, развитие пользовательских интерфейсов для доступа к информации требует возможности расширения списка атрибутивной информации, подлежащей вводу при каталогизации первичных объектов. Например, уже сегодня прослеживается потребность привязки контента к географическим координатам, которая полностью игнорируется действующими правилами и сложившейся практикой каталогизации. И наконец, в-третьих, существует необходимость описания не только информационного контента первичного объекта (заметим, что в основном именно информационный контент первичного объекта описывается сегодня в традиционных библиотеках), но и общего контекста существования первичного объекта с фиксацией всех событий в процессе его существования. Отдельно можно подчеркнуть, что при попытке описания контекста перестают работать все действующие библиотечные правила каталогизации.

В качестве иллюстрации можно привести попытку создать разумное описание цифрового объекта, который является изображением глиняной таблички, найденной в точке с координатами (x_1, y_1) в момент времени t_1 , помещенной в хранилище с координатами (x_2, y_2) в момент времени t_2 , сфотографированной в момент времени t_3 на выездной выставке в точке с координатами (x_3, y_3) . При этом на глиняной табличке описано событие, имеющее место быть в момент времени t_0 в точке (x_0, y_0) . Обязательным требованием к структурированному описанию первичного объекта должно быть требование возможности поиска по всем временным и пространственным характеристикам как контекста, так и контента [10-11].

Следует заметить, что невозможность создания нужного вторичного ресурса в рамках существующих правил ставит под сомнение необходимость последних в их сегодняшней форме. Кстати, другой возможной формы этих правил традиционные библиотеки не знают и в силу многих объективных причин знать не могут, чем заставляют усомниться в своей полезности при создании библиотек электронных.

Нельзя не отметить еще один момент. Переход к хранению и учету цифровых объектов делает неэффективной существующую парадигму каталогизации и создания массивов вторичных ресурсов. Действительно, необходимость во вторичных ресурсах объяснялась многие века разнородностью носителей первичной информации, необходимостью систематизации первичных ресурсов и организации поиска информации, хотя бы методом перебора каталожных карточек в алфавитном или систематическом каталогах. Переход к электронным каталогам существенно расширил возможности поиска в массивах вторичной

информации. Однако, сегодняшние технологии позволяют, во-первых, внедрять метаданные в первичные объекты, и, во-вторых, организовывать поиск по первичному ресурсу. При этом метаданные образуют с первичным объектом единое целое, а функциональность сервисов доступа к массивам информации не страдает. Таким образом, метаданные, описывающие контент, должны создаваться вместе с цифровым объектом и внедряться в него. Метаданные, описывающие контекст должны внедряться в первичный объект в течение всего времени его существования. При этом:

- Отпадает необходимость поддержки системы вторичных информационных ресурсов
- Отпадает необходимость в каталогизации первичного ресурса в современной ее форме
- Упрощается система управления информацией

Глобальная интероперабельность

Наиболее плодотворная идея обеспечения доступа к цифровому контенту включает в себя обязательное наличие интерфейсов доступа к этому контенту из Интернет. Глобальность сети требует обеспечения полного соответствия этих интерфейсов и предоставляемой информации определенным международным стандартам. При этом стандартизации должны подлежать:

- протоколы и интерфейсы доступа к данным;
- поисковые языки и интерфейсы;
- схемы и форматы представления данных;
- правила кодирования информации;
- правила контроля доступа к данным.

Игнорирование этих требований рано или поздно может привести к построению в рамках Интернет очередной вавилонской башни.

Система поиска информации

Как уже отмечалось, в традиционной библиотеке возможности поиска ограничивались поиском по алфавитному или систематическому каталогам для вторичных информационных ресурсов с прямой ссылкой (указания шифра хранения) на соответствующий первичный ресурс. Использование электронных каталогов расширило поисковые возможности, но сохранило основным типом поиска поиск по предопределенным поисковым атрибутам. Это *атрибутивный поиск*, именно этот тип поиска сегодня является основным в традиционных библиотеках, в том числе и в библиотеках цифровых объектов. При этом поиск производится только по массивам вторичной информации,

оставляя открытым вопрос соответствия последней первичным информационным ресурсам. Заметим, что внедрение атрибутивной информации в первичный цифровой объект, как упоминалось выше, могло бы вместе с отказом от привычной процедуры каталогизации существенно упростить технологии атрибутивного поиска.

Другой возможный тип поиска — *поиск по заданным шаблонам* имеет смысл только в массивах первичных цифровых ресурсов.

Наконец, поиск *с привлечением онтологий* является поиском более интеллектуальным, для его реализации требуется дополнительная информация — информация о предметной области, включающая определения терминов, сущностей и связей. Следует отметить, что представление этой дополнительной информации должно соответствовать глобальным договоренностям — международным стандартам, иначе, поиск с привлечением онтологий всегда будет ограничен текущей системой, а интоперабельность не будет реализована. К сожалению, этот тип поиска обсуждается при построении специальных информационных систем, но практически полностью игнорируется в практике построения электронных библиотек.

Контроль доступа

Для интоперабельных цифровых библиотек требуется реализация системы контроля доступа к информационным ресурсам, основанной на международных стандартах. Эта система контроля должна легко сшиваться с глобальными и региональными системами идентификации пользователей для определения их полномочий и ведения финансовых расчетов. Наверное, излишне утверждать, что подобные системы сегодня в электронных библиотеках не существуют.

Безопасность

Наконец, в электронных библиотеках должна обеспечиваться информационная безопасность для хранимой информации, для инфраструктуры и для всех ее пользователей. Можно указать следующие обязательные сервисы электронной библиотеки по обеспечению информационной безопасности [12]:

- Идентификация и аутентификация как основа концепции единого входа;
- Разграничение доступа;
- Протоколирование / аудит;
- Экранирование;
- Туннелирование;
- Шифрование;

- Контроль целостности;
- Контроль защищенности;
- Обнаружение отказов и оперативное восстановление.

Функционирование этих сервисов должно соответствовать международным стандартам и действующим рекомендациям. Наиболее важными из них являются рекомендации¹ серии X рабочей группы № 17 международного телекоммуникационного союза (ITU-T).

В заключение следует отметить, что, несмотря на уже достаточно длинную историю развития электронных библиотек в части хранилищ цифровых объектов и большой опыт создания таких хранилищ, сегодня существует существенная дистанция между парадигмой идеальной цифровой (электронной) библиотеки и реальной практикой создания и эксплуатации таких библиотек.

Библиография

Антопольский А. Б., Вигурский К. В. Концепция электронных библиотек [Электронный ресурс] // Электронные библиотеки: рос. науч. электронный журн. — 1999. — Т. 2, вып. 2. — URL: <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/1999/part2/antopol> (дата обращения: 04.05.2010).

Воройский Ф. С. Электронные и традиционные библиотеки — суть не одно и то же [Электронный ресурс] // Электронные библиотеки: рос. науч. электронный журн. — 2003. — Т. 6, вып. 5. — URL: <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2003/part5/voroisky> (дата обращения: 04.05.2010).

Глухов В. А., Голицына О.Л., Максимов Н.В. Электронные библиотеки. Организация, технология и средства доступа // Научные и технические библиотеки. — 2000. — № 10. — С. 1-7.

Ершова Т. В., Хохлов Ю. Е. Межведомственная программа «Российские электронные библиотеки» [Электронный ресурс] // Электронные библиотеки: рос. науч. электронный журн. — 1999. — Т. 2, вып. 2. — URL:

<http://www.elbib.ru/index.phtml?page=elbib/rus/journal/1999/part2/ershova> (дата обращения: 04.05.2010).

Когаловский М. Р., Новиков Б. А. Электронные библиотеки — новый класс информационных систем // Программирование. — 2000. — № 3. — С. 3-8.

¹ Большая часть которых стандартизована Международной организацией по стандартизации — ISO.

I. МАТЕРИАЛЫ VI МЕЖДУНАРОДНОЙ КОНФЕРЕНЦИИ ...

- Земсков А. И., Шрайберг Я. Л. Электронные библиотеки [Текст] : учеб. пособие для студентов ун-тов и вузов культуры и искусств и др. учеб. заведений. — 3-е изд., испр. и доп. — М. : ГПНТБ России, 2004. — 130 с.
- Зуев Д. С. Модели и принципы построения прототипа программной системы управления вузовской электронной библиотекой // Электронные библиотеки: перспективные методы и технологии, электронные коллекции: труды 11-й Всерос. науч. конф. «RCDL'2009» (Петрозаводск, Россия, 17–21 сентября 2009 г.). — Петрозаводск: ИПМИ КарНЦ РАН, 2009. — С. 203–209.
- Жижимов О.Л., Мазов Н.А. Принципы построения распределенных информационных систем на основе протокола Z39.50. — ОИГГМ СО РАН, Новосибирск: ИВТ СО РАН. — 2004. — 361 с.
- Жижимов О.Л. Мазов Н.А. Об интеграции библиотечно-информационных и геоинформационных технологий // Вычислительные технологии. Совместный выпуск. Вестник КазНУ им. Аль-Фараби. Сер. математика, механика, информатика. — 2008. — Т. 13. — № Ч. II. — С. 97-101.
- Шокин Ю.И., Федотов А.М., Жижимов О.Л. Проблемы интеграции информационных ресурсов // Международная конференция «Математические и информационные технологии» MIT-2009 (VIII конференция «Вычислительные и информационные технологии в науке, технике и образовании»), — Копаоник, Сербия, 27-31 августа 2009 года; Будва, Черногория, 31 августа — 5 сентября 2009 года.
- Жижимов О.Л. Мазов Н.А. Об использовании географических координат при поиске библиографической информации // Научные и технические библиотеки. — 2009. — № 1. — С. 54-60.
- Федотов А.М. Информационная безопасность в корпоративной сети // Проблемы безопасности и чрезвычайных ситуаций. М.: ВИНТИ, 2008, N 2, с.88-101.