

Формирование полнотекстовых коллекций в Интернете: опыт библиотек России

Анализ результатов обследования сайтов российских библиотек; отличительные особенности полнотекстовых собраний; типология коллекций. Перспективы цифровых библиотек

В Российской Федерации практика формирования в интернете полнотекстовых электронных коллекций берет свое начало с середины 1990-х гг. Начиная с этого времени бесплатные цифровые коллекции в российском сегменте Сети создавались и создаются прежде всего энтузиастами-любителями. Несмотря на дилетантизм создателей, именно за счет этих проектов интернет наполнился десятками интереснейших собраний, отражающих как всемирную классику, так и современные произведения практически всех направлений и жанров.

Традиционные российские библиотеки, в очередной раз подтвердив реноме одного из наиболее консервативных общественных институтов, начали работу по формированию цифровых коллекций с опозданием более чем на десятилетие – на рубеже 2004–2005 гг.

К настоящему времени в интернете насчитывается около сотни коллекций самого разного масштаба, созданных отечественными библиотеками. Задача данной работы состоит в обобщении накопленного за несколько лет опыта, определении тенденций и перспектив этого направления деятельности в российских книгохранилищах; ее источниковой базой служат материалы, собранные автором при непосредственном изучении библиотечных интернет-сайтов.

Обследование сайтов российских библиотек свидетельствует, что работой по формированию полнотекстовых собраний занимаются в той или иной мере библиотеки практически всех типов. Федеральные, центральные отраслевые, республиканские, краевые и областные универсальные, вузовские и даже муниципальные библиотеки уже создают или планируют в ближайшей перспективе создание полнотекстовых собраний, источником для которых в подавляющем большинстве являются издания из их фондов.

Отличительная особенность полнотекстовых собраний, создаваемых библиотеками, – их юридическая чистота. Это резко отличает проекты библиотек от уже упоминавшихся полнотекстовых коллекций, создаваемых в нашей стране в рамках частных инициатив отдельными гражданами или сообществами. Библиотеки крайне редко включают в состав своих цифровых собраний произведения, попадающие под действие Части 4 Гражданского кодекса РФ, регламентирующего использование интеллектуальной собственности. Там же, где такие произведения включаются, библиотеки заключают с правообладателями соответствующие договоры.

Типология формируемых коллекций прослеживается достаточно отчетливо. Наиболее ярко она выражена на региональном уровне – в республиканских, краевых и областных библиотеках, которые целенаправленно и последовательно формируют коллекции национального и краеведческого содержания. К числу таких коллекций относятся: Электронная библиотека республики Карелия (<http://elibrary.karelia.ru>), Электронная библиотека Национальной библиотеки Республики Саха (Якутия) (<http://www.nlib.sakha.ru/elib>), «Тысячелетняя Казань» (<http://www.kitaphane.ru/ebooks/elib.shtml>), «Книжные редкости Удмуртии» (http://unatlib.org.ru/index.php?option=com_content&view=article&id=394&Itemid=69), «Белгородчина в книжных памятниках» (<http://www.bgunb.ru/bgunb/publish/book/kh>), «Псковiana» (<http://78.36.232.16/catalog/kray>), Донская электронная библиотека (<http://www.dspl.ru/digit.asp>), Краеведческая электронная библиотека Томской области (<http://tomskhistory.lib.tomsk.ru/page.php?id=6>) и др.

Собрания, формируемые массовыми, прежде всего городскими библиотеками, также имеют ярко выраженную краеведческую направленность, например «Коллекция Югры» (<http://slib.admsurgut.ru/collection/ugra.htm>), размещенная на сайте Сургутской центральной городской

библиотеки им. Пушкина. В состав подобных коллекций входят, как правило, местные издания – книги и периодика (адресные книги, статистические справочники, губернские и епархиальные ведомости и т.д.), а также продукция центральных издательств, содержательно связанная с регионом.

Весьма отчетливо просматриваются и отличительные черты цифровых коллекций, размещаемых на сайтах вузовских библиотек, главная задача которых заключается в информационной поддержке учебного процесса. Именно поэтому в состав вузовских коллекций включается в основном учебная и учебно-методическая литература, при этом авторами размещаемых источников зачастую выступают преподаватели данного вуза. Среди наиболее интересных университетских собраний – цифровые коллекции библиотек Иркутского государственного университета (<http://elib.library.isu.ru>), Горно-Алтайского государственного университета (<http://e-lib.gasu.ru>), Хабаровской государственной академии экономики и права (<http://www.ael.ru/tabid/114/Default.aspx>).

Наибольшим типологическим разнообразием отличаются коллекции федеральных библиотек. Их состав должен в значительной степени диктоваться профилем и составом фондов каждой конкретной библиотеки, однако сегодня их наполнение носит во многом случайный характер. Так РНБ разместила на сайте коллекции старых открыток, посвященных С.-Петербургу и Сестрорецку (<http://leb.nlr.ru>); РГБ параллельно реализует два схожих проекта универсальной направленности, которые в перспективе имеют тенденцию к слиянию – «Открытая русская электронная библиотека» (<http://orel.rsl.ru>) и «Электронная Ленинка» (http://aleph.rsl.ru/F/?func=file&file_name=find-b&local_base=xelb); Государственная публичная историческая библиотека ограничилась публикацией на своем сайте небольшой коллекции предвыборных плакатов и листовок «Выборы в России» (<http://www.shpl.ru/plakat/#>), зато передала в состав Единой коллекции цифровых образовательных ресурсов (<http://school-collection.edu.ru/catalog/rubr/2027e92c-0900-4665-b74f-3b09d45e6554>) большое собрание дореволюционных изданий по всемирной и российской истории.

Особняком среди этих коллекций библиотек России стоит созданная РГБ «Электронная библиотека диссертаций» (<http://diss.rsl.ru>). Это собрание задумывалось и реализуется как коммерческий проект. Благодаря организации в других библиотеках виртуальных читальных залов (ВЧЗ), электронные версии диссертаций из фондов РГБ доступны по подписке для значительного круга читателей. Практика создания ВЧЗ формально позволяет частично обойти существующие правовые ограничения, что, однако, не снимает вопроса о правомерности осуществления библиотеками фактической торговли диссертациями, представляющими интеллектуальную собственность их авторов.

В центральных, региональных и федеральных библиотеках проекты по созданию цифровых коллекций нередко финансируются в рамках федеральных и региональных программ по сохранению и консервации фондов. Такой подход можно признать идеальным, поскольку создание и размещение на сайтах электронных копий уникальных изданий в максимальной степени обеспечивает их сохранность и одновременно открывает неограниченный доступ к редким источникам. Финансирование вузовских цифровых библиотек осуществляется за счет бюджетов учебных заведений, которые справедливо рассматривают эти коллекции как эффективный инструмент информационного обеспечения образовательного процесса.

Практически все показатели электронных коллекций, размещенных на сайтах библиотек, отличаются большим разнообразием, что еще раз подчеркивает их опытный характер. Объем информационных массивов колеблется от полутора десятков названий книг (например, на сайтах Астраханской (<http://aonb.astrakhan.net/html/astakhanica.html>) и Орловской (<http://www.buninlib.orel.ru/elkol.htm>) областных библиотек) до более 4 700 многостраничных номеров местных газет на сайте Иркутской ОУНБ (<http://irabis.irklib.ru:8080/gaz>) и более чем 6 000 документов учебного назначения на сайте научной библиотеки Иркутского государственного университета (<http://elib.library.isu.ru>).

Одна из ключевых характеристик электронных коллекций – формат представленных в них текстовых документов. За свою недолгую историю электронные собрания библиотек России прошли значительный эволюционный путь в этой области. Так, на начальной стадии представленные в них страницы документов нередко размещались в виде обычных иллюстраций в форматах TIFF, GIF или JPEG. Эти форматы, использовавшиеся, в частности, в начальный период в коллекциях «Память Томска» (<http://elib.tomsk.ru>), не позволяли производить поиск по тексту, «листать» страницы и к тому же далеко не всегда обеспечивали достойное качество даже при немалых размерах размещаемых файлов. Параллельно с этим применялись форматы DOC и HTML. Перевод исходных документов в эти форматы требовал немалых временных затрат, включая ручное форматирование, но не обеспечивал исходного вида источника. Типичные примеры

использования этих форматов – полнотекстовая коллекция Камчатской областной библиотеки (<http://library.iks.ru>) и начальная часть Татарской электронной библиотеки (<http://www.kitaphane.ru/ebooks/tatelib.shtml>).

В последние несколько лет прослеживается повсеместная тенденция перехода на форматы PDF и DJVU, которые сегодня признаны оптимальными для создания электронных библиотек, поскольку обеспечивают аутентичный вид исходного документа, листание страниц, увеличение текста, а при желании создателей собрания и поиск по полному тексту источника.

Важнейший параметр любого электронного собрания – способ его программной реализации. Используемым программным обеспечением определяются такие важнейшие характеристики, как поиск и навигация среди документов коллекции, способность к масштабированию, интерфейс электронного собрания. Обследование коллекций позволило установить, что к настоящему времени российские библиотеки практически не применяют профессиональное программное обеспечение, разработанное специально для создания электронных библиотек. Исключения составляют лишь РГБ, использующая «Алеф» в «Электронной Ленинке» (http://aleph.rsl.ru/F/?func=file&file_name=find-b&local_base=xelb), уже упоминавшаяся Иркутская ОУНБ, частично использующая возможности системы ИРБИС для навигации в коллекции местной дореволюционной печати, Челябинская ОУНБ, также применяющая ИРБИС в «Уральской электронной библиотеке» (http://unilib.chel.su:6005/urelib/main_page.htm), и научная библиотека Волгоградского государственного университета, которая использует «Гринстоун» в формируемой ею электронной библиотеке авторефератов (<http://lib.volsu.ru/gsd/cgi-bin/library.exe>).

Некоторые библиотеки пошли по пути разработки собственных программных оболочек. К их числу относятся национальные библиотеки Карелии (<http://elibrary.karelia.ru>), Удмуртии (http://umatlib.org.ru/index.php?option=com_content&view=article&id=394&Itemid=69) и Ямало-Ненецкого автономного округа (http://libraries-yanao.ru/index.php?option=com_content&task=view&id=318&Itemid=75). Такой подход вряд ли может быть признан оправданным, поскольку при высоких интеллектуальных, финансовых и временных затратах эффективность работы с коллекциями во всех случаях весьма невысока. Ряд библиотек, видимо, по неопытности, приспособил для создания электронных коллекций программное обеспечение, предназначенное для ведения блогов или форумов. В результате в коллекции появляется возможность оставлять отзывы о произведении, но полностью отсутствует какой-либо поиск и навигация. Ряд библиотек, среди которых Астраханская областная, Карельская и Якутская национальные библиотеки, требуют от виртуальных читателей прохождения бесплатной регистрации с целью получения статистических сведений о посетителях электронных хранилищ.

Большинство же российских библиотек до настоящего времени ограничиваются выкладыванием на сайт полных текстов, делая ссылки на них с одной из страниц сайта. Поскольку основное число коллекций еще относительно невелико, такой подход пока не вызывает проблем при поиске источников.

На основе проведенного анализа отчетливо вырисовываются перспективы формирования российскими библиотеками сетевых полнотекстовых собраний. Число таких коллекций в перспективе будет неуклонно увеличиваться по мере того, как интернет-сайт будет восприниматься библиотекарями в качестве максимально эффективного инструмента раскрытия фонда и предоставления различных информационных сервисов.

Развитие цифровых коллекций будет продолжаться в уже обозначенном русле – региональные библиотеки расширят формирование национально-краеведческих коллекций, вузовские и академические библиотеки сосредоточатся на профильных собраниях. В первом случае наполнение будет происходить в рамках программ по оцифровке местных раритетных изданий и книжных памятников. Их библиографические своды уже составлены во многих регионах (см. <http://kp.rsl.ru/index.php?f=405>), и следующим этапом работы станет процесс оцифровки этих изданий. Темпы, которыми ведется оцифровка, дают основания полагать, что в ближайшие 3–4 года будут оцифрованы и выложены в Сеть все хранящиеся в региональных библиотеках и сохранившие свое значение редкие и ценные издания, включая местную периодику. Исключение составят лишь библиотеки нескольких регионов, которые традиционно имеют показатели гораздо ниже среднего уровня по стране.

Работа по оцифровке имеет для региональных российских библиотек временный характер и продолжится до

тех пор, пока не будут оцифрованы все дореволюционные и довоенные издания, на которые не распространяется действие законов по охране прав на интеллектуальную собственность. После того, как эта работа завершится, число региональных собраний и их основное наполнение останутся стабильными. Добавления материалов в коллекции можно ожидать лишь за счет оцифровки мультимедийных массивов – имеющих историческую ценность фотографий, аудио- и видеозаписей.

В отличие от библиотек республик, краев и областей, библиотеки вузов и научных учреждений в перспективе будут только наращивать работу по формированию полнотекстовых коллекций. Обладая возможностью заключать с авторами публикуемых произведений соответствующие договоры и движимые требованием более эффективно поддерживать учебный процесс, включая и его дистанционные формы, библиотеки этого типа будут все шире разворачивать работу по формированию собраний учебного и научного профиля. Нюансы заключаются лишь в темпах этого процесса и способах обеспечения читательского доступа. Как и сегодня, некоторые университетские библиотеки будут предоставлять доступ к формируемому ими информационному массиву всем желающим, а некоторые – лишь студентам, преподавателям и сотрудникам данного вуза по выдаваемым им паролям.

Главные технические перспективы цифровых библиотек связаны со все более широким применением специализированного программного обеспечения, которое позволит значительно повысить качественный уровень коллекций. За счет применения специального ПО будут обеспечены быстрый поиск источников, включая поиск по полным текстам, простота навигации в документном массиве вне зависимости от его объема, а также легкое масштабирование собрания и включение в их состав новых изданий. Повышение качественных характеристик окончательно закрепит статус цифровых коллекций как «типовой» составной части фонда библиотеки.

Обобщая сказанное, можно сделать вывод, что текущий этап создания и размещения в интернете цифровых коллекций – пилотный для российских библиотек. В ходе него каждая библиотека опытным путем вырабатывает методику и выстраивает технологию этого процесса. Многие электронные собрания с примитивными системами поиска и навигации напоминают первые библиотеки человеческой цивилизации, имевшие простейшую структуру и потому не нуждавшиеся в унификации представленных в них документов и справочном аппарате.

Пока лишь немногие библиотечные коллекции могут быть по праву отнесены к серьезным информационным ресурсам. Однако их положительный опыт тем более ценен, поскольку указывает верный путь другим библиотекам, по сути предлагая им готовые концептуальные и технические решения, которые могут быть легко тиражированы.

Российские библиотеки постепенно набираются опыта, оттачивают технологические цепочки, более осмысленно подходят к отбору источников. На смену фрагментарным коллекциям со случайно подобранными источниками приходят более фундаментальные собрания, отражающие целостные документные пласты. Все это дает основания полагать, что через 2–3 года библиотечные цифровые коллекции смогут стать заметным явлением в общероссийском масштабе, составляя весомую долю содержательного наполнения российского сегмента Сети, предназначенного для самых взыскательных читателей.