

Российские электронные библиотеки: состояние и проблемы

Словосочетание "электронная библиотека" (ЭБ) за последние десять лет становится все более популярным и повсеместно употребляемым. Это имеет отношение не только к России, но и ко многим другим странам, широко использующим информационные технологии. Неуклонно растет число систем от предельно простых до весьма сложных, в названии которых присутствует упомянутое словосочетание. Ведутся научные исследования и опытные разработки, создаются разнообразные ЭБ, открытые для широкого сетевого доступа пользователей. Спектр участников этих работ весьма широк: от отдельных энтузиастов до объединений университетов, исследовательских центров и крупных фирм. Во многих странах были открыты соответствующие национальные программы (США, Германия, Япония и др.), бюджет каждой из которых колеблется от десятков до сотен млн. долларов.

За высокой популярностью этого словосочетания стоит не только и не столько дань моде, сколько попытка охарактеризовать новый феномен – возникновение принципиально нового класса систем, призванных аккумулировать и распространять информацию в электронной форме. А большой интерес к самим системам данного класса объясняется актуальными потребностями общества и наличием развивающихся возможностей по их удовлетворению. В связи с этим можно сформулировать основные задачи, стоящие перед электронными библиотеками:

- сделать информацию более доступной;
- содействовать сохранению научного и культурного наследия;
- повысить эффективность работы и обучения.

В России первые работы по ЭБ относятся к середине 90-х гг. С самого начала основную роль в этих работах играли энтузиасты-любители, что естественно сопровождалось стихийностью и невысоким качеством. Государственные интересы и поддержка в этой области проявлены крайне слабо. Попытка открытия российской целевой программы федерального уровня (1998 г.) не увенчалась успехом. Открытие двух-трех отраслевых программ заметных результатов не дало. По нескольким успешно начатым и общественно значимым проектам государственная поддержка прекращена, что обусловило невозможность их развития или эксплуатации. Проблематика ЭБ в ФЦП «Электронная Россия» фактически не представлена. На этом фоне следует отметить положительную роль, которую играют Российский Фонд Фундаментальных исследований и Российский Гуманитарный Научный Фонд, поддерживающие в меру своих возможностей ряд проектов ЭБ.

Вместе с тем растет число ЭБ, повышается их уровень, расширяются функциональные возможности этих библиотек, ведутся исследования, сведения о которых в течение нескольких лет сообщаются на Всероссийской конференции «Электронные библиотеки».

С целью ответа на вопросы о составе русскоязычных ЭБ в НТЦ «Информрегистр» в настоящее время проводится соответствующий анализ. Далее представлены некоторые результаты, полученные на первом этапе этой работы.

Систем, ресурсов, которые имеют в названии слова «электронная библиотека» свыше 1 тысячи. Однако в этом числе: интернет–магазины, электронные каталоги обычных библиотек, сетевые репрезентации обычных библиотек, каталоги сетевых ресурсов, библиографические БД и некоторые другие системы. Эти системы и ресурсы очевидно ЭБ не являются.

Далее, после учета «зеркал», изъятия неработающих (не отвечает ссылка) и пустых систем, а также некоторых других, остается менее половины.

Но при этом оказываются неучтенными системы и ресурсы, обозначаемые как:

- библиотеки с определениями «виртуальная», «цифровая», «сетевая», «интернет»;

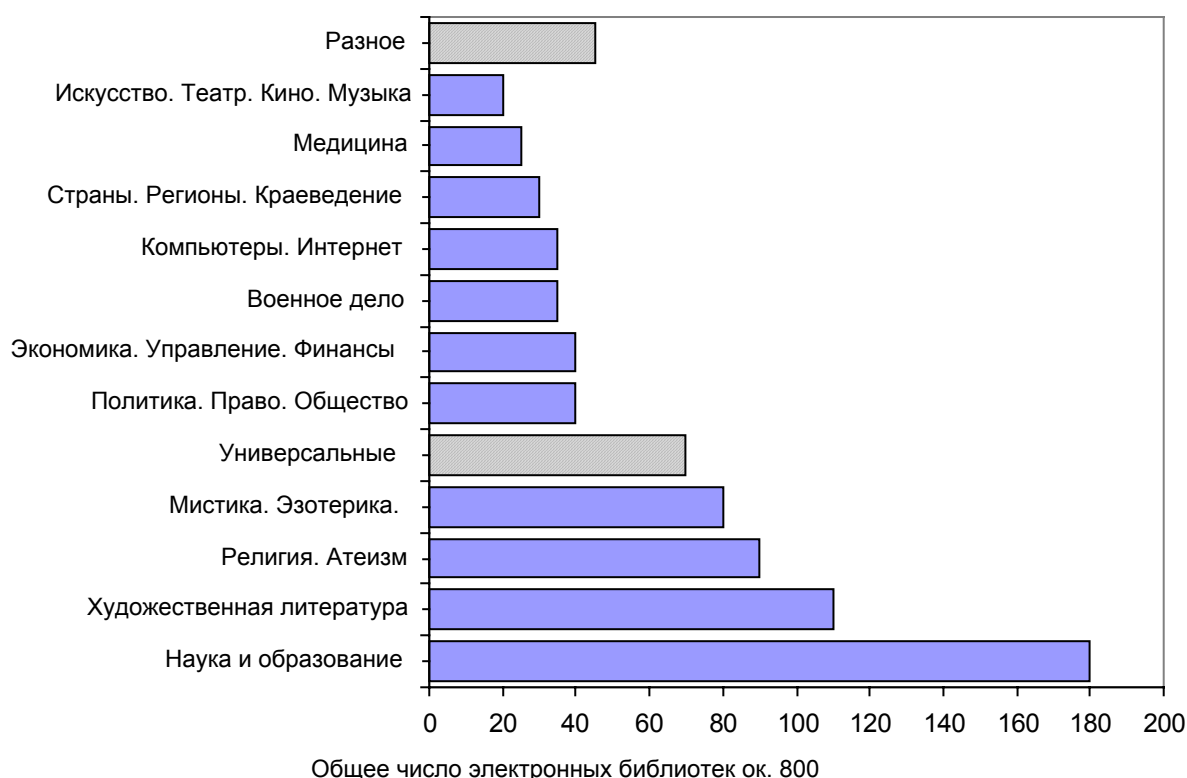
- собрания, коллекции, архивы, библиотеки без уточняющих определений;
- системы, имеющие только собственное имя типа «abc.ru».

Важным показателем, как это не странно, является размер информационного фонда. Вряд ли к классу ЭБ следует относить ресурсы, как бы они не назывались, содержащие 2, 3, 4 документа, а таких «электронных библиотек» насчитывается несколько десятков.

После учета всех этих вариантов получается около 800 систем, которые могут рассматриваться как ЭБ. Число это дает лишь приблизительное значение, поскольку данная область очень динамична: ЭБ создаются и бесследно исчезают, модифицируются, меняют названия и адреса. Среднее время существования ЭБ можно оценить в 3–4 года.

Следующим важным моментом является распределение этого количества ЭБ по тематическим направлениям. Результат представлен на диаграмме.

Русскоязычные электронные библиотеки



Не все русскоязычные ЭБ организуются в России. Около 40 из них ведутся за рубежом, в том числе в Украине – 14, в США – 7, в Белоруссии – 5. Среди этих ЭБ есть весьма популярные, например «Электронная библиотека Александра Белоусенко» (США, <http://www.belousenko.com/>), «Im Werden» (Германия, <http://imwerden.de/>), «Литбиблиотека Бориса Бердичевского» (Израиль, <http://www.borisba.com/litlib/index.html>).

Не менее половины из указанного числа библиотек создаются отдельными лицами или группами энтузиастов. Этим ЭБ свойственны стихийность и субъективность формирования, невысокое качество, ненадежность. К числу таких библиотек следует добавить и целый ряд систем, которые хотя и представляют какие-то организации и учреждения, фактически ведутся на энтузиазме со всеми вытекающими последствиями.

Важным параметром оценки любой ЭБ является качество представления информации. В целом анализ показывает здесь неутешительную картину. Значительное число ЭБ содержат полураспознанные тексты с большим числом ошибок, многие тексты

искажены, отсутствуют описания воспроизводимых произведений и т.д. В 20% библиотек качество информации не выдерживает никакой критики. Системный отбор информации ведется не более чем в 30% библиотек.

Сходную общую оценку можно дать и по качеству обслуживания в ЭБ, в частности, по предоставляемым пользователям функциональным возможностям. Многие ЭБ не имеют системы навигации, невозможно узнать, что содержится в этих библиотеках. Далеко не всегда имеются даже элементарные указатели авторов и произведений. Впрочем, даже когда они есть, проку от них бывает мало. Например, в одной из библиотек в указателе произведений, сгруппированных по авторам, имя писателя Антуана де Сент-Экзюпери встречается четырежды на разные буквы (на «А» – Антуана де Сент-Экзюпери, на «Д» – де Сент-Экзюпери, на «С» – Сент-Экзюпери и на «Э» – Экзюпери) и под каждым вариантом имени стоят разные произведения. Подавляющее большинство библиотек не имеют полнотекстового поиска.

Даже этот предельно краткий обзор показывает серьезные проблемы в области ЭБ. Прежде всего, в настоящее время отсутствует ясное понимание того, что такое электронная библиотека, какие информационные системы следует относить к данному классу, каковы требования и критерии оценок, предъявляемые к ним. Необходимость рассмотрения понятия "электронная библиотека" обусловлена несколькими причинами:

- отсутствием обоснованного и общепринятого определения;
- широким употреблением этого понятия в научной, технической и популярной литературе;
- наделение этого понятия разными авторами различными, иногда противоречивыми значениями.

Совершенно очевидно, что для эффективного исследования, проектирования и эксплуатации некоторого класса систем, он, этот класс, должен быть более или менее точно определен.

Впервые понятие "электронная библиотека" появилось в зарубежных публикациях в середине 80-х гг. С тех пор было несколько десятков определений, двенадцать из которых рассмотрены в [1]. Учитывая накопленный опыт, в указанной работе используется в качестве основного определение, ранее предложенное в [2].

Электронная библиотека — информационная система, позволяющая надежно сохранять и эффективно использовать разнообразные коллекции электронных документов (текстовых, изобразительных, звуковых, видео и др.), локализованных в самой системе, а также доступных ей через телекоммуникационные сети.

Остановимся на некоторых моментах этого определения.

Коллекция понимается достаточно неформально как некоторое собрание, совокупность, множество объектов определенной природы, в данном случае электронных документов.

Документ понимается в интерпретации стандарта ГОСТ 7-60 [3] как зафиксированная на материальном носителе информация с реквизитами, позволяющими ее идентифицировать. Далее, электронный документ, согласно ГОСТ 7.83 [4], рассматривается как документ, представленный в машиночитаемой форме, для использования которого необходимы средства вычислительной техники.

Уточнение о локализации документов в самой системе или о доступности их через телекоммуникационные сети позволяет распространять данное определение как на системы с централизованным, так и с распределенным информационным фондом и говорить о централизованных и распределенных ЭБ.

Следует обратить особое внимание на то, что не любая коллекция электронных документов может рассматриваться как ЭБ, точно так же как не любое собрание книг рассматривается как традиционная библиотека (книжный склад, книжный магазин и т.п.). Необходимым условием квалификации коллекции как ЭБ является ее нацеленность на сохранение и определенные виды эффективного использования информации. Отсюда

следуют определенные требования к системам такого класса. С другой стороны, эти требования конкретизируются на основе определения целевой аудитории – основной части пользователей, на которую рассчитана ЭБ.

Основное требование, предъявляемое к отбору информации – комплектованию информационного фонда, – системность, что означает:

- наличие профиля;
- взаимосвязанность фонда;
- определенность вида(ов) отбираемых документов;
- соответствие научному, техническому, культурному и социальному уровню общества (отрасли, тематического направления и т.п.).

Цель информационного фонда – обеспечить пользователя всей необходимой по указанному профилю информацией и минимизировать его потребность в других источниках.

Представление информации в ЭБ связано с двумя основными вопросами:

- точностью воспроизведения исходных материалов;
- идентификацией информации.

В идеале представляется необходимым аутентичное воспроизведение исходного документа, например, книги. Это означает, что в ЭБ должны быть воспроизведены все без исключения части книги, ее структура, орфография и пунктуация, распределение материала по страницам, иллюстративный ряд, шрифтовое оформление и другие особенности издания. Только в этом случае конечный пользователь может доверять источнику и избежать необходимости перепроверять его по печатному оригиналу.

Решение этой задачи вполне достигается путем представления книги в виде набора графических образов всех ее страниц. Однако эта форма представления информации не обеспечивает все необходимые пользовательские функциональные возможности (проведения поисков в полных текстах документов, экспорт произвольных фрагментов текста и т.п.), для этого необходима алфавитно-цифровая форма текстовой информации. Создание такой электронной формы требует более сложной технологии (включая распознавание, вычитку, разметку и др. операции) и заметно больших финансовых вложений.

Существенным требованием является идентификация электронных документов в ЭБ. Это означает, что каждый документ и, шире, некоторые совокупности документов должны быть максимально точно и унифицированным образом описаны. Эти описания необходимы для:

- информирования пользователей о материале, содержащимся в ЭБ, с каким конкретным документом он работает или может работать;
- навигации в ЭБ;
- поисков документов по их формальным признакам (атрибутный поиск);
- указания на источник при цитировании и экспорте информации из ЭБ.

Поскольку информационный фонд ЭБ составляют документы, в первую очередь, текстовые, то в качестве их идентификаторов целесообразно использовать библиографические описания. Существуют и другие виды описаний. Какие системы идентификации, в конечном счете, установятся в ЭБ, покажет время.

Отсутствие описаний приводит к весьма негативным последствиям и создает множество проблем, не всегда осознаваемых создателями ЭБ. Пожалуй, самым неприятным из этих последствий является то, что пользователь, получая доступ к определенному документу, не знает и не может узнать, с чем он работает. Это, в конечном счете, приводит к тому, что теряется доверие пользователя к данной ЭБ.

ЭБ должна предоставлять конечному пользователю некоторый набор функциональных возможностей для ознакомления с информацией и работы с ней. Состав таких возможностей в каждом конкретном случае определяется в зависимости от целей и назначения ЭБ. Однако представляется целесообразным определить круг тех

пользовательских задач, решение которых было бы весьма желательно для большинства ЭБ. В число таких задач могут входить:

- задачи навигации, обеспечивающие ознакомление со структурой и составом ЭБ, а также доступ к отдельным структурным элементам;
- атрибутный поиск, позволяющий находить произведения и издания по заданным значениям их формальных признаков;
- просмотр информации;
- экспорт информации.

Было бы желательно ввести, особенно для ЭБ большого объема (тысячи и более информационных объектов), полнотекстовый поиск.

Следует отметить, что в существующих ЭБ далеко не всегда реализуются перечисленные функции. Например, в целом ряде систем, имеющих название "электронная библиотека", электронные документы представлены только в архивном формате, что достаточно затрудняет их оперативный просмотр.

В представленных выше результатах анализа понятие «электронная библиотека» трактовалось достаточно широко, и число соответствующих систем и ресурсов составило около 800. Применение к этому множеству перечисленных требований, с целью выделить класс «настоящих» ЭБ, сокращает его объем примерно в десять раз. Оставшиеся системы целесообразно рассматривать как полнотекстовые информационные ресурсы.

Выше были кратко рассмотрены некоторые основные вопросы, связанные с понятием "электронная библиотека". В целом, круг вопросов, требующих ответа, значительно шире. В него входят, в частности, вопросы метаданных, вопросы представления и обработки информации в распределенных ЭБ, вопросы типологии и классификации систем данного класса. Чрезвычайно актуальными остаются правовые вопросы, связанные с формированием и эксплуатацией ЭБ.

Представляется, что одним из наиболее важных направлений исследований в области ЭБ является работа по формированию научно обоснованного понятийного аппарата. Достаточно очевидны и последующие этапы работ в этой области – поиск наиболее эффективных и унифицированных научных, технических и организационных решений по проектированию и ведению ЭБ, а также разработка нормативно-технических документов (рекомендаций, стандартов, классификаторов и др.), фиксирующих эти решения. При этом чрезвычайно важно, чтобы получаемые решения и соответствующие нормативные документы основывались не только на теоретических исследованиях, но и учитывали результаты реальной практической деятельности, реальный практический опыт в области ЭБ.

Литература:

1. Вигурский К.В., Горный Е.А. Развитие электронных библиотек: мировой и российский опыт, проблемы, перспективы // Интернет и российское общество / Под ред. И. Семенова; Моск. Центр Карнеги. – М.: Гендальф, 2002. – С. 158–188. – (<http://pubs.carnegie.ru/books/2002/08is/>).

2. Антопольский А.Б., Вигурский К.В. Концепция электронных библиотек // Электронные библиотеки. — 1999. — Т.2. — Вып.2. — (<http://www.iis.ru/el-bib/1999/199902/antopol/antopol.ru.html>).

3. ГОСТ 7.60–2003. СИБИБД. Издания. Основные виды. Термины и определения.

1. ГОСТ 7.83–2001. СИБИБД. Электронные издания. Основные виды и выходные сведения.