

Новый взгляд на проект электронной библиотеки диссертаций

© Лаврёнова О.А.

Российская государственная библиотека

lavr@rsl.ru

Аннотация

В докладе описываются принципиальные основы, предлагаемые перспективы развития электронной библиотеки диссертаций в РГБ и результаты последних лет. Новая версия библиотеки основана на использовании полных текстов в формате PDF и метаданных для работ и процессов их защиты в форме XML-файлов на базе программного обеспечения ИСИР.

Введение

Проектам электронной библиотеки диссертаций (ЭБД) РГБ посвятила достаточно публикаций [1-5], за 4 с лишним года прошедшие с момента разработки ее концепции, но ни разу не позволила себе «замахнуться» на публичное описание того «идеального» образа данной электронной библиотеки, к которому стремится шаг за шагом. Материальное воплощение этого образа, или модели, что звучит более определенно в контексте данной области деятельности, на каждом этапе ведения проекта выглядит то более, то менее реальным. Обстоятельства различного рода заставляют разработчиков ЭБД развивать проекты то в одном, то в другом направлении, в зависимости от того, какое из них оказывается более перспективным. При этом ранее реализованные фрагменты модели продолжают свое развитие в соответствии с предлагаемыми обстоятельствами. Эти обстоятельства могут иметь правовой, финансовый, личностный, организационный, технологический характер.

1 Исходные положения

1.1. В интересах развития науки к полным текстам таких публикаций, как авторефераты и диссертации, библиотеки, на которые возложено депозитарное хранение (РГБ и ГЦНМБ), должны

обеспечить доступность широкому кругу пользователей, но, естественно, с соблюдением существующего законодательства.

1.2. Диссертации и авторефераты – не только публикации, но и официальные документы, демонстрирующие и подтверждающие основания для присвоения ученых званий их авторам. В отличие от зарубежной практики, когда сам университет решает, достоин ли претендент ученой степени, у нас принято широкое распространение информации о содержании работы до защиты, чтобы дать возможность ученым в различных учреждениях выразить мнение о ней. Всем авторам, руководителям диссертаций, оппонентам требуется на разных этапах подготовки работы убедиться, что она не повторяет содержания других исследований. Кроме того, ВАК должна получить достаточно информации для принятия своего решения. В связи с этим необходимость максимальной доступности полных текстов работ заложена в самом факте появления и существования диссертаций и авторефератов. *Не должны ли эти официальные документы в таком случае публиковаться для всех в открытом удаленном доступе?* Разумеется, речь не идет о диссертациях с грифом «Для служебного пользования».

1.3. Основное содержание и результаты работ должны быть обязательно опубликованы в специальных изданиях до защиты, так что диссертация как таковая, по определению, не отличается принципиальной научной новизной. Этим свойством должно обладать само исследование, как и свойствами актуальности, практической значимости и т.д., что и доказывается в диссертации. *Можно утверждать, что в интересах общества и ВАК стимулировать представление этого доказательства как можно большему кругу специалистов?*

1.4. Диссертация защищается для того, чтобы получить ученую степень, т.е. признание своей научной состоятельности в глазах научной общественности, придать известность своим работам, расширить круг научных контактов, занять более высокую должность, несколько повысить свое благосостояние. *Почему бы ни расширить эти возможности максимально, используя публикацию работ в открытом удаленном доступе?*

1.5. Нелегитимное заимствование идей из текста диссертации бесполезно, ибо они уже опубликованы. Хотя соответствующие издания могут быть доступны далеко не всем, нуждающийся найдет все-таки. Заимствование фрагментов текста без ссылок возможно и, к сожалению, практикуется, причем наличие невероятно ограниченного количества хранилищ диссертаций (по месту защиты, в РГБ и ЦНМБ, а также во ВНИИЦ в виде микрокопий) и небольшого тиража авторефератов существенно затрудняет обнаружение фактов плагиата. Затраты вузов и других научных учреждений, вынужденных регулярно направлять в Москву своих преподавателей, научных сотрудников, аспирантов для изучения фондов диссертаций в РГБ и ЦНМБ, огромны. *Разве не очевидно, что размещение полных текстов диссертаций и авторефератов в открытом теледоступе сделает плагиат не столь легким, сколь бессмысленным, ибо всем желающим эта «дурь видна» будет (особенно при использовании соответствующих программных средств)?*

1.6. *Какие причины противодействия авторов открытой публикации в теледоступе можно признать действительно уважительными?* Что можно придумать, кроме подготовки к изданию монографии на базе текста диссертации? Не пора ли настолько повысить требования к качеству диссертаций, чтобы всем авторам ими можно было гордиться?

1.7. Подавляющее большинство авторов оформляют тексты диссертаций и авторефератов путем ввода в компьютер своими руками или с использованием труда операторов. *Не обидно ли, что большая часть таких файлов далее не используется или используется частично?*

Если кому-то приведенные утверждения представляются тривиальными и не стоящим специального упоминания, пусть попробует дать ответы на поставленные выше вопросы и объяснить, почему до сих пор общество не может решиться на открытую публикацию в теледоступе хотя бы всех авторефератов, а затем и диссертаций как официальных документов в рамках процедуры получения ученой степени (см. п. 1.2.). Только не из финансовых соображений. Например, РГБ находит решение таких проблем.

2 Перспективная модель электронной библиотеки диссертаций

Процесс формирования «идеального» образа ЭБД можно описать так:

- Единственно реальный способ обеспечения широкой доступности диссертаций – это электронная библиотека диссертаций (ЭБД).
- На законном основании можно поместить произведения в открытом доступе только по

договорам с авторами. Авторы хороших диссертаций заинтересованы в широкой известности своих работ и расширении научных контактов; поэтому каждый автор защищенной диссертации должен иметь возможность опубликовать ее на официальном Web-сайте РГБ в сети Интернет. Обеспечиваем такую возможность каждому, но таковых авторов оказывается мало, примерно 1% от защитившихся (причины рассматривались в прежних публикациях [1, 6]).

- Следующий шаг: каждая организация, имеющая совет по защите диссертаций, получает возможность продемонстрировать труды своих диссертантов в открытом доступе, даже не имея собственных технических и материальных возможностей, на основе договоров с авторами и с РГБ. Однако и в этом случае все зависит напрямую от настроения автора.
- Посему становится очевидным, что в настоящее время обеспечение не только полноты, но и даже определенной представительности коллекции диссертаций можно достичь лишь путем сканирования поступающих в фонд диссертаций. В 2004 г. в закон, стоящий на страже прав автора, вносятся поправки, которые позволяют библиотекам предоставлять экземпляры «произведений, выраженных в цифровой форме» во временное безвозмездное пользование, но только в помещениях библиотек [7, ст. 19, п.2] и получать от организаций, управляющих имущественными правами на коллективной основе «от имени всех обладателей авторских и смежных прав, включая и тех, которые не передали организации полномочий» [7, ст.44, п.3], лицензии на использование произведений. На данном основании РГБ с 2004 г. сканирует все поступающие в ее фонды диссертации и предоставляет их пользователям в своих стенах и виртуальных читальных залах РГБ, создаваемых в других организациях по всей стране по договорам. К сожалению, к этим диссертациям невозможен свободный доступ по Интернет в соответствии с Законом РФ «Об авторском праве и смежных правах».
- Поиск полных электронных текстов диссертаций и авторефератов в базах данных РГБ обеспечен в основном на базе электронных каталогов, в которых данные представлены в формате MARC21. Для развития возможностей поиска с частичным использованием фрагментов текста используется язык XML только в экспериментальной ЭБД на базе программного обеспечения ИСИР РАН. Исследуются возможности обеспечения интегрированного поиска данных об электронных диссертациях, в какой бы базе данных они ни находились. Привлекаются новые организации к участию в проектах

развития открытой электронной библиотеки и создания виртуальных читальных залов.

- Для снижения стоимости поддержки ЭБД и, тем самым, надежности ее дальнейшего развития признано целесообразным создавать такие технические и организационные условия, которые постепенно перенесут существенную часть нагрузки по формированию базы данных с персонала ЭБД на авторов и организации, представляющие своих диссертантов и свою науку в электронной библиотеке (путем заполнения специальных форм для формирования метаданных и т.д.). Если весь документооборот между ВАК и советами по защите, как сейчас разрабатывается в рамках проектов других организаций, будет производиться по электронным каналам связи, основная часть метаданных относительно диссертаций и процессов защиты может быть подготовлена непосредственно по месту защиты. Для этого подготовлен проект единого формата представления данных о диссертациях, авторах и организациях в государственной системе проведения защит.
- Однако кардинальное изменение картины доступности диссертационных работ на данном этапе реально только на основе появления государственного порядка представления всеми авторами до защиты авторефератов в электронной форме для включения в электронную библиотеку, а также создание стимулов для авторов, способствующих передаче ими диссертаций в ЭБД.

Насколько полно удастся реализовать эти идеи, не может предсказать никто, но получены уже существенные результаты. Образно говоря, поток информации о диссертациях прорывается к пользователю, обходя и постепенно подтачивая различные «камни преткновения».

Итак, **перспективная модель ЭБД выглядит следующим образом:**

- для каждого автора до защиты в определенной системе, поддерживаемой его организацией, создается некая интерактивная форма, в которую автоматически вносятся данные об организации и Совете по защите диссертаций;
- автор перед защитой диссертации вводит в эту форму определенные данные о своей диссертации и о себе, а также полностью копирует в форму структурные составляющие автореферата;
- автореферат печатается из этой формы, что позволяет автоматически правильно его оформить, а также передается в электронном виде в другие системы по определенному списку рассылки;
- организации, которые будут для этого заранее определены ВАК и широко известны, обязаны установить авторефераты на обозначенных ранее Web-сайтах; при этом определяется

порядок и электронный адрес передачи официальных отзывов ученых;

- после защиты автор и Совет вносят в форму автора остальные данные о рукописи диссертации и процессе защиты; в частности, разносятся по составляющим формы библиографические данные, оглавление и библиографический список. Затем система рассылает определенные группы данных по установленным правилам ВАК электронным адресам (например, в ВАК, РГБ, ЦНМБ, сеть академических библиотек и т.д.); тем самым все организации получают нужные для их систем метаданные;
- метаданные преобразуются в нужные внутренние форматы (например, в MARC для электронных каталогов) и используются в локальных ЭБД и, конечно, в центральной ЭБД РГБ;
- вероятно, длительное время формирование баз данных полных текстов диссертаций будет происходить по существующей технологии (сканирование или передача авторами и организациями), но постепенное изменение правил представления диссертаций на хранение могло бы пойти следующим путем:
 - при представлении диссертации в электронном виде в ЭБД позволить автору печатать только 1 экземпляр для защиты и для депозитарного хранения в РГБ;
 - сделать более строгими правила оформления текста диссертаций и одновременно создать удобные программы интерактивного формирования текстов диссертаций для авторов, позволяющие затем красиво распечатать текст, не зная тонкостей текстовых редакторов (см. зарубежный опыт [6]);
 - ввести в правила обязательное представление электронных диссертаций перед защитой или после нее даже за рубежом решается не так много вузов [6], но к этому тоже следует стремиться;
 - добавить в правила представления диссертаций на депозитарное хранение в РГБ передачу не только печатной диссертации в фонд, но и полного электронного текста (сначала в формате pdf), предоставив автору возможность выбора формы доступа (закрытая или открытая ЭБД);
- данные о доступе к электронным диссертациям в закрытой библиотеке РГБ и ее виртуальных читальных залах, в открытой ЭБД РГБ или на открытых сайтах организаций-соисполнителей проекта должны быть доступны в одной

централизованной системе РГБ с прямым выходом на те полные тексты, которые помещены в свободном доступе.

3 Полученные результаты

3.1 Действующая ЭБД РГБ – результат разработки 3-х проектов [5]:

- Электронная библиотека диссертаций в теледоступе – 2001-2003 гг. (при поддержке РФФИ, проект № 01-07-90310);
- Проект сканирования диссертаций из фонда РГБ для пользователей РГБ в ее стенах и виртуальных филиалах – с 2003 г. (РГБ с 2004 г. тратит и планирует тратить на это 8 млн. руб. в год из своего бюджета);
- Интегрированная библиотека электронных диссертаций - с 2004 г. (при поддержке РФФИ, проект 04-07-90154).

3.2 ЭБД состоит из:

1) закрытой базы данных (БД) отсканированных диссертаций /<http://www.diss.rsl.ru/>, доступной с компьютеров зала РГБ и виртуальных читальных залов ЭБД в других организациях (общий объем базы данных электронной библиотеки отсканированных диссертаций – около 60000; для удаленного доступа создано около 75-ти виртуальных читальных залов в 40 городах страны / <http://diss.rsl.ru/ru/rooms/open/>);

2) открытой БД диссертаций и авторефератов (без ограничения года подготовки, около 550 текстов), состоящей из электронных документов, переданных авторами по договорам с РГБ /http://www.rsl.ru/text/pub/10_4.doc/, переведенных из закрытой БД по договорам с авторами, переданных другими организациями по договорам сотрудничества с РГБ /http://www.rsl.ru/text/pub/10_5.doc/, (обычно по FTP); свободно доступна в электронной библиотеке OREL /<http://orel.rsl.ru/> и экспериментальной ЭБД на основе программного обеспечения ИСИР (Интегрированная система информационных ресурсов) РАН [8].

3.3. С точки зрения организации поиска, созданы два варианта структуры ЭБД:

- структура, основанная на сочетании ЭК (с его возможностями отображения и использования метаданных) и массива полнотекстовых документов в формате PDF;
- структура, основанная на использовании метаданных, организованных средствами языка XML (формат RDF), и полных текстов в формате PDF.

3.4. Для обеспечения дальнейшего развития ЭБД в направлении реализации ее перспективной модели с участием организаций-соисполнителей проекта:

- разработана структура метаданных относительно текста диссертации и процесса ее защиты, включающая максимум данных, необходимых для различных целей (хранения и поиска информации, вывода информации о диссертациях, авторефератах, лицах и организациях, связанных с фактом защиты, для учета самих процессов защиты в советах и ВАК, формирования информационных карточек диссертаций (метаданные структурированы с использованием XML – skeleton);
- для расширения возможностей тематического поиска используются как классификации и ключевые слова, так и тексты оглавлений;
- интерактивная форма ввода метаданных реализована на Web-сайте геологического факультета МГУ;
- разработан проект типовой электронной формы реферата;
- развивается программное обеспечение ИСИР как перспективная технология поддержки базы данных ЭБД и поиска информации;
- налажены контакты с разработчиками системы для ВАК;
- расширяется круг организаций-участников проекта.

Предполагается постоянное расширение круга участников проекта, но **базовое проектирование распределенной системы в настоящее время** ведут вместе с РГБ специалисты из ВЦ РАН, Уральского государственного университета (УрГУ), Ярославского государственного университета (ЯрГУ), МГУ им. Ломоносова (геологический факультет), БЕН РАН, Пушкинского научного центра и Ставропольского государственного университета.

Особенности новой версии ИСИР (2004-2005 гг.), адаптированной для задач ЭБД совместно с разработчиками ВЦ РАН, будут продемонстрированы во время доклада на конференции.

Литература

- [1] Лавренова О.А. Электронная библиотека диссертаций в РГБ // Материалы 9-й Международной конференции “Крым – 2002” “Библиотеки и ассоциации в меняющемся мире: новые технологии и новые формы сотрудничества”. – Судак; Ялта. 2002. – Т.1 - С. 235-239
- [2] Лавренова О.А., Вежневцев А.А. Структура и реализация электронной библиотеки диссертаций в РГБ. // Электронные библиотеки; перспективные методы и технологии, электронные коллекции: Труды пятой всероссийской научной конференции RCDL'2003 (Санкт-Петербург, 29-31 октября 2003 г.) – Санкт-Петербург, НИИ Химии СПбГУ, 2003. – С. 308-314

- [3] Лавренова О.А. Тематический поиск в электронных каталогах и электронных библиотеках // Библиотекосведение. – 2004. - №5, - с. 42-50
- [4] Лавренова О.А., Авдеева Н.В. Интегрированная библиотека электронных диссертаций // Информационные технологии, компьютерные системы и издательская продукция для библиотек: Доклады и тезисы докладов. – М.: ГПНТБ России, 2004. - С. 110-117 («LIBCOM-2004»)
- [5] Лавренова О.А. Проект интегрированной библиотеки электронных диссертаций // Технология информационного общества и культура. Международные конференции и проекты (EVA 2004. 7-я международная конференция «Информация для всех: культура и технология современного общества»). – М.: Центр ПИК, 2004 (CD-ROM).
- [5] Лавренова О.А., Литвинова Н.Н., Сарычева Н.В. Электронные библиотеки диссертаций за рубежом // Библиотекосведение. – 2004. - №1. – С. 85-92; №2. – С.91-96
- [7] Закон РФ «Об авторском праве и смежных правах». 9 июля 1993 года N5351-1 (в ред. Федеральных законов от 19.07.2004 N 110-ФЗ, от 20.07.2004 N 7-ФЗ)
- [8] Кулагин М. В. , Серебряков В. А. , Бездушный А. Н. , Галочкин М. П. , Меденников А. М. , Агошков С. В. Интегрированная Система Информационных Ресурсов (ИСИР) РАН - подход к созданию интегрированных цифровых библиотек// Международная научная конференция "Электронные библиотеки: перспективные методы и технологии, электронные коллекции", Санкт-Петербург, 1999

A new Look on the Project of the Dissertation Digital Library

Olga Lavrenova

The paper deals with the RSL Dissertation Digital Library. The following aspects are discussed: base principals, supposed future development, and last results. The last version of digital library is based on fulltext database of theses and dissertations in PDF and metadata for works and theirs defence process in the form of XML files on the base of ISIR.