

Электронные библиотеки России

МАРКАРОВА Тамара Сергеевна - зам.директора ГНУ ГНПБ им. К.Д.Ушинского Российской академии образования

МОИСЕЕВ Константин Владимирович – исполнительный директор компании «Технологии управляемого хаоса»

АГАФОНОВ Юрий Валерьевич – ведущий программист компании «Технологии управляемого хаоса»

СОЗДАНИЕ ЭЛЕКТРОННЫХ ОБРАЗОВАТЕЛЬНЫХ РЕСУРСОВ В УСЛОВИЯХ ТРАДИЦИОННОЙ ОТРАСЛЕВОЙ БИБЛИОТЕКИ

В современную эпоху кибертехнологии (информационно-коммуникационные технологии) активно применяются не только в сфере информационных услуг, но и в информационном обеспечении науки, образования и почти всех отраслей производства. Это значительно облегчает обработку, сортировку, хранение информации и способствует эффективному обмену информационными ресурсами между различными организациями, странами и пользователями. На сегодняшний день электронные библиотеки приобретают такой же официальный статус, как и традиционные. Более того, многие крупные государственные библиотеки совмещают в себе эти две формы существования. В связи с этим возникла некоторая терминологическая путаница: одни специалисты в области библиотечно-информационных технологий склонны считать электронными библиотеками только те, которые содержат полностью оцифрованные фонды (полные тексты), другие же придерживаются мнения, что к электронным библиотекам можно с полным правом отнести и те, которые располагают только оцифрованными каталогами. На наш взгляд, электронные библиотеки все-таки должны в достаточной мере отвечать современным тенденциям развития информационного общества и предоставлять доступ не только к электронным каталогам, но и к полнотекстовым базам данных, снабженным, в свою очередь, отвечающим современным стандартам и ГОСТам аннотированным библиографическим каталогом с традиционным набором поисковых параметров. Условия доступа к полнотекстовым ресурсам, безусловно, должны соответствовать требованиям Российского законодательства об авторском и смежных правах.

С появлением сети Интернет создалась принципиально новая ситуация распространения и получения научно-педагогической информации. И хотя возможности интернета используются в этом плане еще далеко не в полной мере, объективно вычлняются характеристики реальных достижений в тех или иных научных дисциплинах, проблемных полях и направлениях исследований, формирующее соответствующее пространство знаний, и поэтому увеличиваются способности, расширяется и углубляется ориентация ученых в соответствующих сферах знаний. Особый смысл приобретают организационные возможности интернета, например, в кластеризации потоков работ по определенным дисциплинам, проблемам и т. д.

Приведем некоторые принципы выделения кластеров информационных потоков научных исследований по педагогике:

- принцип учета и индивидуальных потребностей исследователей в области педагогики;
- принцип адекватности информационно-образовательных ресурсов и услуг потребностям пользователей;
- принцип доступности педагогической информации;
- принцип легитимности доступа к педагогической информации;
- принцип взаимосвязи теории с практикой в педагогике;
- принцип открытости педагогическим инновациям
- принцип наглядности и моделирования педагогической информации.

Упорядочение образовательной информации имеет важное значение для повышения качества экспертизы научно-образовательных ресурсов и прогнозирования, обеспечивающего направления и содержание проблемного поля педагогических исследований. А это, в свою очередь, способствует координации и эффективному управлению развитием научных образовательно-информационных потоков.

Именно в этих условиях отраслевые (педагогические) библиотеки становятся активными проводниками внедрения информационных технологий в профессиональные сферы, частично беря на себя задачу по созданию информационных ресурсов и мультимедиапродуктов. Деятельность библиотек оценивается в большей степени, чем раньше, по их способности выполнять свою коммуникативно-образовательную функцию, т.е. осуществлять целевую информационную поддержку научно-образовательной деятельности и удовлетворять информационные потребности специалистов в области педагогической и психолого-педагогической науки.

В настоящее время появляется очень много различных электронных образовательных ресурсов, причем наряду с официальными, лицензированными сайтами и порталами (такими, как Информика, ЦОР,

ФЭБ, некоторые сайты и даже странички отдельных вузов, кафедр, институтов и т.д.) наблюдается обилие не совсем корректных, полуофициальных, дезинформирующих и, зачастую, недолговечных так называемых информационно-образовательных ресурсов, что создает определенные трудности для основных пользователей (потребителей) данных ресурсов: ученых, преподавателей и учащихся (участников научно-образовательного процесса), которые не всегда обладают должной осведомленностью о качестве и статусе того или иного образовательного ресурса. В сложившейся ситуации перед создателями и распространителями информационно-образовательных ресурсов встают следующие задачи:

- осуществление и совершенствование информационной поддержки фундаментальных научных исследований и практической деятельности в области педагогики и педагогической психологии на базе ИКТ;
- обеспечение легитимного доступа пользователям к документам из фондов традиционных библиотек, предоставление которых затруднено в силу ряда причин: отдаленности пользователя от библиотеки, малой экзemplарности документа и т.д.;
- обеспечение легитимного доступа к информации, существующей исключительно в электронной форме;
- целенаправленное информационное обеспечение работников образования путем формирования полнотекстовых баз данных, снабженных аннотированным электронным каталогом со стандартным набором поисковых полей, и предоставления легитимного доступа к ним;
- информационное и научно-методическое сопровождение всех уровней и форм образования и самообразования.

Для осуществления этих задач необходимо создать условия для реализации возможностей функционально-информационного взаимодействия распределенных научно-образовательных ресурсов глобальных и локальных сетей как основы функционирования единого информационного пространства. Одним из важных этапов в воплощении идеи единого информационно-образовательного пространства является, на наш взгляд, создание вариативных моделей социального и межведомственного партнерства в сфере разработок, внедрения и функционирования образовательных ресурсов. Безусловно, в этой связи хочется надеяться на взаимодействие государственного, общественного и коммерческого секторов, о чем неоднократно говорит А.Б. Антопольский [1]. Партнерство в сфере единого информационно-образовательного пространства предполагает:

1. сотрудничество субъектов социального взаимодействия на межведомственном и региональном уровнях;
2. интеграцию ресурсов;
3. координацию ресурсов;
4. обеспечение доступа к образовательным ресурсам на согласованной основе.

Однако интегрирование информации, различные каналы распространения информации, включение информации в сложную и разнотипическую канву медиаформатов требует соблюдения некоторых вполне приемлемых условий существования информационного пространства:

- стандарты электронного взаимодействия;
- взаимосвязь информационных систем различных ведомств;
- координация в области информационных технологий;
- централизованный подход к автоматизации;
- стратегия интеграции единого информационно-образовательного пространства.

Итак, первоочередной задачей педагогических библиотек России в современных условиях является оперативное и исчерпывающее обеспечение образовательных учреждений всех ступеней и уровней в любой точке страны учебной и научно-педагогической информацией.

Одним из важных и приоритетных направлений научно-практической деятельности традиционных библиотек является формирование базовой инфраструктуры, развития сервисов и ресурсов для создания электронных библиотек под своей эгидой. Крупные государственные библиотеки объединяются в консорциумы (Научная электронная библиотека eLibrary, НЭИКОН), ассоциации (ЭБНИТ – Электронные библиотеки и научно-информационные технологии, ЛИБNET – Общероссийская информационно-библиотечная компьютерная сеть, ЭЛБИ – Ассоциация «Электронные библиотеки»).

ГНУ ГНПБ им. К.Д. Ушинского РАО, будучи научно-информационным центром Российской академии образования, членом Ассоциации ЭБНИТ, членом Ассоциации ЭЛБИ, де-факто является также государственным библиотечно-информационным центром образовательной отрасли в целом в России и одной из немногочисленных крупных педагогических библиотек в мире. Современные тенденции развития библиотечного дела, ориентирующие на организацию электронных библиотек, тем более актуальны для ГНПБ как для Библиотеки, обладающей большим и уникальным отраслевым (педагогическим) фондом, но не имеющей достаточной площади для приема читателей, размещения и хранения ресурсов на традиционных носителях, что, в свою очередь, усложняет расширение комплектования фондов. Преодоление названных проблем мы видим в дальнейшем ускорении темпов перевода фонда Библиотеки в машиночитаемую форму и создания новых электронных ресурсов.

Ресурсной основой электронного фонда ГНПБ, по нашему мнению, должны стать:

- оцифрованные материалы отраслевого характера (ключевые части фондов педагогических библиотек, представляющих собой материалы высокой научной ценности, уникальные и оригинальные);
- полнотекстовая база авторефератов и диссертаций по отраслевым (педагогическим) наукам;
- полнотекстовая база отраслевых и межотраслевых зарубежных и отечественных электронных журналов;
- оцифрованные учебные программы и методические пособия для отраслевых специалистов (учителей и преподавателей вузов);
- оцифрованные малотиражные отраслевые издания;
- авторские работы, размещенные в БД по просьбам различных авторов;
- тематические подборки на компакт-дисках;
- образовательные материалы, размещенные на серверах других организаций в интернете.

Наряду с плановой оцифровкой фонда в рамках проектов ЭЛБИ совместно с компанией «ССТ-Технология хаоса», учитывая огромный интерес ученых и учителей-практиков к историческому отечественному и зарубежному педагогическому наследию, Библиотека приступила к созданию полнотекстовых коллекций на компакт-дисках наиболее значимых изданий серии «Из истории отечественной и зарубежной педагогики», «Из истории отечественной и зарубежной психологии». Начали оцифровку с наиболее полного собрания сочинений К.Д. Ушинского (11 томов), отдавая дань великому русскому педагогу и будучи библиотекой, носящей его имя. Готовятся к изданию сочинения В.А. Сухомлинского, А.С.Макаренко, Л.С. Выготского, В.В. Давыдова, Песталоцци. Серию «Из истории учебной литературы» откроет коллекция «Русские буквари» и «Дореволюционные учебники математики» с методическим и культурологическим комментарием специалистов.

Создаваемые полнотекстовые коллекции при помощи программных продуктов компании Технологии управляемого хаоса могут быть размещены как на компакт-дисках, так и в Интернет-библиотеке. Это удастся сделать во многом благодаря применению компанией специального формата DjVu, который оптимизирован для хранения отсканированных документов.

Компанией разработан собственный алгоритм полнотекстового поиска, который обладает возможностями интеллектуального поиска, в том числе ассоциативного поиска, точного поиска, поиска похожего для запросов на естественном языке. Этот алгоритм используется для поиска необходимой информации в создаваемых нами тематических коллекциях.

Другой важной особенностью предложенной технологии является возможность объединения в одной коллекции оцифрованных архивных материалов и современных, подготовленных документов. Это позволяет объединять в одной коллекции как отсканированные первоисточники, так и современные материалы, изначально подготовленные на компьютере.

Для страниц помещаемых в архив материалов сохраняется два представления - графический образ страницы и соответствующий странице распознанный текст. Графический образ используется для точного представления оцифрованных материалов: это и иллюстрации, и таблицы, и, конечно же, специфическая аура, присущая старым книгам. Текстовое представление каждой страницы необходимо для построения полнотекстового индекса, на основе которого реализован механизм полнотекстового поиска.

Следует обратить особое внимание на то, что в большом количестве случаев при создании электронных архивов мы можем опустить один из самых трудоемких процессов – вычитывание автоматически распознанного текста. Современные системы оптического распознавания текста при хорошем качестве исходной картинке обеспечивают качество распознавания порядка 98 %. Учитывая то, что распознанный текст используется исключительно для работы поисковой подсистемы нашего программного обеспечения и никогда не показывается конечному пользователю, как правило, такого качества распознавания вполне достаточно.

Одной из важных особенностей предложенной технологии является возможность объединения в одной коллекции оцифрованных архивных материалов и современных, подготовленных документов. Это позволяет объединять в одной коллекции как отсканированные первоисточники, так и современные материалы, изначально подготовленные на компьютере.

Выбранный формат для хранения помещаемой в архив информации – файлы в формате DjVu. Этот формат специально оптимизирован для хранения отсканированных документов. В нем применяется механизм «многослойности». В одном графическом слое «нарисована» текстовая информация страницы, ряд графических слоев предназначен для сохранения всей нетекстовой информации из образа страницы – например, иллюстрации, фоновые рисунки и т.д. Такой способ представления графической информации позволяет использовать разные алгоритмы сжатия для разных слоев, позволяет достичь отличного качества показа растрового текста при очень небольшом размере файла. Важнейшей особенностью формата DjVu является возможность подкладывать текстовый слой под графический образ страницы. Пользуясь этой особенностью в процессе распознавания текста на отсканированной странице, мы можем сопоставить каждому распознанному слову прямоугольник на графическом слое, на котором нарисовано соответствующее слово. Перечисленные свойства формата DjVu позволяют автоматизировать процесс подготовки архива. Дополнительным плюсом формата DjVu можно считать возможность создания многостраничных документов.

Формат DjVu обладает наилучшим соотношением качества показа страницы документа к размеру файла. Эта особенность формата позволила нам при полноцветном сканировании с качеством 300 dpi сохранять до 10000 страниц на одном CD диске и более 50000 на одном DVD диске. Средний размер одной страницы составляет 60 килобайт. Это также является исключительно важным при размещении создаваемых электронных архивов в интернете, потому как позволяет существенно оптимизировать Интернет-трафик.

В программном продукте ССТ Publisher реализована функция подчеркивания найденной информации прямо по картинке - графическому образу страницы (за счет имеющейся у нас информации о том, в каком прямоугольнике какое слово нарисовано). В программе существует возможность извлекать текстовую информацию, копируя ее прямо с картинки. Можно копировать образ страницы как графический файл, таким образом можно легко копировать химические и математические формулы при помощи любого стандартного графического редактора как часть картинки. Удобный и интеллектуальный поисковый механизм дает быстрый доступ ко всей необходимой информации.

Благодаря такому выбору формата хранения помещаемых в архив материалов, а также благодаря уникальным особенностям созданных программных продуктов удалось серьезно автоматизировать процесс подготовки электронного архива. Это позволило заметно уменьшить затрачиваемое на создание архива время, а также существенно снизить себестоимость этого процесса. Именно это и позволяет процессу выпуска подобного рода электронных коллекций на CD/DVD дисках быть коммерчески оправданным.

Используемая при подготовке электронных архивов процедура является унифицированной как для подготовки электронного издания на CD/DVD дисках, так и для размещения цифровой коллекции в интернете, а также при ее использовании в локальной компьютерной сети.

При этом, было особенно важно, сохранив все преимущества выбранной концепции – сохранение первоначального внешнего вида страниц, интегрированный полнотекстовый поиск – обеспечить автоматизированный механизм публикации подготовленных электронных архивов в интернете.

Большинство существующих проектов Интернет-библиотек создают на основе текстового материала набор html-страниц, либо просто предлагают целиком загрузить файлы в формате PDF или DjVu. В первом случае теряется первоначальный вид издания, да и трудоемкость перевода книги в html-вид подчас весьма немаленькая. Во втором случае возникает необходимость загрузки файла весьма немалого размера даже тогда, когда пользователя интересует лишь небольшая часть книги. Кроме того, в случае, если книга охраняется законом об авторском праве, скачивание ее целиком является нарушением закона.

Предложенная нами концепция Интернет-библиотеки решает обе эти проблемы. На защищенном от несанкционированного доступа сервере создается хранилище книг, документов в формате PDF и DjVu. Доступ к этим файлам напрямую запрещен, но обеспечивается при помощи специального серверного программного обеспечения к каждой отдельной странице книги или документа в отдельности. Для работы с библиотекой пользователю, как правило, достаточно обычного Интернет-браузера. И, конечно же, серверная часть программного обеспечения позволяет реализовывать возможность полнотекстового поиска, обеспечивая универсальность предлагаемого решения.

На наш взгляд, создание полнотекстовых тематических коллекций положит начало формированию интегрированного информационно-педагогического пространства, целевыми функциями которого будет научно-информационное обеспечение педагогического сообщества как России, так и зарубежья.

Литература:

1. Антопольский А.Б. *О социально-экономических моделях электронных библиотек: вопросы стратегии / Интернет и современное общество: Труды X Всероссийской объединенной конференции, Санкт-Петербург, 23-25 октября 2007. - СПб., 2007. - С.3-14*
2. Антопольский А.Б., Майстрович Т.В. *Электронные библиотеки: принципы создания. - М., 2007. - 283 с.*
3. Баврин П.А. *Методика оценки эффективности применения информационных ресурсов в учебном процессе [Электронный ресурс] / П.А. Баврин.- Электрон.дан.- Режим доступа: <http://humanities.edu.ru>.*