

**Н.М. Кондрашова,**

*руководитель департамента библиотек*

*ЗАО «ПроСофт-М», кандидат педагогических наук*

**С.Ф. Гнлязов,**

*зам. руководителя департамента библиотек*

*ЗАО «ПроСофт-М», кандидат физико-математических наук*

## **СОВРЕМЕННЫЕ ПРОМЫШЛЕННЫЕ ТЕХНОЛОГИИ СОЗДАНИЯ ЭЛЕКТРОННЫХ БИБЛИОТЕК**

Отличительной чертой сегодняшнего этапа развития общества является то, что информация в нем существует как в традиционной печатной, так и в электронной форме. Современные информационные технологии позволяют, с одной стороны, приступить к широкомасштабному переводу уже накопленной человечеством информации в электронную форму, с другой, к созданию большого числа новых информационных ресурсов сразу в электронном виде. Информатизация, применение современных информационных систем в сфере образования, науки, культуры ведет к переходу на принципиально новый уровень получения и обобщения знаний, их распространения и использования. В связи с этим все большее значение в деле информационно-библиотечного обслуживания приобретает создание электронных библиотек.

### **Электронная библиотека как продукт производства**

Создание электронной библиотеки представляет собой многоэтапный производственный процесс. Определить оптимальную последовательность этапов, технологические решения на каждом из них, обеспечить безусловный контроль качества конечного продукта в полной мере под силу лишь профессионалам, имеющим солидный опыт работы в данной сфере. Опишем основные этапы производственного процесса на примере создания электронной библиотеки путем перевода в цифровую форму библиотечного книжного фонда.

Производство начинается с **разработки проекта** электронной библиотеки. Как и всякое промышленное изделие, электронная библиотека должна быть тщательно спроектирована. В некоторых случаях целесообразно реализовать пилотный проект, т.е. создать небольшой действующий

фрагмент электронной библиотеки, обладающий основными желаемыми характеристиками.

Обязательно проводится **обследование фонда документов, предназначенного для оцифровки**. Задача обследования — определить характерные особенности массива документов (состав, качество и др.) и сформулировать требования для выбора технологий оцифровки и обработки образов.

Следующий этап - **сканирование бумажных документов**. Он не так прост, как может показаться на первый взгляд. Привычные офисные сканеры для этого не подходят. Оцифровка массивов в десятки, сотни тысяч книг возможна и рентабельна лишь при условии использования промышленной технологии, специализированного оборудования, обученного персонала.

На следующем этапе **отсканированные образы бумажных книг компонуются в электронные книги** в соответствии с требованиями проекта. Одним из наиболее распространенных форматов электронных книг является формат РБР (РoPаЬе Ooсшпеш RоgпШ). Пользователь электронной библиотеки может просматривать такие электронные книги, используя свободно распространяемую программу АаоЬе Кеааег.

Важнейшим этапом является **создание поискового аппарата** электронной библиотеки. В нашем случае поиск может производиться с помощью электронного библиотечного каталога. Для организации поиска по содержанию электронных книг требуется дополнительная обработка отсканированных образов бумажных книг.

Надежное **хранение** большого по объему фонда электронных книг также должно осуществляться с использованием промышленного оборудования и технологии архивации на всех этапах реализации проекта. Хранить необходимо как промежуточные результаты производственного процесса (в частности, массивы цифровых образов документов), так и собственно контент электронной библиотеки. Для этих целей могут быть использованы специальные промышленные архивные накопители: роботизированные библиотеки оптических дисков\*.

Подробное описание каждого из вышеперечисленных этапов работы заслуживает отдельной статьи. Мы остановимся на описании двух этапов: промышленном сканировании библиотечных фондов и создании поискового аппарата электронной библиотеки.

\* Подробную информацию о возможностях и моделях промышленного оборудования, промышленных технологиях для создания электронных библиотек читатель может найти на сайтах [Биr:/Дy\y.e1аg.ги](http://Биr:/Дy\y.e1аg.ги) и [Биr:/\y\y.pro5oЙ-т.ги](http://Биr:/\y\y.pro5oЙ-т.ги)

## Промышленное сканирование библиотечных фондов

Офисное сканирующее оборудование предназначено для обработки небольших объемов документации. Производительность недорогих офисных сканеров ограничена не только невысокой скоростью их работы, но и небольшой допустимой ежедневной нагрузкой. Попытка повысить производительность, например, за счет организации многосменной работы приводит к быстрому износу и выходу из строя оборудования.

Большие неудобства при сканировании книг создает конструкция планшетных сканеров. Сканируемая поверхность разворота книги не является плоской, что неизбежно приводит к искажению электронного изображения. Книгу приходится плотно прижимать к стеклу сканера, что часто приводит к порче переплета. Свет ламп офисного сканера неблагоприятно действует на бумагу.

Оптимальным техническим решением для сканирования книг, журналов и других сброшюрованных документов является использование специальных **планетарных книжных сканеров**. Сканирующий блок такого аппарата не контактирует с книгой. Высокотехнологичные источники света практически не наносят какого-либо вреда бумаге, а в некоторых последних моделях они просто отсутствуют. Для сканирования толстых книг используется специальная подвижная площадка — колыбель, позволяющая точно выровнять поверхность сканирования. Встроенные функции обработки изображений позволяют быстро и качественно выполнить коррекцию цифрового образа, в частности, устранить тень от сгиба и выровнять искривление текста на сгибе. Сканирование книг ведется практически в темпе листания страниц, что позволяет оцифровывать десятки книг за смену.

Для оцифровки расшитых документов применяются **высокоскоростные протяжные сканеры**. Производительность такого оборудования достигает нескольких десятков тысяч сканирований в сутки, причем информация считывается сразу с обеих сторон документа. Сканеры гибко подстраиваются под различные форматы, толщину и цвет бумаги. При этом учитываются и степень износа документов.

Промышленное сканирующее оборудование отличается высоким качеством и надежностью, способно эффективно функционировать много лет.

В тех случаях, когда организация-владелец книжного фонда, предназначенного для оцифровки, не имеет возможности приобрести промышленное сканирующее оборудование, целесообразно воспользоваться услугами специализированной компании, выполняющей такие работы. При этом нет никакой необходимости перемещать книжный фонд, так как промышленные сканеры легко транспортируются и сканирование может быть организовано непосредственно в месте размещения книжного фонда. Таким образом, будет дополнительно гарантирована сохранность книжного

фонда, а организация-владелец получит возможность оперативного контроля за качеством работ по оцифровке.

## **Создание поискового аппарата электронной библиотеки**

Электронные библиотеки создаются ради быстрого и точного поиска нужной информации, поэтому поисковый аппарат должен позволять как минимум: а) найти нужную электронную книгу; б) найти в книге нужную информацию.

Стандартным инструментом поиска книг в электронной библиотеке является электронный каталог, в который загружаются электронные библиографические описания книг. Если книга, подвергаемая оцифровке, уже имеет электронное библиографическое описание, то оно лишь дополняется специальным параметром: гиперссылкой на файл, соответствующий электронной книге. В случае отсутствия электронного библиографического описания оно формируется на основе бумажного библиографического описания (карточный каталог, инвентарная книга, титульные страницы книги).

Для создания электронных библиографических описаний в промышленных объемах (сотни тысяч, миллионы описаний) применяется специальная **технология ретроспективной конверсии**. Бумажные библиографические описания сканируются на промышленном сканере. Цифровые образы библиографических описаний обрабатываются операторами по специальной методике, в результате чего формируется электронное описание в заданном формате, например, в коммуникативном формате IШ8МАК.С. Одним из параметров электронного описания является гиперссылка на электронную книгу.

Операторская обработка цифровых образов библиографических описаний строится по принципу конвейера: разбивка на элементарные операции, специализация операторов на однотипных элементарных операциях, массовое использование готовых шаблонов (электронные словари). Широко используются разнообразные формы автоматического и человеческого контроля, что позволяет гарантировать одинаково высокое качество работ на любом этапе технологического процесса.

Для организации контекстного поиска по тексту электронной книги проводится программное распознавание цифровых образов страниц. Распознанный электронный текст «сшивается» с цифровыми образами страниц, образуя второй, невидимый для пользователя «слой» электронной книги, по которому ведется контекстный поиск. Распознанный электронный текст синхронизирован с цифровым образом текста, поэтому видимый результат поиска всегда соответствует заданному контексту.

Дополнительным инструментом поиска нужной информации может быть поиск по тематическим разделам книги, например, оглавлению. Текст оглавления может быть введен непосредственно в одно из полей электрон-

ного библиографического описания с возможностью контекстного поиска. Для организации такого поиска в электронной книге могут быть сформированы электронные закладки по тексту книги.

## **Проекты создания электронных библиотек**

Хотя процесс создания общедоступных электронных библиотек в России находится на начальной стадии, уже есть примеры успешного использования новейших промышленных технологий и оборудования в этой области. Мы остановимся на двух проектах, в осуществлении которых компания «ПроСофт-М» принимала и принимает самое активное участие.

### ***Проект «Встреча на границах»***

Цель проекта — создание двуязычной (англо-русской) электронной библиотеки, материалы которой повествуют об освоении и заселении Сибири и российского Дальнего Востока, параллельном освоении американского Запада, а также о встрече русских и американцев на Аляске и на северо-западном побережье Америки. Доступ к этой библиотеке возможен по сети Интернет (<http://www.prosoft-m.ru>). Главной целью создания электронной библиотеки является предоставление широкого доступа к разнообразным историческим и культурным документам, в том числе ранее не публиковавшимся.

Инициаторами проекта и основными партнерами являются Библиотека Конгресса США, Российская государственная библиотека и Российская национальная библиотека.

### ***Проект «Электронная библиотека диссертаций РГБ»***

Цель проекта - существенное расширение доступа к фонду диссертаций Российской государственной библиотеки с использованием новейших информационно-коммуникационных технологий.

РГБ располагает уникальным фондом диссертаций. Ежегодно в него поступает около 17000 кандидатских и 8000 докторских диссертаций. Это примерно один железнодорожный вагон печатной продукции. Читальный зал отдела диссертаций РГБ ежедневно посещают 300<sup>00</sup> человек. Большая часть читателей приезжает сюда из регионов России. Ежегодно выдается более 250000 диссертаций. В настоящее время традиционные формы обслуживания читателей диссертационным фондом исчерпали себя.

Начиная с 2002 г. компания «ПроСофт-М» производит промышленную оцифровку бумажных диссертаций и создание электронных диссертаций. На сентябрь 2006 г. количество оцифрованных диссертаций превысило 100000. Доступ к электронной библиотеке диссертаций возможен по сети Интернет (<http://www.prosoft-m.ru>) и через сеть виртуальных читальных залов РГБ.