

Электронный архив ЦНСХБ как основа создания электронной библиотеки и ретроконверсии

М.А. Аветисов

Центральная научная сельскохозяйственная библиотека

Сложившаяся практика автоматизации библиотек основана на использовании систем, обеспечивающих создание и управление записями, содержащими вторичные данные о первоисточниках (библиографические описания, элементы классификации, данные об экземплярах и т.п.). Однако развитие программно-технических средств позволяет создавать автоматизированные библиотечно-информационные системы, обеспечивающие работу с электронными документами и электронными копиями документов и объектов реального мира. Это, в частности, обозначает создание электронной библиотеки, позволяющей размещать в своих "хранилищах" объекты, которые могут быть предоставлены пользователям как в стенах, так и вне стен библиотеки.

При поддержке программы АРИС (программа поддержки сельскохозяйственных реформ всемирного банка развития) и благодаря усилиям фирмы "Интеллектуальные технологии" и совместно с ней Центральная научная сельскохозяйственная библиотека (ЦНСХБ) разработала и в настоящее время приступила к производственной эксплуатации автоматизированной информационной системы (АИС) "Электронный Архив ЦНСХБ". Система рассматривается как часть общей интегральной системы и обеспечивает:

- формирование копий печатных первоисточников или их частей в электронном виде;
- хранение электронных документов;
- хранение, аудио- и видео-объектов;
- поиск документов по реквизитам и по полному тексту документов;
- выдачу копий документов (или отображение соответствующими средствами для аудио- и видео-объектов) для использования другими подсистемами – обслуживания читателей, пользователей Интернет, электронной доставки и т.п.

Система строится на базе технологии клиент-сервер. Хранение информации в настоящее время обеспечивается на сервере, в дальнейшем предусматривается использование CD-библиотеки.

В системе могут быть выделены следующие основные подсистемы:

- подсистема первичного ввода документов, включающую подсистему потокового ввода документов;

- подсистему распознавания тестов;
- подсистемы редактирования;
- подсистемы индексирования;
- подсистемы ввода каталожных карточек;
- подсистема администрирования.

Электронный архив ЦНСХБ обеспечивает работу с объектами, устройство которых достаточно несложно. Каждый объект, хранящийся в архиве, представлен в нем как совокупность следующих частей:

Часть 1 – исходное представление. При этом:

- электронный объект, представляется в архиве в исходной форме;
- печатный первоисточник в виде копии всего объекта или совокупности его частей (отдельные статьи, реклама, оглавление и т.п.) в tiff-формате;
- аудио- и видео-объекты соответственно в виде оцифрованного звука или видеоизображения в принятых форматах.

Часть 2 – распознанное представление текстового объекта.

Часть 3 – информационное описание объекта. Эта часть фактически представляет собой структурированный документ, структура которого определяется настройкой. Для каждого типа документов соответственно свои форматы. Настройка определяет и т, какие элементы и как индексируются.

Часть 4– управленческая информация о состоянии и истории объекта.

В связи с тем, что настоящий доклад на конференции отнесен к секции "Ретроспективная конверсия", далее рассматривается только этот аспект создания и ведения электронного архива.

Объектом для электронного архива является библиографическая карточка. Также как и любой другой объект она сопровождается всеми частями документа электронного архива.

Задача ретроконверсии каталога может быть представлена как последовательность технологических операций по формированию "правильного" документа, который может быть включен, в частности, в традиционный уже электронный каталог.

Решение этой задачи возможно разными путями, но во всех известных нам случаях, это большие трудо- и капиталозатарты. Не располагая этим ресурсами в должном объеме, мы предполагаем решать эту задачу методом последовательных приближений. Это тем более возможно, что АИС "Электронный архив" является открытой системой и подключение соответствующих программ обработки документов достаточно просто.

На первом этапе создается электронная картотека карточного каталога. Под этим мы понимаем технологию, при которой читатель обращается не к обычной

бумажной картотеке, а работая за рабочим местом (в нашем случае в Интранете) осуществляет поиск по "надписям на ящиках" и "разделителях в ящиках", перебирая полученный ряд сосканированных образов карточек.

Подсистема ввода каталожных карточек обеспечивает потоковый ввод карточек. В результате ввода редактор имеет дело с "ящиком", в котором "расставляются" (вводятся клавиатурным вводом) разделители, т.е. ключевые тексты, которые в дальнейшем индексируются. Индексирование готового ящика осуществляется автоматически. Каждой карточке приписываются ключевые слова разделителей всех уровней. После этого пользователь в результате поиска всегда может увидеть все "карточки" отвечающие данному запросу и просмотреть их.

На втором этапе, который осуществляется параллельно с существованием электронной каталожной картотеки (она виртуальная, реально пользователь работает с электронным архивом) осуществляется распознавание, разнесение необходимых текстов по полям и индексирование. Естественно, только самый необходимый минимум. После этого карточка может быть изъята из каталожной картотеки, а ее информационная часть передана в электронный каталог. Кстати, для записей электронного каталога, полученных из электронного архива сохраняется возможность просмотреть и карточку из архива. Передача карточки это установление необходимых связей и создание соответствующего образа в электронном каталоге.

Рабочее место пользователя ЦНСХБ это рабочее место Интранета, в котором ему предоставляются в доступ все ресурсы - электронный архив, электронные каталоги, различные базы данных и т.п.

В заключение следует сказать, что возможно будут и следующие этапы. Принципиально то, что каждая карточка сохраняется в электронном архиве как объект, который можно вновь подвергнуть обработке. Дальнейшие шаги подскажет время.
