

## Подготовка полнотекстовых ресурсов для электронного каталога с учетом многоуровневой каталогизации

**А.С. Карауш**

Заместитель директора по информационным технологиям  
*Муниципальная информационная библиотечная система г.Томска*

Представлены основные этапы и мероприятия по организации в Томской МИБС оптимального технологического процесса по созданию полнотекстовых ресурсов с учетом особенностей распределенной многоуровневой каталогизации. Многоуровневая каталогизация позволяет существенным образом сократить время создания библиографического описания сериального издания. Использование дополнительных возможностей АБИС для распределенного создания полнотекстовых документов позволяет сократить время и затраты.

Работа с электронными полнотекстовыми документами в последнее время стала для библиотек обычным явлением. В некоторых предметных областях уже сейчас преобладает производство информации и знаний в электронном виде. Для библиотек любого уровня как информационных учреждений остается важнейшим вопросом обеспечение хранения информации во всех ее видах. Однако кроме хранения библиотека должна иметь технологии и знания для конверсии информации с одного типа носителя на другой и, по возможности, без потерь. В теории данных конверсии без потерь не бывает, точнее, затраты на обеспечение конверсии без потерь стремятся к бесконечности. Поэтому в реальных условиях требуется найти оптимальное соотношение между качеством полученной конверсией информации и затратами.

Многие библиотеки создают или планируют создавать полнотекстовые информационные ресурсы. Но высокая себестоимость создания (оцифровки) бумажных носителей при нестабильном спросе тормозит процесс создания цифровых коллекций и полнотекстовых баз данных в библиотеках. В ряде случаев, цифровые коллекции содержат полные тексты статей из периодических источников. При использовании современных технологий многоуровневой каталогизации при описании периодических изданий возникает множество вопросов относительно технологических цепочек между отделами (библиотеками), где происходит каталогизация источника и где создается полнотекстовый ресурс на этот источник информации. Если же в библиотечной системе используется распределенная многоуровневая каталогизация, то сложность многократно возрастает, как увеличиваются и дополнительные технологические цепочки внутри организации, отчасти нерациональные.

В МИБС Томска с 1998 года существует распределенная технология создания библиографических описаний на периодические издания с учетом места их получения. При этом существует централизованная автоматизированная технология подписки и комплектования. Регистрация поступления происходит в библиотеке-филиале, которая осуществляет аналитическую роспись данного периодического издания. Соответственно, основные работы по созданию полнотекстового ресурса на основе периодических изданий должны выполняться в библиотеке-филиале. При создании цифровой коллекции, наряду с процессом сканирования требуется провести библиографическое описание источника, при этом использовать или восстановить поля связи многоуровневой каталогизации.

В общем случае основными этапами для создания полнотекстовых ресурсов являются:

1. Отбор литературы (информационных источников) из фондов библиотеки, отвечающих требованиям теме цифровой коллекции и всем юридическим моментам использования для создания цифровой копии (законы об авторских и смежных правах).
2. Подготовка технологического процесса сканирования – получение готового файла, содержащего распознанный или графический образ документа.

3. Подготовка технологического процесса описания полученных и приобретенных полнотекстовых ресурсов в собственном электронном каталоге с учетом современных технологий электронной библиографии.
4. Подготовка составляющих для представления и доступа пользователей к электронному каталогу с полнотекстовыми ресурсами или цифровой коллекции.
5. Изучение использования документов и дополнительного спроса на полнотекстовые ресурсы библиотеки.

Представленная последовательность подробно описана в [1]. Однако стоит остановиться на основных «точках равновесия» технологий. Таковыми являются:

- отсутствие технологической операции редактирования и коррекции при использовании режима замены неуверенно распознанных букв их графическими образами;
- использование возможностей ПО АБИС работы с 856 полем RUSMARC (\*MARC);
- использование многоуровневой каталогизации для удобства работы пользователя;
- наличие четкой технологии доступа к полнотекстовым документам, позволяющей отделу создания баз данных определять уровень доступа на уровне системы автоматизации (без участия отделов автоматизации);
- наличие технологии отслеживания использования полнотекстовых ресурсов (данные пользователя) для определения востребованности (без участия отделов автоматизации).

Нормально работающая технологическая схема данного процесса позволяет создавать ежедневно не менее 150-200 страниц полнотекстовых документов путем сканирования силами одного работника отдела создания баз данных (каталогизации) при использовании вычислительной техники «средней» мощности, например в МИБС Томска это:

- Компьютер – Celeron-533 МГц, ОЗУ 128 Мб, 17” монитор, CD-RW-привод;
- Сканер Acer Prisa-640S (SCSI-2) – реальное время сканирования страницы А4 (300 dpi) и передачи данных в компьютер не более 10 секунд;
- ПО распознавания - ABBYY FineReader 6.

Технологический процесс создания полнотекстовых документов не должен содержать этапы редактирования распознанного документа. Современное программное обеспечение позволяет прятать неуверенно распознанные символы за графические изображения участков страницы. Таким образом, пользователь видит текст книги без ошибок распознавания. При дальнейшем использовании созданных таким образом полнотекстовых ресурсов в качестве исходного материала для создания дайджестов, обзоров и пр. требуются дополнительные затраты времени на редактирование. Кроме этого, пользователь, получив такой документ, должен будет провести дополнительную редакцию.

При создании полного текста статьи из периодического издания и при работе системы многоуровневой каталогизации необходима редакция не только аналитического описания статьи (библиографической записи), но и редакция полей в связанных записях, отражающих содержание периодического издания на предмет ссылок на файл полного текста (856 поле). Т.е. для библиографического описания отдельного номера журнала требуется редакция 461 или 464 полей, используемых автоматизированными системами в качестве полей связи, где необходимо обеспечить существование 856 поля, определяющего местонахождение файла полного текста. Если же система использует свои собственные механизмы обеспечения работы полей связи, то необходимо провести дополнительную настройку конверторов для коммуникационных форматов.

Выполнение этого условия позволит упростить для пользователя распределенный поиск. Так пользователь уже на этапе просмотра содержания найденного периодического издания будет иметь информацию о наличии файла полного текста статьи. Все это существенным образом сократит количество транзакций и поисков по библиографической базе данных.

Разработка технологии уровней доступа к полнотекстовым документам на примере системы «ИРБИС» подробно описана в [2]. Но здесь стоит сказать, что устанавливать и изменять уровни доступа к файлам должны каталогизаторы или библиографы, а не работники отделов автоматизации. Также как необходимо программное обеспечение, которое будет предоставлять данные о количестве обращений к тому или иному файлу полного текста документа.

Библиотека со «средними техническими возможностями», подготовив необходимые документы на основные технологические этапы сканирования и, применив современное программное обеспечение, получает технологический процесс создания полнотекстовых ресурсов с минимальными материальными и производственными затратами. При этом же использование современных технологий и стандартов электронной каталогизации позволяет создать для пользователя дополнительные сервисы, а также обеспечить отслеживание использования создаваемых полнотекстовых ресурсов.

### **Литература**

1. Опыт создания полнотекстовых ресурсов в Муниципальной информационной библиотечной системе города Томска/ Карауш А.С., Левицкая Л.В.// 10-я Международная конференция "Крым 2003" Библиотеки и ассоциации в меняющемся мире: новые технологии и новые формы сотрудничества.", Судак, 8-16 июня, 2003 - М.: Изд-во ГПНТБ России, 2003. - Т. 2.
2. Организация доступа к полнотекстовым документам в системе ИРБИС/ Мешечак Н.А., Карауш А.С., Терехова М.В.// Библиотеки и ассоциации в меняющемся мире: новые технологии и новые формы сотрудничества: Материалы конф. – М., 2002. – Т.1. – С. 120-122.