

**Мультимедиа расширения и средства  
полнотекстового поиска в АБИС «АзЪ»**

**Multimedia Extensions and Full-text Retrieval Facilities  
of the AZ Automated Library Information System**

**Мультимедіа розширення та засоби  
повнотекстового пошуку в АБІС «АЗ»**

*А. В. Тарасов, С. В. Мошкин*  
*ООО «ИнфоКомм», Санкт-Петербург, Россия*

*Alexander Tarasov and Sergey Moshkin*  
*«InfoComm» Company, St. Petersburg, Russia*

*О. В. Тарасов, С. В. Мошкин*  
*ТОВ «ІнфоКомм», Санкт-Петербург, Росія*

Рассматриваются новые функциональные возможности АБИС «АзЪ», связанные с возможностями хранения и управления различными мультимедиа расширениями в связке со стандартным библиографическим описанием документа.

New functionalities of the AZ Automated Library and Information System are a result of storage and maintenance facilities introduced for multimedia extensions in addition to standard bibliographic descriptions of documents.

Розглядаються нові функціональні можливості АБІС «АзЪ», пов'язані з можливостями зберігання та управління різними мультимедіа розширеннями у зв'язці зі стандартним бібліографічним описом документа.

Бурное развитие мультимедиа технологий принципиально изменило формат окружающего нас информационного пространства. Эти инновации, безусловно, не могли не коснуться публичных библиотек. В настоящее время реализовано и реализуется множество проектов по созданию электронных библиотек: от глобальных порталов типа «Европианы» или Президентской библиотеки им. Б. Н. Ельцина до узкоспециализированных, предназначенных для использования одной организацией или даже просто частным лицом. При этом описание документов в таких библиотеках может включать в себя целый комплекс расширений, содержащих тексты, графику, аудио и видео файлы. Мощь этого инструментария не возможно переоценить. Фактически, при достаточном наполнении информационной базы, мы получаем доступ не только к библиографическому описанию документа в том или ином формате, но и к содержанию документа, изображениям помет, авторских знаков, любой другой значимой информации. Например, для многих произведений библиографическое описание можно дополнять записью интервью самого автора произведения с его собственной оценкой своего труда.

Такое существенное усложнение структуры хранимой информации предъявляет следующие требования к современным автоматизированным библиотечным системам (АБИС):

- наличие средств эффективного полнотекстового поиска;
- возможность отображения и воспроизведения графической, видео- и аудио- информации;
- загрузка прикрепленных мультимедиа расширений за разумное время;
- возможность пополнения электронной библиотеки новыми материалами, внесения изменений в прикрепленный текст.

При этом для реализации системы с вышеперечисленным функционалом определяющими являются:

- способ хранения прикрепленной мультимедиа информации;
- способ реализации полнотекстового поиска;

Для организации хранения файлов рассматривались два метода:

- Запись файлов мультимедиа расширений непосредственно в поля таблиц базы данных;
- Файловая система. В базе хранятся только ссылки (пути) к соответствующим файлам.

При разработке системы «АзЪ» были использованы оба способа. Текстовая информация хранится в таблице базы данных в формате rtf. Это открытый формат, обеспечивающий возможность сохранения форматирования текстовых материалов. Для графических файлов, видео и аудиозаписей в базе хранятся только ссылки на пути к файлам. Эти файлы значительно превосходят по объему текстовые, поэтому не эффективно заставлять приложение ждать пока файл видео- или аудиозаписи выгрузится во временную директорию, чтобы начать воспроизведение ролика.

Поскольку АБИС «АзЪ» реализована на базе СУБД MS SQL Server (поддерживаются все версии старше 7. 0), то для нас вопрос организации полнотекстового поиска (ПП) сводился к выбору между реализацией ПП с помощью встроенных средств СУБД (компоненты Microsoft Full-Text Search) и разработкой собственных алгоритмов.

Анализ возможностей и условий использования компоненты Full-Text Search выявил ряд существенных аспектов:

- Компонента Full-Text Search не является бесплатным дополнением. Она входит далеко не во все поставки SQL Server;
- Рано или поздно возникнет проблема обратной совместимости при переходе к старшим версиям СУБД. Это связано, в частности, с использованием компонентой Full-Text Search внутренних таблиц, сохранность структуры и само наличие этих таблиц в новых версиях не гарантируется;
- Полнотекстовый поиск с учетом грамматики реализован только для английского языка. Для всех остальных языков вы вынуждены или искать слова по полному совпадению или подключать платные лингвистические модули;
- Невозможность сохранения, функционала, использующего ПП при переходе на другую СУБД;
- Необходимый процесс слияния фрагментов индекса в единый полнотекстовый индекс занимает значительное время и требователен к ресурсам системы.

Естественно, что организация ПП на уровне приложения с использованием базы данных только как хранилища полнотекстовых индексов решает все выше перечисленные проблемы. При этом, если у вас стоит задача реализации ПП на русском языке, то любой алгоритм, учитывающий словоформы, является более предпочтительным в сравнении с поиском по точному совпадению.

Непосредственно сам механизм ПП АБИС «АзЪ» успешно решает следующие задачи:

- Создание и ведение единого полнотекстового индекса. Причем пополнение индекса производится сразу же при прикреплении очередного текстового документа. Высокая скорость пополнения индекса обеспечивается оригинальным алгоритмом индексации;
- Поддержка словоформ и поддержка поиска словосочетаний (фраз) с учетом расстояния между словами;
- Настраиваемый механизм исключения «шума» при поиске и построении индекса.
- Высокая степень универсальности алгоритмов ПП относительно языка поиска.

В заключении необходимо сказать, что, несмотря на то, что базовым форматом описания документов для АБИС «АзЪ» является RUSMARC, она легко может быть переориентирована на поддержку любого MARC формата. Пользователь может также совсем отключить опцию отображения описания документа в MARC формате и сформировать собственный, удобный ему набор атрибутов. Все это делается с помощью изменения настроек в административном модуле и регистрации в системе соответствующей библиотеки проверки. Такая универсальность открывает широкие возможности для применения системы для создания электронных библиотек, каталогов музейных коллекций, частных архивов.

Описанные решения были использованы при создании электронной библиотеки архива демократического движения Санкт-Петербурга в период 1989-1994 г. г. Это уникальное собрание документов, газет, книг, свидетельств очевидцев, листовок, видео- и аудиозаписей выступлений политиков, отразившие глобальные изменения, происходившие в обществе в постсоветское время.

Решить задачу профессионального описания столь разнородного массива удалось только благодаря полной поддержке АБИС «АзЪ» RUSMARC формата. А эффективная система хранения и представления мультимедиа документов и автоматического пополнения полнотекстового поиско-

вого индекса позволила в данном случае использовать АБИС «АзЪ» как полноценную электронную библиотеку.