

ЭЛЕКТРОННЫЕ АРХИВЫ И БИБЛИОТЕКИ

В наше время, когда активно развиваются всевозможные электронные средства коммуникаций и рабочие инструменты, когда в нашей стране наконец-то принят закон об электронной цифровой подписи, огромное количество – **до 80%** – оперативной и справочной информации по-прежнему остаётся на бумажных носителях и складывается в пыльных библиотеках или архивах. Ручная обработка таких бумажных хранилищ становится узким звеном функционирования многих крупных организаций.

Первым шагом на пути перехода от хранения и поиска бумажных документов к работе с электронным представлением информации является автоматизация каталогов архивов/библиотек на уровне атрибутивного описания имеющихся документов и работа с клиентами хранилища посредством пластиковых карт и соответствующих терминалов. Под *атрибутивным описанием* здесь понимается описание документа в специализированной БД каталога посредством определения набора атрибутов, характеризующих документ, – это, прежде всего, такие элементы, как автор, название и дата документа, а также некоторые другие, которые могут быть информативны для конечного пользователя, осуществляющего поиск; такое описание не обеспечивает возможность полнотекстового поиска информации в документе, однако может частично компенсировать этот недостаток посредством предоставления в описании ключевых слов, характеризующих семантику документа. Работа с клиентами хранилища посредством пластиковых карт и соответствующего оборудования – устройств чтения пластиковых карт и терминалов поиска – позволяет обеспечивать должную оперативность и простоту идентификации клиента и регистрации его операций, а также значительную экономию бумаги (требования, карточки регистрации и т.д.).

На примере нашей Владимирской областной научной библиотеки им. М.А. Горького могу сказать, что даже этот первый шаг даётся с большим трудом, так как он требует, прежде всего, значительных финансовых вливаний и поддержки со стороны государства, а также более высокой квалификации сотрудников при снижении их численности. Двое моих бывших одноклассников, ныне инженеров по специальности 2201 «Вычислительные машины, комплексы, системы и сети» работали некоторое время в этом учреждении и пытались сдвинуть дело с мёртвой точки, в частности, разрабатывалась некоторая система поддержки атрибутивного описания и поиска информации, однако, насколько мне известно, мало что в результате изменилось; максимум электронных услуг, которые предоставляет эта библиотека, сводится к платному поиску реферата на бесплатном сайте в Интернете и сохранению его на дискету или распечатке на принтере. Сам я когда-то был свидетелем того, насколько непрактично использование бумажных форм регистрации клиента, когда при проходе в читальный зал мне выдали контрольный листок регистрации взятых на руки материалов, а так как я ничего там не нашёл и на руки не брал, то на выходе этот листок, на котором была написана только моя фамилия, смяли и выкинули!

Следующим шагом является построение полноценной системы, электронного архива/библиотеки с возможностью эффективного полнотекстового поиска информации с использованием средств ИИ. Такая система представляет собой сложный и достаточно дорогой аппаратно-программный комплекс, включающий как средства создания/наполнения супербанка данных, так и средства обеспечения его должного функционирования и развития. Техничко-экономическая возможность её реализации появилась относительно недавно, как следствие комбинации следующих факторов:

- появились дешёвые носители – библиотеки компакт- и магнитооптических дисков;

- резко снизился показатель «стоимость/производительность» для высокоскоростных вычислительных систем, сетей и устройств;
- получили развитие аппаратно-программные системы, реализующие параллельную обработку запросов;
- повысился уровень интерфейса работы с СУБД;
- появились новые ИТ индексирования сверхбольших массивов данных;
- разработаны и развиваются отечественные технологии и программные продукты распознавания и анализа русскоязычных текстов;
- наметилось направление внедрения средств ИИ, позволяющих моделировать и анализировать большие массивы информации.

Общую идею можно обрисовать следующим образом. Организуется развёртывание высокопроизводительной сети, включающей графические рабочие станции и мощные серверы ввода и обработки информации. Для ввода документов с бумажных носителей низкого качества, используются промышленные сканеры потокового ввода и соответствующие русифицированные программные средства. Система обеспечивает эффективное индексирование и полнотекстовый поиск неструктурированной информации большого объёма. Данные, необходимые для поиска документов, хранятся в высокопроизводительной и отказоустойчивой системе памяти, а графические образы документов – в виде изображений на носителях, характеризующихся длительным временем хранения и дешевизной. Основные функции электронного архива:

- сканирование;
- распознавание и исправление ошибок;
- создание и миграция электронных документов и образов;
- индексирование документов;
- оперативный поиск и отображение документов;

- анализ документов;
- управление функционированием системы.

Для реализации данных функций в электронном архиве должны быть подсистемы ввода, хранения, индексирования, поиска и отображения информации, анализа, управления потоками, администрирования и научно-технического сопровождения.

Ядром электронного архива можно считать именно технологии индексирования и поиска. Сегодня наметились несколько направлений построения электронных архивов в зависимости от используемых в них методов поиска: уже упоминавшегося атрибутивного поиска или полнотекстового поиска, представляющего собой одно из основных преимуществ полноценного электронного архива. Полнотекстовое индексирование, предшествующее поиску, бывает двух видов:

- контекстно-независимое индексирование, не зависящее от естественного языка по причине бинарной или словарной индексации;
- контекстно-зависимое индексирование, позволяющее оптимизировать индексацию и поиск с учётом специфики морфологии и семантики естественного языка.

Очевидно, что рассматриваемая технология стоит весьма дорого и доступна только крупным организациям. Однако и её преимущества значительны: повышение полноты и оперативности отработки запросов к архиву, компактность и надёжность хранения, в том числе обеспечение безопасности конфиденциальной информации, возможность проведения оперативного анализа имеющейся информации.