

**Смысловая обработка документного потока в условиях
корпоративного взаимодействия библиотек Беларуси**
**The Semantic Processing of Documents within the Corporate
Interaction between Libraries in Belarus**

**Смислова обробка документного потоку в умовах
корпоративної взаємодії бібліотек Білорусі**

Т. В. Кузьминич, С. А. Пугачёва
Национальная библиотека Беларуси, Минск, Беларусь

Tatyana Kuzminich and Svetlana Pugacheva
National Library of Belarus, Minsk, Belarus

Т. В. Кузьминич, С. О. Пугачова
Національна бібліотека Білорусі, Мінськ, Білорусь

Рассмотрены особенности смысловой обработки документного потока в библиотеках Беларуси, особенности взаимодействия на начальном этапе внедрения системы корпоративной каталогизации. Особое внимание уделяется опыту работы Национальной библиотеки Беларуси (НББ), в которой ведутся карточные и электронный каталоги, используются классификационные языки (УДК, ББК, ГР НТИ), дескрипторный язык с контролируемой лексикой и язык предметных рубрик, Национальная база данных авторитетных записей.

The peculiarities of semantic processing of documents at Belarus libraries and their interaction on the initial stage of implementation of the corporate cataloging system are analyzed. Special attention is paid to the experience of the National Library of Belarus. The Library maintains card and electronic catalogs, uses classification languages (UDC, Library Bibliographic Classification, The State Rubricator on Sci-Tech Information), descriptor language with controlled vocabulary and the subject headings language as well as the National database of authority files.

Розглянуто особливості смислової обробки документного потоку в бібліотеках Білорусі, особливості взаємодії на початковому етапі впровадження системи корпоративної каталогізації. Особлива увага приділяється досвіду роботи Національної бібліотеки Білорусі (НББ), у якій ведуться карткові та електронні каталоги, використовуються класифікаційні мови (УДК, ББК, ДР НТИ), дескрипторна мова з контрольованою лексикою та мова предметних рубрик, Національна база даних авторитетних записів.

Формирование качественных совместимых информационно-поисковых систем – необходимое условие полноценного использования информации как важного стратегического ресурса современности. Накоплены огромные массивы фиксированной информации, чрезвычайно интенсивным является их пополнение, ускоряются коммуникативные процессы, пользователю сегодня в большей степени важны факты, идеи, концепции, аналитические материалы вне зависимости от места их издания и хранения, языка публикации и материальной конструкции документа. В таких условиях возрастают требования и к структуре самих информационно-поисковых систем, и к лингвистическому обеспечению, используемому в процессе представления различных уровней информации. Необходимо на качественно новом уровне создавать, сохранять, структурировать и распространять различные виды информации (текстовую, графическую, аудиовизуальную и другую), обеспечивая тем самым не только формирование и использование отечественных информационных ресурсов, но и возможности интеграции в мировое информационное пространство, создание режима оптимально точного удовлетворения информационных запросов пользователей. Библиотеками Беларуси накоплен богатый информационный ресурс (приобретенный и сгенерированный), который нуждается в упорядочении и представлении как совокупный. Это касается всех видов ресурсов и в первую очередь записей электронных каталогов (ЭК).

Лингвистические средства, используемые библиотеками Беларуси достаточно разнообразны. Это наследие традиционных (карточных) информационно-поисковых систем, когда их перечень определялся, прежде всего, типом библиотеки. Не последнюю роль играли и субъективные факто

ры. Автоматизация, пришедшая в библиотеки Беларуси в начале 1990-х г.г., увеличила это многообразие: выбирались и/или разрабатывались программные продукты, лингвистические средства, использовались различные форматы представления записи. Такая тенденция сохранилась до настоящего времени, она усложняет процессы обмена вторичной информацией, во многом определяет особенности корпоративного взаимодействия библиотек Беларуси.

В НББ на сегодняшний день продолжают вестись карточные каталоги, ЭК. ЭК пополняется за счет каталогизации входного документного потока в новой АБИС с использованием формата BelMARC¹, представления (конвертирования) записей 13 баз данных, созданных в локальном формате старой АБИС², начиная с 1993 г., полученных в результате рекаталогизации национального документа и создаваемых сегодня на периодические и продолжающиеся издания, аналитику. В ближайшей перспективе (ноябрь 2007 г.) частью ЭК станут записи, полученные в результате ретроконверсии генерального алфавитного каталога (около 4 млн. карточек, отражающих по языковому, видовому признакам фонд, накопленный НББ до 1993 г.) Необходимо признать, что сконвертированный массив не отличался высоким качеством: часть информации была утеряна либо некорректна, практически каждый этап конвертирования сопровождался редактированием словарей и библиографических описаний. Причинами этого можно считать не только несовершенство конверторов, доработка которых проводилась постоянно, но и недостаточно последовательные методические решения, принятые для различных видов документов в базах данных, ошибки каталогизаторов, закономерно возникающие на начальном этапе создания ЭК. В таких условиях особенно уязвимым является лингвистическое обеспечение как важнейший элемент любой поисковой системы, использование которого дает формализованное представление о смысловом содержании документов и должно обеспечивать многоаспектный поиск релевантной информации.

Национальная библиотека Беларуси всегда стремилась иметь развитую структуру лингвистического обеспечения, которая включала бы информационно-поисковые языки (ИПЯ) различных типов: язык библиографического описания, классификационные языки, ИПЯ ключевых слов и предметных рубрик. Это связано с тем, что входной документный поток, поступающий в НББ имеет широкий видовой диапазон, разнообразен состав читателей и, соответственно, широк спектр поступающих запросов. Правильный выбор и качество применяемых в библиотеке лингвистических средств способствует высокому уровню обслуживания читателей. В связи с этим было решено использовать следующие поисковые средства:

- классификационные языки (УДК, ББК, ГР НТИ);
- вербальные языки (дескрипторный язык с контролируемой лексикой и язык предметных рубрик).

Решение о применении в индексировании входного документного потока УДК для основного массива документов, ББК для картографических и нотных изданий было продиктовано традициями ведения карточных каталогов. Работа с этими языками регламентирована ГОСТ 7. 49-84, ГОСТ 7. 59-2003, соответствующими таблицами классификации и вызывает затруднения лишь в принятии решения по отражению содержания «сложных» документов. Соблюдение принципов и методики индексирования, постоянное слежение за изменениями и дополнениями к таблицам, поддержка профессиональных контактов, помогает точно отразить содержание документа с помощью этих ИПЯ.

ГР НТИ стал новым классификационным языком, ранее не используемым в смысловой обработке документов в НББ. Его выбор обусловлен возможностями проведения многоаспектного индексирования путем присвоения документу нескольких равноправных рубрик, и перспективами использования рубрикатора в обменно-информационных процессах.

Наиболее трудным решением был выбор дескрипторного ИПЯ (язык предметных рубрик было решено поддерживать только для ведения карточных каталогов). Он, в отличие от языка предметных рубрик, позволяет осуществлять детальный поиск информации на определенную тему в машиночитаемых библиографических базах данных (БД) с установлением различных отношений между терминами, расположенными в едином алфавитном ряду. Дескрипторный язык, как и ГР

¹ АБИС, разработанная на основе LIBIS в среде СУБД ORACLE 9i.

² Операционная система MS-DOS, в среде ИПС «АСПИД» с использованием СУБД BTRIVE. В качестве формата представления библиографических данных используется локальный формат, не унифицированный и не совместимый со стандартными MARC-форматами.

НТИ, был новым инструментом индексирования в НББ, но освоении методики его создания и формирование массива вызвало значительные сложности. Использование этого языка регламентировано ГОСТ 7. 66-92, но это слишком сложный логический процесс, который требует не только хорошей профессиональной подготовки, но и «лексической интуиции», особого «чувства слова». Потребовались немалые профессиональные усилия со стороны индексаторов и более 10 лет работы, чтобы овладеть процессом координатного индексирования документов и создать свой словарь ключевых слов. При этом анализировался опыт работы крупнейших отечественных и зарубежных информационных центров. Исходный массив ключевых слов, необходимый для построения словаря был получен путем накопления ключевых слов и словосочетаний при свободном индексировании документов. В качестве справочных массивов использовались тезаурусы ИНИОН, ВИНТИ. За 1992-1994 годы словарь насчитывал около 101 000 ключевых слов, отражающих содержание полученных документов. Вместе с тем, клавиатурный набор ключевых слов индексаторами, отсутствие библиотечных методических руководств привели к большому удельному весу ненормализованных терминов, дублированию понятий, орфографическим ошибкам. В 1994 г. в первой редакции вышла «Инструкция по индексированию входного потока документов для формирования массивов электронного каталога» (автор С. В. Ильина), в которой описаны правила отражения содержания документов с помощью ИПЯ, принятых в библиотеке, описана методика координатного индексирования. В дополнение к «Инструкции...» индексаторами и службой ведения ИПЯ принимаются отдельные методические решения, фиксируемые в соответствующей картотеке.

Этот массив стал основой для создания авторитетных записей (АЗ) на тематический предмет и пополнения Национальной БД АЗ в режиме текущей каталогизации. Производя автоматический ввод в процессе координатного индексирования, индексатор извлекает термины, необходимые для отражения содержания данного документа, из массива уже имеющихся в БД АЗ и вносит их в поисковый образ документа (ПОД). В случае отсутствия термина в базе индексатор выписывает его на рабочий листок и вместе с документом передает для дальнейшей нормализации в службу ведения ИПЯ редактору-лингвисту. Если термин принимается для внесения в БД, то он с соответствующей пометой возвращается индексатору. Таким образом ежегодно БД увеличивается на 3, 5 – 4 тысячи новых лексических единиц. Национальная БД АЗ предоставляет качественно новые возможности для осуществления смысловой обработки документов и проведения различных видов поиска. Ее формирование начато в 1998 г. на базе ранее созданных массивов библиографических записей. Были сконвертированы машинные словари ключевых слов в структуру авторитетных файлов, выделены из этого массива другие энтитеты (например, географические названия). Разработана технология формирования АЗ в процессе каталогизации документов и технология ведения Национальной БД АЗ, создано программное обеспечение подготовки и использования АЗ. Национальная БД АЗ сегодня включает наименования следующих энтитетов: названия коллективов (6. 533), предметов (49. 009), имена лиц (17. 561), географические названия (6. 443), торговые марки (69), родовые имена (92), унифицированные заглавия (587), имя/заглавие, в том числе обобщающие (1. 599), место издания (407), виды, формы и жанры документов (2. 671)³ Пополнение базы ведется в основном в режиме текущей каталогизации, причем различными структурными подразделениями – отделами обработки документов, индексирования документов, ведения Национальной БД АЗ. Такой подход продиктован прежде всего жестким регламентом обработки документного потока, уровнем подготовки специалистов, в том числе языковой и технологической и др.

В перспективе эта работа, на наш взгляд, должна осуществляться в текущем режиме корпоративной каталогизации документного потока с использованием не только уже сформированного массива, но и импортирования записей, созданных другими учреждениями, полученных в результате конвертирования иных БД (например, БД НББ «Беларусь в событиях и лицах»).

ЭК НББ станет основой сводного электронного каталога (СЭК) библиотек Беларуси, который будет пополняться на корпоративных началах. И одна из самых сложных проблем – лингвистическое обеспечение СЭК. Библиотекам-учасницам (на начальном этапе это НББ, Центральная научная библиотека Национальной академии наук Беларуси (ЦНБ НАН Беларуси), Республиканская научно-техническая библиотека (РНТБ)) будет предоставлена для использования Национальная БД АЗ. Классификационные индексы (кроме ГР НТИ) библиотеки могут заимствовать вместе с биб

³ Данные на 1.04.07 г.

лиографической записью либо проставлять сами для своих электронных каталогов. Второй вариант взаимодействия в части смысловой обработки документного потока предусмотрен для НББ. На начальном этапе корпоративного взаимодействия это, пожалуй, наиболее правильное решение, т. к. библиотеками не только традиционно используются различные лингвистические средства, но и различные их версии, модификации (например, 4 версия УДК – в НББ, ЦНБ НАН Беларуси, библиотеках вузов БГАТУ, ГГАУ, ВГУ, БНТУ, 5-я версия ГР НТИ – НББ, Брестской областной библиотеке им. Горького, библиотеке Белорусско-российского университета, БНТУ, БГАТУ, 3-я версия УДК – Президентской библиотекой, библиотекой БГМУ и других вузов, ББК (для массовых, областных и научных библиотек – библиотеками БГЭУ, БГМУ, Брестской областной библиотекой). Кроме того, различен уровень квалификации сотрудников и, к сожалению, на сегодняшний день не создана система специальной подготовки, переподготовки, повышения квалификации каталогизаторов, обмена опытом. Необходимо также совершенствование методико-технологического обеспечения. Большая часть этих проблем будет решаться НББ, в которой обрабатывается самый значительный по объему документный поток, создано и поддерживается лингвистическое обеспечение, отвечающее современному уровню требований, накоплен уникальный в стране опыт по формированию авторитетных записей, функционирует отдел корпоративного взаимодействия, в перспективе – Центр корпоративной каталогизации.