

Г. В. Кулькова, В. Н. Белоозеров, О. А. Антошкова

## Разработка и реализация лингвистического обеспечения электронной библиотеки промышленной корпорации

*Рассматриваются особенности электронной библиотеки (ЭБ) как ядра информационной системы крупной промышленной корпорации. Отмечается, что корпоративные ЭБ предназначены для работы с разнородными информационными ресурсами, что требует разработки лингвистического обеспечения как совокупности языковых средств для формирования корпоративных информационных ресурсов и эффективного поиска в них. Описывается разработанный и внедренный комплекс лингвистических средств корпоративной ЭБ, включающий систему метаданных, локальный рубрикатор и предметное индексирование ключевыми словами.*

Для повышения конкурентоспособности крупные промышленные корпорации, как отечественные, так и зарубежные, в настоящее время стремятся объединять свои информационные ресурсы. Наиболее эффективный вид информационной системы, хранящей разнородную информацию и являющейся актуальным средством реализации интеграционных процессов, — электронные библиотеки (ЭБ) [1]. При создании ЭБ необходимо учитывать опыт библиотечной и информационной деятельности в целом, а также общие организационные принципы создания, построения и функционирования автоматизированных библиотечных информационных систем (АБИС), включая требования действующих в библиотечной области международных и российских стандартов.

Электронные библиотеки предназначены для работы с разнородными информационными ресурсами, но для обеспечения взаимодействия пользователя с этими ресурсами необходим единый язык обмена информацией. Таким образом, определяется актуальность разработки единого лингвистического обеспечения (ЛО) как совокупности языковых средств для формирования корпоративных электронных информационных ресурсов и обеспечения эффективного поиска, обусловленная потребностями практического менеджмента на уровне современных крупных промышленных корпораций.

Проблемы определения функций электронной библиотеки в контексте современных информационных технологий рассмотрены в работах многочисленных авторов. Накопленный опыт по разработке лингвистического обеспечения корпоративных информационных систем, в частности, автоматизированных библиотечно-информационных систем наиболее полно представлен в работах [2–4].

Опыт создания электронных библиотек выявил некоторые проблемы. Практически не описано создание электронной библиотеки как ядра информационной системы крупной корпорации. Не разработаны специальные лингвистические средства, учитывающие функциональную специфику электронной библиотеки крупной промышленной корпорации. Для решения этих задач необходимо было исследовать аспекты создания, особенности развития

и функционирования корпоративных информационных систем, рассмотреть варианты их видового деления, описанные в литературе, а также изучить электронные библиотеки как особый класс информационных систем. Нами были проанализированы особенности функционирования следующих информационных систем: (1) информационные системы образования, (2) подсистемы Государственной системы научно-технической информации, (3) библиотечно-информационные системы, в том числе, АБИС на платформе LotusDomino/Notes.

Принципы организации и технология построения и функционирования ЭБ крупной промышленной корпорации были развиты применительно к проектированию информационной системы, создаваемой в Корпоративном университете “Северсталь”.

Понятие корпоративных информационных систем включает широкий класс функций и бизнес-процессов. Для нашего исследования основной интерес представляют корпоративные “системы управления знаниями”, архитектура которых включает ядро — корпоративную “базу знаний”, вокруг которой группируются средства групповой работы, электронная почта, средства диалогового общения, виртуальные клубы и другие технологии, поддерживающие коммуникации внутри организации. Основными функциями систем рассматриваемого класса являются сбор, хранение, поддержание и поиск разнородной информации, необходимой для обеспечения деятельности корпорации, в том числе для организации обучения сотрудников. Среди разнообразных информационных технологий, применяемых для реализации “систем управления знаниями”, наиболее перспективным представляется интранет, основанный на технологиях WWW (Web-технологиях).

Ядро рассматриваемых систем представляет собой электронную библиотеку. Особую важность для нас имеют интеграционные возможности ЭБ применительно к разнородным информационным ресурсам. Среди них выделяются несколько категорий внутрикорпоративных ресурсов: система документооборота, корпоративный портал, научно-техническая библиотека, информационные систе-

мы предприятий корпорации, информационная система Корпоративного университета. ЭБ должна также иметь возможность интегрировать и внешние ресурсы различных типов.

Сравнение с информационными системами, смежными с проектируемой электронной библиотекой Корпоративного университета «Северсталь», в том числе с системами научно-технической информации, образовательными порталами, библиотечными информационными системами, показало, что хотя ЭБ крупной промышленной корпорации не включается однозначно ни в один из этих классов, многие решения и стандарты, разработанные в этих системах, могут быть использованы и здесь. Это особенно существенно, поскольку проектируемая электронная библиотека должна в перспективе вступать во взаимодействие со всеми этими системами.

Настоящее исследование проводилось в рамках заданного программно-технологического комплекса LOTUS. В связи с этим предварительно было рассмотрено его применение для библиотечных информационных систем на примере библиотеки Тверского государственного технического университета. Опыт использования LOTUS показал, что этот комплекс позволяет реализовать основные функции ЭБ крупной корпорации при условии разработки специализированного лингвистического обеспечения, учитывающего функциональность и специфику электронной библиотеки. Однако LOTUS предъявляет дополнительные требования к структуре и составу лингвистического обеспечения, поскольку часть обычных функций реализуется непосредственно средствами LOTUS. Это относится к системе метаданных и к используемым средствам поиска по тексту. Таким образом, разрабатываемые языковые средства должны дополнять и органически сочетаться с готовыми средствами LOTUS.

Ранее нами были сформулированы требования к информационно-поисковым языкам, входящим в состав лингвистического обеспечения корпоративной информационно-поисковой системы [5]. Применительно к электронным библиотекам как особому классу информационных систем они сводятся к следующим характеристикам. Основой интеграции для разнородных ресурсов является система метаданных. Наиболее широкими интеграционными возможностями обладает система Дублинского ядра метаданных, которая и выбрана для применения в корпоративной ЭБ. В то же время часть функций системы метаданных, в частности формальные и административные функции, реализуется средствами программно-технологического комплекса. Для других функций — библиографических — должны быть использованы элементы традиционного языка библиографических данных, учитывающего специфику информационной базы ЭБ.

Практика применения систем метаданных показала, что конечный пользователь Корпорации неудовлетворительно применяет систему метаданных для обработки генерируемых документов, поэтому описание документов при помощи метаданных целесообразно осуществлять централизованно силами персонала библиотеки.

Проектируемая электронная библиотека Корпоративного университета «Северсталь» должна

обеспечивать поиск как по широкотематическим запросам, так и по запросам, сформулированным с точностью до детальных понятий. Эффективным для поиска по широкотематическим запросам является классификационный ИПЯ, для поиска по детальным запросам — вербальные ИПЯ. Следовательно, для лингвистического обеспечения электронной библиотеки требуется наличие хотя бы одного вербального и одного классификационного ИПЯ.

Анализ классификационных языков показал, что в смежных областях, с которыми должна взаимодействовать корпоративная ЭБ, наиболее распространенным является Государственный рубрикатор научно-технической информации (ГРНТИ), обладающий свойствами масштабируемости и разработанным механизмом настройки под задачи конкретных ЭБ. Это послужило причиной применения в разрабатываемой системе ГРНТИ и локального рубрикатора на его основе.

Исследование возможностей применения вербальных языков показало, что наиболее эффективным является сочетание свободного поиска по тексту с использованием контролируемого словаря. Свободный поиск по тексту в рассмотренных выше условиях реализации корпоративной электронной библиотеки обеспечивается средствами принятого программно-технологического комплекса. В связи с этим был сделан вывод о необходимости формирования для ЭБ словаря контролируемых терминов. Первым шагом к построению такого словаря должно быть предметное индексирование ключевыми словами, которое позволяет сформировать лексическую базу и экспериментально проверить качественные характеристики поиска по полному тексту документов.

Таким образом, в результате исследований существующих средств ЛО и заданных условий функционирования ЭБ были определены требования к составу ЛО, его компонентам, установлены те компоненты ЛО, которые необходимо разработать и внедрить в дополнение к уже имеющимся средствам, задаваемым принятым программно-технологическим комплексом.

На основе этих предпосылок были определены цели и задачи электронной библиотеки Корпорации, виды документов и информационные потоки. Был разработан комплекс языковых средств в следующем составе:

- система метаданных,
- локальный рубрикатор,
- вербальный язык предметного индексирования ключевыми словами.

Разработанный комплекс доказал свою работоспособность, о чем свидетельствует значительный рост количества запросов в библиотеку за 2002–2005 гг.

**Система метаданных.** Метаданные включают несколько сотен библиографических реквизитов, объединенных семантикой Дублинского ядра. Кроме того, имеется набор системных и административных реквизитов, которые реализуются средствами LOTUS. Библиографические метаданные органично интегрированы с системой метаданных Lotus Domino Document Manager, в результате чего значительно расширились поисковые и навигационные возможности пользователей системы. Разработаны механизмы для пополнения словарей