

Широков В.А., Ерошенко Т.А.

*Украинский языково-информационный
фонд Национальной АН Украины*

Электронная библиотека Украинского языково-информационного фонда Национальной академии наук Украины

Описана информационная система, содержащая полную информацию из оригинальных источников, организованная по принципу обыкновенной библиотеки, в которой в компьютерном виде хранится не только вторичная информация (библиографические описания), но и первичная – тексты книг, журналов, газет, нотные издания и т. д.

В последнее время отмечается тенденция к неуклонному росту полнотекстовых баз данных [1, 2], которая, очевидно, сохранится и в ближайшие годы, так как для этого существует объективная основа, связанная с постепенным переводом научно-технической,

военной, оперативной и другой информации в машиночитаемую форму.

Логично, чтобы информационные системы, содержащие полную информацию из оригинальных источников, были бы организованы по принципу обыкновенной библиотеки. Реализация этой идеи приводит к понятию электронной библиотеки, в которой в компьютерном виде хранится не только вторичная информация (библиографические описания), но и первичная — тексты книг, журналов, газет, нотных изданий, а также другие документы. С переходом к электронной библиотеке фундаментальное понятие традиционных библиотек — единица хранения — подвергается значительному обобщению. Это связано с тем обстоятельством, что к обычным изданиям и их электронным аналогам все чаще прилагаются аудио- и видеокассеты. Рисунки, графики и другая невербальная информация, включаемая в содержание изданий, также требует специальных средств обработки. Традиционные информационные системы (банки данных, знаний и т. п.) также входят в состав электронных библиотек в качестве отдельных объектов.

Следовательно, проектирование электронных библиотек прежде всего требует решения проблемы форматов хранения и обмена гетерогенной информацией.

Электронные библиотеки так же, как и традиционные, могут быть универсальными и специализированными. Однако, если в традиционной библиотеке специализация определяется только подбором литературы, то в электронной еще и специфическим набором средств, позволяющих осуществлять те или иные манипуляции с объектами хранения.

Попытку создания электронной библиотеки предпринял Украинский языково-информационный фонд Национальной АН Украины. Данная библиотека специализирована и ориентирована на решение проблем украинской компьютерной лексики. Она является двухуровневой и состоит из электронного каталога и корпуса текстов. Технически электронная библиотека реализована на персональных компьютерах IBM PC, соединенных в локальную сеть, которая работает под управлением сетевой операционной системы Novel Netware.

Тексты книг, полученные из издательств, проверяются операторами-корректорами и сравниваются с печатными изданиями для дальнейшего использования читателями и создания лексической картотеки украинского языка, разноплановых словарей, проведения

лексико-семантических исследований. Проверенные электронные тексты книг архивируются и включаются в электронную библиотеку, после чего каталогизаторами составляются библиографические описания на эти книги, которые затем заносятся в электронный каталог (ЭК).

Еще одним источником получения электронных текстов книг, помимо издательств Украины, служит компьютерная сеть EРUBNET. На книги, полученные по этой сети, также составляются библиографические описания. Информация и учет по каждой книге фиксируются в специальном электронном журнале.

На данный момент в электронной библиотеке зарегистрировано около 200 текстов книг на оптических дисках. Одновременно ведется архивный экземпляр дубля.

Кроме того, фонд составляет ретроспективное библиографическое описание словарей. На первом этапе создаются электронные библиографические описания словарей, находящихся в фондах ЦНБ. К настоящему времени в базу введены библиографические описания около 2 тыс. словарей.

ЭК создан средствами ППП CDS/ISIS/М версии 3.0, которая обеспечивает поддержку работы в локальной сети, т. е. предоставляет возможность одновременного доступа к базе данных двух или более пользователей как для поиска данных, так и для ввода. Формат ЭК разработан с учетом всех требований международного коммуникативного формата UNIMARC и соответствует международному стандарту ISO-2709. В ЭК используется оболочка, разработанная специалистами ГПНТБ России А. Бродовским и Л. Очаговой, обеспечивающая дружелюбный для пользователя интерфейс и направляющая поиск по элементам БД, а также по ключевым словам.

Выбор и установка БД, рабочего листа и формата выдачи осуществляется с помощью меню. При этом предусматривается возможность ведения каталогов соответствующих элементов (БД, РЛ, ФВ), содержащих информацию о их назначении (содержании), которая включается в меню и облегчает пользователю выбор этих элементов.

Ввод и корректировка документов (новых, указанных по номерам и/или отобранных в результате поиска) имеет следующие возможности:

— ввод полей списочных значений (т. е. полей, принимающих значения из некоторого конечного списка) с помощью меню; по-

мимо традиционного управления меню предусмотрен доступ к его элементам по ключу (первым буквам);

- ввод полей табличных значений (т. е. повторяющихся полей с подполями) с использованием наглядной табличной формы, которая не требует от пользователя ввода разделителей подполей и повторяющихся полей и обеспечивает максимальное удобство ввода собственно данных. Таким образом конечный пользователь полностью избавлен от необходимости знать такие понятия, как подполя и повторяющиеся поля;

- автоматизированная (оптимизированная) установка РЛ при корректировке документов, т. е. для документов, выбранных для корректировки, используется не один (текущий) рабочий лист, а для каждого документа автоматически выбирается свой РЛ согласно таблице соответствия (вид документа — рабочий лист);

- в случае отмены ввода/корректировки при создании нового документа внутренняя нумерация (MFN) документов не изменяется.

Просмотр документов/результатов поиска обеспечивает следующие возможности:

- отбор документов для просмотра как по номерам (MFN), так и по результатам последнего поиска;

- управление просмотром путем листания вперед, назад, в начало, в конец выдачи; путем мягкого скроллинга вперед и путем прямого доступа к документу по его порядковому номеру в выдаче;

- ручной отбор документов для последующей печати, т. е. среди просматриваемых документов можно произвольно отбирать те, которые необходимо отпечатать; собственно печать производится непосредственно после завершения просмотра;

- просмотр графических данных (т. е. любого иллюстративного материала, подготовленного в формате РСХ и связанного с соответствующим документом).

Библиографическое описание изданий составлено в соответствии с существующими правилами.

В электронной библиотеке предусмотрена возможность использования следующих подсистем:

- подсистемы глобальной корректуры БД, которая предназначена для внесения изменений во все или часть записей базы электронной библиотеки;

- подсистемы "ввод/корректурa", предназначенной для ввода новых и корректуры ранее введенных записей в БД с одновременным контролем их корректности. Кроме этого, в описываемой вер-

сии подсистемы существует возможность поиска записей в БД и переноса информации из найденной записи во вновь вводимую БД. Подсистема разработана на языке PASCAL CDS/ISIS/M. Для использования подсистемы в последующих версиях ППП CDS/ISIS/M может потребоваться повторная компиляция исходного текста подсистемы в соответствующей версии ППП;

подсистемы "Контроль записей в базе данных", предназначенной для проверки корректности данных во всех или в части записей БД.

Для одной БД может быть создано несколько различных таблиц контроля. Таблица контроля является текстовым файлом и может быть создана любым текстовым редактором.

В электронной библиотеке используется контроль на обязательное наличие названия произведения; сверка на дублетность данных; контроль идентификаторов (ISIS, ISBN, ALKOD).

Средствами ППП CDS/ISIS/M создана также и упомянутая выше ретроспективная база данных "Словари", в которой накапливаются библиографические описания словарей, изданных в мире (исходя из доступной нам информации), а также и сами тексты словарей. Использование ППП CDC/ISIS, к сожалению, накладывает существенные ограничения на качество информации, накапливаемой в электронной библиотеке Украинского языково-информационного фонда. Действительно, существующая в настоящее время версия этого ППП функционирует в операционной среде MS-DOS, не поддерживает графического представления информации и не работает с сложными информационными объектами, в которых присутствует гетерогенная информация. Поэтому в настоящее время в фонде ведутся работы по проектированию системы, которая смогла бы работать в современных операционных средах (MS-Windows/NT или XWindows) и манипулировать сложными информационными объектами, включающими текст, графику, изображение и звук и таким образом полностью поддерживать информационные потоки MultiMedia сред.

СПИСОК ЛИТЕРАТУРЫ

1. Сумароков Л.Н. Перестройка и система научно-технической информации // НТИ, сер. 2. 1987. № 9. С. 1-4.
2. Галаревский Р.С. Роль интеллектуальных информационных систем в развитии информатики // НТИ, сер. 2. 1987. № 9. С. 5-8.