

Костышин А.М., Широков В.А., Ерошенко Т.А., Пещак М.М.

*Украинский языково-информационный
фонд, Киев, Украина*

Электронная библиотека Украинского языково-информационного фонда

Быстрое развитие современных средств вычислительной техники, а именно персональных компьютеров линии IBM PC, PS/2 и Apple Macintosh, высокопроизводительных персональных рабочих станций на основе RISC-процессоров, накопителей на оптических дисках большой емкости как с возможностью однократной записи, так и перезаписываемых, позволяет рассмотреть новый подход к библиотекам и библиотечным системам. Во многих крупных современных библиотеках, например Библиотеке Конгресса США, Библиотеке Ватикана библиотечное дело организовано на компьютерной основе. В подобных библиотеках организованы электронные каталоги, обеспечивающие пользователям быстрый и эффективный доступ к материалам, хранящимся в них, быстрый поиск и заказ необходимых изданий. Более того, так как библиотеки развитых стран имеют доступ к глобальным сетям передачи данных, пользователи этих библиотек могут получить сведения о том, где именно находятся интересующие его материалы, заказать их, получить электронные копии этих материалов.

Практические преимущества электронных копий очевидны. Многие библиотеки хранят уникальные издания, которые на руки не выдаются. В то же время электронные копии такого издания становятся доступными каждому пользователю. Следует отметить также, что в последнее время значительное количество печатной продукции издается в электронном виде на компакт-дисках — CD-ROM издания. Емкость таких носителей информации достигает 620–650 Мбайт, что позволяет разместить на них многотомные издания. Кроме того, развитие современных устройств вывода информации (цветные термо и лазерные

принтеры) позволяет получить при необходимости твердые копии, практически не отличимые от оригинала.

При проектировании электронной библиотеки возникает проблема форматов хранения и представления информации, особенно в случае многоязычных изданий и изданий, в которых совмещаются текстовая и графическая информация.

В алфавитно-цифровом режиме отображения информации каждый символ представлен фиксированным знакоместом и его генерация осуществляется контроллером вывода. Причем в силу конструктивных особенностей контроллеров алфавитно-цифровых дисплеев эти символы жестко запрограммированы согласно стандарту ASCII/ISO, а их число ограничено и не может быть больше 256. Причем значительную часть их занимают специальные символы и символы — команды управления устройствами вывода. Более того, в этом режиме дисплей не может воспроизводить графическую информацию.

Стандарт ASCII/ISO не предусматривает символов кириллицы. Существующие в настоящее время модификации набора символов из стандарта ASCII/ISO, разработанные для локальных (ex-СССР) версий различного программного обеспечения, не решают данную проблему, так как в этом случае из стандарта ASCII/ISO изъяты символы, присущие западноевропейским языкам (немецкий, французский и т.д.). Таким образом, как видно из изложенного, алфавитно-цифровой режим хранения и вывода информации накладывает значительные ограничения на ее содержание.

Для решения этой проблемы консорциумом OSF в США была создана организация UNICODE, которая разработала новую систему кодировки и хранения символов национальных алфавитов. Эта система включает в себя не только символы западно- и восточноевропейских алфавитов, но и символы иврита, арабского, греческого и восточных языков (корейский, китайский и японский). Само собой разумеется, что для обеспечения подобной универсальности размер таблицы для хранения символов расширен с 256 до 65 тысяч знакомест. Однако, как упоминалось ранее, такое количество символов не поддерживается алфавитно-цифровыми дисплеями и требует использования графического представления информации и графического режима работы дисплея.

Современные операционные системы персональных компьютеров и рабочих станций, например Microsoft Windows, IBM OS/2, разработки группы X/Open (OSF) и операционные системы компьютеров Apple Macintosh рассчитаны именно на графическое представление информации и позволяют совмещать на экране одновременно графическую и текстовую информацию. Для этих операционных систем разработано большое количество прикладных программных систем (СУБД, информационно-поисковые системы, текстовые процессоры, компьютерные издательские системы), которые в полной мере пользуются преимуществами данного режима.

Однако при использовании графического режима представления информации возникают определенные сложности. Во-первых, компьютеры, базирующиеся на разных аппаратных платформах, по-разному хранят информацию в машинном слове. Во-вторых, большое количество перечисленных выше прикладных программных пакетов имеет свой формат хранения данных и связей между графическими и текстовыми объектами. Вследствие этого каждый такой программный продукт оснащается различными конверторами для импорта информации из других программных систем и экспорта своей информации в эти системы. С целью некоторой стандартизации и для облегчения развития новых программных продуктов независимыми разработчиками, фирмами Microsoft и Apple был выработан платформо-независимый метод хранения сложных документов, принятый в качестве стандарта де-факто, так называемый формат RTF (Rich Text Format).

Основной особенностью данного формата хранения информации является то, что он является практически форматизованным языком описания документа. Иными словами, используя стандартизованный синтаксис этого языка, можно полностью описать сложный документ, включая формат страниц, типы шрифтов (их гарнитуру и кегль), расположение иллюстративного материала, место и формат хранения иллюстративного материала и т.п. Таким образом мы получаем платформо- и системно-независимый сложный документ, который можно обрабатывать на любом компьютере с помощью программного обеспечения, реализующего интерпретацию команд RTE. Следует подчеркнуть, что в последнее время практически все распространенные программные про-

дукты, например текстовые процессоры (Microsoft Word, Word for Windows, Ami Professional, WordPerfect), издательские системы (Ventura Publisher for Windows, QuarkXpress, Aldus PageMaker, FrameMaker) и пакеты прикладной графики обеспечиваются конверторами из RTF в их внутренний формат, независимо от того, на какой аппаратной платформе (IBM PC, Apple Macintosh или рабочие станции Hewlett Packard, Sun или DEC) они функционируют.

В настоящее время в Украинском языково-информационном фонде создана электронная библиотека, в которой предприняты попытки реализации описанной современной библиотечной технологии.

Технически электронная библиотека реализована на персональных компьютерах серии IBM PC AT и PC/386, связанных в локальную сеть, находящуюся под управлением сетевой ОС Novell NetWare V 3.12. На физическом уровне связь осуществляется коаксиальным кабелем, на транспортном уровне используется протокол Ethernet 802.2. К файл-серверу сети подключены рабочие станции пользователей, а также основные устройства накопления и архивирования поступающей информации — накопитель на сменных оптических дисках емкостью 600 Мб и накопитель на бегущей магнитной ленте (стриммер) емкостью 150 Мбайт.

Электронные тексты книг, изданных как на Украине, так и за ее пределами, после предварительной обработки (проверка на наличие ошибок, форматирование текста и архивация) поступают в электронный архив. Доступ к отдельным изданиям архива осуществляется посредством системы электронного каталога, созданного на СУБД CDC/ISIS Rel. 3.0. Выбор данной версии системы обусловлен тем, что она обеспечивает формат электронного каталога (ЭК), разработан с учетом всех требований международного коммуникативного каталога формата UNIMARC. В ЭК предусмотрен "дружественный" интерфейс, помогающий пользователю и направляющий поиск по элементам базы данных, а также по ключевым словам.

В электронной библиотеке использованы программные средства, расширяющие возможности стандартной версии CDS/ISIS (3.0). Выбор и установка базы данных, рабочего листа и формата выдачи осуществляется с помощью меню. При этом предусматри-

вается возможность ведения каталогов соответствующих элементов (БД, РЛ, ФВ), содержащих информацию об 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) включает в себя следующие возможности:

ввод полей списочных значений (т.е. полей, принимающих значения из некоторого конечного списка) с помощью меню. Помимо традиционного управления меню предусмотрен доступ к элементам меню 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 данных. Таким образом, конечный пользователь полностью избавляется от необходимости знать такие понятия, как подполе и повторяющиеся поля;

автоматизированную (оптимизированную) установку 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 в соответствии с некоторой таблицей соответствия (вид документа — рабочий лист);

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

Просмотр документов/результатов поиска обеспечивает следующие возможности:

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

управление просмотром путем листания вперед, назад, в начало или конец выдачи; путем мягкого скроллинга вперед, а также 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.

Библиографическое описание изданий подготовлено согласно правилам библиографии и полностью соответствует стандарту UNIMARC.

В электронной библиотеке предусмотрена возможность использования следующих подсистем.

Подсистема глобальной корректировки баз данных предназначена для внесения изменений во все или часть записей базы электронной библиотеки.

Подсистема "Ввод/корректировки" предназначена для ввода новых и корректировки ранее введенных записей в БД с одновременным их контролем. Кроме этого, описываемая подсистема поддерживает возможность поиска записей в других БД и транспорта информации из них во вновь вводимую БД.

Подсистема "Контроль записей в базе данных" предназначена для проверки корректности данных во всех или 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, AKOD).

Использование СУБД CDC/ISIS, к сожалению, накладывает существенные ограничения на качество информации, накапливаемой в электронной библиотеке Украинского фонда. Действительно, существующая в настоящее время версия этой СУБД функционирует в операционной среде MS-DOS, не поддерживает графического представления информации и не работает со сложными информационными объектами, в которых присутствует смешанная текстово-графическая информация. Поэтому в настоящее время в фонде ведутся работы по проектированию информационно-поисковой системы, которая смогла бы работать в современных операционных средах (MS-Windows/NT или XWindows) и манипулировать сложными информационными объектами, включающими в себя текст, графику, изображение и звук, т.е. полностью поддерживать информационные потоки MultiMedia сред.