

Тамара Єрошенко,

н. с. Українського мовно-інформаційного фонду
НАН України,

Лариса Шевченко,

учений секретар УМІФ, к. ф. н.,

Володимир Широков,

директор УМІФ, к. ф.-м. н.

ЕЛЕКТРОННА БІБЛІОТЕКА В УКРАЇНСЬКОМУ МОВНО-ІНФОРМАЦІЙНОМУ ФОНДІ НАН УКРАЇНИ: ПРОБЛЕМИ СТВОРЕННЯ ТА ВИКОРИСТАННЯ

У статті розглядаються проблеми створення електронних бібліотек у світі, а також ті результати, які одержано в цьому напрямі в Українському мовно-інформаційному фонді НАН України. Висвітлюються структура електронної бібліотеки УМІФ і її технічне та програмне забезпечення, принципи користування нею і ті нові можливості, що відкриваються перед користувачами цієї системи.

Традиційне бібліотечно-інформаційне обслуговування має серйозні обмеження, оскільки бібліотеки орієнтуються переважно на локальних користувачів, унаслідок чого останні мають доступ практично тільки до інформаційних ресурсів конкретної бібліотеки. Крім того, інформаційне обслуговування зведено здебільшого до видачі користувачам матеріалів тільки на традиційних носіях інформації, що є значною перешкодою, якщо врахувати поширення нових інформаційних середовищ.

Бажання подолати ці «лімітування» привело до надання доступу до документів в електронній формі, у тому числі й застосування потенціалу Internet, для чого знадобилося об'єднати традиційні організаційно-технологічні засади бібліотечної справи з можливостями системотехніки цифрових документів. У результаті такого симбіозу і з'явився термін «електронна бібліотека».

У цифровій бібліотеці наявні аналоги всіх основних елементів, характерних для бібліотеки традиційної:

- система каталогів, яка відіграє роль вхідної підсистеми бібліотеки і являє собою її інформаційно-пошуковий апарат;

- «сховище» об'єктів зберігання, де містяться переведені в електронну форму книжки, журнали, газети, інша інформаційна продукція;

- комплекс сервісних та технологічних засобів (реєстрація й обслуговування користувачів, каталогізація, систематизація, облік, статистика, аналіз і зміни у фондах тощо).

Специфіка електронної бібліотеки порівняно з іншими типами інформаційних систем (ІС), найближчими з яких за своїми функціями та структурою є повнотекстові бази даних, полягає в тому, що в такій бібліотеці набагато повніше, більш структуровано й стандартизовано подано всі основні компоненти інформаційно-лінгвістичного забезпечення. Це означає:

- пошуковий апарат (електронний каталог) системи формується на основі узагальнення поняття бібліографічного опису об'єкта зберігання, яке є предметом докладної, доволі розгалуженої уніфікації і стандартизації (ISO 2709, MARC, USMARC, UNIMARC тощо);

- системотехніка бази узагальнених об'єктів зберігання заснована на сучасних форматах подання гетерогенної інформації та відповідному інформаційному, лінгвістичному і програмному забезпеченні Multy & Hypermedia Systems, що дозволяє вводити й обробляти в електронних бібліотеках комп'ютерні файли будь-якої приро-

© Єрошенко Тамара Олександрівна, Київ, 2000

© Шевченко Лариса Леонідівна, Київ, 2000

© Широков Володимир Анатолійович, Київ, 2000

ди (зокрема текстову, графічну, аудіо- та відео-інформацію тощо). Останнє зближує електронні бібліотеки з електронними системами керування документами - Electronic Document Management Systems;

- сервісно-технологічна частина електронної бібліотеки базується на досвіді формування і експлуатації традиційних бібліотек з можливістю реалізації всіх доцільних у цьому випадку «бібліотечних» функцій, а також додаткових, що враховують специфіку автоматизованих ІС.

Електронні бібліотеки, як і традиційні, поділяються на універсальні та спеціалізовані.

Для універсальної характерні:

- якомога повніша реалізація стандарту БО, введення до формату бібліографічного запису якнайповнішої системи інформаційних елементів бібліографії («полів» та «підполів») на основі уніфікованих правил каталогізації;

- широке застосування стандартних елементів інформаційно-лінгвістичного забезпечення (класифікаторів, словників, тезаурусів та ін.);

- достатньо повна система форматів подання гетерогенної інформації в базі узагальнених об'єктів зберігання, а саме така, що забезпечує надійне та ефективно відображення в електронній бібліотеці її об'єктів;

- доволі багатий набір функцій системного сервісу, що не поступається відповідному набору традиційних бібліотек.

Спеціалізовані електронні бібліотеки, як правило, відіграють роль складників більших ІС, до котрих вони входять як підсистеми. З цього випливають певні вимоги до їх функцій та структури. Так, у підсистемі ЕК може застосовуватися обмежений формат БО (що не повинно негативно відбиватися на його відповідності до стандартів). Разом із тим, у ЕК можуть бути реалізовані деякі функції, зумовлені спеціалізацією системи (наприклад, додаткові аналітичні функції - інформаційно-статистичні, семантичні, лінгвістичні, когнітивні, прогнозні тощо; засоби транспорту бібліографічної інформації з бази ЕК до видавничої підсистеми та ін.).

Методи, інформаційні й програмні засоби, що застосовуються для подання узагальнених об'єктів зберігання в спеціалізованій електронній бібліотеці, мають бути зорієнтовані на специфіку згаданих об'єктів. Наприклад, в електронній бібліотеці технічної документації доцільно передбачити програмне забезпечення, характерне для CAD-systems, що надасть її абоненту можливість із санкції системного адміністратора виконувати необхідні операції з конструкторською документацією.

Щодо змісту поняття «електронна бібліотека», то слід застерегти про обмеженість аналогії з тра-

диційними бібліотеками. Термінологічна й певна предметна близькість цих понять здатні призвести до непорозумінь, оскільки йдеться про різні за своєю природою об'єкти. Електронні бібліотеки - це, насамперед, автоматизовані ІС, які проектуються відповідно до законів і стандартів комп'ютерних систем, а технології їх формування, ведення й використання мають дуже мало спільного зі стандартними бібліотечними технологіями.

В УМІФ з 1991 р. працюють над формуванням електронної бібліотеки. Нині тут є книжки в електронній формі, одержані з українських видавництв, а також через Internet; планом передбачено включити до цієї бібліотеки академічні наукові журнали та монографії, які виходять друком в інститутах НАН. Необхідність створення електронної бібліотеки академічних видань викликана специфікою функціонування такого великого продуцента інформації, яким є НАН України, в умовах формування інформаційного суспільства. Електронний ІР становить тим більший інтерес на ринку інформаційних продуктів, чим більшу предметну галузь він охоплює. За умови розвитку необхідних сервісних засобів він, як академічна електронна бібліотека наукової періодики, має всі підстави бути конкурентоспроможним на світовому інформаційному ринку. Про це, зокрема, свідчить досвід OCLC та Європейського інституту фізики.

Роботі зі створення он-лайнової електронної бібліотеки передувала праця над проектом локальної електронної бібліотеки УМІФ, призначеної для виконання праць у галузі комп'ютерної лінгвістики та лексикографії. Ця бібліотека є спеціалізованою і має всі характерні риси електронної такого зразка: вона поєднує власне комп'ютерну бібліотеку та систему обробки текстової інформації, яка складається з двох підсистем - автоматизованої лексичної картотеки й лексикографічної підсистеми.

Електронна бібліотека УМІФ є дворівневою і складається з ЕК та корпусу текстів. Апаратно вона реалізована на персональних комп'ютерах IBM PC, з'єднаних у локальну мережу, яка працює під управлінням сітьової операційної системи Novel Netware.

Електронний корпус формується на оптичному диску типу RICOH ROD-5062F та порівнюється операторами-коректорами з друкованими виданнями з метою подальшого використання як читачами, так і фахівцями-лексикографами для формування лексичної картотеки, укладання різнопланових словників, проведення лексико-семантичних досліджень.

Перевірені електронні тексти книжок архівуються і включаються до електронного кор-

пусу, після чого каталогізатори складають їх бібліографічні описи, які заносяться до ЕК. На сьогодні в електронній бібліотеці зареєстровано понад 400 текстів книг на оптичних дисках.

На цій базі формується лексична картотека української мови, орієнтована на укладання різноманітних словників, здійснення на її основі лінгвістичних досліджень, формування словопокажчиків і конкордансів мови окремих авторів або літературних джерел, підготовки спеціальних допоміжних картотек для лексики того чи іншого типу тощо. Як одна з підсистем лексичної картотеки української мови функціонує пошукова система, створена на підґрунті українського перекладу Біблії. Вона являє собою їх повний електронний текст, словник лексем (словоформ), що зустрічаються у цьому тексті, архів запитів до системи, який поповнюється згідно з побажаннями користувача і може зберігати до тисячі відповідей на запити, довідковий файл.

Лексикографічна орієнтація електронної бібліотеки спонукає до формування ретроспективної словникової підбібліотеки. На першому етапі створюється база бібліографічних описів словників (введено понад дві тисячі), які знаходяться у фондах НБУВ. На цей час у базу ЕК введено більше двох тисяч БО словників.

ЕК створено засобами пакета прикладних програм (ППП) CDS/ISIS/M версії 3.0, що забезпечує підтримку роботи в локальній мережі, тобто дає можливість одночасного доступу до БД двом або більше користувачам як для пошуку даних, так і їх введення.

Формат ЕК був розроблений з урахуванням вимог міжнародного комунікативного формату UNIMARC і відповідає міжнародному стандарту ISO-2709. В ЕК використовується оболонка, розроблена фахівцями ДПНТБ Росії.

Застосування СУБД CDS/ISIS, на жаль, накладає суттєві обмеження на якість інформації електронної бібліотеки. Існуюча версія цієї СУБД функціонує в операційному середовищі MS-DOS, не підтримує графічного відбиття інформації і не працює зі складними інформаційними об'єктами, в яких присутня гетерогенна інформація. Зараз проводяться роботи з модернізації системи, що буде спроможна працювати в сучасних операційних середовищах (MS-Windows/NT або XWindows), маніпулювати гетерогенними об'єктами, повністю підтримувати інформаційні потоки MultiMedia середовищ.

Зупинимося докладніше на викладі систе-

мотехніки блоку узагальнених об'єктів зберігання. Сучасні технології та системи керування електронними документами забезпечують: а) перекладення документів з традиційних носіїв в електронну форму; б) індексацію документів; в) експорт до бази електронних документів (електронний архів); г) реструктуризацію та конвертацію форматів зберігання й маніпулювання гетерогенною інформацією; д) створення інтерфейсу користувача.

Технологія перекладення паперових документів в електронну форму має декілька операцій.

Перша - підготовка якісної паперової копії, що здебільшого є необхідною, оскільки використання сучасних засобів репрографії дозволяє підвищити контрастність та прочитуваність оригіналів.

Друга операція забезпечує створення електронної копії (образу) оригіналу в електронному архіві. Вона полягає у скануванні документа та збереженні його цифрового образу. Для застосувань важливо зберігати саме електронну копію документа, оскільки при наступних її перетвореннях можуть бути втрачені важливі подробиці.

Третя операція перетворює цифровий образ документа на т. зв. «мультивалентний документ», який містить множину форматів, що забезпечують подальшу роботу з ним (редагування, пошук, індексацію та ін.).

Індексація дає можливість сформувати пошуковий образ документа. Існує два основні варіанти її індексації. Перший ґрунтується на методах автоматичної специфікації. Він зводиться до автоматичного аналізу тексту документа, розбиттю його на зони, які відповідають атрибутам (полям) бібліографічної БД або вхідному формату ЕК, та автоматичному введенню структурованого тексту до БД. Реалізація цього методу потребує ретельного дослідження типів вхідних документів, їх формалізації й створення досить складної системи алгоритмів розпізнання (специфікації) документа, близької до систем штучного інтелекту.

Другий варіант індексації базується на властивостях природної мови і містить підсистему лінгвістичної обробки початкового тексту, яка забезпечує вилучення т. зв. стоп-слів, лематизацію словоформ і побудову списку всіх слів, що є в документі із посиланням на унікальне ім'я (ідентифікатор) документа, у котрому вони зустрічаються. Експорт до бази зберігає і сам документ, і його пошуковий образ у БД.

1. Сумароков Л.Н. Перестройка и система научно-технической информации // Научно-техн. информация, сер. 2. - 1987. - № 9. - С.1-4.
2. Гуляревский Р.С. Роль интеллектуальных информационных систем в развитии информатики // Там же. - С.5-8.
3. UNIMARC: Universal MARC format. 2nd ed. - 1980. - XII. - 131 p.

4. Международный стандарт ISO-2709. Документация. Структура записи для обмена библиографической информацией на магнитной ленте. - М.: Изд-во стандартов, 1984.
5. Широков В.А. Інформаційна теорія лексикографічних систем. - К.: Довіра, 1998. - 330 с.