

# До питання про формат UNIMARC

(критерій для вибору програмних засобів для ведення бібліографічних баз даних)

Однією з головних проблем, пов'язаних з технологією зберігання й обробки інформації на ЕОМ та забезпечення сумісності між автоматизованими інформаційними системами є стандартизація формату запису бібліографічних даних на магнітних носіях.

Відсутність єдиного стандарту призводить до того, що бібліотеки України працюють з програмними продуктами, які обробляють інформацію, подану в різних форматах - USMARS, UNIMARC, МЕКОФ або взагалі в довільному форматі. Це ускладнює обмін даними між бібліотеками, робить бібліографічні бази даних, створені в довільному форматі, недоступними для користувачів при роботі в мережі або потребує значних зусиль для переконвертації даних з одного формату в інший.

Для впорядкування діяльності у сфері автоматизованої обробки інформації необхідна стандартизація формату запису бібліографічних даних у машинопрочитуваній формі, подібно до існуючих у традиційній бібліотечно-бібліографічній діяльності правил та схем опису видань, які з різною мірою повноти визначають набір елементів, використовуваних при описі видання.

Робочою групою Міжнародної федерації бібліотечних асоціацій та установ (IFLA) у межах Програми універсального бібліотечного обліку та Міжнародної програми MARC розроблено Міжнародний комунікативний формат UNIMARC. Його завданням є встановлення позначок, індикаторів та кодів підполів, що надаються бібліографічним записам у машинопрочитуваній формі, визначення структури та наповнення формату.

В Україні розроблено проект міждержавного стандарту «Документація. Формат для обміну бібліографічними даними на магнітних носіях». На жаль, перша редакція має ряд недоліків.

Проект стандарту відповідає формату UNIMARC, що стосується позначок полів, але не розкриває наповнення формату: не наведено специфікації елементів даних (підполів), їх назву, ознаку обов'язковості, немає пояснення змісту, способу та прикладів запису. Хоча пропонується проект стандарту поданий як міждержавний, у ньому відсутній блок полів міжнародного використання, який є у форматі UNIMARC. Отже, ця перша спроба стандартизації формату не може бути використана в практичній роботі.

Поки немає вітчизняного стандарту, доцільно використовувати існуючу структуру формату UNIMARC, включаючи позначки полів та підполів.

Проблема форматного подання бібліографічної

інформації тісно пов'язана з проблемою вибору програмних засобів для ведення електронного каталога [1].

Формат UNIMARC, що покладений в основу проекту стандарту «Документація. Формат для обміну бібліографічними даними на магнітних носіях», висуває ряд специфічних вимог до програмних засобів для ведення бібліографічних баз даних:

I. Повинні бути засоби для обміну даними з іншими системами (підтримка файлів формату ISO 2709).

Оскільки основним призначенням UNIMARC є забезпечення обміну бібліографічними даними в машинопрочитуваній формі, необхідно щоб система, яка містить базу даних, формувала файл обміну, що відповідає положенням Міжнародного стандарту ISO 2709, визначає структуру записів, що містять бібліографічні дані. Стандарт ISO 2709 встановлює, що кожний бібліографічний запис, підготовлений для обміну, повинен містити:

1. Маркер запису, що має 24 символи.

2. Довідник, що складається: з трьохцифрової позначки для кожного поля даних, довжини поля та позиції початкового символу, який належить до першого поля даних.

3. Поля даних змінної довжини, відділені одне від одного розділювачем поля.

II. Система повинна надавати можливість вміщувати в полі два та більше елементів даних (підполів).

У форматі UNIMARC визначальними змісту є коди, що ідентифікують елементи даних або надають додаткову інформацію про елементи даних і складаються з позначок, індикаторів та ідентифікаторів підполів.

Елемент даних - це найменша частина інформації, що підлягає точній ідентифікації. Елемент даних визначається ідентифікатором підполя - кодом, що складається з двох символів: розділювача \$ (символ, встановлений по ISO 2709) та коду підполя (цифрового або буквеного, наприклад, поле 200 НАЗВА ТА ВІДОМОСТІ ПРО ВІДПОВІДАЛЬНІСТЬ [2] містить такі підполя: \$a - основна назва; \$b - загальне позначення матеріалу; \$c - основна назва твору іншого автора; \$d - паралельна основна назва; \$e - відомості про назву; \$f - перші відомості про назву; \$g - подальші відомості про назву; \$h - номер частини; \$z - мова паралельної основної назви; \$v - позначення тому.

III. Має бути можливість зробити поле повторюваним.

У форматі UNIMARC вказується на повторюваність у описі кожного поля даних та у визначеннях підполів [2].

У випадках, коли елемент даних може зустрітися понад один раз у даному запису, системою має бути сформована така кількість полів, яка необхідна для вміщення всіх повторювань елементів даних. Такий тип поля називається повторюваним. Усі повторення мають у запису однакову позначку.



На наш погляд, найповніше відповідає цим вимогам міжнародно визнана система CDS/ISIS/M, розроблена в ЮНЕСКО і призначена для структурованих нечислових баз даних [3].

Ця система підтримує сітьовий режим, що дуже важливо при організації роботи бібліотек. Система висуває мінімальні вимоги до апаратних засобів - функціонує на комп'ютерах IBM-PC/XT/AT та сумісних з ними і вимагає об'єм пам'яті 512 Кб. Існує версія системи для комп'ютерів серії VAX, що працюють під керуванням операційної системи VAX/VMS. Важливим є й те, що ЮНЕСКО підтримує систему CDS/ISIS, супроводжує та вдосконалює її, готуючи нові версії та додаткові засоби, що розширюють її можливості. У Києві існує регіональний центр ЮНЕСКО з розповсюдження системи CDS/ISIS.

Важливою особливістю системи є її призначення для управління полями змінної довжини, що дає змогу раціонально використовувати дисковий простір комп'ютера і вільно визначати максимальну довжину кожного поля.

Для полегшення і спрощення процесу взаємодії користувача з системою CDS/ISIS/M, італійською асоціацією документальних служб, бібліотек та архівів розроблено систему HEURISKO [4], що розповсюджується ЮНЕСКО і має дружній інтерфейс кінцевого користувача для організації пошуку інформації, перегляду та експорту баз даних, підготовлених у системі CDS/ISIS/M. Система дуже зручна для використання читачами як електронний каталог у бібліотеках. Крім зручного меню пошуку, система HEURISKO дає змогу експортувати зроблений запит

на магнітний носій (дискету) у файл формату ISO 2709, що дає можливість користувачеві поповнити свою власну інформаційно-пошукову систему результатами, одержаними в режимі ретроспективного пошуку (пошук умасиві за певною темою) або в режимі вибіркового розповсюдження інформації (регулярний пошук за заданою темою тільки в масиві поточних надходжень).

Передача по модемному зв'язку файлів імпорту-експорту у форматі стандарту ISO 2709 дасть змогу бібліотекам обмінюватися інформацією, поповнюючи свої бази даних за умов, що ці бази спроектовані в єдиному форматі.

Оскільки кожна бібліотека самостійно вирішує за допомогою яких програмних засобів здійснюватиметься ведення бібліографічних баз даних, відділи автоматизації повинні передусім керуватися вимогами, які відповідають вимогам формату UNIMARC, а не зручністю інтерфейсу або розповсюдженістю тієї чи іншої інформаційно-пошукової системи. При визначенні структури бібліографічної бази даних необхідно забезпечити відповідність позначок полів та кодів підполів вимогам формату UNIMARC. Наприклад, у Державній бібліотеці України для юнацтва експлуатується база даних з аудіовізуальних матеріалів; розроблена в системі CDS/ISIS/M.

Використання в бібліотеках України програмних засобів, що не відповідають сформульованим вимогам, ускладнить інтеграцію електронних каталогів різних бібліотек у єдиний електронний каталог, вимагатиме значних коштів і часу на конвертацію даних, підготовлених для обміну.

1. Сенченко Н.И. Библиотеки и компьютеры / Под ред. В.В.Васильева. - Киев: Наук. думка, 1990. - 216 с., ил.
2. Руководство по UNIMARC: Руководство по применению международного коммуникативного формата UNIMARC / Пер. с англ. авт. коллектива под руководством А.И.Земскова, Я.Л.Шрайберга - М., 1992. - 320 с.: табл.
3. Пакет прикладных программ CDS/ISIS/M: версия 2.3/ Меж-

- дународный центр научно-технической информации. - М., 1991. - Вып. 72. - 256 с.: табл. (Метод. материалы и документация по пакетам прикладных программ).
4. HEURISKO: Дружественный интерфейс конечного пользователя CDS/ISIS / Пер. на рус. язык; Ип-т программных систем АН Украины - Киев, 1993. - 51 с., ил.

Олексій  
Бодня

## Обробка документальних потоків (на основі розпізнавання бібліографічної інформації)

**В**тілення в життя Національної програми інформатизації неминуче приведе до суттєвого зростання обсягів електронної інформації. Від джерел інформації до споживача йдуть інформаційні або документальні потоки [1]. Зрозуміло, що інформаційні потоки мають дійти до кожного конкретного споживача, незважаючи на колосальні обсяги загального документообігу.

У бібліотеках серед загальних документальних потоків чільне місце займають документи з бібліографічною інформацією [2]. Вона може розповсюджуватися серед користувачів у форматі

найпоширеніших баз бібліографічних даних, у форматі ISO [3] чи інших міжнародних комунікативних форматах [4,5], а також текстовими файлами. Одним з варіантів текстових файлів буде файл бібліографічних записів з сукупністю описів різних видань (статей) за загальноприйнятими правилами таких описів. Функціонально він може вміщувати в собі якийсь бібліографічний покажчик, каталог бібліотеки чи певну його частину, реферативний журнал, індекс цитування та ін.

Чимало таких документів уже безпосередньо готуються на машинних носіях. За наявності відповідного програмного забезпечення у користувача не викликає жодних проблем і перенос інформації з паперових носіїв у файли EOM з допомогою сканерів.