

Ольга Ісаєва, Євген Сидиченко

РЕАЛІЗАЦІЯ КОМП'ЮТЕРНОЇ ОБРОБКИ ДОКУМЕНТІВ МОВАМИ З ДІАКРИТИЧНИМИ ЗНАКАМИ (з досвіду роботи)

Багатомовність документів, що комплектуються НБ України імені В.І.Вернадського, зумовила необхідність вирішення ряду питань при впровадженні нової електронної технології:

- *відпрацювання авторитетного лінгвістичного визначення мов складання бібліографічних записів з використанням універсального для України й сумісного з міжнародним кодування мов;*
- *доцільність організації бази даних (зокрема «загальної» або «мовної») документів, виданих мовами, що використовують латинський алфавіт тощо.*

Одним з головних є питання про можливість реалізації занесення до ЕК інформації про видання мовами, що включають літери з діакритичними знаками, зі зручним і коректним пошуком і дотриманням обов'язкових технологічних умов обробки документів від комплектування до місць їх зберігання.

Як підтверджує обмін досвідом між фахівцями з різних бібліотек, зокрема й зарубіжних, ця проблема досі не знайшла оптимального розв'язання, бо зумовлена, насамперед, програмними засобами, що працюють у середовищі DOS, і необхідністю застосування комп'ютерів застарілих типів, використовуваних у бібліотеках, а також відсутністю міжнародного стандарту, єдиного для латиномовних країн і держав, у основу національних алфавітів яких покладено кириличний алфавіт.

Кардинальним вирішенням проблеми багатомовності могло б стати використання двобайтових наборів символів UNICODE. Але такий перехід потребує модернізації апаратних засобів (комп'ютерів) і відповідного програмного забезпечення.

З тієї латиномовної літератури, яка надходить у бібліотеку, у алфавітах 37-ми мов є діакритичні знаки (німецька, французька, іспанська, португальська та ін).

До того ж, у деяких мовах (угорська, в'єтнамська тощо) літери з діакритичними знаками входять у словникові алфавіти, і враховуються при розташуванні словникових статей. В інших, наприклад, у французькій, італійській і деяких інших літери з діакритикою складають додатковий фонетичний набір алфавіту і при розстановці словникових статей не враховуються.

Що стосується практики ведення алфавітних каталогів зарубіжних видань, то ніякі діакритичні знаки не враховуються при розстановці каталожних карток. Готичний шрифт видання передається латинським, грецький алфавіт транслітерується латинськими літерами.

Каталог закордонних видань у НБУВ (окремо книжковий та для періодичних видань) можна охарактери-

зувати як «загальний» для всіх мов, алфавітизація в якому здійснюється в межах стандартного нерозширеного латинського набору символів. Існують методичні рекомендації щодо доцільності й можливої практики ведення «мовних» каталогів - окремих для кожної мови або груп мов (див. «Описание книг на иностранных языках» / автор-упорядник Р.С.Гіляревський).

При впровадженні нової технології відбувається процес суміщення традицій та іновацій: нові можливості, що надаються «комп'ютером», мають пов'язуватися з традиціями ведення алфавітних масивів, адже співіснування у визначеному часовому відрізку старої (карткової) та електронної технологій - об'єктивна реальність.

Пропонована модель ЕК закордонних видань може бути побудована за мовним принципом з урахуванням розбіжностей кодових сторінок алфавітів у різних країнах.

Надане користувачеві апаратне забезпечення, IBM комп'ютерів задовольняє розширеному фонетичному алфавіту тільки 8-ми мов: німецької, іспанської, італійської, каталонської, шведської, естонської, фінської, ретророманської з застереженням наявності для деяких мов тільки маленьких літер з діакритичними знаками.

Вивчаючи досвід бібліотек, у яких процес комп'ютеризації почався значно раніше, ніж в Україні, ми познайомилися з простою методикою реалізації «діакритики». У Мюнхенській національній бібліотеці (Bayerische Staatsbibliothek) при складанні бібліографічних описів у комп'ютерному режимі діакритичні знаки заносяться перед відповідними літерами, а під час друку, завдяки спеціальній обробці, діакритичні знаки з літерами-основами (два символи) замінюються на літери з діакритикою (один символ). Від цієї методики вирає традиційна каталожна картка, але перегляд запису на екрані досить незручний і важкий для сприйняття. Самі німецькі фахівці вважають цей метод «далеким від оптимального».

У НБУВ почалось відпрацювання іншого шляху для розв'язання проблеми «діакритики». Було вирішено, що для забезпечення у майбутньому взаємного обміну

інформацією з країнами далекого й близького зарубіжжя без трудомісткого процесу перекодування знаків можливе використання кодових сторінок з національними алфавітами (розширений фонетичний набір). Для цього було послано запити в національні бібліотеки і великі інформаційні центри з проханням надати інформацію про кодування національних алфавітів (при наявності декількох державних мов - декількох варіантів цих таблиць).

Наявність необхідних кодових сторінок не стала б в остаточному підсумку вирішенням питання «діакритики», як і можливість розташування діакритичних знаків перед літерами-основами.

У процесі складання бібліографічного запису, маючи об'єктом опису документ мовою з діакритичними знаками, бібліотекар повинен, заздалегідь ідентифікувавши мову документа з присвоєнням коду мови, занести інформацію у відповідний робочий лист. Окремі бібліографічні елементи, занесені мовою опису, окремо й сукупно складуть пошуковий образ документа. Текст бібліографічного запису можна умовно розбити на два блоки: бібліографічні елементи, що заносяться мовою документа, і бібліотечно-бібліографічні елементи (шифри зберігання типу «Колекція», літерна частина індексів ББК, службові позначки про додаткові записи та ін.), що вносяться українською мовою.

Отже, користувач має в розпорядженні стандартні технічні засоби зі стандартною клавіатурою (латино-кириличною), програмне забезпечення, що працює в середовищі DOS: на екрані може бути відображений набір, що не перевищує 256 символів, а це свідчить про неможливість одночасного використання символів (великих і малих літер) з різних наборів - кириличного й розширеного фонетичного латинського набору. Екран у режимі перегляду бібліографічного запису супроводжений також службовими написами й підказками українською мовою, а сам запис має бути представлений мовою складання опису і в тому ж вигляді відображений на паперовому еквіваленті - друкованій карточці.

Отже, основна проблема «діакритики» - неможливість реалізації одночасного відображення на екрані розширених наборів символів латинського шрифту з кирилицею. Метод вирішення питання базується на можливості по чергово використовувати різні набори символів у різних режимах комп'ютерної обробки літератури.

1. Режим введення бібліографічних даних. Записи заносяться до ЕК мовою документа й українською. При цьому використовується набір символів та їх кодування, прийняте в країні, де ця мова традиційно вважається національною. Перемикання між набором символів з повним фонетичним набором і набором символів ДСТУ 2018 (українське національне кодування) здійснюється оперативним, без виходу з активного режиму програми, вручну «гарячою» клавішею. При формуванні точок доступу символи з однаковими машинними кодами, залежно від того, яку інформацію вони несуть (якою мовою занесені), перекодовуються по-різному.

2. Режим пошуку. Інформаційно-значущі бібліографічні елементи згруповані в алфавітних показниках, наприклад: «Особі: автори, редактори, укладачі та ін.», «Серії» «Вихідні дані» тощо. Тематичний пошук реалізується завдяки індексам ББК і їх розшифруванню у вигляді предметних рубрик (українською мовою). Виходячи з практики ведення традиційних карткових каталогів закордонних видань без врахування «діакритики», що створює більш комфортні й спрощені умови проведення пошуку, можна вважати припустимим при формуванні пошукових термінів (точок доступу) всі літери з діакритичними знаками перетворити в літери стандартного нерозширеного латинського набору, наприклад:

$$\begin{array}{l} \grave{O}, \acute{O}, \hat{O}, \bar{O}, \ddot{O} ==> O \\ \grave{A}, \acute{A}, \hat{A}, \bar{A}, \ddot{A} ==> A \end{array}$$

Дане перетворення відрізняється від методу транслітерації, в якому кожній літері з діакритикою ставиться у відповідність унікальне сполучення літер з необхідністю однозначного зворотного перетворення сполучення літер у літеру з діакритичним знаком. Отже, в режимі пошуку користувач повністю задовольняє стандартний набір ДСТУ 2018. Пошук здійснюватиметься із застереженням «без діакритики».

3. Режим перегляду запису. Бібліографічний запис подається на екран з відповідними діакритичними знаками. Перемикаючи набори символів натисканням «гарячої» клавіші, користувач має можливість переглядати запис або екранні підказки. Зворотне перемикання здійснюється повторним натисканням тієї ж клавіші. Режим перегляду є стартовим для друку каталожної картки з «діакритикою». Перемикання «гарячою» клавішею реалізується, наприклад, при використанні DISPLAY.SYS (зі стандартного набору DOS) для завантаження наборів символів з «діакритикою» і KEYRUS.COM (версії 7.3 розробки Д.О.Гуртяка) для українського набору.

Пропоноване проектне вирішення надалі дасть змогу безболісно перейти до технології, яка використовується в країнах з розвинутою індустрією інформатики.

При друкуванні карток можуть виникнути деякі ускладнення, але при використанні пакета програм «ALLEGRO», що дає можливість працювати на рівні кожного окремого символу, налагодження принтерів для друку будь-яких знаків - завдання, яке можливо розв'язати.

У НБУВ для наповнення ЕК з 1994 р. використовується програмний продукт «ALLEGRO», розроблений у бібліотеці Брауншвейгського університету (ФРН). Програма включає модулі комплектування, каталогізації, обслуговування, управляє такими бібліотечними функціями, як облік примірників, можливість друку покажчиків, обмін записів у різних форматах та ін. Гнучкість програми дозволяє, не порушуючи бібліотечної традиції, вносити необхідні корективи під час адаптації конкретної програми до умов функціонування в кожній конкретній бібліотеці.

З ФОНДІВ НБУВ

les Misérables

Факсиміле заголовка
рукопису В.Гюго
«Знедолені»