

УДК 004.352:017]:02



Юлія Яковлева,
науковий співробітник Центру
бібліотечно-інформаційних технологій НБУВ,
кандидат технічних наук

Сканований каталог дисертацій на сайті Національної бібліотеки України ім. В. І. Вернадського

Наведено досвід реалізації проекту ретроспективної конверсії в Національній бібліотеці України імені В. І. Вернадського на прикладі створення імідж-каталогу дисертацій.

Ключові слова: бази даних, бібліотека, імідж-каталог дисертацій, ретроспективна конверсія.

Ретроконверсія — переведення паперових карткових каталогів в електронний вигляд — є важливим етапом автоматизації бібліотеки. Яким саме методом буде розв'язане завдання ретроконверсії — вирішує кожна з них на стадії його постановки. Будь-яка помилка при прийнятті рішення може відбитися як на вартості проекту ретроконверсії, так і на його ефективності. Метою пропонованої статті є аналіз досвіду бібліотек для визначення рекомендацій щодо розроблення проектів ретроспективної конверсії бібліографічних каталогів.

Електронний каталог Національної бібліотеки України імені В. І. Вернадського, який на сьогодні містить більше 400 тис. записів, наповнюється новими надходженнями, починаючи з 1994 року, але старші фонди досі не знайшли відображення у нових каталогах, що значно ускладнює пошук. З метою переведення паперових карткових каталогів основного зберігання в електронний вигляд у 2007 році Центром комп'ютерних технологій НБУВ було розпочато проект ретроспективної конверсії традиційних каталогів та картотек.

У бібліотечному світі вже нагромаджено досвід ретроконверсії карткових та друкованих каталогів, який було враховано при розробленні проекту ретроспективної конверсії каталогів НБУВ. Звичайно, розглядають п'ять можливих способів конверсії карткових каталогів [1—5]: 1) рекаталогізація, яка передбачає створення нового бібліографічного запису при наявності книги; 2) ретроспективне переведення шляхом уведення з клавіатури інформації з каталожної картки; 3) корпоративна ретроконверсія, тобто імпортування машиночитаних баз даних з зовнішніх джерел; 4) сканування традиційних каталогів без подальшого оброблення; 5) сканування каталогів з подальшим переведенням графічних образів у текстовий і форматуванням відповідно до прийнятого формату машиночитаної каталогізації.

Реалізація ретроконверсії шляхом ручного введення дає змогу отримати найкращий результат, але прийнятна тільки в невеликих бібліотеках (або окремих спеціалізованих фондах) і рідко використовується у великих книгозбірнях з огляду на декілька причин, зокрема високу вартість конвертованого запису, тривалість процесу конверсії, значну кількість помилок уведення тощо.

Спосіб корпоративної ретроконверсії передбачає запозичення необхідних бібліотеці готових бібліографічних записів в електронній формі з великих бібліотечних фондів, наприклад, OCLC. Цей спосіб впроваджено у Росії, де основним фондом для запозичення записів виступила Російська державна бібліотека, загальна кількість бібліотечних записів

якої становить більше 2,5 мільйона. Цей спосіб уможливило суттєве підвищення ефективності ретроконверсії, але потребує серйозної організації всіх робіт.

Сканування без подальшого оброблення зображень — це досить простий та дешевий спосіб, при якому пошук забезпечується шляхом перегляду графічних зображень каталожних карток. При цьому, звичайно, втрачаються всі можливості автоматизованого пошуку за пошуковими полями бази даних. Наразі при ретроконверсії каталогів великого обсягу найчастіше використовується спосіб, що базується на використанні сканерів і програм розпізнавання символів і має незаперечні переваги. Однак ефективність цього способу безпосередньо залежить від фізичного стану каталогу, особливостей оформлення бібліотечних карток та роздільників, пов'язаних, наприклад, з історичними змінами стандартів бібліографічного опису, традиціями ведення системи каталогів у конкретній бібліотеці тощо. Такий метод застосовує Державна публічна науково-технічна бібліотека Росії (<http://gpntb.ru/>), яка впровадила технологію розпізнавання компанії ABBYY, вбудовану в новий продукт сімейства ІРБІС- "ІРБІС Імідж-каталог" [7]. Спеціалісти ДПНТБ Росії обрали продукт ABBYY FineReader Engine, виходячи з високої якості технологій розпізнавання компанії ABBYY. Це дало змогу реалізувати повнотекстовий пошук з ранжируванням результатів, враховуючи близькість слів, морфологію та інші особливості тексту (рис. 1).

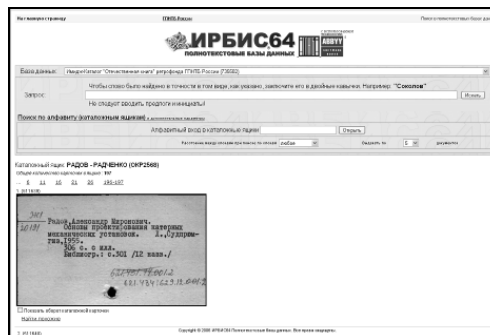


Рис. 1

Надзвичайно вдалий приклад колег ДПНТБ надихнув фахівців Білоруської сільськогосподарської бібліотеки (<http://belal.by/Cataloguer.html>) на використання продукту "ІРБІС Імідж-каталог" для створення власного імідж-каталогу, що включає електронні копії графічних образів алфавітного каталогу (до 1993 року) (рис. 2).

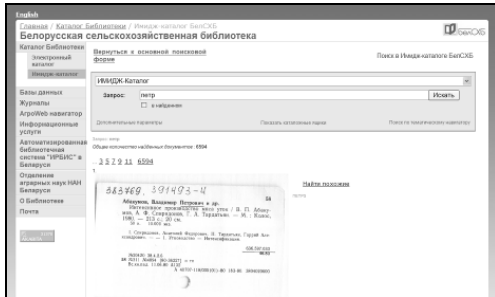


Рис. 2

Якщо розпізнавання графічних карток виявляється недоцільним, то можливий комбінований варіант, що поєднує процес отримання графічних зображень карток і ручне введення бази даних роздільників, тобто спеціальних каталожних карток, що відділяють один від одного масиви карток. Такий спосіб реалізовано в багатьох бібліотеках, зокрема в Російській національній бібліотеці (<http://www.nlr.ru/>), у Державній публічній історичній бібліотеці Росії (<http://www.shpl.ru/>), сканований предметний каталог якої містить більше 900 тис. графічних образів, у бібліотеці Варшавського технічного університету (Польща). Цей спосіб також реалізовано в науковій бібліотеці Московського державного університету, в алфавітному імідж-каталозі якої відображені надходження бібліотеки з 1755 по 2004 рік (<http://www.nbmgu.ru/>). У технології представлення імідж-каталогу МДУ заслуговує уваги можливість використання як послідовного пошуку, так і надзвичайно гнучкого пошуку за роздільниками (рис. 3).

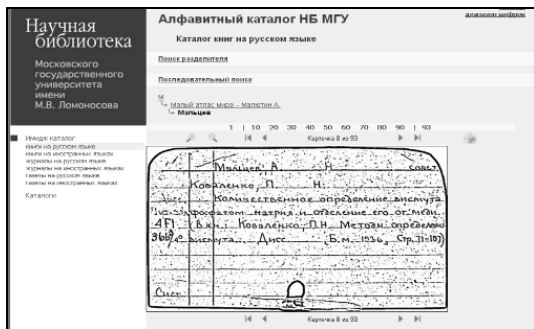


Рис. 3

Послідовний пошук за роздільниками в своєму імідж-каталозі впровадила також Національна бібліотека Білорусії, де створено електронні версії її генеральних каталогів: алфавітних — білоруською, російською та українською мовами, алфавітних каталогів мікрофішів дисертацій, нот, систематичного каталогу картографічних видань. Імідж-каталоги Національної бібліотеки Білорусії містять 2,5 млн карток (<http://image.nlb.by/>) (рис. 4).

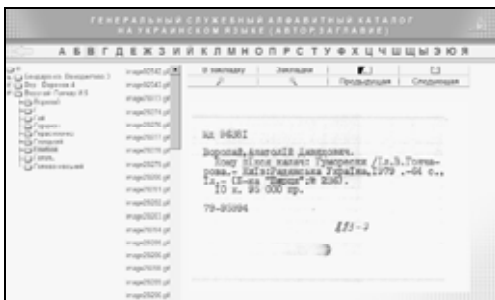


Рис. 4

Цікавий приклад створення імідж-каталогу фонду рідкісних книг та рукописів реалізовано в Бібліотеці Вроцлавського університету (Польща), в якому представлено оцифровану версію карткового каталогу, що містить фотографії

титульних аркушів старовинних та рідкісних книг, створеного у 1947 р. (<http://www.bu.uni.wroc.pl/katalog/>) (рис. 5).



Рис. 5

Специфіка фондів Національної бібліотеки України імені В. І. Вернадського та їх каталогів така, що виконання ретроконверсії тільки за рахунок безпосереднього введення з клавіатури або імпортування неможливе. Бібліотека розпочала власний проект з ретроспективної конверсії шляхом сканування традиційних карткових каталогів, оскільки це разове завдання, а високошвидкісні сканери поточного введення не дешеві та складні в обслуговуванні, то для створення масиву графічних зображень карток доцільним вбачається залучення сторонніх організацій. Для виконання робіт з двостороннього потокового сканування масиву карток у НБУВ було залучено фірму PIRIT Systems. З метою відпрацювання технології ретроконверсії фахівці бібліотеки обрали каталог дисертацій з фондів НБУВ, що містить приблизно 97 тис. карток та більш ніж 1200 роздільників (рис. 6).



Рис. 6. Пошук у сканованому каталозі дисертацій на сайті НБУВ (http://www.nbu.gov.ua/db/scan_dis.html)

Аналіз отриманого масиву карток показав, що якість графічних зображень недостатня для запровадження оптичного розпізнавання тексту, оскільки велика кількість карток написана від руки або надрукована на дуже темному папері. У зв'язку з цим пошук у сканованому каталозі (імідж-каталозі) запропоновано проводити за роздільниками. База даних роздільників створена за допомогою АІБС ІРБІС [8].

Для представлення імідж-каталогу на веб-сайті НБУВ (<http://www.nbu.gov.ua>) розроблено власне програмне забезпечення за допомогою PHP — однієї з найпоширеніших мов, що використовуються у сфері веб-розробки. Можливість швидкого пошуку роздільника (у нашому випадку — прізвища автора) та перегляду списку всіх роздільників реалізовані за допомогою програмного модуля WWW ISIS, що входить до пакету прикладних програм CDS/ISIS [9]. Таке сполучення веб-орієнтованого програмного забезпечення і

