

**Т. Кузьмінч,**

заст. директора з інформаційних ресурсів Національної бібліотеки Білорусі

**С. Пугачова,**

зав. відділу індексування документів Національної бібліотеки Білорусі

## **Змістове опрацювання документного потоку в умовах корпоративної взаємодії бібліотек Білорусі**

Розглянуто особливості змістового опрацювання документного потоку у бібліотеках Білорусі, особливості взаємодії на початковому етапі впровадження системи корпоративної каталогізації. Особливу увагу приділено досвіду роботи Національної бібліотеки Білорусі, у якій ведуться карткові та електронні каталоги, використовуються класифікаційні мови (УДК, ББК, ДР НТІ), дескрипторна мова з контрольованою лексикою і мова предметних рубрик, Національна база даних авторитетних записів.

**З**містове опрацювання документів – головний складник достовірної і вичерпної інформації, зафіксованої у цих документах і наданої споживачам.

У бібліотеках Білорусі упродовж багатьох років традиційно склалася підвищена відповідальність до представлення інформації в електронному і карткових каталогах. Накопичено багатий інформаційний ресурс (придбаний і згенерований), який потребував упорядкування та представлення як сукупний. Це стосується усіх видів ресурсів і першою чергою – записів електронних каталогів (ЕК).

Лінгвістичні засоби, використовувані бібліотеками Білорусі, доволі різноманітні. Це спадщина традиційних (карткових) інформаційно-пошукових систем, коли їх перелік визначався, насамперед, типом бібліотеки. Не останню роль відіграва-

ли і суб'єктивні чинники. Автоматизація, що прийшла в бібліотеки Білорусі на початку 1990-х років, збільшила це різноманіття: вибиралися і розроблялися програмні продукти, лінгвістичні засоби, використовувалися різноманітні формати представлення запису. Така тенденція збереглася дотепер, вона ускладнює процеси обміну вторинною інформацією і багато в чому визначає особливості корпоративної взаємодії бібліотек Білорусі.

На сьогодні у Національній бібліотеці Білорусі (НББ) поряд із ЕК продовжують вестися карткові каталоги. ЕК поповнюється за рахунок каталогізації вхідного документного потоку в новій АБІС із використанням формату BELMARC<sup>1</sup>, конверту-

---

<sup>1</sup> АБІС, розроблена на основі LIBIS у середовищі СУБД ORACLE 9i.

вання записів із 13 баз даних, створених у локальному форматі старої АБІС<sup>2</sup> починаючи з 1993 р., отриманих у результаті рекаталогізації національного документа, а також створюваних сьогодні на періодичні видання і продовжувані видання, а також аналітику.

У найближчій перспективі (листопад 2007 р.) частиною ЕК стануть записи, отримані в результаті ретроконверсії генерального алфавітного каталога, які нараховують близько 4 млн карток, що відбивають за мовною і видовою ознаками фонд, накопичений НББ до 1993 р.

Необхідно визнати, що сконвертований масив не відрізнявся високою якістю: частина інформації була загублена або некоректна; практично кожний етап конвертування супроводжувався редагуванням словників і бібліографічних описів. Причинами цього можна вважати не тільки недосконалість конверторів, доробка яких проводилася постійно, але і недостатньо послідовні методичні рішення, прийняті для різноманітних видів документів у базах даних, помилки каталогізаторів, що закономірно виникають на початковому етапі створення ЕК.

Лінгвістичне забезпечення, що є найважливішим елементом будь-якої пошукової системи, у таких умовах стає найбільш уразливим.

Національна бібліотека Білорусі завжди намагалася мати розвинену структуру лінгвістичного забезпечення, що включала б інформаційно-пошукові мови (ІПМ) різноманітних типів. Це пов'язано з тим, що вхідний документний потік, який надходить до НББ, має широкий видовий діапазон, склад читачів різноманітний і, відповідно, є широким спектром запитів, що надходять.

Правильний вибір і якість застосовуваних у бібліотеці лінгвістичних засобів сприяє високому рівню обслуговування читачів. У зв'язку з цим було вирішено використовувати такі пошукові засоби:

- класифікаційні мови (УДК, ББК, ДР НТІ);
- вербальні мови (дескрипторна мова з контрольованою лексикою і мова предметних рубрик).

Рішення про застосування в індексуванні вхідного документного потоку УДК для основного масиву документів, ББК для картографічних і нотних видань було продиктовано традиціями ведення карткових каталогів.

<sup>2</sup> Операційна система MS-DOS, у середовищі ІПС «АСПІД» з використанням СУБД VTRIVE. Як формат представлення бібліографічних даних використовується локальний формат, неуніфікований і не сумісний зі стандартними MARC-форматами.

ДР НТІ став новою класифікаційною мовою, яка раніше не використовувалася у змістовому опрацюванні документів у НББ. Її вибір був зумовлений перспективами використання рубрикатора в обмінно-інформаційних процесах.

Робота з цими мовами регламентована Держстандартами (ГОСТ 7.49-84, ГОСТ 7.59-2003), відповідними таблицями класифікації і спричиняє труднощі лише у прийнятті рішення щодо відбиття змісту «складних» документів. Дотримання принципів і методики індексування, постійне відслідковування змін і доповнень до таблиць, підтримування фахових контактів допомагає точно відбити зміст документа за допомогою цих ІПМ.

Найбільш важким рішенням був вибір дескрипторної ІПМ. Вона, на відміну від мови предметних рубрик, дає змогу здійснювати детальний пошук інформації на визначену тему в машиночитаних бібліографічних базах даних (БД) із установленням різноманітних відношень між термінами, розташованими в єдиному алфавітному ряді.

Дескрипторна мова, як і ДР НТІ, був новим інструментом індексування в НББ. Освоєння методики його створення і формування масиву лексичних одиниць (ЛО) спричиняло значні труднощі. Робота з дескрипторною мовою регламентована ГОСТом 7.66-92, але його використання – занадто складний логічний процес, який потребує не тільки відповідної фахової підготовки, але і «лексичної інтуїції», особливого «відчуття слова».

Знадобилися чималі фахові зусилля з боку індексаторів і понад 10 років роботи, щоб опанувати процесом координатного індексування документів і створити свій словник ключових слів. При цьому аналізувався досвід роботи найбільших вітчизняних і зарубіжних інформаційних центрів. Вихідний масив ключових слів, необхідний для побудови словника, був отриманий шляхом накопичення ключових слів і словосполучень при вільному індексуванні документів. Як довідкові масиви використовувалися тезауруси ІНІСН, ВІНІТІ.

За 1992–1994 роки словник нараховував близько 101 000 ключових слів, що відбивають зміст отриманих документів. Водночас, клавіатурний набір ключових слів індексаторами, відсутність методичних керівництв призвели до великої питомої ваги ненормалізованих термінів, дублювання понять, орфографічних помилок.

У 1995 р. вийшла «Інструкція з індексування вхідного потоку документів для формування масивів електронного каталога», у якій описані пра-

вила відбиття змісту документів за допомогою ІПМ, прийнятих у бібліотеці, описано методику координатного індексування. На додаток до «Інструкції...» індексаторами і службою ведення ІПМ приймаються окремі методичні рішення, що фіксуються у відповідній картотеці.

Накопичений масив дескрипторів став основою для створення авторитетних записів (АЗ) на тематичний предмет і поповнення Національної БД АЗ у режимі поточної каталогізації.

Виконуючи автоматичне уведення в процесі координатного індексування, індексатор витягає терміни, необхідні для відбиття змісту даного документа, із масиву вже наявних у БД АЗ і вносить їх у пошуковий образ документа (ПОД). У разі відсутності терміна в базі індексатор виписує його на робочий аркуш і разом із документом передає для подальшої нормалізації у службу ведення ІПМ редактору-лінгвісту. Якщо термін приймається для внесення в БД, то він із відповідною позначкою повертається індексатору. Таким чином, тільки за ентитетом тематичний предмет БД АЗ щорічно збільшується на 4–5 тисяч нових лексичних одиниць.

Національна БД АЗ надає якісно нові можливості для здійснення змістового опрацювання документів і проведення різноманітних видів пошуку. Її формування розпочато в 1998 р. на базі раніше створених масивів бібліографічних записів. Машинні словники ключових слів були сконвертовані у структуру авторитетних файлів, із цього масиву виділено різноманітні ентитети. Розроблено технологію формування АЗ у процесі каталогізації документів і технологію ведення Національної БД АЗ, створено програмне забезпечення підготовки і використання АЗ.

Національна БД АЗ сьогодні включає найменування таких ентитетів: імена колективів (7479), тематичний предмет (52 751), імена осіб (22 635), адміністративно-територіальна, або географічна назва (7030), торгову марку (товарний знак) (87), родові імена (101), уніфікований заголовок (618), ім'я/заголовок (1606), місце видання (407), форму, жанр або фізичні характеристики документа (2774)<sup>3</sup>. Поповнення бази ведеться в основному в режимі поточної каталогізації, причому різними структурними підрозділами – відділами опрацювання документів, індексування документів, ведення Національної БД АЗ. Такий підхід продик-

тований насамперед жорстким регламентом опрацювання документного потоку, рівнем підготовки спеціалістів, у тому числі мовної і технологічної.

У перспективі ця робота, на нашу думку, має здійснюватися в поточному режимі корпоративної каталогізації документного потоку з використанням не тільки вже сформованого масиву, але й імпортування записів, створених іншими установами, отриманих у результаті конвертування інших БД.

ЕК НББ стане основою зведеного електронного каталога (ЗЕК) бібліотек Білорусі, який поповнюватиметься на корпоративних засадах.

Одна із найскладніших проблем – лінгвістичне забезпечення ЗЕК. Бібліотекам-учасницям (на початковому етапі це НББ, Центральна наукова бібліотека Національної академії наук Білорусі, Республіканська науково-технічна бібліотека) буде надано для використання Національну БД АЗ.

Передбачається два варіанти взаємодії в частині змістового опрацювання документного потоку. Класифікаційні індекси (крім ДР НТІ) бібліотеки можуть запозичати разом із бібліографічним записом або проставлять самостійно для своїх електронних каталогів. Другий варіант передбачений для НББ. На початковому етапі корпоративної взаємодії це, мабуть, найбільш правильне рішення, тому що бібліотеками не тільки традиційно використовуються різноманітні лінгвістичні засоби, але і різноманітні їх версії, модифікації.

Проблемою на шляху корпоративної взаємодії є також різний рівень кваліфікації співробітників і, на жаль, відсутність, на сьогодні, системи спеціальної підготовки, перепідготовки, підвищення кваліфікації каталогізаторів, обміну досвідом. Необхідно також удосконалити методико-технологічне забезпечення.

Більша частина цих проблем буде розв'язуватися НББ, у якій обробляється найбільший за обсягом документний потік, створено і підтримується лінгвістичне забезпечення, яке відповідає сучасному рівню вимог, накопичено унікальний у країні досвід з формування авторитетних записів, функціонує відділ корпоративної взаємодії, у перспективі – Центр корпоративної каталогізації.

Насамкінець можна сказати, що і результати наших досягнень, і складність наших проблем, і очікування наших перспектив однаково корисні нам для подальшого просування вперед у справі забезпечення доступу до достовірної і вичерпної інформації.

<sup>3</sup> Дані на 1 червня 2007 р.