

УДК 025.4

Інна Лобановська,

зав. відділу наукового опрацювання
й каталогізації документів
ДНПБ України ім. В. О. Сухомлинського

Лінгвістичне забезпечення електронного каталогу Державної науково-педагогічної бібліотеки України імені В. О. Сухомлинського

Розглядаються принципи лінгвістичного забезпечення електронного каталогу Державної науково-педагогічної бібліотеки України ім. В. О. Сухомлинського, що має забезпечити його ефективне використання. Описано інформаційно-пошукові мови, які використовуються для розкриття змісту документів і забезпечення їх релевантного пошуку.

В останні роки відбулася й продовжується глобальна інформатизація такої сфери діяльності як освіта. Особливо яскраво це виявилось в бібліотечно-інформаційній справі, де за досить короткий термін змінилося вже декілька поколінь технічних і програмних засобів. Процес інформатизації охопив усі рівні бібліотек – від державних до шкільних. В епоху побудови суспільства знань рівень бібліотечного сервісу вийшов за межі окремих бібліотек. Створення й упровадження автоматизованих бібліотечно-інформаційних систем (АБІС) стало необхідною реальністю.

Державна науково-педагогічна бібліотека України імені В. О. Сухомлинського як головний науково-інформаційний підрозділ Академії педагогічних наук України є державним бібліотечно-інформаційним і методичним центром освітянської галузі країни. Традиційно бібліотеки – це найважливіша складова освітнього процесу як у широкому (загальнокультурному) так і у вузькому (навчальному) сенсі цього слова. Сьогодні в умовах інформаційного буму надзвичайно актуальною є проблема створення оптимального інформаційно-освітнього простору, а також всебічного й грамотного використання інформаційних ресурсів з науковими та освітніми цілями.

З 2003 року в нашій бібліотеці поряд з традиційними картковими каталогами й картотеками ведеться електронний каталог (ЕК). З моменту свого створення ЕК бібліотеки зазнав різноманітних змін з ряду причин: перехід на іншу АБІС „ІРБІС”, введення нового стандарту бібліографічного описування документів, удосконалення лінгвістичних засобів опрацювання інформаційних матеріалів. У зв'язку з упровадженням у практику роботи інформаційно-комунікаційних технологій (ІКТ), перед працівниками відділу наукового опрацювання й каталогізації документів постало питання щодо розроблення нових і удосконалення старих способів аналітико-синтетичного опрацювання інформаційних матеріалів.

Мета статті – обґрунтування необхідності комплексного підходу для лінгвістичного забезпечення електронного каталогу бібліотеки з використанням різних інформаційно-пошукових мов (ІПМ) для максимально повного розкриття змісту документів і якнайповнішого задоволення інформаційних запитів користувачів бібліотеки.

Становлення й розвиток інформаційно-пошукових систем (ІПС), як традиційних, так і автоматизованих, нерозривно пов'язано з якістю лінгвістичного забезпечення (ЛЗ), що є їх невід'ємною частиною. Будь-яка ІПС містить такі елементи: інформаційний масив; інформаційно-пошукова мова, якою перекладається вхідна інформація та запити; правила такого перекладу (індексування); критерії видачі, тобто правила порівняння перекладу запиту на ІПМ вхідної інформації, що визначають відбір інформації для видачі за запитом. Поняття лінгвістичного забезпечення ширше поняття інформаційно-пошукової мови, оскільки крім ІПМ містить методики індексування документів і запитів на них, інструкції та методики їх ведення та використання, а також засоби підтримання ІПМ в автоматизованих системах.

Лінгвістичні засоби забезпечують індексування документів і запитів, проведення ефективного інформаційного пошуку в інформаційно-пошукових системах, формують і структурують бази даних (БД), визначають їхні тематичні межі, сприяють сумісності ЕК різних бібліотек для обміну даними. Як правило, до складу ЛЗ входить декілька інформаційно-пошукових мов, кожна з яких виконує свою функцію в ІПС. Цим пояснюється необхідність досягнення їхньої сумісності в ІПС, що дозволяє формувати інформаційні запити тією ІПМ, якій користувач надає перевагу.

Основними видами ІПМ для розкриття змісту документів є: класифікаційна, предметизаційна та дескрипторна.

Класифікаційна мова призначена для індексування документів та інформаційних запитів за допомогою понять і кодів якої-небудь класифікаційної системи (Універсальна десяткова класифікація, Бібліотечно-бібліографічна класифікація, Десяткова класифікація Дьюї, Міжнародна патентна класифікація тощо). В основі класифікаційних мов лежить систематична класифікація понять, тобто класифікація, що відбиває смислові відношення між поняттями. Класифікаційні мови призначені перш за все для формалізування логічних зв'язків слів природної мови.

Усі ІПМ класифікаційного типу характеризуються рядом властивостей, що утруднюють їх використання під час інформаційного пошуку, особливо з використанням технічних засобів. До таких властивостей можна віднести: попередню координацію (зв'язок) слів і словосполучень у рубриці, практичну неможливість повного й детального розроблення схеми класифікування й підключення її до АІБС, складність оновлення й доповнення, трудомісткість використання під час індексування тощо.

Усе це призвело до створення мов вербального типу. Ці мови використовуються для представлення лексичних одиниць слова і вираження природної мови в їхній орфографічній формі. На відміну від класифікаційних мов, які використовуються для систематизації документів, вербальні мови орієнтовані не на позначення класифікаційних галузей, до яких можна віднести зміст індексованих документів, а на позначення в пошуковому образі документів (ПОД) або в пошуковому образі запитів (ПОЗ) конкретних об'єктів або предметів змісту цих документів. До мов вербального типу належать предметизаційна й дескрипторна мови.

Предметизаційна мова призначена для індексування документів за допомогою рубрикаторів (словників предметних рубрик), а в разі їх відсутності

– за методикою предметизації. В основу предметизаційної мови покладено абетковий перелік предметних рубрик (ПР), який є коротким формулюванням теми природною мовою. Предметизаційна мова є штучною системою і має бути побудована стандартизовано й одноманітно (нормалізовано). Це допомагає користувачеві в пошуку інформації. У той же час предметні рубрики не придатні для глибокого, вірогідного пошуку за темою в машиночитних бібліографічних базах даних. За ПР неможливо знайти який-небудь аспект теми, неможливо перерахувати всі власні назви предметів, про які йдеться в документі тощо. Крім того, досить проблематично отримати готовий список ПР у друкованому вигляді. Інколи деякі побічні теми документа не описуються ПР. Особливо це стосується наукових збірників, матеріалів конференцій, симпозіумів тощо, де трапляються статті, доповіді, що лише віддалено стосуються основної тематики збірника, але становлять інтерес для користувачів конкретної бібліотеки. Найімовірніше, такі побічні теми не будуть відображені ні класифікаційними індексами, ні ПР, але можуть бути описані ключовими словами (КС).

Дескрипторна мова слугує для координатного індексування документів і запитів за допомогою тезаурусів (словників дескрипторів) або визначенням ключових слів з використанням природної мови. В основу дескрипторних ПМ покладено абетковий перелік лексичних одиниць. Комплекс ключових слів є свого роду лексичною моделлю наукового тексту. Функціональна значимість КС визначається тим, що вони є одним з найоптимальніших способів класифікування, зберігання й передавання інформації. Відбиваючи розвиток і термінодинаміку визначеної наукової галузі, комплекс КС є ще й системою відслідковування й розповсюдження сучасної термінології.

КС доцільно використовувати для додаткового розкриття змісту документа на більш глибокому рівні, тобто використовувати ПР для описування основних предметів документа та їхніх аспектів, а КС – для подальшої деталізації, а також описування побічних тем документа. Тоді в ПОД будуть міститися ПР (нормалізована лексика) і КС у редакції автора. Такий підхід зручний для всіх категорій користувачів: як для тих, хто приблизно знає, що хоче знайти (для цього зручно використовувати ПР), так і для тих, кого може зацікавити щось дуже специфічне (пошук відбувається за КС). Особливу цінність список КС може набути на останньому етапі пошуку, коли за запитом знайдено якийсь масив документів. Переглянувши списки КС знайдених документів, можливо одразу визначити потрібні документи, навіть не переглядаючи їх. Таким чином, використання КС збільшує пошукові можливості ЕК. КС слід розглядати як додатковий засіб для збільшення повноти індексування та представлення різних за значимістю тем документа, не описаних іншими ПМ.

Кожна із зазначених ПМ має свої переваги й недоліки У зв'язку з необхідністю забезпечити пошук як за широкотематичними запитамі, так і за запитамі, сформульованими з точністю до детальних понять, ефективними для пошуку за широкотематичними запитамі є тільки класифікаційні ПМ. Для пошуку за детальними запитамі – тільки вербальні ПМ. Виходячи з цього, у складі лінгвістичного забезпечення інформаційно-пошукової системи має бути хоча б одна вербальна і одна класифікаційна ПМ.

Плануючи і створюючи лінгвістичне забезпечення, перш за все необхідно виходити з різноманітних потреб користувачів. Практика роботи бібліотек доводить, що користувачі частіше зацікавлені вести пошук не за класифікаційними системами, а за запитамі тематичного, предметного характеру. При цьому предметні рубрики менш популярні, ніж ключові слова, тому що в бібліотеках немає єдиної системи предметизації, відсутній універсальний словник предметних рубрик, користувачі не володіють методикою предметизації, що ускладнює формулювання пошукового образу запиту, адекватного пошуковому образу документа, не в усіх ЕК реалізується доступ користувачів до внутрішніх словників предметних рубрик тощо. Тому здається доцільним розширити можливості тематичного пошуку за допомогою використання ключових слів, які для користувачів є найбільш зрозумілим пошуковим засобом, за допомогою якого вдається знайти велику кількість релевантних документів.

В умовах відсутності в Україні єдиного науково-дослідного, науково-методичного та інформаційного центру з розроблення й ведення лінгвістичних засобів, бібліотеки на практиці стикаються з цілою низкою складних проблем, не вирішених на теоретичному рівні. До них, наприклад, належить проблема вибору ІПМ. Під час організації електронного каталогу неминуче виникає запитання: якою ІПМ скористатися для розкриття предметного змісту документа: мовою предметних рубрик, мовою ключових слів або їх поєднанням? Незважаючи на принципові відмінності цих мов, практика їх застосування ускладнюється через невирішені проблеми. Ні в теорії предметизації, ні в теорії координатного індексування немає чітких рекомендацій відносно вибору таких параметрів, як довжина лексичної одиниці (слово або словосполучення), сталість словосполучень, прямий чи інверсований порядок наведення лексичних одиниць у словосполученнях тощо. На практиці це призводить до суб'єктивізму, тягне за собою неможливість порівняння результатів семантичного опрацювання документів і запитів, знижує показники повноти й точності інформаційного пошуку.

Проблема нормалізації лексичних одиниць посилюється через те, що на практиці немає засобів контролю лексики, рівною мірою доступної всім бібліотекам. Ефективність реалізації мови предметних рубрик різко падає через відсутність виданих нормалізованих списків ПР. Відсутність інформаційно-пошукових тезаурусів, які б підтримувалися в робочому стані, були придатні для поповнення й внесення змін, враховували галузеву специфіку суттєво ускладнюють процес координатного індексування, а окремі спроби їх створення не вирішують цих проблем.

ДНПБ України ім. В. О. Сухомлинського використовує як лінгвістичні засоби декілька ІПМ, які забезпечують виконання її різноманітних функцій. Паралельне використання декількох ІПМ сприяє швидкому й різноманітному доступу користувачів до інформації найбільш відомою їм інформаційно-пошуковою мовою. У бібліотеці для якнайширшого розкриття змісту документів і надання їх користувачам наукове опрацювання документів здійснюється в рамках моделі „Комплексне індексування”. Для цього використовуються такі ІПМ:

- Класифікаційні мови (УДК, ББК);

➤ Вербальні мови (мова предметних рубрик і мова ключових слів з неконтрольованою лексикою).

Різноманітність інформаційно-пошукових мов, які складають структуру лінгвістичних засобів, пояснюється типо-видовою різноманітністю й тематичною складністю вхідного документопотоку, традиціями та особливостями ДНПБ України ім. В. О. Сухомлинського, навичками працівників і запитам користувачів бібліотеки, а також тим, що жодна з сучасних ІІМ не може в повному обсязі забезпечити повноту й ефективність пошуку релевантної інформації. Комплексне застосування ІІМ дозволяє здійснювати опрацювання політематичного вхідного документного потоку бібліотеки, найбільш повно розкриває семантичний зміст документів. У процесі аналітико-синтетичного опрацювання відбувається створення комплексного пошукового образу документа.

Основною класифікаційною мовою є Універсальна десяткова класифікація (УДК). Частково використовується також ББК (для розстановки документів і ведення систематичного каталогу та систематичної картотеки статей у філії бібліотеки). Вибір УДК як основної інформаційно-пошукової мови ІІС обумовлено такими чинниками:

- * УДК використовується для організації переважної більшості традиційних систематичних каталогів і картотек бібліотеки;

- * УДК є однією з найпоширеніших класифікаційних систем у світі, що постійно розвивається й удосконалюється;

- * УДК є обов'язковою для індексування всіх наукових публікацій, використання в науково-технічних видавництвах, редакціях науково-технічних журналів, в органах науково-технічної інформації, наукових і науково-технічних бібліотеках країни;

- * УДК має хорошу пристосовуваність для автоматизованих технологій;

- * Саме за УДК є можливість здійснювати галузевий пошук інформації в АБІС „ІРБІС”;

- * Використання УДК дозволяє зберегти в ЕК усі позитивні якості традиційних каталогів, які увібрали в себе багаторічний досвід бібліотекарів і бібліографів, і привнести нові можливості пошуку інформації за допомогою використання інформаційно-комунікаційних технологій.

Виходячи з цього, галузевий пошук інформації в ЕК здійснюється через формулювання нормалізованих тематичних рубрик, узятих з таблиць УДК, виражених за допомогою класифікаційних індексів. Пошукова система, побудована за галузевим систематичним принципом, дозволяє відповісти на запити, пов'язані з галуззю в цілому, її окремими частинами або конкретним питанням, темою, якщо вони цікавлять користувача тільки у зв'язку з даною галуззю.

Для систематизації документів у ДНПБ України ім. В. О. Сухомлинського використовуються україномовні таблиці УДК, підготовлені Книжковою палатою України ім. Івана Федорова (К., 2000, зі змінами й доповненнями), а також розроблені фахівцями бібліотеки на їхній основі з урахуванням власних потреб робочі таблиці УДК з окремих галузей знань.

Для розширення пошукових можливостей ЕК бібліотеки застосовується предметний пошук, який на противагу галузевому є міждисциплінарним, міжгалузевим, комплексним. За таким запитом користувачі отримують комплекс усіх документів з предмета, проблеми, теми незалежно від галузей знань, до яких належать усі ці документи. Основою для створення мови предметних рубрик став перелік рубрик абетково-предметного покажчика (АПП) до Центрального систематичного каталогу бібліотеки, організованого за УДК. Використовується також методика традиційної предметизації. У найближчих планах роботи відділу – розроблення авторитетних файлів заголовків предметних рубрик.

Для подальшої деталізації ПОД і ПОЗ використовується мова ключових слів. Це обумовлено такими чинниками:

- ✓ Універсальний тематичний склад інформаційних ресурсів бібліотеки, що робить вельми проблематичним використання жорстко нормованої лексики в ході їх наукового опрацювання;

- ✓ Наявність семантичних труднощів індексування документів і запитів жорстко нормованими термінами для рівня підготовки користувачів ЕК;

- ✓ Відносна простота методики використання лексики ключових слів, близької до природної мови;

- ✓ Можливість доповнення й оновлення списку КС, оскільки до абеткового переліку можна вводити будь-яку лексичну одиницю, необхідну для індексування.

У створенні записів для ЕК ДНПБ України ім. В. О. Сухомлинського застосовується метод вільного координатного індексування.

Вільне координатне індексування означає індексування ключовими словами, вибраними безпосередньо з повного тексту документа й представленими в ПОД у термінології автора без нормалізації, з мінімальним контролем над лексикою і без урахування того, які КС уже використовувалися раніше для індексування близьких за змістом документів. На відміну від методики індексування з використанням тезауруса, де ПОД складається на основі КС документа, але представлених у вигляді стійких понять, зв'язаних з тезаурусом, під час вільного індексування словник КС створюється одночасно з формуванням бази.

Основними особливостями пропонованої моделі смислового опрацювання інформації та формування ПОД є максимальна повнота індексування інформації, можливість описання сфери інтересів користувачів у вигляді переліку основних понять (ключових слів), які предметно незалежні й можуть бути не регламентовані тезаурусами та іншими нормативними словниками. Індексування змісту документа оцінюється з точки зору як поточних, так і передбачуваних потреб користувачів, тому в ПОД відображаються всі найважливіші поняття, що розглядаються в документі. З літературних джерел відомо, що до 51 % інформації втрачається через те, що індикатор не ввів у ПОД необхідний термін.

Слід враховувати, що неконтрольоване і некероване застосування КС призводить до значних втрат у характеристиках повноти й точності пошуку. Для усунення недоліків у використанні КС, фахівцями відділу наукового опрацювання й каталогізації документів розроблено методичні рекомендації,

що визначають режим ведення й удосконалення ПІМ ключових слів. У методичних рекомендаціях докладно відображено правила описування основного смислового змісту документа за допомогою вільного координатного індексування. Згідно з ними в процесі індексування здійснюється інтелектуальний аналіз документа, у результаті якого з нього вибираються поняття, що перетворюються потім у терміни індексування. Методичні рекомендації було розповсюджено серед бібліотек освітянської галузі, де вони дістали схвальні відгуки. Ця технологія становить значний інтерес для бібліотек будь-яких типів і видів, незалежно від складу їхніх фондів і рівня автоматизації, оскільки вона універсальна, орієнтована на користувачів як локальних, так і глобальних телекомунікаційних мереж і призначена сприяти поліпшенню якості смислового пошуку.

З огляду на все вище сказане, можна зробити висновок, що сучасні інформаційно-комунікаційні технології, що впроваджуються в бібліотечне виробництво, загострили проблему доступу користувачів до інформаційних ресурсів. Бібліотека як особливий вид виробництва, що характеризується специфічною технологією та засобами перетворення інформації, повинна мати розвинену структуру лінгвістичного забезпечення, що включає комплекс ПІМ різних типів і відповідні методики індексування. Найважливішими функціями лінгвістичних засобів бібліотечної технології є забезпечення якості інформаційних продуктів і послуг, що виробляються й надаються бібліотеками, а також забезпечення доступу споживачів інформації до інформаційних ресурсів бібліотек у зручній для них формі. На вибір засобів ЛЗ суттєвий вплив мають тип бібліотеки, типологічний склад і галузева структура вхідного документного потоку, специфіка завдань, які стоять перед бібліотекою, склад користувачів і особливості інформаційних запитів, диференціація функцій конкретних ПІМ, обумовлених їхньою внутрішньою структурою.

Література

1. Алешин Л. И. Поиск документов: анализ и новые возможности / Л. И. Алешин // Библиография. – 2002. – №4. – С. 14–22.
2. Воройский Ф. Индексирование документов в АИБС / Ф. Воройский // Библиотека. – 1996. – № 9. – С. 42–44.
3. Воройский Ф. С. Информационные ресурсы АРБИКОН – пора подумать об их качестве / Воройский Ф. С., Острая С. А. // Науч. и техн. б-ки. – 2005. – № 2. – С. 76–83.
4. Воройский Ф. С. Некоторые пути повышения качества поисковых характеристик электронных каталогов / Ф. С. Воройский // Библиотеки и ассоциации в меняющемся мире: новые технологии и новые формы сотрудничества : материалы 3-й Междунар. конф. „Крым – 96”. – 1996. – Т. 1. – С. 286–289.
5. Воройский Ф. С. Основные принципы обеспечения информационного поиска в корпоративных электронных каталогах / Ф. С. Воройский // Информационные технологии, компьютерные системы и издательская продукция для библиотек : Восьмая Международ. конф. и Выставка LIBCOM-2004. 15–19 нояб. 2004 г. – Режим доступа: <http://www.gpntb.ru/libcom4>. – Загл. с экрана.

6. Жарикова Л. А. Возможности тематического поиска в электронном каталоге ГПНТБ СО РАН / Л. А. Жарикова, Г. А. Скарук, А. В. Бобко, Г. И. Сергеева // Науч. и техн. б-ки. – 2002. – № 8. – С. 30–38.
7. Зайцева Е. М. Лингвистическое обеспечение автоматизированных библиотечно-информационных систем: современные требования и направления развития / Е. М. Зайцева // Науч. и техн. б-ки. – 2000. – № 3. – С. 54–57.
8. Моторина А. Н. Лингвистическое обеспечение предметного поиска в электронном каталоге / А. Н. Моторина, А. В. Шкиркова // Библиотека вуза и информатизация общества: возможности и перспективы : Зональная конференция, 13–15 апреля 1998 г. / Южно-Урал. гос. ун-т. – Челябинск, 1998.
9. Оранская Л. И. Некоторые особенности использования дескрипторного поискового языка в библиографической информационно-поисковой системе универсальной библиотеки / Л. И. Оранская // Науч. и техн. б-ки. – 1997. – № 9. – С. 13–23.
10. Пименов Е. Н. Нормативность и некоторые проблемы разработки тезаурусов и других лингвистических средств ИПС / Е. Н. Пименов // НТИ. Сер. 1. Организация и методика информ. работы. – 2000. – № 5. – С. 7–16.
11. Пименов Е. Н. О факторах, влияющих на индексирование: индексирование и предметная область / Е. Н. Пименов // НТИ. Сер. 1. Организация и методика информ. работы. – 2000. – № 2. – С. 15–23.
12. Пирумова Л. ИПЯ – язык, который надо знать : о роли лингвистического обеспечения в развитии информационно-поисковых систем библиотек / Лидия Пирумова // Библиотека. – 2006. – № 10. – С. 61–65.
13. Скарук Г. А. Тематический поиск в электронном каталоге: проблемы лингвистического обеспечения / Г. А. Скарук // Библиотековедение. – 2001. – № 3. – С. 48–57.
14. Смагина Н. Н. Лингвистическое обеспечение электронного каталога : консультация специалиста / Смагина Н. Н. // Вузовские б-ки Алтай. края : сборник. – Барнаул, 2003. – Вып. 3. – С. 40–42.
15. Шутько Е. В. Аналитическая обработка текстовой информации и смысловой поиск в электронных базах данных: проблемы противоречий / Е. В. Шутько // Университетская библиотека: проблемы и решения : пятая Всерос. науч.-практич. конф. – Екатеринбург, 2003. – С. 62–68.
16. Шутько Е. В. О некоторых проблемах лингвистического обеспечения создания баз данных и поиска в электронном каталоге / Е. В. Шутько // Университетская библиотека на рубеже тысячелетий: выбор пути : третья Всерос. науч.-практич. конф. – Екатеринбург, 2001. – Режим доступа: <http://consensus./conferences/unilib2000>. – Загл. с экрана.