

- поиск информации по любому элементу библиографической записи;
- использование авторитетных данных при создании и поиске информации;
- интеграцию ресурсов библиотеки в информационный простор университета, региона и сеть Интернет;
- создание информационного комфорта для пользователей.

УДК 025.4.06

Губренко Ірина Юрїївна,

завідувачка сектором відділу інформаційних технологій
та комп'ютерного забезпечення НБ ХДУХТ

конт.тел.: 349-45-35

e-mail: biblio-hduht@mail.ru

«СТВОРЕННЯ ГАЛУЗЕВОГО ТЕЗАУРУСА В ЛІНГВІСТИЧНОМУ ЗАБЕЗПЕЧЕННІ ЕЛЕКТРОННОЇ БІБЛІОТЕКИ ХАРКІВСЬКОГО ДЕРЖАВНОГО УНІВЕРСИТЕТУ ХАРЧУВАННЯ ТА ТОГІВЛІ»

Активне впровадження в практичну роботу бібліотек автоматизованих технологій ставить високі вимоги до лінгвістичного забезпечення автоматизованих інформаційних бібліотечних систем (АІБС), тому воно знаходиться в стані постійного вдосконалення і є найбільш проблемним аспектом роботи електронної бібліотеки (ЕБ).

У зв'язку із зростанням обсягу документальної інформації організація її швидкого і ефективного пошуку стає все більш актуальною проблемою. При організації автоматизованого пошуку документальної інформації найважливішою є проблема формалізації змісту документа та запиту. При цьому необхідно, щоб пошук здійснювався відповідно до змісту всього тексту документа чи за його пошуковим образом, а як запит найчастіше виступають окремі ключові слова або їх логічні комбінації. Саме на такій технології заснована дія тезауруса, основу якого становлять логічні зв'язки.

Як показує досвід нашої бібліотеки, користувачі частіше зацікавлені вести пошук не за класифікаційними системами, а за запитамі тематичного, предметного характеру. Тому нам здається доцільним розширити можливості тематичного пошуку на основі використання тезауруса, бо він для користувачів є найбільш зрозумілим пошуковим засобом, за допомогою якого вдається знайти велику кількість релевантних документів за тематикою та галузевим профілем Харківського державного університету харчування та торгівлі (ХДУХТ).

Тезаурус – це термін, який широко використовується в галузі інформатики як складова частина інформаційно-пошукових систем. Можна виділити два визначення тезауруса:

1. Тезаурус – це словник, що відображає семантичні відношення між лексичними одиницями дескрипторної інформаційно-пошукової мови (дескрипторами) і призначений для пошуку слів за їх смисловим змістом.

2. Тезаурус – це контрольований словник термінів предметної області, який створюється для поліпшення якості інформаційного пошуку в певній предметній сфері.

І в тому, і в іншому випадку мова йде про словник, що має полегшити пошук необхідної інформації. У словниках можливі два способи розташування слів: за близькістю їх літерного складу і за смисловою близькістю.

Тезауруси створюються саме за смисловою близькістю. Якщо словники, скоріше, розкривають значення будь-якого слова, то тезауруси створюються таким чином, щоб пошукове слово виражало певне якість поняття на основі закладених в ньому взаємозв'язків. Тобто якщо у словнику на основі слова шукають його смисл, то у тезаурусі за заданим смислом здійснюється пошук слова, яке виражає цей смисл.

Тезаурус є невід'ємною частиною пошукової системи і являє собою ієрархічну мережу понять, що відповідають тим чи іншим значенням окремих слів або текстових виразів. Для опису зв'язків між поняттями використовується традиційна для тезаурусів система семантичних відносин. Всі семантичні зв'язки мають ваговий коефіцієнт.

На сучасному етапі розвитку ЕБ відбувається розширення сфер використання автоматичних тезаурусів. При цьому тезауруси є складовою частиною сучасних систем підготовки текстів, здійснюючи лінгвістичну підтримку процесу підготовки й обробки текстів на природній мові.

Серед найбільш перспективних напрямів розвитку автоматичних тезаурусів необхідно визначити такі:

1. Одержання довідки за пошуковим словом. Вказавши слово як джерело запитання, користувач у відповідь отримує потрібний фрагмент словника, який вміщує лінгвістичну інформацію про це слово.

2. Контекстні заміни на вимогу користувачів. У нашому випадку тезаурус підбирає замість одного словосполучення інше, котре користувач визнає як таке, що більш відповідає контексту за смислом або стилем.

У сучасних умовах розвитку інформаційного суспільства користувачам бібліотек необхідні допоміжні засоби орієнтації у потужному потоці інформації, тому необхідні спеціальні методи інформаційної обробки документів, організації швидкого та ефективного пошуку. Саме інформаційно-пошукові тезауруси, розроблені в автоматизованому режимі, призначені для розв'язання цієї проблеми.

Стан галузевого тезауруса на початку виконання проекту.

Упровадження в 2005 р. у практику роботи бібліотеки ХДУХТ комп'ютерних технологій значно розширило межі послуг, що надаються читачам. На основі технологічної розробки АІБС «УФД/Бібліотека» було закладено фундамент для створення ЕБ. Для надійного функціонування ЕБ необхідно належним чином організувати лінгвістичне забезпечення. Тому було створено кілька пошукових БД на основі таблиць УДК, ББК, словників, енциклопедій та «Інформаційно-пошукового тезауруса НІБ України». Отже, для більш досконалого пошуку документів потрібно було створити галузевий інформаційно-пошуковий тезаурус.

Підготовчий етап з розробки галузевого тезауруса. Розділи «Плоди та овочі» і «Рослини та прянощі».

Першим кроком на шляху, який пройшла бібліотека ХДУХТ, до автоматизованого обслуговування користувачів було створення ЕК та лінгвістичного забезпечення до нього.

Основу формування ЕК бібліотеки ХДУХТ становить автоматизована інформаційно-бібліотечна система (АІБС) «УФД/Бібліотека». АІБС «УФД/Бібліотека» надає широкі можливості користувачам. Такими можливостями є звернення до різноманітних фільтрів та різнобічний пошук інформації на основі лінгвістичного забезпечення за допомогою мов описування та маніпулювання даними. Це мови запитів, засоби та методи індексування, словники. На основі класифікаційних мов здійснюють вибірку за ББК та УДК, а за допомогою мови запитів користувачі знаходять авторів та назви документів, тип літератури, вид документа, дату його видання, місце видання тощо. Найвужчий пошук в АІБС «УФД/Бібліотека» проводиться за допомогою ключових слів, вони якнайповніше розкривають зміст документа. Створення інструменту для швидкого пошуку потрібної інформації – головний напрям розвитку ЕБ ХДУХТ. Саме тому відділом інформаційних технологій та комп'ютерного забезпечення бібліотеки було вирішено створити галузевий тезаурус ХДУХТ. Робота над тезаурусом почалась з розділів «Плоди та овочі» і «Рослини та прянощі», де міститься термінологія галузевих напрямів університету – «Товарознавство плодоовочевих та смакових товарів» і «Ресторанний бізнес».

Робота над галузевим тезаурусом саме з цих розділів почалась не випадково. Колективом авторів – провідних науковців та фахівців нашого університету – були видані книги: Плодоовочеві та смакові товари: довідник / Укл. В. М. Михайлов, Л. Д. Льовшина, Г. І. Дюкарева, О. В. М'ячиков.– Х.:ХДУХТ, 2006.– 262с.; Льовшина Л. Д. Товарознавство плодоовочевих товарів, пряно-ароматичних рослин та прянощів: навч. посіб.: РМОіНУ/ Л. Д. Льовшина, В.М. Михайлов, О.В. М'ячиков.– К.: Ліра-К, 2010.– 388с. Термінологію із цих видань положено в основу тезауруса.

Галузевий тезаурус ХДУХТ розроблявся з урахуванням:

- завдань, що стоять перед галузями «Товарознавство плодоовочевих та смакових товарів» і «Ресторанний бізнес»;
- інформаційних потреб користувачів бібліотеки з урахуванням напрямів навчального процесу, стану виконання читацьких запитів.

Створення докладного галузевого тезауруса – це значна за обсягом робота, що вимагає багато часу і залучення широкого кола фахівців. Тезаурус має вигляд словника, в якому в певному порядку розташовані терміни і зафіксовані зв'язки між ними.

Як правило, тезаурус складається з алфавітного і систематичного покажчика. Алфавітний покажчик містить алфавітний перелік дескрипторних статей. Дескрипторна стаття може мати таку структуру:

- заголовний дескриптор;
- ключові слова з класу еквівалентності;

- дескриптори, що підпорядковуються заголовному;
- дескриптори, асоційовані з заголовним.

Асоціації можуть бути таких видів: причина – наслідок; сировина-продукт; процес – об'єкт; процес – суб'єкт; властивість – носій – властивості, функціональна подібність тощо.

Систематичний показчик служить для розкриття, обліку та контролю парадигматичних відносин між дескрипторами. Для кожного інформаційно-пошукового дескриптора повинні бути знайдені відповідні родові та видові дескриптори. Причому попередньо необхідно сформулювати критерії, що визначають число рівнів ієрархії. Визначення ступеня відображення парадигматичних відносин залежить від вимог, що пред'являються до системи. Чим ширше ці відносини відображені, тим вище повнота видачі, але одночасно відбувається збільшення словникового запасу, ускладнюється система посилань і може статися так, що зменшиться точність пошуку.

Процес створення галузевого тезауруса ХДУХТ мав декілька етапів. І на кожному етапі опанувались нові форми роботи та набувався чималий практичний досвід.

Перший і головний етап – пошук джерел та накопичення галузевої інформації, на основі якої і створюється словник. Створення словника припускає визначення термінології. Тому було розглянуто найбільш інформативні документи щодо галузі товарознавства плодоовочевих та смакових товарів, таких як – «Товарознавство плодоовочевих товарів, пряно-ароматичних рослин та прянощів» і «Плодоовочеві та смакові товари». Вибиралися слова, що вживаються в цих джерелах.

Другий етап – формування пошукових слів. Зі словників були взяті пошукові слова. При відборі цих слів враховувалися інформативність слова та роль слова в цій предметній сфері.

Третій етап – виділення дескрипторів і виявлення підпорядкованості та асоціативних відносин.

Четвертий етап – введення термінів до тезауруса. При введенні термінів враховувалась додаткова граматична інформація: орфографія, особливі функції слова чи словосполучення.

Заключний, п'ятий етап – це обов'язковість перевірки: усі терміни були враховані, проаналізовані та введені до тезауруса у вигляді, що дає можливість задіяти максимальну кількість алгоритмів, реалізованих у програмі.

При створенні галузевого тезауруса ХДУХТ інформацію закладено спочатку за алфавітним принципом (це дає можливість зручніше та швидше заносити інформацію до БД), а потім за галузевим відношенням термінів. Тезаурус створювався двома мовами, щоб в умовах реальної двомовності в країні якнайповніше задовольнити запити користувачів бібліотеки.

На цей час текстовий файл тезауруса становить собою контрольовану систему рубрик, пов'язаних між собою ієрархічними і асоціативними зв'язками. Галузевий тезаурус ХДУХТ є структурою, яка постійно доповнюється даними: в міру необхідності (за поданням практиків – працівників відділу інформаційних технологій і комп'ютерного забезпечення та інформаційно-

бібліографічного відділу) він поповнюється новими, більш дрібними рубриками із встановленням ієрархічних та асоціативних зв'язків.

Таким чином, створенням галузевого тезауруса ХДУХТ зроблено значний внесок у розвиток бібліотеки і удосконалено інформаційне забезпечення навчального процесу. Створення галузевого тезауруса для розвитку лінгвістичного забезпечення і зміцнення пошукових можливостей ЕБ в інформаційному просторі ХДУХТ оптимально забезпечило навчальний і науковий процеси університету. Розглянуто знакову структуру галузевого терміну за навчальними та науковими напрямками ХДУХТ та виявлено засоби відображення терміносистеми у галузевих словниках різних типів та інших документах: у традиційному каталозі відповідні відсилочні картки з помітками «Дивись також», словники і посібники, видані університетом ХДУХТ, створені раніше у автоматизованому режимі ключові слова, словник «Харчування – ХДУХТ», шифри УДК та ББК; розроблено принципи побудови автоматизованого галузевого тезауруса на основі Міждержавного стандарту «Тезаурус інформаційно-пошуковий одномовний», реалізовано чотири типологічні словарні функції – систематизуючу, довідкову, навчальну і наукову – та відповідні їм лексично-графічні параметри; організоване бібліотечне обслуговування з використанням сучасних можливостей ЕБ на основі удосконалення лінгвістичного забезпечення – створення галузевого тезауруса.

Література

1. Архангельская В. А. Информационно-поисковый тезаурус по экономике и демографии [Текст] / В. А. Архангельская, С. В. Базарнова. – НТИ, 2001. – №7. – С 24-32.
2. Жмайло С. В. К разработке современных информационно-поисковых тезаурусов [Текст] / С. В. Жмайло. – НТИ. – 2004. – № 1. – С. 23-31.
3. Тезаурус информационно-поисковый одноязычный. Правила разработки: структура, состав и форма представления: Межгосударственный стандарт 7.25. – Минск: Межгосударственный совет по стандартизации, метрологии и сертификации, 2001.
4. Шемакин Ю. И. Тезаурус в автоматизированных системах управления и информации [Текст] / Ю. И. Шемакин. – М: Военное изд-во Министерства обороны СССР, 1974. – 192 с.

УДК 025.34:004.65

Єпіфанова О. В., директор бібліотеки

Савельєва В. В., заст. директора бібліотеки

Наукова бібліотека

Східноукраїнський національний університет ім. Володимира Даля

Адреса: квартал Молодіжний, 20А, м. Луганськ, 91034, Україна

Тел.: (0642) 41-52-61

E-mail: library@snu.edu.ua

Сайт: <http://library.snu.edu.ua>