

БАЗА ДАНИХ АНГЛОМОВНОЇ ТА УКРАЇНОМОВНОЇ ТЕРМІНОЛОГІЇ У СФЕРІ ГЕОІНФОРМАЦІЙНИХ СИСТЕМ

У статті запропоновано розробку бази даних англomовної та українomовної термінології у сфері геоінформаційних систем. Розроблено математичну модель вибору виду системи управління базою даних (СУБД) за багатьма критеріями, на основі якої робиться висновок про необхідність використання реляційної СУБД Firebird.

Постановка завдання. Подальша інтеграція України до Європи, розвиток взаємовідносин з іншими країнами у сфері геоінформаційних технологій та участь нашої країни у ряді спільних проєктів привели до необхідності взаємного обміну науково-технічною та геопросторовою інформацією з іншими країнами. При цьому також виявилися деякі розбіжності при тлумаченні певних англomовних і українomовних термінів, що стосуються обробки геопросторової інформації. Частково цю проблему було вирішено шляхом прийняття ряду нормативно-правових документів зі стандартизації та уніфікації термінів. Однак на даний час при оформленні науково-технічної документації існують ще певні труднощі з правильним перекладом технічних термінів, які призводять до неоднозначності розуміння поданих матеріалів.

Виникає ряд аналогічних труднощів і під час адаптації спеціалізованого програмного забезпечення обробки геоінформаційних даних (ArcGIS, MapInfo, ArcInfo, GeoDraw, ArcView) та технічної документації на нього. Зважаючи на те, що програмісти, які виконують адаптацію, користуються, як правило, загальними словниками, тому переклад не завжди правильний з технічного та математичного погляду.

Перераховані проблеми призводять до необхідності створення єдиного спеціалізованого тлумачного англо-російсько-українського словника геоінформаційних термінів. Оскільки геоінформаційні системи безпосередньо належать до класу комп'ютерних інформаційних технологій і їх застосування без обчислювальних засобів майже неможливе, то найбільша необхідність використання такого словника виникає під час роботи за комп'ютером. Це і зумовлює необхідність розробки тлумачного словника як спеціалізованої бази даних. Такий підхід дозволяє досягнути ряду переваг, а саме: впорядковане зберігання інформації, її захист, можливість доступу до термінів за різними видами запитів, легке нарощування словника та можливість застосування в подальшому елементів штучного інтелекту.

Таким чином, розробка тлумачного словника геоінформаційних термінів як спеціалізованої бази даних є актуальним науковим завданням.

Аналіз останніх публікацій і досліджень. Розробка бази даних спеціальних термінів має два аспекти: філологічний та технічний. Філологічний пов'язаний з відбором основних термінів та їх тлумаченням, а технічний – з організацією та управлінням базою даних.

Аналіз публікацій з філології показав, що, незважаючи на ряд споріднених наукових досліджень [1, 2], семантико-тематична структура англomовної термінології у сфері геоінформаційних технологій достатньою мірою не визначена і вимагає подальших

досліджень. Ця проблема є предметом окремої публікації і в даній статті розглядається лише концептуально як частина загальної методики.

Аналізу різних проблемних питань розробки баз даних (БД) присвячена велика кількість наукових праць як вітчизняних [3 – 6], так і зарубіжних вчених [7 – 10]. Їх досягнення мали великий вплив на розвиток усіх сфер діяльності і стали основою фундаментальних досліджень у галузі систем підтримки прийняття рішень та штучного інтелекту (бази знань). Однак, незважаючи на значну кількість робіт з даної проблематики, у науковій літературі недостатньо повно розроблено математичне обґрунтування вибору ефективної системи управління базами даних за багатокритерійним принципом.

Формулювання цілей статті. Метою дослідження є обґрунтування розробки бази даних англійської та української термінології у сфері геоінформаційних систем та обґрунтування математичної моделі вибору СУБД для побудови системи інформаційного забезпечення і технологій баз даних.

Для досягнення поставленої мети необхідно: визначитися з типом та кількістю даних, що підлягають впорядкуванню; обрати типологію бази даних; проаналізувати основні типи і види СУБД; розробити математичну модель вибору СУБД; вибрати СУБД для інформаційного забезпечення підтримки прийняття рішень з ефективного перекладу геоінформаційних термінів.

Виклад основного матеріалу. Практично всі основні ідеї сучасних інформаційних технологій базуються на концепції баз даних. Згідно з цією концепцією основою інформаційних технологій є дані, які повинні бути організовані у бази даних з метою адекватного відображення змін сучасного світу і задоволення інформаційних потреб користувачів.

Одним з основних понять у теорії баз даних є поняття інформації. Під нею розуміють будь-які дані про певні події, процеси й об'єкти. Для інформаційних технологій, які використовуються при мовному перекладі, дані – це інформація про слова, словоформи, словосполучення, синоніми та пояснення до них. Вона може бути подана у дискретному або фіксованому вигляді і є зручною для зберігання, обробки комп'ютерними системами, а також для передачі по каналах зв'язку.

Сукупність взаємопов'язаних даних такого об'єму, який дозволяє використовувати їх належним чином для певного програмного забезпечення у конкретній предметній галузі, називають базою даних.

У науковій літературі [5, 6] існує така класифікація БД:

за структурою інформації, яку вони зберігають (документографічні; фотографічні; числові; повнотекстові; лексикографічні);

за призначенням інформації (БД ділової інформації, інформації для спеціалістів, масової інформації та інші);

за способом доступу (БД, що доступні через мережі; БД, що існують у комунікативних форматах; локальні БД);

за типом логічної моделі даних (файлові, реляційні, об'єктно-орієнтовані та об'єктно-реляційні БД).

Важливим елементом БД є додатки. Вони являють собою окремі програми або комплекси програм, які реалізують автоматизацію вирішення прикладних завдань обробки

даних. Оскільки одні і ті ж дані можуть використовуватися для вирішення багатьох завдань, то і додатків до однієї і тієї ж бази даних може бути багато. Усі додатки, що працюють з однією і тією ж базою даних, повинні функціонувати коректно, не заважаючи один одному і враховуючи всі зміни, які вносяться іншими. Така координація їх роботи здійснюється системою управління базами даних. Основною функцією будь-якої СУБД є підтримка незалежності, цілісності і непротиворічності даних в умовах колективного користування. Крім того, СУБД – це програмне забезпечення, за допомогою якого можна: визначати базу даних та її структуру; організувати виконання різноманітних, нефіксованих заздалегідь запитів; здійснювати підтримку цілісності і забезпечувати безпеку даних; контролювати процеси спільного доступу до даних; відновлювати інформацію баз даних.

У середовищі СУБД виділяють декілька компонентів: дані, користувачі, апаратне забезпечення, програмне забезпечення, процедури. Схематично ці компоненти показані на рис. 1.



Рис. 1. Компоненти середовища СУБД

Зрозуміло, що основним компонентом середовища СУБД є дані, а інші елементи даної структури існують для того, щоб користувачі отримували інформацію на необхідному рівні. Серед загальних характеристик СУБД можна виділити такі:

всі СУБД включають мову визначення даних, за допомогою якої можна задати базу даних, її структуру, типи даних, а також засоби завдання обмежень для інформації, яка зберігається;

СУБД дозволяють вставляти, видаляти, оновлювати і здобувати інформацію із баз даних за допомогою мови управління даними, яка називається мовою запитів. Вона дозволяє користувачу формувати різні за змістом запити до бази даних. Сучасні СУБД користуються єдиною мовою – SQL (Structured Query Language – мова структурованих запитів). Мова SQL дозволяє визначати схему БД і маніпулювати даними. Вона має всі необхідні засоби для роботи з БД, починаючи від її створення і до забезпечення базового інтерфейсу користувачів з базами даних;

більшість СУБД можуть працювати на комп'ютерах з різною архітектурою і під різними операційними системами, причому на роботу користувача при доступі до даних тип платформи значення практично не має.

У той же час зрозуміло, що кожна із СУБД має свої специфічні характеристики. Вони відрізняються функціональним набором можливостей, підтримують різний режим роботи з БД, мають різні засоби захисту інформації, технологію обробки даних тощо. У цілому в науковій літературі СУБД класифікують так, як це показано на рис. 2.

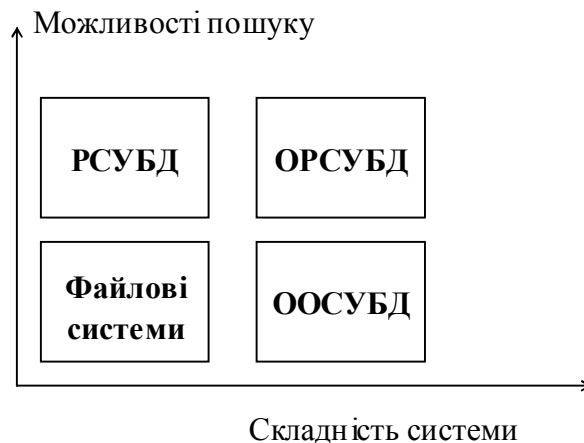


Рис. 2. Класифікація СУБД

Дана класифікація відносить у нижній лівий квадрат програмне забезпечення, яке використовується для обробки простих даних, які не передбачають вимог щодо виконання запитів до даних. До них відносяться такі пакети для обробки текстів, як Word, WordPerfect і Framemaker, які використовують базову операційну систему для отримання доступу до найважливіших функціональних можливостей будь-яких СУБД.

У верхньому лівому квадраті розміщене програмне забезпечення, яке використовує прості дані, що вимагають виконання складних запитів. Практика показує, що для використання баз даних, які працюють з великою кількістю слів та словосполучень, найбільш прийнятним типом систем управління БД є реляційні СУБД (РСУБД).

До загальних переваг РСУБД можна віднести такі: проста структура таблиць та легкість їх читання; простий користувацький інтерфейс; наявність великої кількості інструментів для кінцевих користувачів (макросів, скриптів); простота зміни й додавання нових даних і записів; прямий доступ до даних, що забезпечує їх швидку й ефективну обробку; оптимізованість для запитів й аналізу та інші.

Природно, що у реляційній моделі є також ряд недоліків, до яких належать: обмежена гнучкість керування запитом й даними; повільний послідовний доступ; труднощі моделювання складних відносин даних, оскільки для цього часто необхідні кваліфіковані програмісти БД; необхідність подання складних відносин у вигляді процедур у кожній програмі, що звертається до бази даних; зниження ефективності через необхідність починати спочатку збирати структуру даних щоразу, коли відбувається звертання до них.

Практично всі існуючі реляційні СУБД належать до так званих мережних СУБД, які призначені для підтримки режиму роботи з базою даних великої кількості користувачів і підтримки можливості децентралізованого збереження даних.

У нижньому правому квадраті знаходиться програмне забезпечення, що обробляє складні дані, які також не мають значних вимог виконання запитів до них. Для таких додатків найбільше підходять об'єктно-орієнтовані СУБД (ООСУБД). ООСУБД, будучи більш новою розробкою в порівнянні з реляційними моделями, забезпечують можливість багатобічного й комплексного опису реального світу й створення структури даних, які користувачі зможуть легко зрозуміти. У спрощеному вигляді різницю між цими типами моделей можна уявити так: об'єкт зберігає інформацію про себе (всі свої атрибути) всередині себе, а не в безлічі взаємозалежних таблиць.

Об'єктно-орієнтована СУБД має ряд переваг: інкапсуляція; підтримка рівнів генералізації, об'єднань й асоціацій; інтегрування з методами імітаційного моделювання; високий рівень цілісності даних та інші. Серед недоліків ООСУБД слід зазначити: складність розробки та побудови; критичність до вибору об'єктів; труднощі при обміні з іншими видами даних; повільніше виконання.

ООСУБД підвищують рівень абстракції людини, що, у свою чергу, прискорює пошук необхідної інформації, однак, їх широке впровадження вимагає великих капіталовкладень, тому перехід до об'єктно-орієнтованих БД є перспективним етапом розвитку технологій баз даних.

Новітньою розробкою в галузі логічних моделей даних є об'єктно-реляційна СУБД (ОРСУБД), що подана у верхньому правому квадраті на рис. 2. Вона надбудовує об'єктно-орієнтовану структуру над міцною реляційною базою даних. Реляційна база даних розширюється за рахунок програмного забезпечення, що включає об'єктно-орієнтовані лінії поведінки, але дані при цьому не інкапсульовані. Інформація в базі даних подана в таблицях, але в деяких з обов'язкових колонок можуть бути дані більш складної структури – дані абстрактного типу. В ОРСУБД широко використовуються такі об'єктно-орієнтовані компоненти, як система типів, що розширюється користувачем, інкапсуляція, наслідування, поліморфізм, динамічне пов'язування методів, використання складових, а також підтримка ідентичності об'єктів. Однак поки що РСУБД залишаються основним типом систем управління базами даних. Ними користуються для вирішення складних завдань різного напрямку діяльності, замінити їх практично не можливо не тільки тому, що для цього необхідна велика кількість ресурсів, але й тому, що для якісної заміни необхідно розробити програмний продукт, який має безперечні переваги над РСУБД.

Таким чином, з чотирьох названих типів СУБД для вирішення завдання розробки бази даних термінології у сфері геоінформаційних систем доцільним є вибір РСУБД.

Серед найбільш розповсюджених РСУБД можна назвати такі: MySQL, MsSQL, Firebird, Sybase, Interbase. Для вибору виду РСУБД, що буде використовуватися для управління БД геоінформаційної термінології, необхідно провести оцінку їх ефективності за сукупністю таких основних критеріїв: функціональна повнота; швидкодія; надійність; об'єм використовуваної пам'яті; вартість; складність використання; захищеність (табл. 1). Критерії можуть бути як кількісні, так і якісні.

Таблиця 1

Критерії ефективності СУБД

№ з/п	Критерії	MySQL	MS SQL Server	Firebird	Sybase	Interbase
1	2	3	4	5	6	7
1	Функціональна повнота	дуже висока	середня	висока	низька	дуже низька
2	Швидкодія	1024 Кб/с	512 Кб/с	768 Кб/с	256 Кб/с	128 Кб/с
3	Об'єм пам'яті, що використовується	5,8 Гб	7 Гб	6,4 Гб	5,6 Гб	5 Гб

1	2	3	4	5	6	7
4	Надійність	середня	дуже висока	висока	дуже низька	низька
5	Захищеність	висока	дуже висока	середня	низька	дуже низька
6	Складність використання	низька	середня	дуже низька	дуже висока	висока
7	Вартість установа та налаштування	2600 грн.	3000 грн.	2400 грн.	3100 грн.	3100 грн.

Для побудови математичної моделі введемо такі позначення: змінну $x_i = \{0;1\}$, яка набуває два значення: $x_i = 1$, коли вибрано i -й вид РСУБД; 0 – у протилежному випадку.

Кожний вид РСУБД характеризується низкою показників: F_i – функціональна повнота i -го виду РСУБД; H_i – швидкодія i -го виду РСУБД; V_i – об’єм пам’яті, що використовується i -м видом РСУБД; N_i – надійність i -го виду РСУБД; K_i – захищеність i -го виду РСУБД; S_i – складність використання i -го виду РСУБД; Z_i – вартість i -го виду РСУБД (вартість наведено на момент написання статті).

Деякі з наведених показників визначаються якісно, тому їм слід надати кількісні значення за якоюсь шкалою, наприклад пронормувати в інтервалі від 0 до 1.

Задамо часткові критерії ефективності, а також обмеження.

Необхідно визначити вид РСУБД за наведеними критеріями з урахуванням обмежень.

Математична модель має такий вигляд.

Часткові критерії:

максимальна функціональна повнота

$$F = \max \sum_{i=1}^n F_i \cdot x_i, \quad (1)$$

максимальна швидкодія

$$H = \max \sum_{i=1}^n H_i \cdot x_i, \quad (2)$$

максимальна надійність

$$N = \max \sum_{i=1}^n N_i \cdot x_i, \quad (3)$$

максимальний об’єм пам’яті, що використовується

$$V = \max \sum_{i=1}^n V_i \cdot x_i, \quad (4)$$

мінімальна вартість

$$Z = \min \sum_{i=1}^n Z_i \cdot x_i, \quad (5)$$

мінімальна складність використання

$$S = \min \sum_{i=1}^n S_i \cdot x_i, \quad (6)$$

максимальна захищеність

$$K = \max \sum_{i=1}^n K_i \cdot x_i. \quad (7)$$

У формулах (1 – 7) n визначається кількістю програмних продуктів РСУБД, що тестуються. Для даних, поданих у табл. 1, $n=5$.

Обмеження:

необхідно обрати тільки один вид РСУБД

$$\sum_{i=1}^n x_i \geq 1, \quad (8)$$

швидкодія РСУБД повинна бути не меншою від заданої $H_{зад}$

$$\sum_{i=1}^n H_i x_i \geq H_{зад}, \quad (9)$$

надійність РСУБД повинна бути не меншою від заданої $N_{зад}$

$$\sum_{i=1}^n N_i x_i \geq N_{зад}, \quad (10)$$

об'єм пам'яті, що використовується РСУБД, повинен бути не меншим від заданого $V_{зад}$

$$\sum_{i=1}^n V_i x_i \geq V_{зад}, \quad (11)$$

вартість закупівлі РСУБД повинна бути не більшою від заданої $Z_{зад}$

$$\sum_{i=1}^n Z_i x_i \leq Z_{зад}. \quad (12)$$

При введених позначеннях та обмеженнях (1 – 12) отримаємо задачу вибору РСУБД, що належить до задач багатокритерійного лінійного дискретного програмування з булевими змінними.

Розглянемо метод розв'язання багатокритерійної задачі. Оскільки часткові критерії мають різні одиниці вимірювання, діапазони зміни, значення екстремумів, то їх слід привести до одних відносних величин за рахунок введення функції корисності часткових критеріїв, що має вигляд:

$$R_j(x) = \frac{k_j(x) - k_{jz}}{k_{jk} - k_{jz}}, \quad (13)$$

де $k_j(x)$, k_{jz} , k_{jk} – відповідно поточне, найгірше та найкраще значення j -го часткового критерію.

У разі, коли $n = 7$, що відповідає даним табл. 1, узагальнений критерій матиме вигляд:

$$W_j = \max \frac{1}{7} \sum_{j=1}^7 \lambda_j R_j(x), \quad (14)$$

де λ_j – вагові коефіцієнти, що відповідають умові

$$\sum_{j=1}^7 \lambda_j = 1. \quad (15)$$

Моделювання й результати аналізу. Проведемо вибір виду РСУБД для БД геоінформаційної термінології. Розрахунки наведені у табл. 2.

Таблиця 2

Значення функцій корисності часткових критеріїв ефективності РСУБД

№ з/п	Критерії	MySQL	MS SQL Server	Firebird	Sybase	Interbase	Ваговий коефіцієнт
1	Функціональна повнота	1	0,5	0,75	0,25	0	0,08
2	Швидкодія	0,5	1	0,75	0	0,25	0,16
3	Об'єм пам'яті, що використовується	1	0,5	0,75	0,25	0	0,10
4	Надійність	0,75	1	0,5	0,25	0	0,09
5	Захищеність	0,75	1	0,5	0,25	0	0,05
6	Складність використання	0,75	0,5	1	0	0,25	0,20
7	Вартість установа та налаштування	0,75	0	1	0,25	0,5	0,32
	Узагальнюючий показник	0,755	0,49	0,845	0,16	0,25	

З проведеного дослідження видно, що для ефективного управління БД геоінформаційних термінів необхідно використовувати РСУБД Firebird.

До переваг РСУБД Firebird можна віднести і те, що вона надає контрольований доступ до бази даних за допомогою: системи безпеки, яка перешкоджає несанкціонованому доступу до інформації БД; системи підтримки цілісності бази даних, яка забезпечує непротиворічний стан інформації, що зберігається; системи управління паралельною роботою додатків, яка контролює процеси їх спільного доступу до БД; системи відтворення, яка дозволяє відтворити БД до попереднього непротиворічного стану; доступного користувачам каталогу, що має опис інформації, яка зберігається у базі даних.

Висновки. Таким чином, актуальним завданням сьогодення є розробка бази даних англійської та української термінології у геоінформаційній сфері, характеристики якої залежать від виду та типу системи управління. Аналіз типів систем управління БД показав доцільність використання РСУБД.

Для обґрунтування вибору виду РСУБД розроблено математичну модель багатокритерійного вибору, яка на відміну від існуючих дозволяє приймати рішення з урахуванням кількісних та якісних критеріїв.

Проведене дослідження за допомогою математичної моделі показало, що для ефективного управління БД геоінформаційних термінів необхідно використовувати РСУБД Firebird.

Подальших досліджень потребують філологічні питання аналізу семантико-тематичної структури англійської термінології у сфері геоінформаційних технологій для наповнення БД та питання формування ключових запитів.

СПИСОК ЛІТЕРАТУРИ

1. Куделько З. Б. Англійська терміносистема ринкових взаємин: синтагматичні та парадигматичні особливості : автореф. дис. на здобуття наук. ступеня канд. філол. наук : спец. 10.02.04 / З. Б. Куделько / Львівськ. нац. ун-т ім. І. Франка. – Львів, 2007. – 20 с.
2. Гутиряк О. І. Англійська термінологія маркетингу: структура та семантичні характеристики : автореф. дис. на здобуття наук. ступеня канд. філол. наук : спец. 10.02.04 / О. І. Гутиряк / Київськ. ун-т ім. Т. Шевченка. – К., 1999. – 18 с.
3. Горев А. Эффективная работа с СУБД / А. Горев, Р. Ахаян, С. Макашаринов. – СПб. : „Питер”, 1997. – 700 с.
4. Малыхин М. П. Базы данных: основы проектирования, использование. – [2-е изд. перераб. и доп.] / М. П. Малыхин. – СПб. : БХВ – Петербург, 2007. – 528 с.
5. Карпова Т. Базы данных: модели, разработка, реализация / Т. Карпова. – СПб. : Питер, 2001. – 304 с.
6. Хоменко А. Д. Базы данных : [учебн. для высш. учебн. заведений] / А. Д. Хоменко, В. М. Цыганков, М. Г. Мальцев; под ред. проф. А.Д. Хоменко. – СПб. : КОРОНА принт, 2000. – 416 с.
7. Гарсиа-Молина Г. Системы баз данных. Полный курс / Г. Гарсиа-Молина, Д. Д. Ульман, Д. Уидом; пер. с англ. – М. : издательский дом „Вильямс”, 2003. – 1088 с.
8. Тиори Т. Проектирование структур баз данных. Т. 1 / Т. Тиори, Дж. Фрай – М. : Мир, 1985. – 400 с.
9. Дейт К. Дж. Введение в системы баз данных. – [6-е изд.] / К. Дж. Дейт; пер. с англ. – М. : издательский дом „Вильямс”, 1999. – 848 с.
10. Коннолли Т. Базы данных: проектирование, реализация и сопровождение. Теория и практика. – [2-е изд.] / Т. Коннолли, К. Бегг, А. Страчан; пер. с англ. – М. : издательский дом „Вильямс”, 2001. – 1120 с.
11. Швецов В. И. Базы данных : учеб. пособие / В. И. Швецов, А. Н. Визгунов, И. Б. Мееров. – Нижний Новгород : ННГУ, 2004. – 217 с.

Подано 12.06.09

О. А. Машков, И. В. Пулеко, А. Г. Мокина
БАЗА ДАННЫХ АНГЛОЯЗЫЧНОЙ И УКРАИНОЯЗЫЧНОЙ ТЕРМИНОЛОГИИ
В СФЕРЕ ГЕОИНФОРМАЦИОННЫХ СИСТЕМ.

В статье предложено разработку базы данных англоязычной и украиноязычной терминологии в сфере геоинформационных систем. Разработано математическую модель выбора вида системы управления базой данных (СУБД) по множеству критериев, на основе которой сделан вывод о необходимости использования реляционной СУБД Firebird.

O. A. Mashkov, I. V. Puleko, A. G. Mokina
THE DATABASE OF THE ENGLISH AND UKRAINIAN TERMINOLOGY IN SPHERE
OF GEOINFORMATION SYSTEMS.

The development of a database of the English and Ukrainian terminology is given in the article. The mathematical model of a choice of a database management system is developed according to many criteria, on the basis of which the conclusion about the necessity to use the relational database management system „Firebird” is drawn.