

## ОСОБЛИВОСТІ ПРОЕКТУВАННЯ СТРУКТУРИ БАЗИ ДАНИХ ДЛЯ СТВОРЕННЯ ЕЛЕКТРОННИХ БАГАТОМОВНИХ АКАДЕМІЧНИХ ГАЛУЗЕВИХ СЛОВНИКІВ (ТЕЗАУРУСІВ)

На сьогодні база даних є однією з основних складових будь якого складного проекту, адже працюючи з великими обсягами даних виникає потреба у швидкому отриманні інформації та зручній роботі з нею. Створення бази даних слід починати з її проектування (розробки). У результаті проектування має бути визначена структура бази, тобто склад таблиць, їхня структура та логічні зв'язки. Процес проектування можна здійснювати двома підходами. За першого підходу спочатку визначають основні задачі, для розв'язання яких створюється база, та потреби цих задач у даних. За другого підходу визначають предметну область (сферу), здійснюють аналіз її даних і встановлюють типові об'єкти предметної області.

Для реалізації бази даних для початку потрібно обрати систему керування базами даних. Було проведено аналіз серед деяких існуючих систем, результати наведено у таблиці 1.

Таблиця 1 – Порівняння систем керування базами даних

Характеристика	MS SQL Server	MongoDB	MySQL
Модель даних	Реляційна	Документно-орієнтована	Реляційна
Адміністративне керування	8	6	7
Графічні інструменти	7	6	8
Процедури, що зберігаються	9	7	7
Можливості програмування	9	8	6
Мова SQL	8	0 (Відсутня)	6
Легкість у використанні	4	6	9
Підтримка індексації	7	4	6

При оцінці використовувався діапазон цілих чисел від 0 до 10. Наш вибір зупинився на MySQL, здебільшого через її відносну простоту і доступність в реалізації та легкість у використанні.

Особливостями створення бази даних для електронних багатомовних академічних галузевих словників (тезаурусів) є те, що кожне слово має різне написання на різних мовах, так як і галузь до якої воно належить. Тому було вирішено створити окремі таблиці, що будуть з'єднувати слова та галузі, написані різними мовами, за змістом: «word\_id\_table» та «area\_id\_table».

Так як різні мови мають різні морфологічні ознаки, описати їх статично в самій базі даних є неможливим. Тому найкращим рішенням, для виходу з даної ситуації, буде додати до цієї частини структури певну гнучкість та динамічність. Зручно було б зробити так, щоб при додаванні нової мови також додавалися б її морфологічні ознаки, за це відповідає таблиця «morph\_feature», яка має для зв'язку з мовою зовнішній ключ «lang\_id». Таким чином адміністратор може додавати, редагувати та видаляти з певної мови її ознаки не змінюючи самої структури бази даних. Значення морфологічної ознаки записується при додаванні нового слова в таблицю «morph\_word», яка зовнішнім ключем «morph\_id» поєднана з назвою ознаки. Дана таблиця не потребує внутрішнього ключа, так як заміну записів зручніше здійснювати шляхом видалення та додавання.

Визначення слів до словника потрапляють з першоджерел. З метою захисту авторських прав потрібно зберігати посилання на джерела. Також на одне визначення може бути декілька посилань. Було створено таблицю «source» з зовнішнім ключем «source\_ref», яка має поле для запису сторінок «page\_string». Для розмежування доступу створена таблиця «admin». Зберігати ідентифікатор адміністратора можна при будь-яких змінах в таблицях, він зберігається в поле «admin\_id» таблиці «word» лише при додаванні або останній зміні слова.

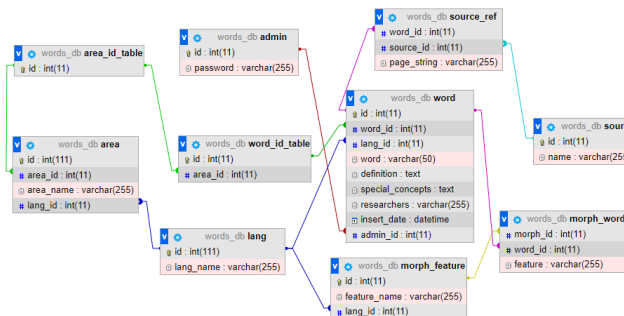


Рис.1. Спроектована структура бази даних словника

Від успішності та якості розробленої структури бази даних безпосередньо залежить надійність та зручність її використання. Використовуючи вищевказані підходи та методи можна досягти вкрай гідних результатів у проектуванні структури та логіки бази даних для електронних багатомовних академічних галузевих словників (тезаурусів).