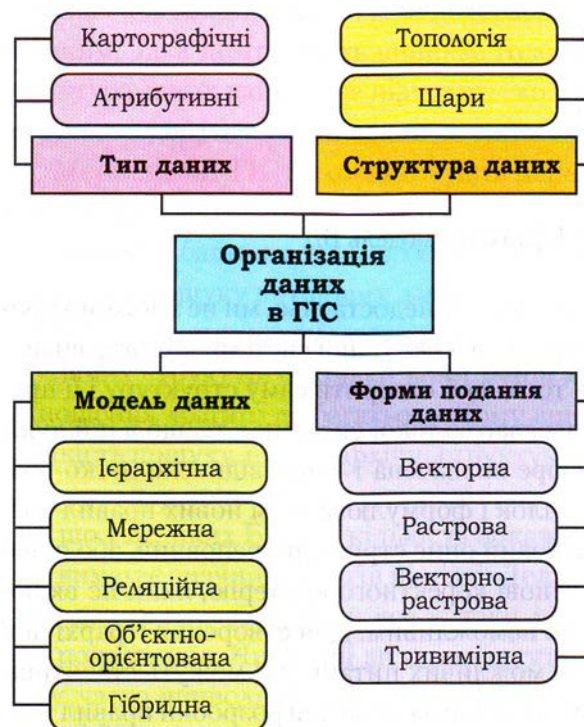


Тема 3.6. Бази даних

Бази даних є обов'язковими компонентами ГІС, яка завжди має два їх типи - *картографічні (графічні)* та *атрибутивні (тематичні) бази даних*. У *графічних базах* зберігається те, що прийнято називати топографічною основою. *Тематичні дані* містять в собі інформацію з так званого навантаження карти і додаткових даних, що належать до просторових, але не можуть бути прямо нанесені на карту - це опис територій або інформація, що міститься у різних звітах.

База даних (БД) - сукупність даних, організованих за певними правилами, що встановлюють загальні принципи опису, зберігання і маніпулювання даними.

Основна ідея організації структури бази даних полягає в тому, щоб максимально нормалізувати їх, тобто розбити на смислові та функціональні групи. При організації бази даних розрізняють (див. мал.1):



Мал. 1. Організація даних в ГІС

- тип даних; (картографічні і атрибутивні (описові));
- структуру даних (топологія і шари);
- моделі даних (ієрархічна, мережка, реляційна, об'єктно-орієнтована, гібридна);
- форму подання просторових даних (векторна, растрова, векторно-растрова, тривимірна).

Логічна структура елементів бази даних визначається обраною моделлю. Хоча постійно створюються все нові реалізації структур баз даних, розглянемо основні типи: ієрархічна (деревоподібна), мережна, реляційна (таблична), об'єктно-орієнтована структури баз даних.

Ієрархічна структура даних. У багатьох випадках існує взаємозв'язок між даними, що називається відношенням «один до багатьох». Це відношення має на увазі, що кожен елемент даних має прямий зв'язок з деяким числом так званих «нащадків», і, звичайно, кожен такий нащадок, у свою чергу, може мати зв'язок зі своїми нащадками і т. д.

Головною характеристикою ієрархічної структури, що ілюструється таксономічним деревом, є прямий взаємозв'язок між однією гілкою та іншою. Розгалуження засноване на формальних ключових ознаках, які визначають просування по цій структурі від однієї гілки до іншої і всі елементи розташовуються у послідовності від вищого до нижчого (мал. 2).



Рис. 2. Ієрархічна модель БД

Якщо інформації про ключову ознаку недостатньо, ми не зможемо просуватися по дереву. Насправді, природа ієрархічної системи вимагає визначення кожного відношення для того, щоб створити саму структуру і її правила розгалуження. Головною перевагою такої системи є те, що в ній дуже легко шукати, оскільки вона добре визначена і може відносно легко розширюватися додаванням нових гілок і формулюванням нових правил розгалуження. Але якщо наш початковий опис структури неповний, або якщо ми хочемо рухатися по ній на основі коректного критерію, який не включений у структуру, то пошук стає неможливим. Для створення ієрархічної структури необхідне знання всіх можливих питань, які можуть ставитись, оскільки ці питання використовуються як основа для розробки правил розгалуження або ключів.

Якщо навести приклад з бібліотеки, то ми можемо шукати інформацію відштовхуючись від теми, імені автора, назви і навіть за діапазоном номерів у каталозі.

Виникають ситуації, коли недостатньо певного запиту, це не так вже й рідко трапляється при роботі з інформацією в ГІС. Одна з найбільш складних речей - передбачити всі можливі запити користувача. БД ГІС містить безліч типів інформації і різних тематичних карт. Однією з найцікавіших особливостей ГІС є те, що ми можемо спробувати виконати пошук або досліджувати взаємозв'язки, які не передбачалися до реалізації системи. На жаль, ієрархічна структура не дуже підходить для цього через її жорстку ключову структуру. Крім цього суворого обмеження ієрархічна структура часто породжує великі за об'ємом файли. Це вимагає значного збільшення пам'яті для зберігання даних та зростання часу доступу до них.

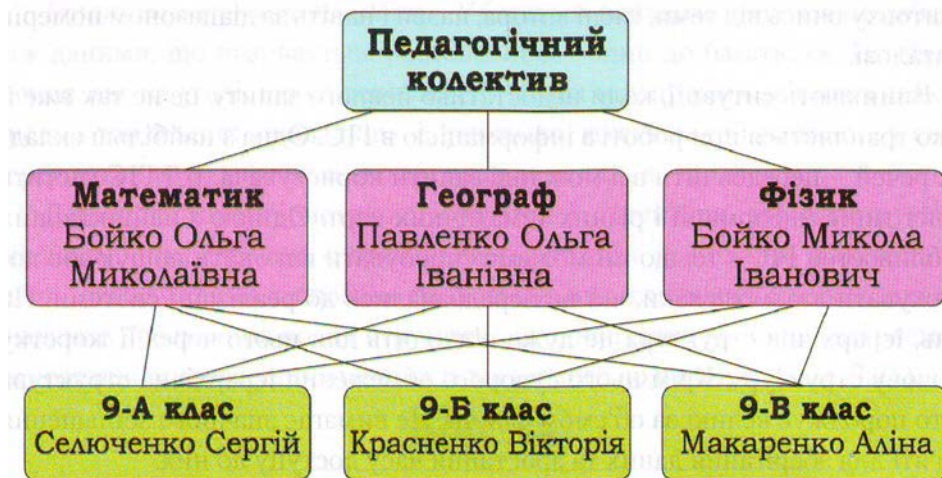
Мережні структури. Як бачимо, можливості швидкого пошуку, що виконується в ієрархічній структурі даних, визначаються структурою самого дерева. Атрибутивні і просторові дані можуть зберігатися в різних місцях, що потребує встановлення великої кількості зв'язків між графічною і атрибутивною частинами БД. У такому разі потенційне число розгалужень і пов'язаних з ними ключів ієрархічної структури може стати дуже великим. Така незграбність виникає головним чином тому, що ієрархічна структура даних найбільше підходить, коли між елементами даних потрібно встановлювати зв'язки «один до одного» або «один до багатьох».

Мережні БД ГІС використовують відношення «багато до багатьох», в якому один елемент може мати багато атрибутів, при цьому кожен атрибут пов'язаний з багатьма елементами. Тому мережна БД є більш гнучкою у здійсненні пошуку потрібних даних.

Мережні структури зазвичай розглядаються як удосконалення ієрархічних структур, оскільки вони менш жорсткі і можуть представляти відношення «багато до багатьох». Тому вони допускають набагато більшу гнучкість пошуку, ніж ієрархічні структури. Також на відміну від ієрархічних структур вони зменшують надмірність даних. Їх головним недоліком є те, що у великих БД ГІС кількість показників може стати дуже великою, а це вимагає значних обсягів пам'яті. Додатково, хоча зв'язки між елементами даних гнучкіші, вони все ж таки мають бути визначені за допомогою показників. Численні зв'язки можуть перетворитися на складну мережу, часто приводами до плутанини, втрачених або помилкових зв'язків (мал. 3).

Реляційні бази даних. Недоліків значної кількості показників можна уникнути, використовуючи ще одну структуру баз даних - реляційну. У ній дані зберігаються як упорядковані записи або рядки значень атрибутів. Атрибути

об'єктів групуються в окремих рядках у вигляді так званих відношень (relations), оскільки вони зберігають своє місцеположення в кожному рядку і безумовно пов'язані один з одним. Кожна колонка містить значення одного атрибута для всього набору об'єктів.



Мал. 3. Мережна модель БД

Наприклад, може бути колонка з номерами квадратів (один атрибут). В іншій колонці може бути додаткова інформація, у третій - дата збирання даних, у четвертій - номер майданчика.

Реляційна модель даних являє собою набір двовимірних таблиць, які складаються зі стовпчиків (полів) і рядків (записів), а також мають ім'я, унікальне в межах даної БД. Реляційна модель даних, як правило, складається з кількох таблиць, які зв'язуються між собою ключами або ключовими стовпчиками. З мал. 4 видно, що таким ключовим стовпчиком є «Код учасника».

Реляційні системи засновані на наборі математичних принципів, що називаються реляційною алгеброю, або алгеброю відношень. Оскільки реляційна алгебра ґрунтується на теорії множин, кожна таблиця відношень функціонує як множина, і перше правило свідчить, що таблиця не може мати рядок, який має повний збіг з будь-яким іншим рядком. Оскільки кожен рядок унікальний, одна або кілька колонок можуть використовуватися для визначення критерію пошуку. Так, прикладом використання однієї колонки для визначення критерію пошуку може бути вибір унікального особистого номера соціального страхування, номера телефону, домашньої адреси і т. д., наявних в інших колонках тієї ж таблиці при виборі певного імені з першої колонки. Такий критерій пошуку називається первинним ключем (primary key) для пошуку значень в інших колонках бази даних. Кожен рядок таблиці повинен мати унікальне значення в колонці первинного ключа, інакше ми не зможемо однозначно ідентифікувати об'єкти.

База даних «Юний краєзнавець», таблиця «Учасники»

Код учасника	Прізвище	Ім'я	Клас	Адреса	Телефон
1	Петров	Олег	8-А	вуд. Польова, 17	2-17-36
2	Степаненко	Ольга	9-Б	вуд. Вікторова, 5	2-12-11
3	Смаль	Роман	10-А	вуд. Гончара, 68	3-20-17

База даних «Юний краєзнавець», таблиця «Знаряддя»

Код знаряддя	Код учасника	Знаряддя
10	2	Палатка
20	1	Ліхтарик
30	3	Компас

Мал. 4. Реляційна модель БД

Реляційні системи цінні тим, що дають нам можливість збирати дані в достатньо прості таблиці. При необхідності ми можемо стикувати рядки з однієї таблиці з відповідними рядками з іншої таблиці, використовуючи механізм, що називається реляційним з'єднанням (relational join). Оскільки реляційні системи переважають в ГІС то цей процес дуже поширений.

Для встановлення реляційних з'єднань, кожна таблиця повинна мати хоча б одну загальну колонку з іншою таблицею, з якою потрібно встановити таке з'єднання.

Об'єктно-орієнтовані моделі застосовують, якщо геометрія певного об'єкта здатна охоплювати кілька шарів, а атрибути таких об'єктів можуть передаватися іншим об'єктам (успадковуватися). Для їх обробки застосовують специфічні методи.

Для обробки даних, які розміщені в таблицях, необхідні додаткові відомості про дані, їх називають метаданими. Метадані - дані про дані: каталоги, довідники, реєстри та інші форми опису наборів цифрових даних.

Узагальнення

База даних - сукупність даних, організованих за певними правилами, що встановлюють загальні принципи опису, зберігання і маніпулювання даними.

Метадані - відомості про дані: каталоги, довідники, реєстри та інші форми опису наборів цифрових даних.

Запитання і завдання

1. Дайте визначення бази даних.
2. Назвіть основні типи структур баз даних.
3. Чим відрізняються картографічні бази даних від усіх інших?