

# Математичне моделювання електротехнічних систем

## Лекція 3

## **Типи наборів даних**

Виділятимемо наступні типи даних: • дані, що складаються із записів • табличні дані • транзакційні дані • графічні дані • хімічні дані.

**Дані, що складаються із записів.** Найбільш дані, що часто зустрічаються, – дані, що складаються із записів (record data). Приклади таких наборів даних: табличні дані, матричні дані, документальні дані, транзакційні або операційні.

**Табличні дані** – дані, що складаються із записів, кожна з яких складається з фіксованого набору атрибутів.

**Транзакційні дані** є особливим типом даних, де кожен запис, що є транзакцією, включає набір значень.

Приклад транзакційної бази даних, що містить перелік покупок клієнтів магазину, приведений на рисунку.

| <b>ІД</b> | <b>Найменування</b>     |
|-----------|-------------------------|
| 1         | Вода, сік, молоко       |
| 2         | Пиво, вода              |
| 3         | Пиво, сік, молоко, квас |
| ...       | ...                     |

Рисунок 2.2 – Приклад транзакційних даних

Графічні дані. Приклади графічних даних: WWW-дані; молекулярні структури; графи

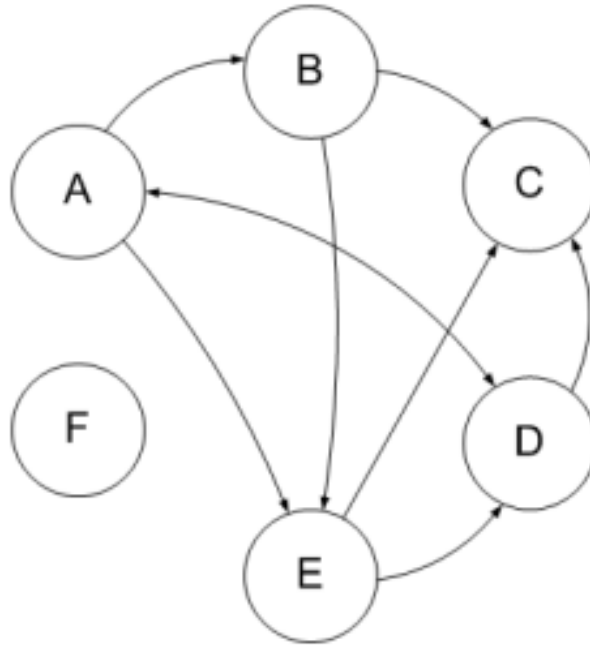


Рисунок 2.3 – Приклад графа

За допомогою карт, наприклад, можна відстежити зміни об'єктів в часі і просторі, визначити характер їх розподілу на площині або в просторі. Перевагою графічного представлення даних є велика простота їх сприйняття, чим, наприклад, табличних даних. Приклад карти, картою Кохонена, що є, представлений на рисунку.

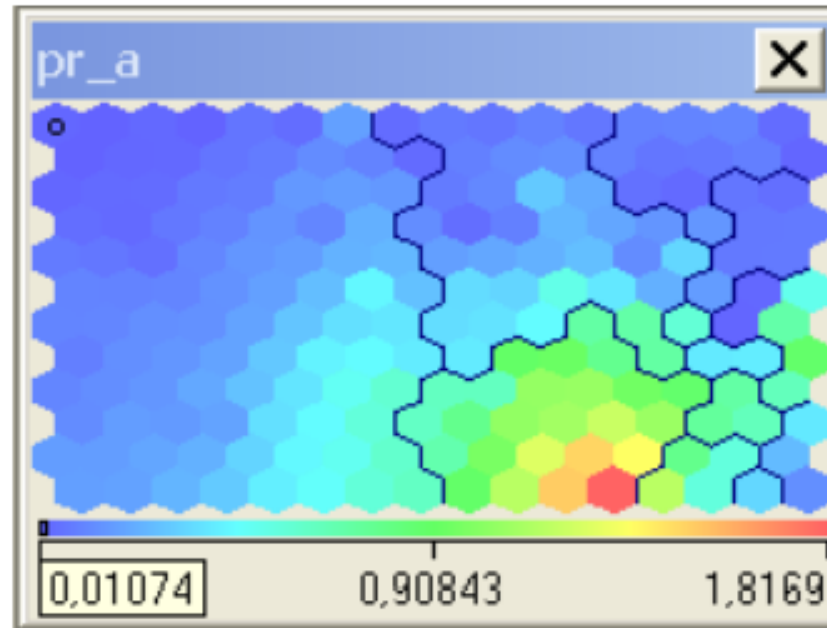


Рисунок 2.4 – Приклад даних типу "Карта Кохонена"

Хімічні дані. Хімічні дані є особливим типом даних. Приклад таких даних: Benzene Molecule:  $C_6H_6$

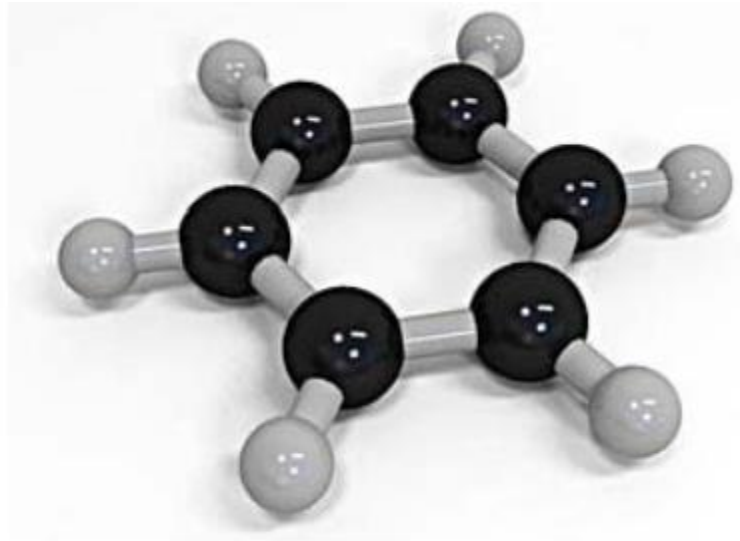


Рисунок 2.5 – Приклад хімічних даних

Згідно опитувань на сайті Kdnuggets, найбільше число опитаних аналізує дані з "плоских" таблиць (26%), реляційних таблиць (24%), далі йдуть тимчасові ряди (14%) і дані у вигляді тексту (11%), решта аналізованих типів даних в порядку убутання: web-контенти, XML, графіка, аудіо, відео і ін.

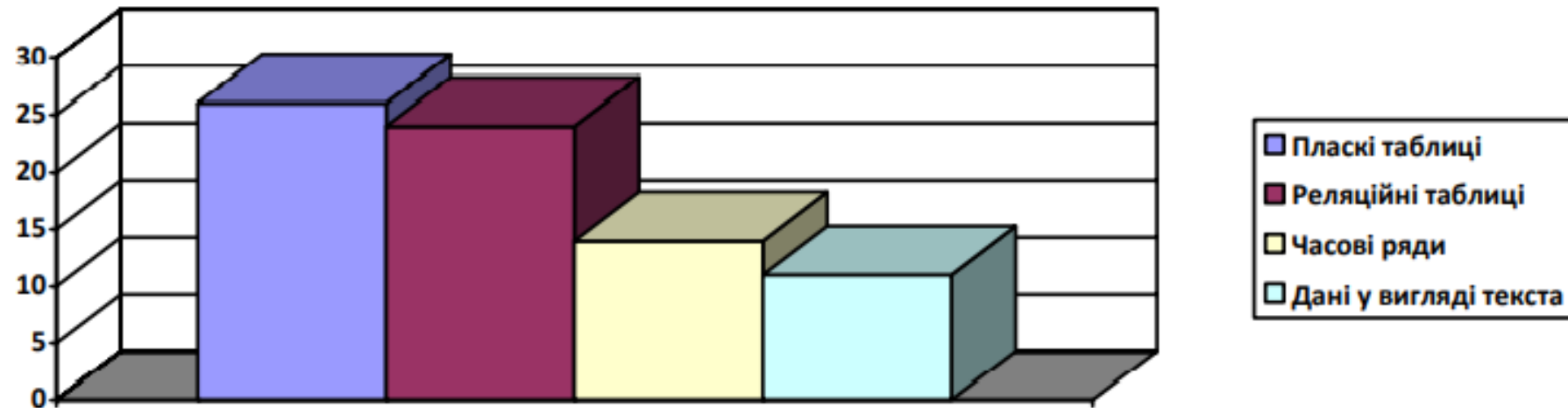


Рисунок 2.6. –Розподіл типів даних

## Формати зберігання даних

Можливі чотири аспекти роботи з даними: визначення даних, обчислення, маніпулювання і обробка (збір, передача і ін.). При маніпулюванні даними використовується структура даних типу "файл". Файли можуть мати різні формати. Найбільш поширені формати, представлені на рисунку.

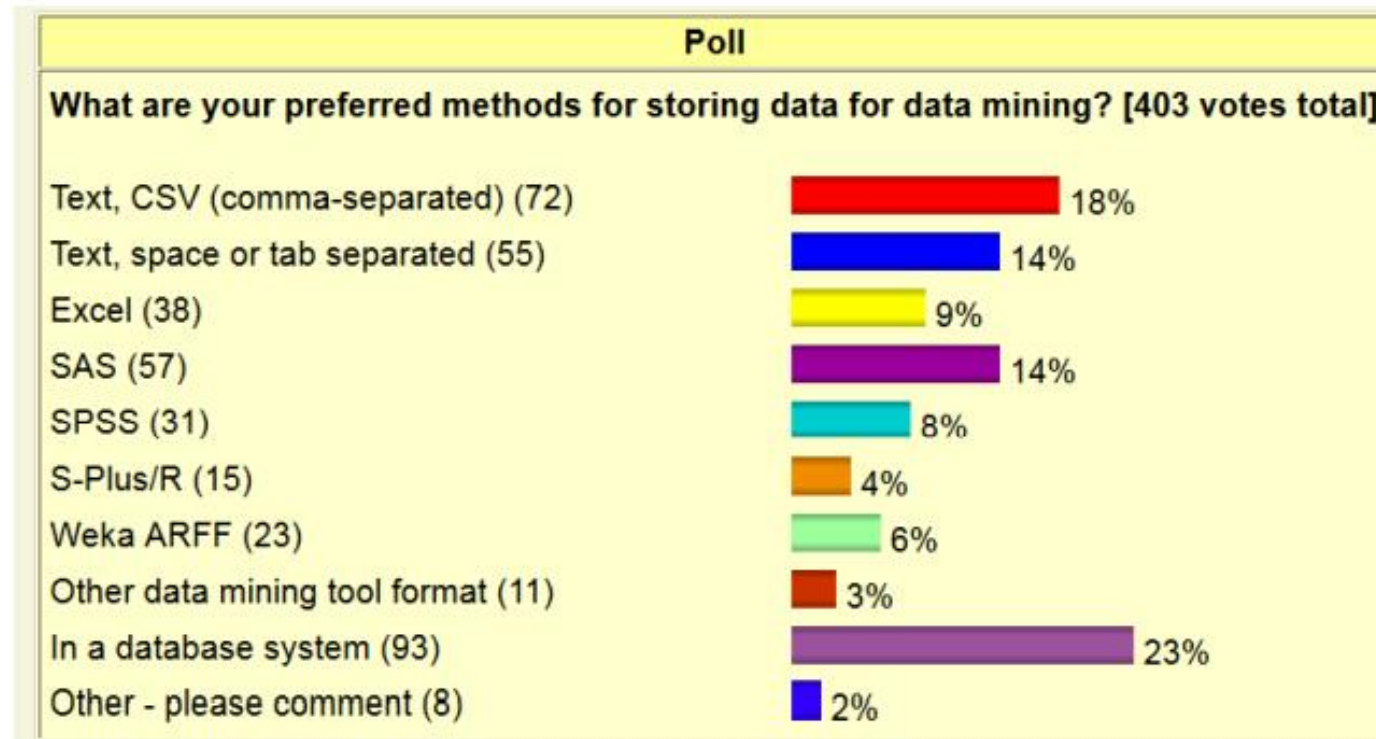


Рисунок 2.7 – Найбільш поширені формати зберігання даних



Найбільше число опитаних (23%) вважають за краще зберігати дані у форматі тієї бази даних, яку використовують. У форматі Text, CSV – 18%, 14% опитаних зберігають дані у форматі SAS; у форматі Excel – 9%, SPSS -8%, S-Plus – 4%, Weka ARFF – 6%, в інших форматах – 2%.

За результатами опитувань, найбільш поширеним форматом зберігання даних виступають бази даних.

## 2.4 Бази даних. Основні положення

Для розуміння організації даних в базі даних необхідне знання основних положень теорії баз даних. Розглянемо деякі положення цієї теорії.

База даних (Database) – це особливим чином організовані і такі, що зберігаються в електронному вигляді дані.

Також така організація даних передбачає полегшити їх пошук і доступ до них для одного або декількох застосувань і наявність мінімальної надмірності даних.

Метою створення баз даних є побудова такої системи даних, яка б не залежала від програмного забезпечення, вживаних технічних засобів і фізичного розташування даних в ЕОМ. Побудова такої системи даних повинна забезпечувати несутеречливу і цілісну інформацію. База даних в простому випадку представляється у вигляді системи двовимірних таблиць.

## **Системи управління базами даних (СУБД)**

Система управління базою даних – це програмне забезпечення, контролююче організацію, зберігання, цілісність, внесення змін, читання і безпеку інформації у базі даних.

Система управління реляційними базами даних (Relational Database Management System) – це СУБД, заснована на реляційній моделі даних.

У реляційній моделі даних будь-яке представлення даних зводиться до сукупності реляційних таблиць (двовимірних таблиць особливого типу).

**Програмні засоби СУБД** включають систему управління, що забезпечує уведення-виведення, обробку і зберігання інформації, створення, модифікацію і тестування бази даних. Внутрішніми мовами програмування СУБД є мови четвертого покоління (C, C++, Pascal, Object Pascal). За допомогою мов БД створюються додатки, бази даних і інтерфейс користувача, що включає екранні форми, меню, звіти.

## **До баз даних, а також до СУБД пред'являються такі вимоги:**

1. Висока швидкодія;
2. Простота оновлення даних;
3. Незалежність даних;
4. Можливість мультикористувацького використання даних;
5. Безпека даних;
6. Стандартизація побудови і експлуатації БД;
7. Адекватність відображення даних відповідної наочної області;
8. Доброзичливий інтерфейс користувача.

**Висока швидкодія** передбачає малий час відгуку, тобто малий проміжок часу від моменту запиту до бази даних до моменту реального отримання даних.

**Незалежність даних** – це можливість зміни логічної і фізичної структури бази даних без зміни уявлень користувачів.

**Безпека даних** – це захист даних від навмисного або ненавмисного порушення секретності, спотворення або руйнування. Безпеку включає два компоненти: цілісність і захист даних від несанкціонованого доступу.

**Цілісність даних** – стійкість даних, що зберігаються, до руйнування і знищення, пов'язаним з несправностями технічних засобів, системними помилками і помилковими діями користувачів.

**Цілісність даних передбачає:**

- відсутність неточно введених даних, захист від помилок при оновленні баз даних;
- неможливість видалення зв'язаних даних різних таблиць;
- збереження даних при збоях техніки і ін.

**Захист даних** від несанкціонованого доступу передбачає обмеження доступу до певних даних бази і досягається введенням мір безпеки: розмежування має рацію доступу до даних різних користувачів залежно від виконуваних ними функцій і/або посадових обов'язків; введенням захисту у вигляді паролів.

**Стандартизація** забезпечує спадкоємність поколінь конкретної СУБД, спрощує взаємодію баз даних одного покоління СУБД з однаковими і різними моделями даних.



## Класифікація видів даних

Якими можуть бути дані? Нижче наведено декілька класифікацій.

Реляційні дані – це дані з реляційних баз (таблиць).

Багатовимірні дані – це дані, представлені в кубах OLAP.

Відмірювання (dimension) або вісь – в багатовимірних даних – це збори даних одного і того ж типу, що дозволяє структурувати багатовимірну базу даних.

По критерію постійності своїх значень в ході рішення задачі дані можуть бути:

- змінними;
- постійними;
- умовно-постійними.

**Відносна шкала** (ratio scale) – шкала, в якій є певна точка відліку і можливі відношення між значеннями шкали.

Приклад такої шкали: ціна на картоплю в супермаркеті вище в 1,2 разу, чим ціна на базарі.

Відносні і інтервальні шкали є числовими.

Для цієї шкали застосовні тільки такі операції: рівно ( $=$ ), не рівно ( $\neq$ ), більше ( $>$ ), менше ( $<$ ), операції складання ( $+$ ) і віднімання ( $-$ ), множення ( $*$ ) і ділення ( $/$ ).

**Змінні дані** – це такі дані, які змінюють свої значення в процесі рішення задачі.

**Постійні дані** - це такі дані, які зберігають свої значення в процесі рішення задачі (математичні константи, координати нерухомих об'єктів) і не залежать від зовнішніх чинників.

**Умовно-постійні дані** - це такі дані, які можуть іноді змінювати свої значення, але ці зміни не залежать від процесу рішення задачі, а визначаються зовнішніми чинниками.

Дані, поклад від тихий функцій, які сморід виконують, можуть бути **довідковими, оперативними, архівними.**

Слід розрізняти дані за період і точкові дані. Ці відмінності важливі при проектуванні системи збору інформації, а також в процесі вимірювань.

- дані за період;
- точкові дані.

**Дані за період** характеризують деякий період години.

Прикладом даних за період можуть бути: прибуток підприємства за місяць, середня температура за місяць.

**Точкові дані** представляють значення деякої змінної в конкретний момент часу. Приклад точкових даних: залишок на рахунку на перше число місяця, температура у вісім годин ранку. Дані бувають первинними і вторинними.

**Вторинні дані** - це дані, які є результатом певних обчислень, застосованих до **первинних даних**.

## Метадані

**Метадані (Metadata)** – це дані про даних.

До складу метаданих можуть входити: каталоги, довідники, реєстри.

Метадані містять зведення про склад даних, змісті, статусі, походженні, місцезнаходженні, якості, форматах і формах уявлення, умовах доступу, придбання і використання, авторських, майнових і суміжних з ними правах на дані і ін.

**Висновок.** У лекції були розглянуті поняття даних, об'єктів і атрибутів їх характеристики, типи шкал, поняття набору даних і його типи. Описані можливі формати зберігання даних. Введені поняття бази даних, системи управління базами даних, метаданих