



МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
“КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ  
імені ІГОРЯ СІКОРСЬКОГО”

Факультет прикладної математики  
Кафедра програмного забезпечення комп'ютерних систем

## Теоретичне завдання

з дисципліни “Інформаційно-пошукові системи та сервіси ”

Виконав  
студент V курсу  
групи КП-31мп

Грищенко Олександр  
Володимирович

Зарахований  
“ \_\_\_\_ ” “ \_\_\_\_ ” 2024 р.  
Викладачем

Погорелов Володимир  
Володимирович

Київ 2024

# 1. Моделі і методи зберігання даних:

Моделі і методи зберігання даних визначають способи організації та зберігання інформації для подальшого доступу та використання. Основні поняття та підходи до зберігання даних включають:

**1. Модель даних:** Визначає, як дані будуть структуровані і організовані в базі даних. Наприклад, реляційна модель використовує таблиці, а документальна модель може використовувати дерево об'єктів.

**2. Методи зберігання:** Визначають способи, якими дані будуть зберігатися фізично. Це може включати в себе вибір системи управління базами даних (СУБД), використання файлової системи або інші технології зберігання.

Різні моделі і методи зберігання відповідають різним потребам та характеристикам даних. Наприклад, для структурованих даних часто використовують реляційні бази даних, тоді як для неструктурованих даних можуть застосовуватися документальні бази даних чи інші формати.

**Реляційні бази даних (РБД):** Використовуються таблиці для зберігання даних, зв'язків між якими визначаються ключами.

**Нереляційні бази даних (NoSQL):** Використовуються для зберігання та отримання даних, які не вимагають жорсткої структури, наприклад, документи, ключі-значення тощо.

**Файлові системи:** Дозволяють організувати та зберігати дані у вигляді файлів в ієрархічній структурі.

**Хмарні системи зберігання:** Надають можливість зберігати дані в хмарних сервісах, доступ до яких забезпечується через мережу Інтернет.

**Кешування даних:** Використовується для збереження копій часто використовуваних даних у швидкодії, щоб прискорити їхнє отримання.

**Реплікація та бекапи:** Методи для забезпечення надійності та доступності даних шляхом створення копій або дублювання інформації.

**Секрети та шифрування:** Використовуються для захисту конфіденційної інформації шляхом застосування методів шифрування та збереження ключів.

## **2. Класифікація інформаційних систем і місце серед них інформаційно-пошукових систем**

Інформаційні системи (ІС) можна класифікувати за різними критеріями, такими як функціональність, масштаб, відомча приналежність. Інформаційно-пошукові системи (ІПС) є підмножиною інформаційних систем, і їхнє місце в класифікації може залежати від конкретного підходу до класифікації.

### **Класифікація інформаційних систем за функціональністю:**

1. Операційні інформаційні системи: Спрямовані на проведення операційного обліку та керування поточною діяльністю підприємства.
2. Стратегічні інформаційні системи: Забезпечують вирішення стратегічних завдань та прийняття рішень на вищому рівні управління.
3. Тактичні інформаційні системи: Використовуються для вирішення конкретних завдань середнього терміну.
4. Оперативні інформаційні системи: Забезпечують вирішення конкретних завдань оперативного обліку та керування на кількох рівнях.

### **Класифікація за масштабом:**

1. Персональні інформаційні системи: Призначені для індивідуального використання.
2. Групові інформаційні системи: Використовуються групою користувачів в організації.
3. Підприємницькі (корпоративні) інформаційні системи: Охоплюють велику частину або всю діяльність організації.
4. Міжорганізаційні інформаційні системи: Об'єднують кілька організацій для спільної діяльності.

### **Місце ІПС в класифікації:**

Інформаційно-пошукові системи можуть бути частиною будь-якого з вищезазначених типів інформаційних систем, забезпечуючи функції пошуку, аналізу та обробки інформації для задоволення потреб користувачів у пошуку даних.

### 3. Організація пошуку. Пошукові машини.

#### Організація пошуку:

Організація пошуку включає в себе ряд технічних та алгоритмічних аспектів, спрямованих на ефективне виявлення та відображення результатів по запиту користувача. Основні етапи організації пошуку включають:

1. Індексація: Автоматизоване створення індексу або бази даних, яка містить інформацію про веб-сторінки та їх зміст.
2. Сполучення інформації: Пошукові системи використовують різноманітні алгоритми для збору, сполучення та обробки інформації з різних джерел.
3. Оцінка релевантності: Аналіз отриманих даних та визначення рівня релевантності для кожного результату.
4. Подання результатів: Представлення відсортованих та оцінених результатів користувачу.

#### Пошукові машини:

Пошукові машини (Search Engines) є ключовим елементом організації пошуку в Інтернеті. Деякі з найвідоміших пошукових машин включають: Google, Bing, Yahoo, Yandex, DuckDuckGo

Пошукові машини використовують складні алгоритми для ранжування та представлення результатів у вигляді списку посилань на веб-сторінки, які найбільше відповідають запиту користувача.

Пошукові машини працюють за принципом індексації та ранжування веб-сторінок. Основні етапи їх роботи:

1. Індексація: Машина автоматично переглядає та індексує веб-сторінки, зберігаючи дані про їх зміст та ключові слова.
  2. Сполучення інформації: Пошукова машина збирає інформацію з різних джерел, включаючи індексовані сторінки, архіви та інші джерела.
  3. Оцінка релевантності: Алгоритми оцінюють релевантність веб-сторінок до певного запиту користувача, враховуючи фактори, такі як ключові слова, авторитет сторінки та інші.
  4. Ранжування: Результати ранжуються за ступенем їхньої відповідності запиту та інших факторів. Найбільш релевантні сторінки виводяться першими.
- Користувачі отримують список відсортованих результатів, які найкращим чином відповідають їхньому пошуковому запиту.

## 4. Створення і типи індексів

Створення індексів - це процес створення структури даних для ефективного пошуку та доступу до інформації в базі даних. Індеси роблять пошук даних більш ефективним, але можуть впливати на швидкість вставки, оновлення та видалення даних. Основні типи індексів:

### 1. Одноколоночний (однорівневий) індекс:

Створюється на одному стовпці таблиці.

Ефективний для швидкого пошуку по конкретному полю.

### 2. Складений (багаторівневий) індекс:

Створюється на двох або більше стовпцях таблиці.

Підтримує ефективний пошук для комбінованих умов.

### 3. Унікальний індекс:

Гарантує унікальність значень в індексованих стовпцях.

Застосовується до полів, які не повинні містити дублікатів.

### 4. Повний текстовий індекс:

Використовується для індексації текстових даних, щоб швидко знаходити слова або фрази.

### 5. Хеш-індекс:

Використовує хеш-функції для визначення позначок для рядків таблиці.

Ефективний для пошуку за точним значенням, але не підтримує діапазонні запити.

### 6. Кластерний індекс:

Таблиця фактично впорядкована за значенням індексованого стовпця.

### 7. Некластерний індекс:

Дані в таблиці індексуються окремо від їх фактичного порядку.

Вибір конкретного типу індексу залежить від конкретних потреб та характеристик бази даних.

## 5. Проблеми індексування

Індексування в базах даних допомагає поліпшити швидкість пошуку, але також може стикатися з рядом проблем:

### 1. Великий обсяг пам'яті:

Деякі типи індексів вимагають значних обсягів пам'яті, особливо при роботі з великими обсягами даних.

### 2. Погіршення швидкості вставки, оновлення та видалення:

Індекси підвищують швидкість пошуку, але можуть призвести до уповільнення операцій вставки, оновлення та видалення даних.

### 3. Втрати швидкості при розподіленому зберіганні:

У розподілених базах даних індексація може призвести до погіршення швидкості обміну даними між різними серверами.

### 4. Погіршення продуктивності при використанні багатостовпцевих індексів:

Багатостовпцеві індекси можуть стати менш ефективними, коли вони мають велику кількість стовпців.

### 5. Ризик конфліктів індексів:

Деякі операції можуть призвести до конфліктів індексів, особливо при використанні паралельних операцій.

### 6. Проблеми з унікальністю:

Використання унікальних індексів може стати проблемою, коли дані не завжди унікальні або коли є потреба додавати дубльовані значення.

### 7. Потреба в обслуговуванні:

Індекси потребують регулярного обслуговування та оновлення, що може впливати на продуктивність системи під час цих операцій.

## 6. Запити до пошукових машин

Запити до пошукових машин - це текстові або мовні команди, які вводить користувач в пошуковий рядок для знаходження конкретної інформації. Такі запити розуміються та обробляються пошуковими системами для представлення користувачеві результатів, що найбільше відповідають його запиту. Основні види запитів до пошукових машин:

### 1. Ключові слова:

Простий запит, що включає ключові слова, пов'язані з тим, що ви шукаєте. Наприклад, "штучний інтелект", "автоматизація тестування".

### 2. Фразовий запит:

Запит, який включає слова в конкретному порядку. Наприклад, "нові технології розробки ПЗ".

### 3. Логічні операції:

Використання операторів, таких як AND, OR, NOT для точнішого визначення взаємозв'язків між словами. Наприклад, "штучний інтелект AND медицина".

### 4. Пошук за допомогою параметрів:

Використання параметрів для обмеження результатів пошуку, таких як "site:" для обмеження пошуку на конкретному веб-сайті або "filetype:" для пошуку конкретного типу файлів.

### 5. Пошук за визначенням:

Введення запиту у формі питання для отримання конкретної відповіді. Наприклад, "Які переваги використання штучного інтелекту в розробці ПЗ?".

### 6. Пошук за часом:

Використання параметрів часу для обмеження результатів за конкретний період.

### 7. Пошук за місцем:

Включення локаційних аспектів у запиті, таких як "погода у Києві".

### 8. Пошук зображень, відео, новин тощо:

Використання специфічних параметрів для пошуку конкретних типів контенту.

Запити до пошукових машин можуть бути дуже різноманітними, і їхнє правильне формулювання допомагає отримати більш точні та релевантні результати.

## 7. Якість роботи пошукачів

Якість роботи пошукових систем визначається ключовими аспектами:

### 1. Релевантність результатів:

Пошукові системи повинні надавати результати, які найкраще відповідають запитам користувачів. Це включає в себе точність та актуальність інформації.

### 2. Швидкість пошуку:

Якість вимірюється швидкістю, з якою пошукова система повертає результати. Швидкість роботи пошуку важлива для задоволення потреб користувачів у невеликому проміжку часу.

### 3. Повнота індексації:

Пошуковий індекс повинен охоплювати якнайбільше сторінок та джерел інформації в Інтернеті, щоб надати користувачам широкий спектр результатів.

### 4. Адаптивність до змін:

Пошукові системи повинні адаптуватися до змін у веб-просторі, оновлюючи свій індекс та алгоритми для врахування нових трендів та технологій.

### 5. Локалізовані результати:

Здатність пошукової системи враховувати контекст індивідуального користувача, такі як географічне положення чи особисті уподобання, для надання більш персоналізованих результатів.

### 6. Спростування спаму:

Ефективність системи у виявленні та усуненні спаму та неправдивої інформації з результатів пошуку.

### 7. Коректність та етика:

ПС повинні дотримуватися етичних стандартів та правил використання особистої інформації, забезпечуючи конфіденційність користувачів.

### 8. Можливості фільтрації та уточнення:

Інструменти для фільтрації та уточнення результатів пошуку, щоб користувачі могли швидше та ефективніше знаходити потрібну інформацію.



## 8. Посилальне ранжування (Page Rank)

Посилальне ранжування (PageRank), є алгоритмом ранжування веб-сторінок, який винайшов Ларрі Пейдж та Сергій Брін, засновники Google. Цей алгоритм використовується для оцінки важливості сторінок в Інтернеті на основі їх взаємозв'язків.

### Принцип роботи:

#### 1. Вихідні дані:

Кожна сторінка отримує початковий ранг, який може бути різним для різних пошукових систем.

#### 2. Передача ваги через посилання:

Якщо сторінка А посилається на сторінку В, то це означає, що сторінка А віддає частину свого рангу сторінці В.

#### 3. Перерозподіл рангів:

Ранги всіх сторінок перераховуються, враховуючи вагу, яку вони отримують від інших сторінок.

#### 4. Зациклювання:

Процес повторюється кілька разів, поки ранги не стабілізуються.

#### 5. Вага джерела:

Якщо сторінка отримує посилання від іншої сторінки з високим рангом, то це посилання має більше ваги.

### Переваги:

- PageRank допомагає виявляти авторитетні сторінки в Інтернеті, забезпечуючи більш точні результати пошуку.
- Враховує структуру Інтернету, а не лише ключові слова, що поліпшує якість пошуку.
- Враховує внутрішню та зовнішню важливість сторінок.

### Недоліки:

- PageRank може бути викритий на певні види маніпуляцій.
- Він не враховує контекст або зміст сторінок, фокусуючись виключно на взаємозв'язках між сторінками.
- У світі соціальних мереж та динамічного вмісту інтернет-простору інші методи ранжування стають важливішими.

## 9. Поняття інформації як категорії, дані і знання

Поняття інформації включає кілька взаємопов'язаних категорій: дані та знання.

### 1. Дані:

Дані - це факти, які можна виміряти або зафіксувати. Це сировинний матеріал, який ще не має значення або контексту.

Характеристики: Дані можуть бути числовими, текстовими, аудіо-, відео-файлами тощо.

Приклад: Набір числових значень, таких як "25", "37.5", "42", становить дані без контексту.

### 2. Інформація:

Інформація виникає, коли дані отримують контекст або отримують оцінку.

Характеристики: Інформація має сенс, організована та спрямована на розуміння або вирішення конкретного завдання.

Приклад: "Середній вік користувачів - 30 років". Тут дані про вік стали інформацією, оскільки їх оцінено та контекстуалізовано.

### 3. Знання:

Знання - це вищий рівень інформації, яке здатне використовуватися для прийняття рішень або вирішення проблем.

Характеристики: Знання включає в себе зрозумілість, досвід та можливість використання інформації для досягнення конкретних цілей.

Приклад: "Користувачі віком до 30 років більш схильні використовувати мобільні додатки для покупок". Тут знання вибудоване на інформації та дозволяє робити висновки та передбачення.

Таким чином, інформація є проміжним етапом між даними та знанням, де контекст та оцінка надають даним сенс та цінність. Знання виникає, коли ця інформація стає корисною та придатною для використання в конкретних ситуаціях.

## **10. Програмне та апаратне забезпечення для організації пошуку інформації в мережі інтернет**

Для організації пошуку інформації в мережі Інтернет використовуються як програмне, так і апаратне забезпечення.

### **1. Програмне забезпечення:**

Пошукові машини (Search Engines): Це програми, які сканують та індексують веб-сторінки для подальшого пошуку користувачів. Приклади: Google, Bing, Yahoo.

Браузери: Деякі веб-браузери також мають вбудовані пошукові можливості та інструменти для зручного пошуку в Інтернеті.

Алгоритми пошуку: Розроблені алгоритми, які визначають порядок виведення результатів пошуку на основі релевантності та інших чинників.

### **2. Апаратне забезпечення:**

Сервери пошукових систем: Великі потужні серверні ферми, які використовуються для зберігання індексів та надання результатів пошуку.

Комп'ютери та мережеве обладнання: Апаратні засоби, які використовуються для введення та обробки даних під час сканування та індексації веб-сторінок.

### **3. Інше:**

Кеш-системи: Використовуються для збереження копій попередньо знайдених результатів пошуку для швидкого доступу до них без повторного сканування.

Комп'ютерні алгоритми та програми для обробки мови: Допмагають в розумінні семантики запитань користувачів та поліпшенні релевантності результатів пошуку.

Програмне та апаратне забезпечення спільно використовуються для ефективного та швидкого здійснення пошуку великої кількості інформації в мережі Інтернет. Технології продовжують розвиватися для поліпшення швидкості, точності та зручності процесу пошуку.