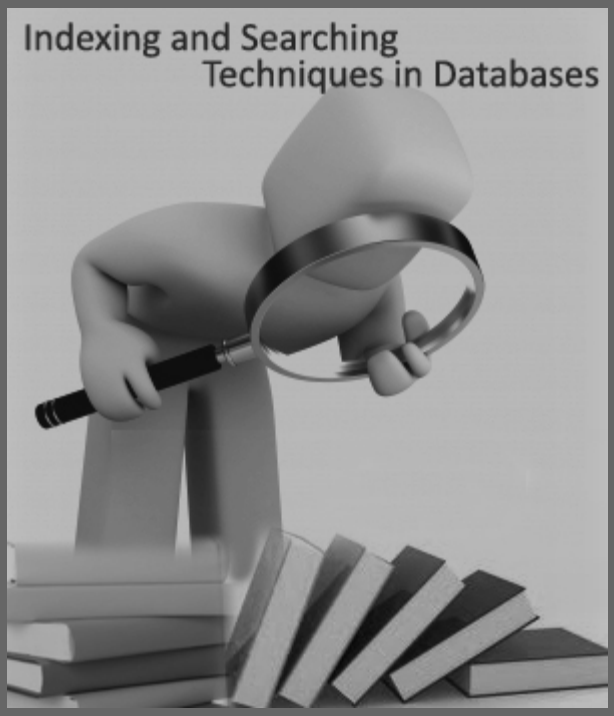

Course Name **Indexing and Searching Techniques in Databases**

Department of Computer Science and Engineering

Instructor **Dr. Arnab Bhattacharya**

Next page

Module 1: Basics and Background

Lecture 1: Basic Database Queries

Prev topic Next topic Prev page
Next page

The Lecture Contains:

General setting

- Database D of n objects

General setting

- Database D of n objects
- Each object has k dimensions

$$O_i = \{O_{i,1}, \ldots, O_{i,k}\}$$

- No particular database design assumed

General setting

- Database D of n objects
- Each object has k dimensions

$$O_i = \{O_{i,1}, \ldots, O_{i,k}\}$$

- No particular database design assumed
- Query Q is an object with the same k dimensions

$$Q = \{Q_1, \ldots, Q_k\}$$

Exact match query or Point query

- Point query: Given D and Q, return all objects $O_i \in D$ such that

$$O_i = Q$$

Exact match query or Point query

- Point query: Given D and Q, return all objects $O_i \in D$ such that

$$O_i = Q$$

- The concept of equality (operator =) needs to be precisely defined

For example, $O_i = Q$ iff all the d dimensions individually are the same,

i.e. $O_i = Q \Leftrightarrow Vj = \{1, - - - -, k\}, O_{i,j} = Q_j$

Exact match query or Point query

- Point query: Given D and Q, return all objects $O_i \in D$ such that

$$O_i = Q$$

- The concept of equality (operator =) needs to be precisely defined

  For example, $O_i = Q$ iff all the d dimensions individually are the same,

  i.e. $O_i = Q \Leftrightarrow \forall j = \{1, - - - -, k\}, O_{i,j} = Q_j$

- Equality can be defined over a subset of the k dimensions as well

  For example, $O_i = Q \Leftrightarrow O_{i,1} = Q_1$

Extended setting

- Database D of n objects
- Each object has k dimensions

$$O_i = \{O_{i,1}, \ldots, O_{i,k}\}$$

- Query Q is an object with the same k dimensions

$$Q = \{Q_1, \ldots, Q_k\}$$

Extended setting

- Database D of n objects
- Each object has k dimensions

$$O_i = \{O_{i,1}, \ldots, O_{i,k}\}$$

- Query Q is an object with the same k dimensions

$$Q = \{Q_1, \ldots, Q_k\}$$

- Distance function $d(O_i, O_j)$ between two objects

Extended setting

- Database D of n objects
- Each object has k dimensions

$$O_i = \{O_{i,1}, \ldots, O_{i,k}\}$$

- Query Q is an object with the same k dimensions

$$Q = \{Q_1, \ldots, Q_k\}$$

- Distance function $d(O_i, O_j)$ between two objects
  - If k dimensions are defined, distance function d can be, for example, Euclidean, Manhattan, etc.
  - Can be over a subset of k dimensions

Extended setting

- Database D of n objects
- Each object has k dimensions

$$O_i = \{O_{i,1}, \ldots, O_{i,k}\}$$

- Query Q is an object with the same k dimensions

$$Q = \{Q_1, \ldots, Q_k\}$$

- Distance function $d(O_i, O_j)$ between two objects
  - If k dimensions are defined, distance function d can be, for example, Euclidean, Manhattan, etc.
  - Can be over a subset of k dimensions
  - Otherwise, d can be a metric distance

Extended setting

- Database D of n objects
- Each object has k dimensions

$$O_i = \{O_{i,1}, \ldots, O_{i,k}\}$$

- Query Q is an object with the same k dimensions

$$Q = \{Q_1, \ldots, Q_k\}$$

- Distance function $d(O_i, O_j)$ between two objects
  - If k dimensions are defined, distance function d can be, for example, Euclidean, Manhattan, etc.
  - Can be over a subset of k dimensions
  - Otherwise, d can be a metric distance
  - If d is not a metric distance, then the problem is much harder

Extended setting

- Database D of n objects
- Each object has k dimensions

$$O_i = \{O_{i,1}, \ldots, O_{i,k}\}$$

- Query Q is an object with the same k dimensions

$$Q = \{Q_1, \ldots, Q_k\}$$

- Distance function $d(O_i, O_j)$ between two objects
  - If k dimensions are defined, distance function d can be, for example, Euclidean, Manhattan, etc.
  - Can be over a subset of k dimensions
  - Otherwise, d can be a metric distance
  - If d is not a metric distance, then the problem is much harder
- Point query: Given D and Q, return all objects $O_i \in D$ such that $d(O_i, Q) = 0$

Similarity search

- Range query: Given D, Q, d and range r , return all objects $T = \{O_i\}$ such that

$$d(O_i, Q) \leq r$$

Similarity search

- Range query: Given D, Q, d and range r , return all objects $T = \{O_i\}$ such that

$$d(O_i, Q) \leq r$$

- Nearest neighbour (kNN) query: Given D, Q, d and number of nearest neighbour k, return k objects $T = \{O_i\}$ such that

$|T| = k$ and for any $O_i \in T \text{ and } O_j \notin T, d(O_i, Q) \leq d(O_j, Q)$

Similarity search

- Range query: Given D, Q, d and range r , return all objects $T = \{O_i\}$ such that

$$d(O_i, Q) \leq r$$

- Nearest neighbour (kNN) query: Given D, Q, d and number of nearest neighbour k, return k objects $T = \{O_i\}$ such that

$|T| = k$ and for any $O_i \in T \; and \; O_j \notin T, d(O_i, Q) \leq d(O_j, Q)$

- Window query: Given D and ranges Q_j^{low}, Q_j^{high} in each dimension $j = 1 \ldots k$, return all objects $T = \{O_i\}$ such that

$$\forall_{j=1 \ldots k}, \; Q_j^{low} \leq O_{ij} \leq Q_j^{high}$$

Example of similarity search

Object	Dim 1	Dim 2
O_1	3	7
O_2	3	6
O_3	5	8
O_4	9	2
Q	4	5

d is Manhattan distance.
Range search: $r = 2$
kNN search: $k = 2$

Example of similarity search

Object	Dim 1	Dim 2	d(.,Q)
O_1	3	7	3
O_2	3	6	2
O_3	5	8	4
O_4	9	2	8
Q	4	5	-

d is Manhattan distance.
Range search: $r = 2$
kNN search: $k = 2$
Range search answer: $T = \{O_2\}$
kNN search answer: $T = \{O_2, O_1\}$