Prasad V. Potluri Siddhartha Institute of Technology:: Vijayawada.
Department of Computer Science and Engineering

I/II M.Tech. (CSE) (Second Semester)

17CSCS2T4 INDEXING and SEARCHING in LARGE DATABASES Credits: 4

Lecture: 4 Periods/week **Internal Assessment: 40 Marks**
Semester end examination: 60 Marks

Course Description

This course presents well-known database searching and indexing techniques. It focuses on similarity search queries, showing how to use distance functions to measure the notion of dissimilarity. This course describes low-dimensional index structures, memory-based index structures, and hierarchical disk-based index structures. The course also outlines useful distance measures and index structures that use the distance information to efficiently solve similarity search queries in High Dimensional Spaces.

Course Outcomes

At the end of this course students will be able to:

CO1: Summarize the need for indexing and searching, and their heuristics.

CO2: Outline various low-dimensional index structures

CO3: Illustrate various hierarchical disk-based index structures

CO4: Compute the distance measure between any two objects

CO5: Summarize the indexing methods used in high-dimensional spaces

Unit-1

Basics: Database Queries: Exact Search, Similarity Search, Join, Errors, Error Parameters, ROC Curve Low-Dimensional Index Structures: Hashing: Static Hashing, Dynamic Hashing, Locality Sensitive Hashing (LSH), Multi-Dimensional Hashing, Space-Filling Curves. Memory-Based Index Structures: Index Structures, Binary Search Tree (BST), Quadtree, K-D-Tree, Range Tree, Voronoi Diagram, Tries, Suffix Tree, Bitmap Index.

Unit-2

Disk-Based Index Structures: Hierarchical Structures: B-Tree and B+-Tree, K-D-B-Tree, General Framework, R-Tree, Splitting, R*-Tree, R+-Tree

Unit-3

Distances: Distance Functions: Metric Spaces, Lp Norm, Quadratic Form Distance, Cosine Similarity, Statistical Distance Measures, Bhattacharyya Coefficient, Distances between Sets of Objects, Earth Mover's Distance, Edit Distance.

Unit-4

High-Dimensional Spaces: Curse of Dimensionality, Expected Nearest Neighbor Distance, Expected Number of Page Accesses, Curse of Dimensionality, High-Dimensionality Structures: X-Tree, Pyramid Technique, IMinMax, VA-File, A-Tree, IQ-Tree

Text Book:

1. Arnab Bhattacharya, "Fundamentals of Database Indexing and Searching" by , CRC Press, 2015

Reference:

1 Cui Yu, "High Dimensional Indexing", Lecture Notes in Computer Science, LNCS 2341, Springer 2002.

2 Donald E. Knuth, "The Art of Computer Programming", Volume 3: Sorting and Searching (2nd Edition)