Fall Term, 2002

CSC 2508S: Information: Quantification, Representation, and Manipulation

Information Sheet

Instructor:

Prof. K.C. Sevcik
Bahen Centre for Information Technology, Room BA 4260
sevcik@cs
978-6219

Lectures:

Time: Mondays and Wednesdays, 1:10 to 2:25
Place: Lorne Pratt Building, Room LP 378

Requirements:

Assignments

Several short assignments in which you will be asked to exercise your understanding of the algorithms and techniques presented in class, or to suggest an appropriate representation for a certain type of information in light of the way in which it is to be manipulated.

Project

A report (15 to 25 pages) on a topic of your choice related to the course material. The report can be either a survey of some papers on a particular subject, or a description of some original investigation or experimentation.

Presentation

You will be asked to present the initial results of your project to the class (in 15 to 20 minutes).

Test

An "open book" and "open notes" test at the end of the term.

Discussion

Participation in class discussions is strongly encouraged.

Marking Basis:

Assignments	30%
Project	30%
Presentation	10%
Term Test	25%
Class Discussion	5%
TOTAL	100%

Course Material:

There is no text for the course, but references to relevant books and journal and conference papers, some recent and some older, will be provided. Informal lecture notes will be distributed as the course progresses.

[see next page]

Unifying Themes:

- Concise Encodings
- Approximate Representations (e.g., histograms, transforms)
- Aggregation and Summarization (e.g., data cubes)
- Hierarchical, Progressive Representations

Tentative List of Topics:

I. Quantification of Information
- A. Discrete Information Theory
- B. Data Compression (lossless and lossy)
- C. Encoding Techniques

II. Varieties of Information
- A. Bits, Integers, Real Numbers, and Strings
- B. Text and Documents
- C. Computer Programs
- D. Files of Records (e.g., Relational Databases)
- E. Figures, Graphs and Maps
- F. Sounds and Music
- G. Natural Images and Video Sequences
- H. Game and Activity Sequences
- I. Odors, Tastes, etc.
- J. Concepts, Knowledge Bases

III. Representation of Information
- A. Dominant Single-key File Structures (B-trees vs. hashing)
- B. Dynamic File Structures (extendible hashing vs. spiral storage)
- C. Multiple-key File Structures (combined indices, inverted files, transposed files, k-d-B-trees, quad trees, grid files, etc.)
- D. Spatial Data Structures (R-trees, X-trees, Filter Trees, etc.)
- E. Data Structures for High Dimensionality Spaces (VA-file, SS-tree, SR-tree, A-tree, IQ-tree)

IV. Manipulation of Information
- A. Point and Range Queries
- B. Nearest Neighbour(s) Queries
- C. Spatial Joins
- D. Clustering

V. Relationship to Human Memory Structures
- A. Characteristics of Human Memory
- B. A Computational Model of Human Memory (Parallel Sparse Distributed Memory)