

NEW COURSE

Topics in Databases: Warehousing, Filtering and Data Mining

CIS 750, Section 001 - Spring 2001
Time: Wednesday, 4:40-7:20pm; Place: Tuttleman 1A

Instructor: Zoran Obradovic

303 Walkman Hall, zoran@joda.cis.temple.edu, phone: 215 204 6094

Office hours: Wednesday 2-3pm and by appointment

Goals:

Traditional database management systems were optimized for an efficient online transaction processing. It has been recognized that in addition to query and report tools, high-level decision making also profits from an efficient support for verifying or disproving user specified hypotheses about relationships in data as well as for finding previously unknown and potentially interesting relationships automatically. The course is devoted to information system environments enabling efficient indexing and advanced analyses of current and historical data for strategic use in decision making. Data management will be discussed in the content of data warehouses/data marts; Internet databases; Geographic Information Systems, mobile databases, temporal and sequence databases. Constructs aimed at an efficient online analytic processing (OLAP) and these developed for nontrivial exploratory analysis of current and historical data at such data sources will be discussed in details. The course is structured to provide ample opportunity for participants to learn about a growing new research area, and scout around for promising research topics by a hands-on experience on problems in financial engineering, e-commerce and bioinformatics areas studied at the Center for Information Science and Technology at Temple University.

Topics will be tailored to interests of the participants. Content will include:

- **Foundations** (Cluster Systems; Parallel and distributed databases; File organizations, Tree-structured indexing, Hash-based indexing)
- **Data Warehousing and Data Marts** (Components, Design, Implementation, Mappings, Metadata, Amortized indexing)
- **Decision Support** (OLAP Technology, Bitmap and join indexes, Views, Computing on Demand, Online Aggregation)
- **Data Modeling** (Model functions, Representation, Multi-models integration, Assessing statistical significance, Error vs. complexity trade-off, Curse of dimensionality)
- **Internet and Distributed Databases** (HTML and beyond, Semistructured data models, Text indexing, Web mining, Collaborative filtering, Distributed data mining)
- **Spatial Data Management** (Geographical information systems, Spatial indexing, R Trees, DBSCAN, partitioning heterogeneous spatial data)
- **Spatial-Temporal Databases** (SQL extensions, Dependencies analysis, Similarity search, Residual analysis, Financial and Bioinformatics applications)

Prerequisites: The course will be self-contained. Elementary computer programming skills and an undergraduate course in databases or data structures are sufficient. Otherwise, the instructor's permission is needed.

Grading: Homework, reading assignments and an individual research project.