Information Organization and Retrieval

CIS*4430 (Winter 2018)

Instructor: *Fei Song* (MacLachlan 215, ext. 58067)
Email: fsong@uoguelph.ca
Course Website: https://courselink.uoguelph.ca/
Office Hours: Mondays and Fridays, 3:00 – 4:00 pm, in MacLachlan 215

Overview

The main objective of this course is to examine the implementation details of database management systems, as opposed to data modeling and user-oriented aspects covered in CIS*3530. The emphasis is on various methods for storing and retrieving data in relational databases, including data organization, indexing, hashing, and query processing. The course may also cover some advanced topics, such as textual databases, distributed databases, XML databases, data warehousing, and data mining. In some sense, the course serves as a "finishing" touch in computer science, since many technical concepts from different sub-fields are integrated within a common context, including file structures, operating systems, language implementations, database operations, and the evaluations of algorithms.

Evaluation

Four Assignments	52% (due on Feb/2, Mar/2, Mar/23, Apr/6, respectively)
Midterm	18% (scheduled on Feb. 16 during the class)
Final Exam	30% (To be scheduled by the Registrar's office)

Note (1) Each assignment includes both paper-and-pencil questions and programming exercises, and should be done individually.

Note (2) Late submissions for assignments are acceptable, but there will be a deduction of marks by 10% for one day late, 25% for two days late, and 50% for three days late. No marks will be given for submissions that are more than three days late.

Note (3) All major functionality of the programming exercises should be coded independently unless approved by the instructor. In such cases, the re-used code should be documented clearly; otherwise, it is considered as plagiarism and will be not tolerated at the University of Guelph. Please refer to the rules about Academic Misconducts in Section VIII: Undergraduate Degree Regulations and Procedures of the Undergraduate Calender at https://www.uoguelph.ca/registrar/calendars/undergraduate/current/.

Textbook and References

(Textbook) Hector Garcia-Molina, Jeffrey D. Ullman, and Jennifer Widom. *Database Systems: The Complete Book*. Second Edition. Pearson/Prentice-Hall, 2009.

(Reference) Abraham Silberschatz, Henry F. Korth, and S. Sudarshan. *Database System Concepts*. Sixth Edition. McGraw Hill, 2010.

(Reference) Ramez Elmasri and Shamkant B. Navathe. *Fundamentals of Database Systems*. Sixth Edition. Pearson/Addison Wesley, 2010.

Topics to be Covered:

Part I: Review of Database Systems and Storage Devices
- Overview of a database system architecture
- Magnetic and optical disks

Part II: File Structures
- Sequential structures and sequential-style processing: field and record structures and co-sequential processing algorithms
- Indexing and hashing structures: indexed-sequential, B-trees, B+trees, tries, hashing, and self-adjustable hashing structures
- Multidimensional structures: hash-like structures, tree-like structures, and bitmap indexes.

Part III: Implementation Issues
- Overview of relational algebra
- Query compilation and optimization

Part IV: Textual Databases and Natural Language Processing.
- String search techniques: string matching, signatures, and PAT trees
- Vector-space model: document indexing, document clustering, and query processing
- Statistical natural language processing

Part V: Advanced Topics (Time Permitting)
- Distributed relational databases
- XML databases
- Data warehousing
- Data mining