Welcome and Introduction

Welcome to CMPT 454

- Who am I?
- Overview of the course
- Administrative issues

Who Am I?

- **Bin Zhou**, last year Ph.D. Candidate
- Research direction: databases, data mining, information retrieval, and machine learning
- Supervisor: Dr. Jian Pei
- Thesis committee: Dr. Martin Ester and Dr. Ke Wang
- Homepage:

  http://www.cs.sfu.ca/~bzhou/personal

  or

  Google “Bin Zhou”, and the 1st one, you got it!

What Did We Learn from 354?

- Queries, queries, queries
  - SQL, relational algebra, relational calculus
  - XML and XQuery
- Relational database design
  - ER design
  - Relational database design
- Application design and development
  - Interfaces to other programming environment
- Summary: How to use (relational) databases

What Does 354 Not Tell?

- What is inside database systems?
  - How is data organized?
  - How are queries conducted?
  - How to process multiple simultaneous queries?
- How to handle more complex data and queries?
  - Text data
  - Keyword search
- How to search massive data?
  - The internet and web search engines

Content of CMPT 454

- Four units
  - Transaction management
  - Data storage and query answering
  - Similarity search and advanced queries
  - Data mining and information retrieval

Units (1)

- Transaction management
  - Transactions
  - Concurrency control
  - Recovery system
  - Distributed databases
- Data storage and query answering
  - Data storage and disk structure
  - Index techniques
  - Query processing
  - Query optimization

Units (2)

- Similarity search and advanced queries
  - Spatial indexes
  - Nearest neighbour search
  - Ranking queries
  - Skyline queries
- Data mining and information retrieval
  - Modeling
  - Ranking, indexing, and searching
  - The internet and web search engines

Grading Scheme

- Reading and to do list for every lecture (no hand-in)
  - You need to read the textbook before and after lectures
  - To do list helps you to get a better understanding
- **4 assignments (40%)**, no course projects, no programming tasks
- Two exams
  - **Midterm exam (20%)**: in early March, covers transaction management, data storage and query answering
  - **Final exam (40%)**: covers all the materials
- No alternative marking scheme
- Grades based on the curve

Textbook and References

- Textbook: "**Database Systems - The Complete Book**", 2nd edition, chapters 13-20, 22, 23

- References:
  - **Database Management Systems**, 3rd edition, by Raghu Ramakrishnan and Johannes Gehrke. McGraw-Hill, 2003.
  - **Database System Concepts**, 5th edition, by Abraham Silberschatz, Henry F. Korth, and S. Sudarshan. McGraw-Hill, 2005.
  - **Introduction to Information Retrieval**, by Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze, Cambridge University Press, 2008.

Lecture Notes

- The textbook does not cover all materials in this course
  - Attending the classes is extremely important!
- Assignments and exams will be based on lecture notes
  - Notions and terminology in lecture notes should be followed.
- Check the course homepage regularly. Information will be updated time to time.

Office Hours

- Office hours
  - TBA
- Emails
  - Mailing list: cmpt-454@sfu.ca
  - Instructor: bzhou@cs.sfu.ca
  - TA: TBA
- Special meetings by appointment only

Questions and Suggestions?