CPSC 504: Data Management

Rachel Pottinger

Course Introduction

2024/01/08

What is this class about?

- Understanding how people design database management systems
- Data management *research*
- Managing your data

What is data management research?

- Research about managing data including:
  - Traditional (relational) database management systems
    - What they are, how to make them work
  - Other kinds of databases
    - Object-oriented, XML, No SQL
  - Other data management applications
    - OLAP, data mining, etc.

This class is a research seminar

A seminar is "a small group of advanced students … under the guidance of a professor who meets regularly with them to discuss..." [dictionary.com]

- 1 or 2 papers to read for most classes. I'll provide:
  - An explanation why we're reading the paper
  - Necessary background beforehand
  - Suggestions on how to read papers where necessary
- Most days students present papers and lead discussions. You'll present once and lead discussion once. I'll provide:
  - The high level goals of reading papers
  - A set of suggested discussion questions
  - Feedback on your plans and answers to questions
  - Possibly a preliminary suggested set of slides
- Sometimes I'll present the papers and lead discussion
- You will learn about database *research*. If you're here to learn how to create a database/do database programming, this is the wrong class for you.

But I haven't taken any database classes!

- You do not need to have taken a database class

- Assuming you have a solid basis in CS (i.e., have majored in CS or closely related field or are a CS graduate student), I will provide you with all background material you need

- Everyone who is not in the class yet but wants to get in, come and see me at the end of class.

What *are* the prerequisites for this class?

- A background in computer science
  - Doesn't necessarily have to have majored in computer science, but you need a smattering of systems, theory, data structures, etc.
- Ability to read and respond to 1 – 2 research papers a class
- Ability to do a project (not necessarily implementation based) either in a group or on your own
- Ability and willingness to present papers and lead discussions
- Willingness to discuss your own ideas and questions in class

Where does my grade come from?

- Analyzing the readings – 30%
  - Post a summary/analysis on Canvas
- Presenting/leading class discussions – 20%
  - One person presents the content
  - One person leads discussion
  - You'll sign up for *different* days for this
- Course project – 30%
  - See website
  - Doesn't have to have an implementation
- One homework – 5%
- In-class participation – 15%

A brief discussion of plagiarism

- Please see discussion of how to cite things on project page

- To cite an idea, cite the source at the end on the sentence/idea.

- To cite more than a few words, use quotation marks.

- To make sure that you don't plagiarize, always add in citations where appropriate as you are working on your paper.

- Using ChatGPT or a similar system on any work that you turn in also constitutes plagiarism.

COVID-19: Why do I care?

- A lot of people seem to think (or wish) that COVID-19 is over.

- It isn't.

- Additionally, people with severe medical conditions can have very bad outcomes and also not respond to vaccines or be able to take vaccines.

- Please protect yourself and those around you who may have a medical condition or live with someone who does by taking precautions, including getting vaccinated and wearing a mask in crowded indoor spaces.

- It's up to you if you wear a mask in this class, but if you are sick PLEASE STAY HOME.

My reason being so cautious

Any administrative questions?

Introduce your partner

Discuss the answers to the following questions with your partner for next 5 minutes:

- What is your name?
- Where are you from?
- What is your database/data management background?
- What do you want to get out of this class?

Grab a card and a marker and write your name on the card

Some reasons to use a database:

- Large amounts of data
- Structured data
- Persistent data
- Valuable data
- Performance requirements
- Concurrent access to data
- Restricted access to data

What data is stored in databases?

This space intensionally left blank

What data do you have?

Here is some data *I* have:

- Papers I've read
- Addresses
- Job search data
- Experiments I've run
- Grades
- Books I own
- Powerpoint slides
- Research notes
- E-mail
- Drafts of research papers and notes from students

Class outline

- Crash course in databases
- Standard Relational databases
- New Relational Databases
- Other data models
- Management of other data
- Advanced topics/Student request potpourri

To do:

Course website: http://www.cs.ubc.ca/~rap/teaching/504/2023W2

- Mailing list: mail majordomo@cs.ubc.ca with "subscribe cpsc504" in the body
- Check out Canvas (see course webpage)
- Think more about data you have
- Think about which topics you'd like to present/lead discussion on (first come, first served)
- Read the project description, and think about projects
- If there are any topics you'd like to see covered that aren't, let me know
- If you are not registered in the class and want to be, come talk to me