COMP9311:
Database Systems

Term 2 2019

Week 1 Monday (Intro and Data Modelling)

By Helen Paik, CSE UNSW

Disclaimer: the course materials are sourced from previous offerings of COMP9311 and COMP3311

Who's who in COMP9311

- Lecturer-in-Charge (LiC)
  - Helen Paik
  - Office: K17 401A, Ext: 54095
  - Drop-in consultations: Monday 3:30-5:30pm (Room 401A)
  - Email: **cs9311@cse.unsw.edu.au**
- Tutors
  - Owen Riddy (CSE postgrad student)
  - Zhuang Li (CSE postgrad student)
  - Sahil Punchhi (CSE postgrad student)
- Course Web site
  - http://www.cse.unsw.edu.au/~cs9311

Why Study Databases?

Every significant modern computer application has Large Data.

This needs to be:

- stored (typically on a disk device)
- manipulated (efficiently, usefully)
- shared (by many users, concurrently)
- transmitted (all around the Internet)

Red stuff handled by databases;

brown by networks.

Challenges in building effective databases: efficiency, security, scalability, maintainability, availability, integration, new media types (e.g., music, video), ...

Why Study Databases?

Intelligent Transportation

Business

Natural

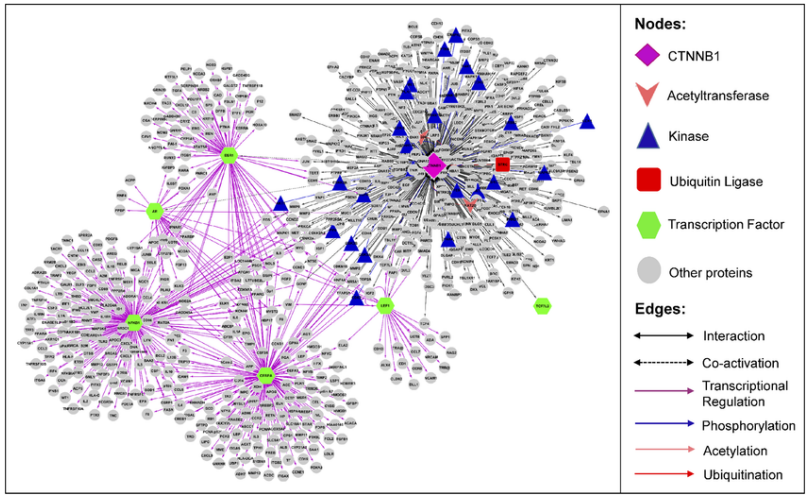
Public Health

Modern Military

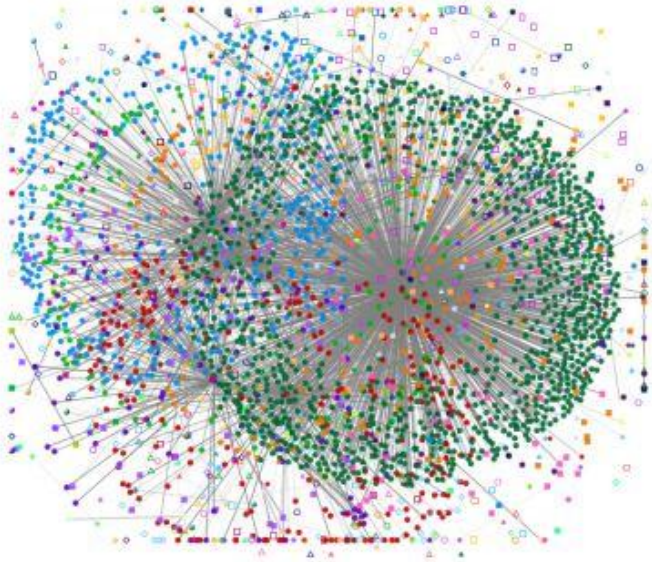
Tourism Development

4

Why Study Databases?



Beta-Catenin Biological Network



Web Graph

Social Network

Databases: Important Themes

The field of *databases* deals with:

- *data* ... representing application scenarios
- *relationships* ... amongst data items
- *constraints* ... on data and relationships
- *redundancy* ... one source for each data item
- *data manipulation* ... declarative, procedural
- *transactions* ... multiple actions, atomic effect
- *concurrency* ... multiple users sharing data
- *scale* ... massive amounts of data

What is Data?

Data (according to Elmasri/Navathe)

- Known facts that can be recorded and have implicit meaning

Example:

- *Students – name, DOB, courses enrolled, etc.*
- *Movies – title, year published, director, etc.*

What is Database?

Database (Elmasri/Navathe)

- ... a collection of **related** data ...

Data items alone are relatively useless

We need the data to have some structure

Example:

- a student records database
  - Contains information identifying students, courses they are enrolled in, results from past courses, ...
- IMDb movie database
  - Contains information about the movies, actors, theatres showing the movies, other movies the director also directed in, …

What is Database Management System

Database can be manipulated by a **database management system (DBMS)**

According to Elmasri/Navathe:

- *DBMS*: ... a collection of programs that enables users to create and maintain a database ...
- *Database system (in short)*: ... the database and DBMS together ...

Database related courses in CSE

COMP9311 introduces foundations & technology of databases

- skills: how to build database-backed applications
- theory: how do you know that what you built was good

After COMP9311 you can go on to study ...

- COMP9315: how to build relational DBMSs (write your own PostgreSQL or Oracle)
- COMP9318: techniques for data mining (discovering patterns in DB)
- COMP9319: Web data compression and search (dealing with a large amount of Web data)
- COMP6714: information retrieval, web search (dealing with text data)
- COMP932[1|2|3]: service-oriented computing, which relies on DB background

Syllabus Overview

Data modelling and database design (Week 1 to Week 3)

- ER model, ODL, ER-to-relational
- Relational model (relational algebra), mapping of ER to relational model

Database application development (Week 3 to Week 6)

- SQL, views, stored procedures, triggers, aggregates
- PostgreSQL: psql (an SQL shell), PLpgSQL (procedural)
  - » PostgreSQL – lab, assignment environment
  - » SQLite: sqlite3 (an SQL shell), final exam environment
- Programming language access to databases (Python, Database connectivity)

Syllabus Overview

Formal database design theory and database system architecture (Week 6 – Week 9)

- Normalisation, functional dependencies
- DBMS architecture: client/server, file system, relational engine
- Storage and indexing, data access operations
- Query processing: translation, optimisation, evaluation
- Transaction processing: transactions, concurrency control, recovery

Future of Databases (Week 10)

- Limitations of RDBMS's, modern/future technologies

- Assignment 1 on Weeks 1-3 topics
- Assignment 2 on Weeks 3-6 topics
- Quizzes and Final Exam on weekly labs, all weeks.

Final mark = A1+A2+Quizzes+Final Exam

Teaching/Learning Approaches

Stuff that is available for you:

- *Texts*: describe most syllabus topics in detail
- *Lectures*: describe all syllabus topics in some detail, with exercises and examples
- *Lab exercises (starting Week 2, every week)*: guides you through the practical skills on the database application programming part of the course
- Extra video tutorial materials on some topics

What we do to support your learning:

- Face-to-face support sessions
  - We run consultations in the lab (more tutor time available around assignment weeks)
  - Weekly LiC consultations (every Monday after the class)
- Online support
  - course *Forums* (under WebCMS3)
  - send an email to cs9311@cse.unsw.edu.au (regularly monitored)

Note: the scheduled lab class times are no longer valid. You don't need to attend (see class timetable)

Teaching/Learning Approaches

Things that we expect you to **do**:

- Follow the lecture content (i.e., weekly topics)
- Theory *exercises*: exercise/example questions (try them yourself)
- *Prac* work: lab-like exercises

You will show us your progress/learning outcomes through:

- *Assignments*: extended practical exercises
- *Quizzes*: weekly progress check
- Final exam: your learning outcomes on **both** practical and theoretical topics

Assignments

Two assignments, which are critical for *learning*

- Data modelling, Relational modelling and SQL definitions, due Week 5 Thursday 5pm
- Queries/Procedures, SQL/PLpgSQL, due end week 9 Thursday 5pm

All assignments are done *individually*, and ...submitted via WebCMS3

Some are automarked (so you must follow the specification exactly)

- plagiarism-checked (copying solutions ⇒ **referred to the student ethics officer, you are likely to get 0 for the course, or 0 for the component**)
- rent-a-coder monitored (buying solutions ⇒ **referred to the student ethics officer, this is considered a highly serious offence by the university**)

Quizzes

- 9 quizzes (planned), all together worth 10 marks
- cover material covered in each week
- aim to make you review materials weekly
- done via WebCMS3 in your own time
- starting Week 1 – check the course schedule
  - Released every Wednesday, Due the following Monday morning
  - You can ask the tutors if you got any questions wrong

Exam

3-hour **lab-based** exam during exam period

Comprising a mixture of

- SQL, PLpgSQL, design exercises, analyses
- Prac part: SQL using SQLite
- All questions: typed in and submitted online
- Only PostgreSQL/SQLite documentation is accessible during exam

Sample exams will be available in the course website in due course …

Supplementary Exam and Special Consideration

Everyone gets **exactly one chance** to pass the Exam

If you attend the Exam:

- I assume that you are fit/healthy enough to take it (Fit-to-Sit rule at UNSW)
- no 2nd chance exams, even with a medical certificate

All Special Consideration requests:

- must *document* how *you* were affected
- must be submitted to UNSW (see the course outline)
- Supplementary Exams are held shortly after the exam period; don't leave town

Textbook (there are many good ones!)

Elmasri, Navathe
Fundamentals of Database Systems (7th ed, 2016)

Other good references:

Garcia-Molina, Ullman, Widom
Database Systems: The Complete Book (2nd ed, 2008)

Ramakrishan, Gehrke
Database Management Systems (3rd ed, 2003)

Silberschatz, Korth, Sudarshan
Database System Concepts (6th ed, 2010)

Kifer, Bernstein, Lewis
Database Systems: Application-Oriented Approach (2nd ed, 2006)

Earlier editions of texts are ok

NOTE: Typically database topics are taught in an order that is considered suitable by the course designer (e.g., assignment topics). Each week, I will point out which chapters of the textbook are relevant.

Database System Management

Two example DBMSs for prac work:

- SQLite (open-source, free, no server needed)
- PostgreSQL (open-source, free, full-featured)

Comments on using a specific DBMS:

- the primary goal is to learn SQL (a standard)
- the specific DBMS is not especially important
- but, each DBMS implements non-standard features
- we will use standard SQL as much as possible
- PG docs describe all deviations from standard

Further Reading Material

The on-line documentation and manuals provided with:

SQLite are reasonably good (http://www.sqlite.org/docs.html)

PostgreSQL are very good (http://www.postgresql.org/docs/)

Python on Database Programming are similarly comprehensive

Some comments on these technologies/language specific books:

- tend to be expensive and ***short-lived***
- many provide just the manual, plus some examples
- generally, anything published by O'Reilly is useful
- Aside: once you understand the concepts, the manual is sufficient