COMP9318 Course Introduction

Table of Contents

1. Pre-requisites
2. Course philosophy and teaching strategies
3. Course aims
4. Learning outcomes
5. Administrative Components
6. Assessment
7. Reading e-mail
8. Academic honesty and plagiarism
9. Course Material
10. Further information
11. Continual Course Improvement

Course Code	COMP9318
Course Title	Data Warehousing and Data Mining
Units of Credit	6
Course Website	https://www.cse.unsw.edu.au/~cs9318/21t1/
Handbook Entry	https://www.handbook.unsw.edu.au/postgraduate/courses/2021/COMP9318/

1. Pre-requisites

The formal pre-requisite for this course is COMP9020 (Foundation of Computer Science) and COMP9024 (Data Structures and Algorithms) and COMP3311/9311 Database Systems.

The knowledge that we assume from COMP9020 is:

- the ability to prove or disprove propositions mathematically.
- basic maths including set and relation theory, recursion, and probability.
- design and analysis of algorithms.

The knowledge that we assume from COMP9024 is:

- experience with python programming, and an understanding of a range of in-memory searching data structures (e.g., binary search trees, 2-4 trees, and hash-tables).

The knowledge that we assume from COMP3311/9311 is:

- experience with relational data model and SQL query language, and an understanding of data structures and algorithms to enable efficient and scalable management of massive amount of data (e.g., B+-tree indexes).

2. Course philosophy and teaching strategies

The learning foci in this course are primarily lectures (theoretical knowledge) and projects (practical knowledge). The course will have an emphasis on problem solving for real applications.

Students will learn the main contents of the course through lectures. Tutorials are available to assist students to obtain in-depth understanding of course materials and develop problem solving skills by working on tutorial questions.

3. Course aims

This course aims to introduce the foundation of data warehousing, the theories of various data mining techniques and explore the practice of developing data mining applications. This course is one of the advanced database course series. Other advanced database courses include:

- COMP9313 (Big Data Management),

- COMP9315 (DBMS Implementation),
- COMP9319 (Web Data Compression and Search),

The course is designed to be practical. As such, real-life examples of data mining issues and applications will also be used throughout the course.

4. Learning outcomes

Students successfully completing this course will be able to:

- understand the whole process of data mining and knowledge discovery from databases
- understand the data models and query processing mechanisms used in Data Warehouses
- understand various data mining techniques and their variants
- develop solutions for real problems using existing data warehousing and data mining technologies.
- appreciate the past, present and future of data warehousing and data mining technology

The learning outcomes are closely related to UNSW graduate attributes 1 — 6. For example, bonus marks might be given in the assessment in order to encourage independent learning and critical thinking.

5. Administrative Components

See the course homepage (http://www.cse.unsw.edu.au/~cs9318) for (up-to-date) information regarding **Course Staff**, **Course Schedule**, and **Course Resource List**.

As the course is continuously developing together with the data warehousing and data mining fields, the course schedule is subject to change too. Please read the introduction slides (lecture notes of the first week) for the course schedule in the current offering.

6. Assessment

The assessment will have the following components:

- Written assignment (full mark: 25): This component helps review the concepts introduced in lectures,
- Labs (full mark: 25): This component helps you gain better understanding of the algorithms and technologies by implementing them,
- Programming project (full mark: 50): This component gives you the opportunities to apply the technologies to solve real problems,
- Written final exam (full mark: 100): This component assesses the various facts-and-knowledge level learning outcomes.

```
Final Mark = 2 * (assn + labs + proj) * FinalExam / (assn + labs + proj + FinalExam)
```

Note that:

- we will use the average of your best 3 out of 5 labs as your mark for labs;
- you also need to score at least 40 (out of 100) in the exam to pass the course. Otherwise, your score will be soled determined by your exam mark.

Both written and programming assignments are helpful to achieve deep understanding of the course materials and develop problem solving abilities.

Relationship to the Learning Outcomes:

- All assignments and the final examination are used to assess whether you have achieved the objectives of the subject.
- The labs and programming assignment, in particular, give you the opportunities to engage in real problems, and helps you to develop in-depth understanding of the course materials and develop problem-solving skills.

Grading Criteria: Grading criteria for each assessment will be detailed in the specification.

Late submission: Assignments/projects submitted late are subject to late penalties, which are specified in the assignment/project specifications.

Assignment submission: Assignment submission procedure is described in the assignment specification document, which will be linked to this page when the assignment specification becomes available. Generally assignments are submitted electronically using the give program running on the School's computer systems (in labs, and on servers). Details are in the assignment specifications.

7. Reading e-mail

You should check your school e-mail frequently in case of announcements relating to this course. We assume that you read e-mail sent to your CSE account by the next working day during teaching sessions.

- How to forward your CSE E-mail to another e-mail address
- CSE E-mail guidelines

8. Academic honesty and plagiarism

Copying assignments is unacceptable. Assignments will be checked. The penalties for copying range from receiving no marks for the assignment, through receiving a mark of 00 FL for the course, to expulsion from UNSW (for repeat offenders). Allowing someone to copy your work counts as plagiarism, even if you can prove that it is your work.

There are several on-line sources to help you understand what plagiarism is and how it is dealt with at UNSW:

- Plagiarism and Academic Integrity
- UNSW Plagiarism Procedure

Make sure that you read and understand these. Ignorance is not accepted as an excuse for plagiarism. In particular, you are also responsible that your assignment files are not accessible by anyone but you by setting the correct permissions in your CSE directory and code repository, if using. Note also that plagiarism includes paying or asking another person to do a piece of work for you and then submitting it as your own work.

UNSW has an ongoing commitment to fostering a culture of learning informed by academic integrity. All UNSW staff and students have a responsibility to adhere to this principle of academic integrity. Plagiarism undermines academic integrity and is not tolerated at UNSW. Plagiarism at UNSW is defined as using the words or ideas of others and passing them off as your own.

If you haven't done so yet, please take the time to read the full text of

- UNSW's policy regarding academic honesty and plagiarism

The pages below describe the policies and procedures in more detail:

- Student Code Policy
- Plagiarism Policy Statement
- Plagiarism Procedure
- Student Misconduct Procedure

9. Course Material

- All lecture notes, tutorial and lab notes will be posted on the course web site. These notes summarises the major contents and help you to understand the materials when you read the textbook later. You definitely need to read the corresponding chapters in the textbook to gain a full understanding of the course materials.
- [RLU13] is available free online (thanks to the authors!)
- [JPT10] is available free available using an UNSW IP.
- We've got several copies of [HK00] on **standard loan** and **two hour loan** in the library.

```
[RLU13] Mining of Massive Datasets (ver 2.1), Anand Rajaraman,
Jure Leskovec, and Jeffrey David Ullman, 2013. Available at
  http://mmds.org/

[JPT10] Multidimensional Databases and Data Warehousing,
Christian S. Jensen, Torben Bach Pedersen, Christian Thomsen.

[HK00] Data Mining: Concepts and Techniques, Jiawei Han and
Micheline Kamber. Kaufmann Publishers, August 2000. ISBN:
1-55860-489-8.
```

- Reference books for this course are:

```
[WF00] Data Mining : Practical Machine Learning Tools and
Techniques with Java Implementations, Ian H. Witten, Eibe
Frank. Morgan Kaufmann, 2000. ISBN: 1558605525.

[TSK05] Introduction to Data Mining, Pang-Ning Tan, Michael
Steinbach, Vipin Kumar. Addison-Wesley, 2005. ISBN: 0321321367.
```

[Aggarwal15] Data Mining: The Textbook, Charu C. Aggarwal. Springer, 2015.

- Useful Resources
  - KD Nuggets
  - The OLAP Report

10. Further information

- Students enrolled in COMP9318 are expected to attend all classes
- The use of School of Computer Science and Engineering computing laboratories is subject to rules described in the Yellow Form, which you acknowledge (electronic) receipt of when you receive your computing account. The Yellow Form also outlines what to do in case illness or misadventure that affects your assessment, and supplementary examinations procedures within the School of Computer Science and Engineering.
- Information on UNSW Occupational Health and Safety policies and expectations
- Equity and Diversity issues

11. Continual Course Improvement

Each year feedback is sought from students and other stakeholders about the courses offered in the School and continual improvements are made based on this feedback. UNSW's Course and Teaching Evaluation and Improvement (CATEI) Process (http://www.unsw.edu.au/learning/pve/catei.html) is one of the ways in which student evaluative feedback is gathered. Significant changes to courses and programs within the School are communicated to subsequent cohorts of students.

There were no serious issues noted in the previous evaluation. We are making incremental refinements to preserve the already high standard of the course.

Yifang Sun

Last updated 2021-02-08 22:40:11 AEDT