

---

# About COMP 9318 (2008 s1)

Wei WANG

The University of New South Wales

[weiw@cse.unsw.edu.au](mailto:weiw@cse.unsw.edu.au)

---

① Logistics

---

# INTRODUCTION

## Lecturer-in-charge

Dr. Wei Wang

School of Computer Science and Engineering

Office: K17 507

E-mail: [weiw@cse](mailto:weiw@cse)

Ext: 9385 7162

<http://www.cs.unsw.edu.au/~weiw>

## Research Interest:

- DB + IR (e.g., google-style keyword search in RDBMS)
- Similarity search and join (e.g., plagiarism detection)
- Indexing and query processing

---

# COMP 9318

## Course Info

- Homepage: <http://www.cse.unsw.edu.au/~cs9318>
- **Email:** cs9318 AT cse.unsw.edu.au  
**IMPORTANT: only** accepts emails from @cse.unsw.edu.au Or @unsw.edu.au.
- Lectures:
  - 1800 – 2100 THU, Civil Engineering G1 (K-H20-G1)
- Tutorials: 4 tutorials (after each “section” of the course)
- Labs: n/a.
- Consultations: 1600 – 1700 MON, K17-507

---

# ASSESSMENT

## Overview

- 2 quizzes + 1 (written) assignments + 1 (programming) projects + 1 final exam

## Assignments and Quizzes

- ass1: Various topics (100%)
- q1 — q2: learn to use (open source) softwares (100%)
- **Penalty for late submissions: 0 mark if late !!!**

## Projects

- proj1:
  - Implement a *clustering Search Engine* in Java (90%)
  - Innovative features (10%)

*Alternatively*, there are alternative R&D projects you can work on, or just propose your own — discuss with me *no later than W4*. Example projects include literature review and implementation of proposed

---

systems in other research papers.

→ **Penalty for late submissions:** 10% soft reduction for the 1st day,  
then 30% soft reduction per day !!!

---

## FINALLY ...

### Exam

- If you are ill on the day of the exam, **do not attend** the exam — I will **not** accept medical special consideration claims from people who have already attempted the exam.

### Final Mark

- Aggregated by **Harmonic Mean**:

$$q = \frac{1}{2} \cdot \sum_{1 \leq i \leq 2} q_i$$

$$t = \frac{1}{3} \cdot (ass1 + q + proj1)$$

$$final\_mark = \frac{t \cdot exam}{0.5 \cdot t + 0.5 \cdot exam}$$

Example:

---

<i>t</i>	<i>exam</i>	<i>final_mark</i>
100	40	57.1
70	70	70.0

---

# RESOURCES

## Text Book

- Han *et al*, *Data Mining: Concepts and Techniques*, 1st/2nd edition, Kaufmann Publishers.

## Reference Books

- Witten *et al*, *Data Mining : Practical Machine Learning Tools and Techniques with Java Implementations*, 1st/2nd edition, Morgan Kaufmann.

## Software

- Pentaho (Mondrian, Kettle)
- Weka

## Reading Materials

- Papers from DM conferences/journals, white papers, surveys, etc.
- All available from the course Web page.

---

## COURSE SCHEDULE (TENTATIVE)

### Theme: Text Mining

---

Week	Contents	Assignments	Tutorials
1	Course overview + Introduction		
2	Data warehousing and OLAP		
3	Data warehousing and OLAP		
4	Data Cleaning	q1	tut1
5	Data Cleaning & Information Retrieval		
B	—	proj1	—
6	Clustering analysis	ass1, q1 due	tut3
7	Clustering analysis		
8	Classification and prediction		tut2
9	Text Classification	q2	
10	Association Rule Mining	ass1 due,	tut4
11	Association Rule Mining		
12	Review	proj1 due, q2 due	tut5

---

---

## LEARNING OUTCOMES

- Be familiar with major data mining tasks and be able to formulate a practical problem into a DM problem
- Understand the algorithmic challenges and pitfalls in solving DM problems.