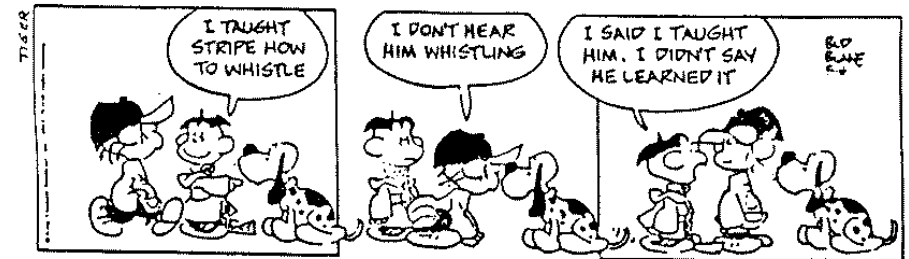# Principles of Knowledge Discovery in Data

Fall 2002

Dr. Osmar R. Zaïane

University of Alberta

Class and Office Hours

Class:
Tuesdays and Thursdays from 11:00 to 12:20

Office Hours:
Tuesdays from 15:00 to 16:00

Course Requirements

- Understand the basic concepts of database systems
- Understand the basic concepts of artificial intelligence and machine learning
- Be able to develop applications in C/C++ or Java

Course Objectives

To provide an introduction to knowledge discovery in databases and complex data repositories, and to present basic concepts relevant to real data mining applications, as well as reveal important research issues germane to the knowledge discovery domain and advanced mining applications.

Students will understand the fundamental concepts underlying knowledge discovery in databases and gain hands-on experience with implementation of some data mining algorithms applied to real world cases.

Evaluation and Grading

There is no final exam for this course, but there are assignments, presentations, a midterm and a project.
I will be evaluating all these activities out of 100% and give a final grade based on the evaluation of the activities.
The midterm has two parts: a <u>take-home exam</u> + <u>oral exam</u>.
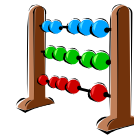
- Assignments (4)   16%
- Midterm   25%
- Project   39%
  - Quality of presentation + quality of report + quality of demos
  - Preliminary project demo (week 12) and final project demo (week 16) have the same weight
- Class presentations   20%
  - Quality of presentation + quality of slides + peer evaluation

More About Evaluation

Re-examination.

None, except as per regulation.

Collaboration.

Collaborate on assignments and projects, etc; do not merely copy.

Notes and Textbook

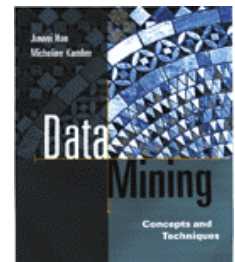## Course home page:
http://www.cs.ualberta.ca/~zaiane/courses/cmput695/

We will also have a mailing list for the course (probably also a newsgroup).

Textbook:
Data Mining: Concepts and Techniques
Jiawei Han and Micheline Kamber
Morgan Kaufmann Publisher, 2001
ISBN

Other Books

- Principles of Data Mining
  - David Hand, Heikki Mannila, Padhraic Smyth,
    MIT Press, 2001, ISBN 0-262-08290-X
    546 pages
- Data Mining: Introductory and Advanced Topics
  - Margaret H. Dunham,
    Prentice Hall, 2003, ISBN 0-13-088892-3
    315 pages
- Dealing with the data flood: Mining data, text and multimedia
  - Edited by Jeroen Meij,
    SST Publications, 2002, ISBN 90-804496-6-0
    896 pages
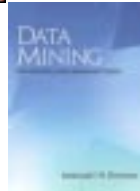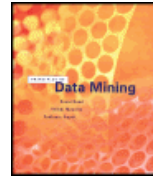
Course Web Page

Course Content, Slides, etc.

On-line Resources

- Course notes
- Course slides
- Web links
- Glossary
- Student submitted resources
- U-Chat
- Newsgroup
- Frequently asked questions

Presentation Schedule

	Presentation
🟩	Presentation
🟥	Review

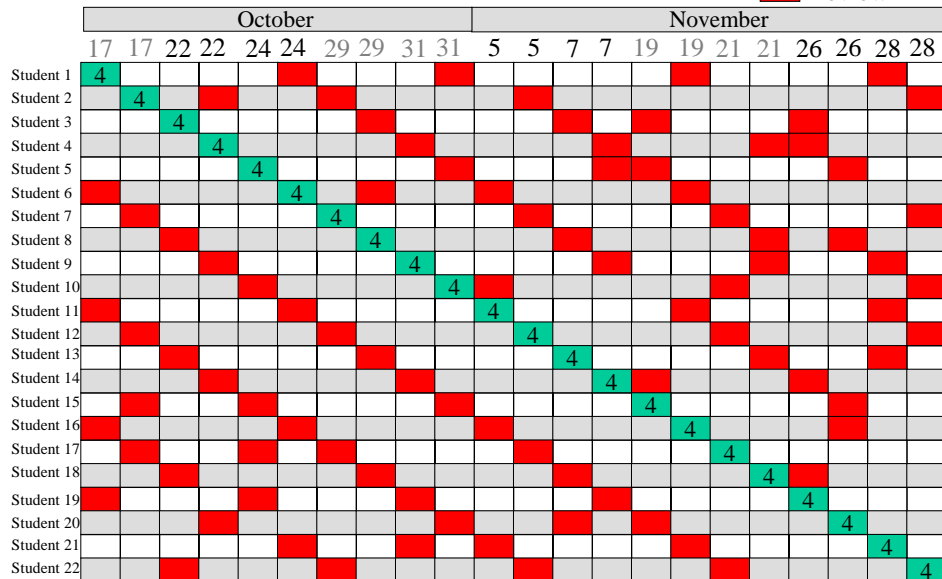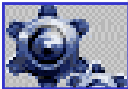|	October										November											
	17	17	22	22	24	24	29	29	31	31	5	5	7	7	19	19	21	21	26	26	28	28
Student 1	4																					
Student 2		4																				
Student 3			4																			
Student 4				4																		
Student 5					4																	
Student 6						4																
Student 7							4															
Student 8								4														
Student 9									4													
Student 10										4												
Student 11											4											
Student 12												4										
Student 13													4									
Student 14														4								
Student 15															4							
Student 16																4						
Student 17																	4					
Student 18																		4				
Student 19																			4			
Student 20																				4		
Student 21																					4	
Student 22																						4

List of Students

- <u>Student 1:</u> An, Zhibin — October 17
- <u>Student 2:</u> Atherton, Michael James — October 17
- <u>Student 3:</u> Cai, Zhipeng — October 22
- <u>Student 4:</u> Chen, Joyce Hui — October 22
- <u>Student 5:</u> Ding, Meng — October 24
- <u>Student 6:</u> Guo, Yuhong — October 24
- <u>Student 7:</u> Hou, Guiwen — October 29
- <u>Student 8:</u> Li, Wenxin — October 29
- <u>Student 9:</u> Malenfant, Rene Michael — October 31
- <u>Student 10:</u> Mocofan, Marian Leonid — October 31
- <u>Student 11:</u> Nulahmet, Mnawer — November 5
- <u>Student 12:</u> Pei, Yaling — November 5
- <u>Student 13:</u> Shi, Zhigang — November 7
- <u>Student 14:</u> Sun, Lisheng — November 7
- <u>Student 15:</u> Tu, Xin — November 19
- <u>Student 16:</u> Wang, Yang — November 19
- <u>Student 17:</u> Wu, Yaohua — November 21
- <u>Student 18:</u> Xing, Zhenchang — November 21
- <u>Student 19:</u> Yap, Peter Kai Yue — November 26
- <u>Student 20:</u> Zhang, Jingyue — November 26
- <u>Student 21:</u> Zhang, Qiongyun — November 28
- <u>Student 22:</u> Zou, Shoudong — November 28

Papers will be announced and assigned at a later date

Projects

	Choice	Deliverables
	Implement data mining project	Project proposal + 10' proposal presentation + project pre-demo + final demo + project report
	Write survey paper (*or research paper*)	Survey proposal + 10' proposal presentation + paper presentation + survey paper (20-30 pages)

Examples of survey topics:
- Web usage mining
- Knowledge discovery from unstructured or semi-structured data on the WWW
- Text mining
- Data mining from non-traditional databases (OODB/deductive DB).
- Spatial data mining
- Multimedia data mining
- Clustering
- Classification
- Association rule mining
- Datacube construction
- Datawarehousing

Examples of data mining projects will be posted on the course web site.

More About Projects

Either for the implementation project or the survey paper, students should write a project proposal (1 or 2 pages).

- project topic;
- implementation choices;
- approach;
- schedule.

All projects are demonstrated at the end of the semester. **December 17 and 19** to the whole class.
Preliminary project demos are private demos given to the instructor on **week November 18-22**.

Implementations: C/C^{++} or Java,
OS: Linux, Window NT/98 , or other systems.

Course Schedule

(Tentative, subject to changes)

There are 14 weeks from Sept. 5th to Dec. 4th.
First class starts September 10th and classes end November 28th.

	Tuesday	Thursday
Week 2:	Sept. 10: Introduction	Sept. 12: **C1**- Into DM
Week 3:	Sept. 17: **C2**- DW	Sept. 19: **C3**-**C4**- DM ops
Week 4:	Sept. 24: **C5**- Char. R.	Sept. 26: **C6**- Asso. Rules
Week 5:	Oct. 1: **C7**- Classific.	Oct. 3: **C8**- Clustering
Week 6:	Oct. 8: **C8**- Clustering	Oct. 10: **C9**- Web Mining
Week 7:	Oct. 15: C10- Spa+MM	Oct. 17: Papers 1 & 2
Week 8:	Oct. 22: Papers 3 & 4	Oct. 24: Papers 5 & 6
Week 9:	Oct. 29: Papers 7 & 8	Oct. 31: Papers 9 & 10
Week 10:	Nov. 5: Papers 11&12	Nov. 7: Papers 13 & 14
Week 11:	No class	No class
Week 12:	Nov. 19: Papers 15&16	Nov. 21: Papers 17&18
Week 13:	Nov. 26: Papers 19&20	Nov. 28: Papers 21&22
W 14-15:	No class	No class
Week 16:	Dec. 17: Final Demos	Dec. 19: Final Demos

Away (out of town)
To be confirmed
November 14th
December 3rd
Dec 2-7 : ICCE
Dec 9-12: ICDM

Due dates
-Midterm **week 8 or 9**
-Project proposals **week 5**
-Project preliminary demo
  week 12
- Project reports **week 16**
- Project final demo
  week 16

Course Content

- Introduction to Data Mining
- Data warehousing and OLAP
- Data cleaning
- Data mining operations
- Data summarization
- Association analysis
- Classification and prediction
- Clustering
- Web Mining
- Multimedia and Spatial Mining
- *Other topics if time permits*