

CSE 450: Web Mining Seminar

Fall 2005

Tue/Thur 1:10-2:25pm Maginnes 110

Instructor: **Dr. Brian D. Davison**

Dept. of Computer Science & Engineering

Lehigh University

davison@cse.lehigh.edu

<http://www.cse.lehigh.edu/~brian/course/webmining/>

Course Objectives

- To become proficient at reading technical papers
- To gain knowledge of important current web mining research
- To gain experience presenting technical material
- To learn to write critical reviews of research papers
- To explore a research project in some depth and write a technical paper summarizing that work

Web Mining Seminar

- No textbook – all papers
- No exams – paper critiques (20%), presentations (20%), participation (20%), and projects (40%)
- Prerequisite – either:
 - CSE345/445 WWW Search Engines
 - CSE430 Textual Data Mining.
- With special permission, a data mining course may be sufficient.

Seminars are less formal

- We have a small class
- Introduce yourselves!

Possible Topics to Cover

- web search
- web usage mining
- text mining
- information extraction
- link analysis
- recognizing search engine spam

Possible Paper Sources

- World Wide Web Conferences (WWW)
- Information Retrieval and Database Confs.
 - SIGIR, ECIR, CIKM, VLDB, SIGMOD, ICDE
- Data Mining Conferences
 - SIGKDD, ICDM, SDM, PKDD
- Other workshops, symposia
 - WebDB, WebKDD, AIRWeb, etc.
- Other conferences
 - ICML, ECML, UM, CHI, AAI, IJCAI, etc.
- Also journal articles

Initial Course Approach

- Except for today, each paper will have:
 - a student presentation
 - usually 20 minutes max, at most 8 slides
 - a student critic (devil's advocate)
 - 5 minutes to say why the paper should never have been published, or at least why it is not useful now
 - class discussion of topics in paper
 - determine consensus for overall paper score

Approach, cont.

- In most classes
 - we will plan to present and discuss one or two papers.
 - you will need to write one review/critique of the paper of your choice (if multiple papers are scheduled).
- The paper presentation and best review will be posted online.

Semester Research Project

- Individual, or groups of two (members will be asked to grade each other).
- Should be the beginning of what could be a publishable project.
 - On some aspect of web mining.
- Write a scientific paper at the end.
 - Publish as a technical report if not more.
- Present project at the end of the semester.

Why read scientific papers?

- Also, how to read scientific papers?

Two questions

- Why read scientific papers?
 - Stay or become current in a technical field
 - Get ideas to improve or refute papers
 - To explain or teach the concepts to others
- How to read scientific papers?
 - Skim to decide whether worthwhile
 - Determine credibility
 - Relate to your work
 - Read in detail
 - Challenge assumptions, arguments, methods, statistics, data
 - Take notes
 - Consider how to use approach in your work
 - Write summary

Why review papers?

Why review papers?

- Comes naturally from a critical reading
- To contribute to the peer-review process
- To learn about new work before published
- To have a voice in determining what gets published
- To get a better feel for the threshold for acceptance

How to write good papers?

How to write good papers?

- Read good and bad papers, and learn to distinguish them
- Provide strong motivation for work
 - Explain why exciting or important
- Demonstrate expertise
 - Connect work to foundational and recent papers
- Clearly present argument or experimental work
 - Provide sufficient detail to reproduce equivalent results
- Show significance of results
- Defuse potential criticisms
- Describe clearly the contributions of the paper
- *Tell 'em what you'll tell 'em; tell 'em; tell 'em what you told 'em*

Homework

- Read meta papers for today
 - Actually quite good
- Read (and study) overview paper for next class
- Propose a paper for us to read
 - Browse through a few conference proceedings from the past few years
 - Send me URL and bibliographic reference for your preferred paper
 - Due by email, anytime Monday