

Syllabus
CS433-CS533 Information Retrieval
Spring 2014

Instructor Information

Name : Weiyi Meng
Office : Q 08, Engineering Building
Telephone : 777-4311
Fax : 777-4729
Email: meng@binghamton.edu
Web: <http://www.cs.binghamton.edu/~meng/meng.html>

Course Description

Indexing and data structures for storing and searching the index. Boolean, statistical, inference nets and knowledge-based models. Thesaurus construction. Query expansion. Natural language and linguistic techniques. Evaluation. Distributed information retrieval. Information integration and fusion. Dissemination of information. Summaries, themes and reading tours. Hypertext. Internet tools. Intelligent agents. Digital libraries.

Learning Objectives

- Learn basic concepts, models, methods and principles of information retrieval.
- Learn measures and techniques to evaluate IR systems.
- Learn basic Web search techniques.
- Learn fundamental techniques to implement IR systems.

Prerequisite and Co-requisite

- Prerequisite: CS375 Design and Analysis of Algorithms

When and Where

- Time : 2:50pm – 4:15pm, Tuesday and Thursday
- Classroom : S2 145

Office Hours

- 4:30pm – 5:30pm, Tuesday, Thursday or by appointment

TA Information

- Name: Tianlin Li
- Office: P17 Engineering Building

- Office Hours: Noon – 1:30pm, Wednesday, Friday or by appointment
- Email: tli16@binghamton.edu

Textbook

- Introduction to Information Retrieval by C. Manning, P. Raghavan and H. Schutze, Cambridge University Press, 2008.
- The entire book is available at <http://nlp.stanford.edu/IR-book/> for free.

Reference Books

- Search Engines: Information Retrieval in Practice by Bruce Croft, Donald Metzler, and Trevor Strohman, Addison-Wesley, 2009.
- Information Retrieval: Implementing and Evaluating Search Engines by S. Buttcher, C. Clarke and G. Cormack, MIT Press, 2010.
- Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data by B. Liu, Springer, Second Edition, 2011.

Lecture Notes and Papers

- Lecture Notes (PowerPoint Slides) and selected relevant research papers will be posted in the course folder on blackboard.

Lecture Topics

The following chapters (excluding chapters 10, 14-17) in the text book will be covered:

Chapter

- | | |
|----|--|
| 01 | Boolean retrieval |
| 02 | The term vocabulary & postings lists |
| 03 | Dictionaries and tolerant retrieval |
| 04 | Index construction |
| 05 | Index compression |
| 06 | Scoring, term weighting & the vector space model |
| 07 | Computing scores in a complete search system |
| 08 | Evaluation in information retrieval |
| 09 | Relevance feedback & query expansion |
| 10 | XML retrieval |
| 11 | Probabilistic information retrieval |
| 12 | Language models for information retrieval |
| 13 | Text classification & Naive Bayes |
| 14 | Vector space classification |
| 15 | Support vector machines & machine learning on documents |
| 16 | Flat clustering |

17	Hierarchical clustering
18	Matrix decompositions & latent semantic indexing
19	Web search basics
20	Web crawling and indexes
21	Link analysis

Grading Policy

- Midterm exam 20%
- Final exam 20%
- Homework Assignments 25%
- Project 30%
- Class Participation 5%

Class participation includes attendance and participation of class discussions. Student attendance is required and will be checked regularly by the instructor. Missing each class will result in a penalty of 0.5 point unless compelling reason for missing the class can be presented in writing to the instructor. Class participation will also be graded by how actively a student participates in class discussions.

- Late penalty. Late homework and project assignments will be penalized at a rate of 5% per day until the hard deadline (no assignment will be accepted after the hard deadline).

Academic Honesty

Academic honesty and integrity are expected of every student. Dishonesty and cheating in all academic work related to this course, when discovered, will be severely punished. Please read the **Student Academic Honesty Code** at <http://www2.binghamton.edu/watson/advising/pdfs/honesty-policy.pdf>.

Students must do their assignments/projects/exams by themselves. For the project report, students must write it using their own languages. All referenced works (including ideas, algorithms, programs, tables, figures, open source tools, etc.) must be clearly cited within the main body of the report and their full citations must be listed at the end of the report. Students' own contributions (new ideas, algorithms, programs, etc.) must be clearly identified.

Classroom Etiquette

- Cell phone: Cell phones must be turned off or in vibrate alert mode.
- Computer: Laptop/notebook computers are not allowed during class.