IIST 433/533: Information Storage and Retrieval (3) Fall 2010

Dr. Ozlem Uzuner
Email: ouzuner@albany.edu
Phone: 518-442-4687
Course time: Thursdays 2:45 PM – 5:35 PM
Class location: LC011
Office hours: Wednesdays 2pm-5pm and Thursdays 10am-12pm, or by appointment.
Office location: 114A Draper Hall

Course Description

This course provides an introduction to current practices in information retrieval (IR). It is intended to prepare you to understand the underlying theories and algorithms of modern IR systems and to introduce the methodology for the design and evaluation of IR systems. Topics covered include fundamental key concepts in information storage and retrieval, document representation, query language/operation, matching mechanisms and formal retrieval models, output presentation, indexing and searching, user interfaces, and the evaluation of information retrieval system effectiveness. In addition, we will investigate the inner workings of retrieval systems and search engines. Familiarity with computers and some programming experience are highly desirable, but not necessary.

Expected Outcomes

Students who successfully complete IIST 433/533 will have gained:

- knowledge of the variety and functionality of IR systems, and of the structures and techniques implemented in such systems;
- understanding of theories and models of IR, and of the principles of IR system design;
- skills in the critical analysis and evaluation of the performance of IR systems, and in the select and use of systems that contribute effectively and efficiently to the satisfaction of information needs in specific contexts.

Textbooks:

Required: Christopher D. Manning, Prabhakar Raghavan and Hinrich Schutze, Introduction to Information Retrieval, Cambridge University Press. 2008. Available online.

Optional: Charles T. Meadow, Bert R. Boyce, and Donald H. Kraft: Text Information Retrieval Systems, Third Edition. Academic Press. 2007.

Lecture Topics and Reading Assignments *All reading assignments must be completed prior to the following week's lectures. The lecture and reading lists are subject to change.*

Manning/Raghavan/Schutze = MRS

1. September 2 –Topics: Introductions, housekeeping.
Reading: None

2. September 16 –Topics: Basic Concepts, Dictionaries, inverted files, postings
Reading: MRS Chapter 1, 2

3. September 23 –Topics: The dictionary and postings lists, Term Weighting
Reading: MRS Chapter 2, 6

4. September 30 –Topics: Term Weighting, Similarity, Ranking and the Vector Space model
Reading: MRS Chapter 6, 7

5. October 7 –Topics: Similarity, Ranking and the Vector Space model, Index Construction
Reading: MRS Chapter 7, 4

6. October 14 – Topics: Wild cards, stemming, and spelling
Reading: MRS Chapter 3

7. October 21 – Midterm

8. October 28 – Topics: Index Compression, Query Expansion
Reading: MRS Chapter 5, 9

9: November 4 – Topics: xml Retrieval, Probabilistic Information Retrieval
Reading: MRS Chapter 10, 11

10: November 11 – Topics: Evaluation
Reading: MRS Chapter 8

11. November 18 – Topics: Web Crawling, Links and anchor text
Reading: MRS Chapter 20, 21

12. December 2 – Presentation of final research projects.

Requirements

Readings

Students are expected to read the assigned materials before coming to the class.

Attendance/participation

Students are expected to attend all the class sessions and fully participate in the class activities.

Assignments

Homework assignments are given in the form of problem sets. Each problem set will include essay-type questions, questions designed to show understanding of specific concepts that may involve calculations, and hands-on exercises involving existing IR engines. Each student should complete each assignment independently and hand-in the work on time.

Points will be deducted for late assignments.

Mid-term Exam

We will have an in-class exam on the topics covered during the first part of the semester.

Final Project

The final course project will be an extensive research paper providing an exhaustive look at an area of IR. Oral presentation of research findings will be made to the class.

Grading

Tasks	**Percentage**
Assignments	30%
Midterm Exam	30%
Final Project	30%
Class Participation	10%

Scale

A	95-100
A-	90-94
B+	85-89
B	80-84
B-	75-79
C+	70-74
C	65-69
C-	60-64
D	50-59
E	0-49

Policies

Students will not be excused from any due date of assignments, projects or exam for any reason. Late assignments will receive a half letter grade reduction for each day late. A late final project or final exam will be penalized a full letter grade for each day late.

Plagiarism and cheating will result in a failing grade for the course, and will be referred to the Office of Judicial Affairs according to the policies set forth in the current University at Albany Undergraduate Bulletin or University at Albany Graduate Bulletin, whichever is appropriate to the student.

Reasonable accommodations will be provided for students with documented physical, sensory, systemic, cognitive, learning and psychiatric disabilities. If you believe you have a disability requiring accommodation in this class, please notify the Director of Disabled Student Services (Campus Center 137, 442-5490).