IS 634 Information Retrieval

Last updated May 05, 2014 (subject to change)

Faculty Instructor: Y. F. Brook Wu, Ph.D.
Office: GITC 5500
Office Hours: Monday 4:00pm – 5:40 pm
E-mail: wu at njit dot edu

Classroom: CKB206
Class Meets: Monday 6pm – 8:55pm
Class Site: please go to moodle.njit.edu and login with your UCID. You will find IS 634, if you are enrolled in this class.

Objective

This course seeks to provide theoretical foundation as well as hands-on experience in information retrieval systems. Students will first learn from analyzing the results of an experimental system to gain insights into issues in the retrieval system design. Students also will gain experience with design, implementation and evaluation of a web-based retrieval system. Students must have taken programming languages and databases before enrolling in the course.

Catalog Description

IS 634 - Information Retrieval (3 credits)
Prerequisites: IS 631 or CS 631; working knowledge of an object-oriented programming language. Modern information retrieval systems, such as web search engines, empower users to easily access information on the web. The course covers the concepts and principles of information retrieval systems design, including web crawling, automatic indexing, vector space modeling, retrieval algorithms, digital libraries, text mining, information extraction, and document warehousing. These techniques are essential for building web systems, text databases, document processing systems, and other advanced information management systems.

Materials Covered during the Semester

- Architecture of a search engine
- Crawling and processing web pages
- Automatic indexing and term's weighting methods
- Link analysis (e.g. page rank, hub and authority)
- Retrieval models (Boolean, Probabilistic, and Language models)
- Search interfaces
- Search evaluation: system-oriented and user-oriented
- Text mining: document classification and clustering
- Social search: personalized search and recommender systems

Course Learning Goals (students are expected to learn the following):

- Architecture of a search engine
- Relationship between automatic indexing (term, frequency, and location based) and retrieval results
- Retrieval models and ranking of search results
- Search evaluation: both user and system oriented

Textbook

Search Engines: Information Retrieval in Practice
By Croft, Metzler, and Strohman.
Publisher: Addision-Wesley
ISBN-13: 978—0-13-607224-9

NJIT Honor Code (http://www.njit.edu/academics/integrity.php, http://www.njit.edu/academics/pdf/academic-integrity-code.pdf) is strictly enforced.

Grading Scheme

1. Participation: attendance and in-class activity 10%
2. Assignment: 10 %
3. Two Retrieval Experiments: 20%
4. Semester Project: 30%
    - Proposal
    - Project and presentation
5. Two Exams: 30%

Total 100%

Schedule (subject to change)

Weeks	Topics	Materials
Sept 08	Welcome and course logistics IR past, present, and future	
Sept 15	Search engine and information retrieval Architecture of search engines Assignment out	Ch 1, 2
Sept 22	Crawls and feeds	Ch 3
Sept 29	Processing text	Ch 4
Oct 06	Processing text (cont)	Ch 4
Oct 13	Retrieval Experiment 1 out **Midterm Exam**	

Oct 20	Ranking with Indexes	Ch 5
Oct 27	Class Activity	
Nov 03	Ranking with Indexes (cont.) Discussions on crawler and indexing designs	
Nov 10	Queries and interfaces Instructions on Retrieval Exp 2	Ch 6
Nov 17	Queries and interfaces (cont) Retrieval models	Ch 7
Nov 24	Retrieval models (cont) Evaluating search engines	Ch 7, 8
Dec 01	Social search	Ch 10
Dec 08	**Final Project Presentations** **Discussions on Retrieval Exp 2**	
Dec 13-19	**Final Exam Period (specific date/time/room TBD)**	