

IS 634 Information Retrieval

Last updated August 23, 2016 (subject to change)

Faculty Instructor: Y. F. Brook Wu, Ph.D.
Office: GITC 5105
Office Hours: Thursday 1:00pm – 2:30 pm
E-mail: wu@njit.edu

Teaching Assistant: Mr. Yang Liu
Office: GITC 5601
Office Hours: Tue and Thu 1-2:30pm
E-mail: yl558@njit.edu

Classroom: CKB317

Class Meets: Thursday 10-1pm

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.

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

- 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

Textbook

Search Engines: Information Retrieval in Practice

(<http://www.search-engines-book.com/>)

By Croft, Metzler, and Strohman.

Publisher: Addison-Wesley

ISBN-13: 978—0-13-607224-9

NJIT Code of Student Conduct

(<https://www5.njit.edu/doss/policies/conductcode/article5.php>) **is strictly enforced.**

Grading Scheme

1. Participation: attendance and in-class activity 10%
 2. Assignment 1: evaluating search engines 10 %
 3. Programming Assignments 1. Crawling (10%) and 2. Indexing (10%)
 4. Semester Project 25%
 - Proposal on implementing a new or customizing an open source search engine
 - Project and Presentation
 5. Two Exams: midterm (15%) and final (20%)
- 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: “evaluating search engines” out	Ch 1, 2
Sept 22	Crawls and feeds First programming assignment out	Ch 3
Sept 29	Processing text	Ch 4
Oct 06	Processing text (cont)	Ch 4
Oct 13	Midterm Exam (in-class, close-book) Demo of Programming Assignment 1	

Oct 20	Discussions on midterm exam Ranking with Indexes	Ch 5
Oct 27	Class Activity: Design of an IR architecture Programming Assignment 2 out	
Nov 03	Ranking with Indexes (cont.) Discussions on the design of IR Architecture	
Nov 10	Queries and interfaces Demo of Programming Assignment 2	Ch 6
Nov 17	Queries and interfaces (cont) Retrieval models	Ch 7
Nov 22 (Tuesday)	Retrieval models (cont) Evaluating search engines (Different Schedule because of Thanksgiving Holiday: http://www.njit.edu/registrar/calendars/2016Fall.php)	Ch 7, 8
Dec 01	Social search	Ch 10
Dec 08	Final Project Presentations Final Exam Out (Take Home and Open Book, due Dec 10)	