IS 634 Information Retrieval  
Last updated August 23, 2017 (subject to change)

Faculty Instructor: Y. F. Brook Wu, Ph.D.  
Office: GITC 5105  
Office Hours: Thursday 11:00am – 12:30 pm  
E-mail: wu@njit.edu

Teaching Assistant: Mr. Yang Liu  
Office: GITC 5601  
Office Hours: Tue and Thu 11am-12:30pm  
E-mail: yl558@njit.edu

Classroom: CKB223  
Class Meets: Thursday 1-4pm  
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 601. 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

NJIT Code of Student Conduct
(https://www5.njit.edu/doss/policies/conductcode/article5.php) is strictly enforced.

Special Accomodation
If you need accommodations due to a disability please contact Chantonette Lyles, Associate Director of Disability Support Services, Fenster Hall Room 260 to discuss your specific needs. A Letter of Accommodation Eligibility from the Disability Support Services office authorizing your accommodations will be required.

Grading Scheme

1. Participation: attendance and in-class activities (12%)
2. Assignment: evaluating search engines (8 %)
3. Programming Assignments: 1. Crawling (10%) and 2. Indexing (10%)
4. Semester Project (25%)
   o Proposal on implementing a new or customizing an open source search engine
   o Project and Presentation
5. Two Exams: midterm (15%) and final (20%)
Total 100%

Schedule
(last updated August 23)

<table>
<thead>
<tr>
<th>Weeks</th>
<th>Topics</th>
<th>Materials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sept 07</td>
<td>Welcome and course logistics; IR past, present, and future</td>
<td></td>
</tr>
<tr>
<td>Sept 14</td>
<td>Search engine and information retrieval Assignment: “evaluating search engines” out</td>
<td>Ch 1</td>
</tr>
<tr>
<td>Sept 21</td>
<td>Architecture of search engines; Crawls and feeds; Programming Assignment 1 out</td>
<td>Ch 2, 3</td>
</tr>
<tr>
<td>Sept 28</td>
<td>Crawls and feeds (cont)</td>
<td>Ch 3, 4</td>
</tr>
<tr>
<td>Date</td>
<td>Event</td>
<td>Chapter(s)</td>
</tr>
<tr>
<td>------------</td>
<td>-----------------------------------------------------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Oct 05</td>
<td>Processing text (cont)</td>
<td>Ch 4</td>
</tr>
<tr>
<td></td>
<td>Demo of Programming Assignment 1</td>
<td></td>
</tr>
<tr>
<td>Oct 12</td>
<td>Ranking with Indexes</td>
<td>Ch 4, 5</td>
</tr>
<tr>
<td></td>
<td>Programming Assignment 2 out</td>
<td></td>
</tr>
<tr>
<td>Oct 19</td>
<td>Midterm review</td>
<td>Ch 5</td>
</tr>
<tr>
<td></td>
<td>Class Activity: Design of an IR architecture</td>
<td></td>
</tr>
<tr>
<td>Oct 26</td>
<td><strong>Midterm Exam (in-class, close-book)</strong></td>
<td></td>
</tr>
<tr>
<td>Nov 02</td>
<td>Discussions on midterm exam</td>
<td>Ch 6</td>
</tr>
<tr>
<td></td>
<td>Discussions on the design of IR Architecture</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Demo of Programming Assignment #2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Queries and interfaces</td>
<td></td>
</tr>
<tr>
<td>Nov 09</td>
<td>Queries and interfaces</td>
<td>Ch 6</td>
</tr>
<tr>
<td>Nov 16</td>
<td>Queries and interfaces (cont)</td>
<td>Ch 6, 7</td>
</tr>
<tr>
<td></td>
<td>Retrieval models</td>
<td></td>
</tr>
<tr>
<td>Nov 21</td>
<td>Retrieval models (cont)</td>
<td>Ch 7, 8</td>
</tr>
<tr>
<td>(Tuesday)</td>
<td>Evaluating search engines</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>(Different Schedule because of Thanksgiving Holiday: <a href="http://www.njit.edu/registrar/fall-2017-academic-calendar/">link</a>)</em></td>
<td></td>
</tr>
<tr>
<td>Nov 30</td>
<td>Evaluating search engines (cont)</td>
<td>Ch 8</td>
</tr>
<tr>
<td>Dec 07</td>
<td>Social search</td>
<td>Ch 10</td>
</tr>
<tr>
<td></td>
<td><strong>Final Project Presentations</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Final Exam Out (Take Home and Open Book, due Dec 10)</strong></td>
<td></td>
</tr>
</tbody>
</table>