This course homepage is accessible from http://www.cs.ust.hk/~dlee/4321/

COMP 4321 Search Engines for Web and Enterprise Data [3-0-1:3]

Fall 2014

Course and Instructor/TA Information

Instructor:	Prof. Dik Lun Lee
Email:	dlee@cse.ust.hk
Office:	3534 (Lift 25/26)
Office Hours:	Tue/Thu 2:00pm-2:45pm * Emails are the best way to get quick response from me. If meeting is needed, try to make an appointment with email.
Lectures:	Tue/Thu 3PM - 4:20PM
Lecture Room:	Rm 6602, Lift 31-32
TA:	**Pengfei Zhao (ericzhao@cse.ust.hk)**
	Xun Zhao (xzhaoag@ust.hk)
Lab LA1:	Monday 12 noon - 12:50 pm, Rm 4214 (Lift 19)
Lab LA2:	Monday 6 pm - 6:50 pm, Rm 4214 (Lift 19)

TA/Lab Homepage: Labs and Lab Schedule, Homework, Exams, Scores, Term Project **[University Calendar]**

Course Outcome

On successful completion of this course, students are expected to be able to:
(1) Design and implement a complete and functional search engine.
(2) Test and evaluate the effectiveness of a search engine.
(3) Identify the limitations of search engine technologies and develop solutions to meet application requirements.

Course Outline

1. Introduction and course overview
2. Business models
3. Information retrieval models and Inverted Files
4. Web-based information retrieval
5. Pattern matching and extended Boolean model
6. Retrieval effectiveness, benchmarking
7. Document preprocessing
8. Query expansion and relevance feedback
9. Document clustering
10. Signature files

Detailed Course Topics [Use your student ID as both username and password]

Text and Reference Materials

Course Description

Text retrieval models, vector space model, document ranking, performance evaluation; indexing, pattern matching, relevance feedback, clustering; web search engines, authority-based ranking; enterprise data management, content creation, metadata, taxonomy, ontology; semantic web, digital libraries and knowledge management applications.

Course Objective

After completing the course, students will have acquired:

1. Core techniques for building search engines
2. Technologies and business models employed in modern web-based search engines
3. Hands-on experience in building a complete web-based search engine including spider, data storage and search modules
4. Knowledge in the future trends and applications of information retrieval Web and Enterprise applications and digital libraries.

Pre-requisites/Background needed:

COMP 151/151H (prior to 2009-10) or COMP 171/171H (prior to 2009-10) or COMP 2012/2012H

Policy on Academic Misconduct

Homework/lab assignments must be done individually. Collaboration between students is strictly forbidden. Any violation will be passed to the Department's Undergraduate/Postgraduate Studies Committee for assessment. The result may lead to dismissal from the University.
Term project must be done by the individual group. No sharing of code and copying of code from previous projects are allowed.