

**City University of Hong Kong**

**Information on a Course  
offered by Department of Computer Science  
with effect from Semester B in 2014 / 2015**

---

---

**Part I**

**Course Title:** Algorithms and Techniques for Web Searching

**Course Code:** CS5286

**Course Duration:** One Semester

**Credit Units:** 3

**Level:** P5

**Medium of Instruction:** English

**Prerequisites:** *(Course Code and Title)*

Nil

**Precursors:** *(Course Code and Title)*

Nil

**Equivalent Courses:** *(Course Code and Title)*

Nil

**Exclusive Courses:** *(Course Code and Title)*

Nil

**Part II**

**Course Aims**

This course teaches principles, tools and techniques for Internet information retrieval. Students will be able to develop automated access software to webpages, to analyze their link structures, to index them according to their contents, to rank them with respect to specific queries, to apply information retrieval tools, computational linguistics and hyperlink graph structural understandings to improve the searching results. Students will also be given chances to create their own innovative ideas for web searching in the group project.

## Course Intended Learning Outcomes (CILOs)

*Upon successful completion of this course, students should be able to:*

No.	CILOs	Weighting (if applicable)
1.	create an automated software agent for web searching;	20%
2.	analyze webpage through hyperlink structures;	20%
3.	index webpages by different document models for different purposes, and remember those models;	20%
4.	explain answering queries in decreasing orders of relevance, and evaluate the relevance;	20%
5.	creatively apply web search tools.	20%

## Teaching and Learning Activities (TLAs)

*(Indicative of likely activities and tasks designed to facilitate students' achievement of the CILOs. Final details will be provided to students in their first week of attendance in this course)*

Teaching pattern:

*Suggested lecture/tutorial/laboratory mix: 2 hrs. lecture; 1 hr. tutorial.*

This course will focus on introducing the fundamental and state-of-the-art techniques in Internet search, with a focus on text file search.

The topics to be covered will be three major components: 1.) hyperlink structures and their analysis, 2.) information retrieval and content analysis, 3.) Internet user behavior analysis.

CILO No.	TLAs	Hours/week (if applicable)
CILO 1, 2, 3, 4	<b>In-class exercise</b> – Students are asked to practice what they learned in lectures to solve some related problems. Solutions by volunteers or selected students will be discussed in class. It supports ILOs #1, 2, 3, 4.	
CILO 2,3,4	<b>Written Quiz</b> – This assignment gives students an opportunity to demonstrate their understanding of the course materials, and also serves a check point in the middle of term for the progress. It supports ILOs #2, 3, 4.	
CILO 1,3	<b>Programming Quiz</b> – The quiz reinforces students' learning in the tutorial sessions of the programming part. It supports Course ILOs #1, 3.	
CILO 5	<b>Project</b> – Two different types of projects are designed that allow students to choose from: either	

	an essay based one, or a programming based one. Software-based assignment gives students an opportunity to create innovative applications for web searching. The assignment will be documented in a report. The essay writing project will be about a study on a specific aspect of web search. This activity supports Course ILO #5.	
CILO 1,2,3,4	<b>Examination</b> – Students will be tested on their overall understanding of the topics covered in CILO 1, 2, 3, 4.	

### Assessment Tasks/Activities

(Indicative of likely activities and tasks designed to assess how well the students achieve the CILOs. Final details will be provided to students in their first week of attendance in this course)

CILO No.	Type of Assessment Tasks/Activities	Weighting (if applicable)	Remarks
CILO 1	Create an automated software agent for web searching. It is be evaluated in in-class exercise, programming quiz and examination.		
CILO 2	Analyze webpage through hyperlink structures. It is be evaluated in in-class exercise, written quiz and examination.		
CILO 3	Index webpages by different document models for different purposes, and remember those models. It is be evaluated in in-class exercise, programming quiz, written quiz and examination.		
CILO 4	explain answering queries in decreasing orders of relevance, and evaluate the relevance; It is be evaluated in in-class exercise, written quiz and examination.		
CILO 5	Creatively apply web search tools. It is be evaluated in project.		

**Grading of Student Achievement:** Refer to Grading of Courses in the Academic Regulations for Taught Postgraduate Degrees.

*Examination duration: 2 hours*

*Percentage of coursework, examination, etc.: 50% CW (20% for quizzes and 30% for project); 50% Exam*

*Grading pattern:* Standard (A+AA-...F)

For a student to pass the course, at least 30% of the maximum mark for the examination must be obtained.

A formative assessment will be made on the students' ability to apply tools and knowledge to different situation. The equal weighting of coursework and examination assessment is due to the emphasis on both the practicality and the theory of market design models.

## **Part III**

### **Keyword Syllabus**

Search engines; Search techniques, Web crawlers; Web Structure Mining; Page rank; micro communities; Web content mining; inverted file index; vector space model; Web pages categorization, Web Usage Mining, user tracking and profiling, Web personalization; Privacy issues; information delivery, information filtering and anti-spam, web reputation and anti-phishing.

### **Syllabus**

1. Build an automated software agent for web searching: Search engines, Google, Yahoo, Baidu; web crawlers.
2. Analyze webpage through hyperlink structures: Web structure mining, page rank, micro communities.
3. Index webpages by different document models for different purposes: web content mining, lexical analysis, inverted indexing, Boolean model, permutation indexing, soundex indexing.
4. Answer queries in decreasing orders of relevance: vector space model, similarity, clustering, user tracking and profiling, rank aggregation, Web reputation.
5. Creatively apply web search tools: information delivery, anti-spam, anti-phishing.

**Recommended Reading  
Text(s)**

*Essential Text*

*Lecture notes of Lecturer  
and*

*C.D. Manning, P. Raghavan, H. Schutze, "Introduction to Information Retrieval",  
Cambridge University Press, 2008, New York*

*Supplementary Reading*

*In addition, a collection of readings from journal literature and conference  
proceedings will be used*

**Online Resources**