**City University of Hong Kong**
**Course Syllabus**

**offered by Department of Computer Science**
**with effect from Semester A 2017/18**

---

## Part I Course Overview

| | |
|---|---|
| **Course Title:** | Algorithms and Techniques for Web Searching |
| **Course Code**: | CS5286 |
| **Course Duration:** | One semester |
| **Credit Units**: | 3 credits |
| **Level**: | P5 |
| **Medium of Instruction:** | English |
| **Medium of Assessment:** | 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 Details

### 1. Abstract

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.

### 2. Course Intended Learning Outcomes (CILOs)

*(CILOs state what the student is expected to be able to do at the end of the course according to a given standard of performance.)*

| No. | CILOs | Weighting (if applicable) | Discovery-enriched curriculum related learning outcomes (please tick where appropriate) | | |
|---|---|---|---|---|---|
| | | | *A1* | *A2* | *A3* |
| 1. | Create an automated software agent for web searching. | | | ✓ | |
| 2. | Analyze webpage through hyperlink structures. | | ✓ | | |
| 3. | Index webpages by different document models for different purposes, and remember those models. | | ✓ | | |
| 4. | Explain answering queries in decreasing orders of relevance, and evaluate the relevance. | | ✓ | | |
| 5. | Creatively apply web search tools. | | | | ✓ |
| | | | | | |

*A1: Attitude*
*Develop an attitude of discovery/innovation/creativity, as demonstrated by students possessing a strong sense of curiosity, asking questions actively, challenging assumptions or engaging in inquiry together with teachers.*

*A2: Ability*
*Develop the ability/skill needed to discover/innovate/create, as demonstrated by students possessing critical thinking skills to assess ideas, acquiring research skills, synthesizing knowledge across disciplines or applying academic knowledge to self-life problems.*

A3: *Accomplishments*
*Demonstrate accomplishment of discovery/innovation/creativity through producing /constructing creative works/new artefacts, effective solutions to real-life problems or new processes.*

## 3. Teaching and Learning Activities (TLAs)

*(TLAs designed to facilitate students' achievement of the CILOs.)*

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

| TLA | Brief Description | CILO No. | | | | | Hours/week (if applicable) |
|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | |
| In-class exercise | 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. | | ✓ | ✓ | ✓ | | |
| Programming assignment | The students are asked to program different search engines based on algorithms, techniques and engineering tricks learnt in the course. The assignments give students an opportunity to demonstrate the understanding of course materials, while exploiting popular APIs for web searching. | ✓ | | | | | |
| Project | Two different types of projects are designed that allow students to choose from: either an essay based one, or a programming based one.<br>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. | | | | | ✓ | |
| Examination | Students will be tested on their overall understanding of the topics covered in CILO 1, 2, 3, 4. | ✓ | ✓ | ✓ | ✓ | | |

## 4. Assessment Tasks/Activities (ATs)

*(ATs are designed to assess how well the students achieve the CILOs.)*

| Assessment Tasks/Activities | CILO No. | | | | | Weighting | Remarks |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | | |
| Continuous Assessment: 50% | | | | | | | |
| Assignment | ✓ | ✓ | ✓ | | | 30% | |
| Project | | | ✓ | ✓ | ✓ | 20% | Can be group project of 2-3 members |
| Examination^: 50% (duration: 2 hours) | | | | | | | |
| | | | | | | 100% | |

^ 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.

## 5. Assessment Rubrics

*(Grading of student achievements is based on student performance in assessment tasks/activities with the following rubrics.)*

| Assessment Task | Criterion | Excellent (A+, A, A-) | Good (B+, B, B-) | Fair (C+, C, C-) | Marginal (D) | Failure (F) |
|---|---|---|---|---|---|---|
| 1. Assignment | The ability to implement and assess the effectiveness of different algorithms and techniques. | High | Significant | Moderate | Basic | Not even reaching marginal level |
| 2. Project | The ability and creativity in applying appropriate algorithms and techniques for real-world applications. | High | Significant | Moderate | Basic | Not even reaching marginal level |
| 3. Examination | The extent to which the students can understand the algorithms and techniques, apply them appropriately for different applications, and evaluate their performances. | High | Significant | Moderate | Basic | Not even reaching marginal level |

**Part III Other Information** (more details can be provided separately in the teaching plan)

## 1. Keyword Syllabus

*(An indication of the key topics of the course.)*

Search engines; web crawlers; web structure mining; page rank; micro communities; web content mining; inverted file index; vector space model; web advertisement; relevance feedback; near-duplicate search; probabilistic information retrieval; language models; topic modelling; multimedia search; deep learning for information retrieval; performance evaluation.

Syllabus

1. Build an automated software agent for web searching: data preprocessing, indexing, search algorithms and techniques.

2. Analyze webpage through hyperlink structures: Web structure mining, page rank, micro communities.

3. Index webpages by different document models for different purposes: inverted indexing, Boolean model, vector space mode, near-duplicate search, probabilistic information retrieval, topic modelling, deep learning..

4. Answer queries in decreasing orders of relevance: similarity, query expansion, relevance feedback, performance evaluation, user behaviour analysis.

5. Creatively apply web search tools: Query log mining, distributed information retrieval, document clustering and classification..

## 2. Reading List

### 2.1 Compulsory Readings

*(Compulsory readings can include books, book chapters, or journal/magazine articles. There are also collections of e-books, e-journals available from the CityU Library.)*

| | |
|---|---|
| 1. | C.D. Manning, P. Raghavan, H. Schutze, "Introduction to Information Retrieval", Cambridge University Press, 2008, New York. |

### 2.2 Additional Readings

*(Additional references for students to learn to expand their knowledge about the subject.)*

| | |
|---|---|
| 1 | Ricardo Baeza-Yates and Berthier Ribeiro-Neto, "Modern Information Retrieval", Addison-Wesley, 1999. |
| 2. | A collection of readings from journal literature and conference proceedings will be used |