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:	Information Retrieval
Course Code:	CS4485
Course Duration:	One semester
Credit Units:	3 credits
Level:	B4
Proposed Area: *(for GE courses only)*	☐ Arts and Humanities ☐ Study of Societies, Social and Business Organisations ☐ Science and Technology
Medium of Instruction:	English
Medium of Assessment:	English
Prerequisites: *(Course Code and Title)*	CS3402 Database Systems or CS3462 Introduction to Database Systems
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

(A 150-word description about the course)

This course aims to provide a broad view and detailed knowledge of all key topics in modern information retrieval (IR). Basic concepts such as retrieval evaluation, query languages, query operations, indexing and searching are introduced. Some advanced topics including parallel and distributed IR, and multimedia IR are discussed.

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.	Describe the basic concepts and models in information retrieval.		✓		
2.	Apply basic techniques in user relevance feedback, indexing and searching.		✓		
3.	Design and build a simple text retrieval systems. Such a system includes (1) storage of documents (converting documents into vectors); (2) ranking algorithms and relevance feedback algorithms.				
4.	Read research papers in this area.		✓		
** If weighting is assigned to CILOs, they should add up to 100%.*		100%			

[#] Please specify the alignment of CILOs to the Gateway Education Programme Intended Learning outcomes (PILOs) in Section A of Annex.

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

TLA	Brief Description	CILO No.				Hours/week (if applicable)
		1	2	3	4	
Lecture	Lectures will be given to introduce the basic concepts, information retrieval models, basic techniques in user relevance feedback, indexing and searching; (2) some exercises (using concept examples) will be given in the tutorial to illustrate the concepts, methods and models. For the first few weeks in the semester, small assignments will be given.	✓	✓	✓	✓	
Tutorial	Exercises will be given during tutorials.	✓	✓			
Assignment	Small assignments will be given in the first few weeks of the semester.	✓	✓			
Midterm test	Midterm will be given to test the basic concepts, models, basic techniques in user relevance feedback, indexing and searching.					
Mini-project	The students will do a min-project (50% of coursework) to build a small system. The system should contain storage of documents (converting documents into vectors); ranking algorithms and (perhaps) relevance feedback algorithms. Students are encouraged to select their own methods. A working system must be delivered in order to pass. At the end of the semester, students should demonstrate the system (40% for the project) to the lecturer and hand in a report (60% for the project) describing the system that they built. The report should clearly describe the methods used for each step. More marks will be given for selecting advanced methods. If the students can propose their own methods, extra marks should be given.			✓	✓	

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		
Continuous Assessment: 50%						
Assignment	✓	✓			10%	
Midterm test	✓	✓			10%	
Mini-project			✓	✓	30%	
Final Examination^: 50% (duration: 2 hours)						
** The weightings should add up to 100%.*					100%	

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

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. Assignments	Question has a score	High	Significant	Moderate	Basic	Not even reaching marginal levels
2. Mini-project	Ability to complete the system, creating new ideas/methods	High	Significant	Moderate	Basic	Not even reaching marginal levels
3. Midterm	Question has a score	High	Significant	Moderate	Basic	Not even reaching marginal levels
4. Final exam	Question has a score	High	Significant	Moderate	Basic	Not even reaching marginal levels

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.)

Retrieval evaluation. Query language. Query operations: user relevance feedback, automatic local analysis, automatic global analysis. Indexing and searching: indexing method, searching method. Parallel and distributed IR: MIMD and SIMD architectures, collection partitioning, source selection. Multimedia IR: models, query language. Searching the web, libraries and bibliographical systems. Digital libraries.

Syllabus

1. Retrieval evaluation
   Recall, precision, and alternative measures for the quality of queries.

2. Query languages
   Keyword-Based querying. Pattern matching. Structure queries and query protocols.

3. Query operations
   User relevance feedback including query expansion and term rewriting for vector models. Probabilistic models and evaluation of relevance feedback strategies. Automatic local analysis: query expansion through local clustering and local context analysis. Automatic global analysis: based on similarity thesaurus and statistical thesaurus.

4. Indexing and searching
   Indexing methods include inverted files, suffix trees, and signature files. Searching: exact matching methods including KMP, Boyer-Moore algorithm, etc. Approximate pattern matching including string matching allowing errors, regular expressions and extended patterns.

5. Parallel and Distributed IR
   Parallel IR: MIMD and SIMD architectures. Distributed IR: Collection partitioning, source selection, query processing and web issues.

6. Multimedia IR
   Models: Multimedia data support in commercial DBMSs, MULTOS data model. Query languages: request specification, conditions on multimedia data, uncertainty, proximity and weights in query expressions. Indexing and searching: spatial access methods, a generic multimedia indexing approach, one-dimensional time series, two-dimensional color images, and automatic feature extraction.

7. Others
   Searching the web. Libraries and bibliographical systems. Digital libraries.
   (This part is optional.)

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.)

2.2 Additional Readings

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