CS4485: INFORMATION RETRIEVAL

Effective Term
Semester A 2022/23

Part I Course Overview

Course Title
Information Retrieval

Subject Code
CS - Computer Science
Course Number
4485

Academic Unit
Computer Science (CS)

College/School
College of Engineering (EG)

Course Duration
One Semester

Credit Units
3

Level
B1, B2, B3, B4 - Bachelor's Degree
R8 - Research Degree

Medium of Instruction
English

Medium of Assessment
English

Prerequisites
CS3402 Database Systems

Precursors
Nil

Equivalent Courses
Nil

Exclusive Courses
Nil

Part II Course Details

Abstract

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.

Course Intended Learning Outcomes (CILOs)

	CILOs	Weighting (if app.)	DEC-A1	DEC-A2	DEC-A3
1	Describe the basic concepts and models in information retrieval.		x		
2	Apply basic techniques in user relevance feedback, indexing and searching.		x		
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.		x		

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 real-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.

Teaching and Learning Activities (TLAs)

	TLAs	Brief Description	CILO No.	Hours/week (if applicable)
1	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.	1, 2, 3, 4	3 hours/week

2	Tutorial	Exercises will be given during tutorials.	1, 2	8 hours/semester
3	Assignment	Small assignments will be given in the first few weeks of the semester.	1, 2	
4	Midterm test	Midterm will be given to test the basic concepts, models, basic techniques in user relevance feedback, indexing and searching.		
5	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.	3, 4	

Assessment Tasks / Activities (ATs)

	ATs	CILO No.	Weighting (%)	Remarks (e.g. Parameter for GenAI use)
1	Assignment	1, 2	10	
2	Midterm test	1, 2	10	
3	Mini-project	3, 4	30	

Continuous Assessment (%)

50

Examination (%)

50

Examination Duration (Hours)

2

Additional Information for ATs

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

Assessment Rubrics (AR)

Assessment Task

Assignments

Criterion

Question has a score

Excellent (A+, A, A-)

High

Good (B+, B, B-)

Significant

Fair (C+, C, C-)

Moderate

Marginal (D)

Basic

Failure (F)

Not even reaching marginal levels

Assessment Task

Mini-project

Criterion

Ability to complete the system, creating new ideas/methods

Excellent (A+, A, A-)

High

Good (B+, B, B-)

Significant

Fair (C+, C, C-)

Moderate

Marginal (D)

Basic

Failure (F)

Not even reaching marginal levels

Assessment Task

Midterm

Criterion

Question has a score

Excellent (A+, A, A-)

High

Good (B+, B, B-)

Significant

Fair (C+, C, C-)

Moderate

Marginal (D)

Basic

Failure (F)

Not even reaching marginal levels

Assessment Task

Final exam

Criterion

Question has a score

Excellent (A+, A, A-)

High

Good (B+, B, B-)

Significant

Fair (C+, C, C-)

Moderate

Marginal (D)

Basic

Failure (F)

Not even reaching marginal levels

Part III Other Information

Keyword Syllabus

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

·   Retrieval evaluation
    Recall, precision, and alternative measures for the quality of queries.
·   Query language
    Keyword-Based querying. Pattern matching. Structure queries and query protocols.
·   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.
·   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.
·   Parallel and Distributed IR
    Parallel IR: MIMD and SIMD architectures. Distributed IR: Collection partitioning, source selection, query processing and web issues.
·   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.
·   Others
    Searching the web. Libraries and bibliographical systems. Digital libraries
    (This part is optional.)

Reading List

Compulsory Readings

	Title
1	Nil

Additional Readings

	Title
1	Nil