Module (Course Syllabus) Catalogue

2022-2023

College/ Institute	Erbil Technical Engineering College	
Department	Information System Engineering	
Module Name	Information Retrieval	
Module Code	IR802	
Degree	Technical Diploma ☐ Bachelor ■ High Diploma ☐ Master ☐ PhD ☐	
Semester	8	
Qualification	PhD In Computer Engineering	
Scientific Title	Lecturer	
ECTS (Credits)	6	
Module type	Prerequisite ☐ Core ■ Assist. ☐	
Weekly hours		
Weekly hours (Theory)	(2)hr Class	(84)Total hrs Workload
Weekly hours (Practical)	(2)hr Class	(78)Total hrs Workload
Number of Weeks	14	
Lecturer (Theory)	Shahab Wahhab Kareem	
E-Mail & Mobile NO.	Shahab.kareem@epu.edu.iq	
Lecturer (Practical)	Mina Farooq	
E-Mail & Mobile NO.		
Websites		

Course Book

Course Description	Search engine systems, such as Google and Bing, are now an essential tool of everyone's life to deal with explosively growing online information (e.g., web pages, tweets, video, images, news articles, forum discussions, and scientific literature). In this course, you will learn the underlying technologies of modern information retrieval system, and obtain hands-on experience by using existing information retrieval toolkits to set up your own search engines and improving their search accuracy. This is a undergraduate-level introductory course for information retrieval. It will cover algorithms, design, and implementation of modern information retrieval systems. Topics include: retrieval system design and implementation, text analysis techniques, retrieval models (e.g., Boolean, vector space, probabilistic, and learning-based methods), search evaluation, retrieval feedback, search log mining, and applications in web information management.			
Course objectives	• Understand the theoretical basis behind the standard models of IR (Boolean, Vector-space, Probabilistic and Logical models), • Understand the difficulty of representing and retrieving documents, images, speech, etc., • Be able to implement, run and test a standard IR system, • Understand the standard methods for Web indexing and retrieval, • understand how techniques from natural language processing, artificial intelligence, human-computer interaction and visualization integrate with IR, and • Be familiar with various algorithms and systems			
Student's obligation	Homework assignments will be a mix of paperwork and machine problems. Written homework should be finished individually, discussions with peers or instructor is allowed, but copying or any other type of cheating is strictly prohibited. You will be given one week to finish the written homework. Some of the machine problems are designed for teamwork and due day may vary. Any late submission will incur a 15% penalty for that assignment.			
Required Learning Materials				
Evaluation	**Task**	**Weight (Marks)**	**Due Week**	**Relevant Learning Outcome**

	Paper Review				
	Assignments	Homework	5	5	
		Class Activity	2	7	
		Report	5	8	
		Seminar			
		Essay			
		Project	5	11	
	Quiz		8	4	
	Lab.		10	6	
	Midterm Exam		25	7	
	Final Exam		40	12	
	Total		100		

Specific learning outcome:	• Students will get the understanding different Information retrieval model. • Students will get to know about evaluation methods of the information retrieval model. • Students will get to know the challenges associated with each topic.
Course References:	• ***Introduction to Information Retrieval,*** by Christopher Manning, Prabhakar Raghavan and Hinrich Schutze, Cambridge University Press. 2009. **[Note: This book is available online from the above link; In the reading assignments it is referred to as the "IR Book"]**. • Research and Reference Articles (provided online). •

Course topics (Theory)	**Week**	Learning Outcome
An introduction to information retrieval.	1, and 2	Discuss the course book and describe information Retrieval, the goals of An introduction to information retrieval.
Boolean retrieval and inverted index.	3	• An example information retrieval problem • A first take at building an inverted index • Processing Boolean queries • The extended Boolean model versus ranked retrieval
Term vocabulary and postings lists.	4	
Dictionaries and tolerant retrieval.	5	• Hardware basics • Blocked sort-based indexing • Single-pass in-memory indexing • Distributed indexing • Dynamic indexing
Index construction.	6, 7	

<table>
<tr><td></td><td></td><td>• Other types of indexes</td></tr>
<tr><td>Index compression.</td><td>8</td><td rowspan="2">• Statistical properties of terms in information retrieval
• Dictionary compression
• Postings file compression
• Parametric and zone indexes
• Term frequency and weighting
• The vector space model for scoring
• Variant tf-idf functions</td></tr>
<tr><td>Scoring, term weighting and the vector space model.</td><td>9, and 10</td></tr>
<tr><td>Computing scores in a complete search system.</td><td>11, and 12</td><td>explain how a generic searching algorithm works demonstrate how depth-first search will work on a graph demonstrate how breadth-first search will work on a graph predict the space and time requirements for depth-first and breadth-first searches</td></tr>
<tr><td>Evaluation in information retrieval.</td><td>13,14</td><td>devise an useful heuristic function for a problem demonstrate how best-first and A ∗ search will work on a graph predict the space and time requirements for best-first and A ∗ search</td></tr>
</table>

Practical Topics	**Week**	**Learning Outcome**
Introduction to **Classes and Functions**	1,2	*Identify the classes and function*
Files and exceptions	3,4	*Read, write file*
Understanding Object-Oriented Basics	5,6	*Identify the Object-Oriented Basics*
Object-Oriented basics	7,8	*Identify the keyword search*
index	9,10	Retrieve the data from computers
searching	11,12	Type of search
Machin Learning	13,14	Implement the clustering

Questions Example Design

(a) Explain how vector space concepts can be used to calculate the similarity between two documents.

(b) You have the collection of documents that contain the following index terms:

D_1: alpha bravo charlie delta echo foxtrot golf

D_2: golf golf golf delta alpha

D_3: bravo charlie bravo echo foxtrot bravo

D_4: foxtrot alpha alpha golf golf delta

(i) Use an incidence matrix of terms to calculate a similarity matrix for these four documents, with no term weighting.

(ii) Use a frequency matrix of terms to calculate a similarity matrix for these documents, with weights proportional to the term frequency and inversely proportional to the document frequency.

-

Extra notes:

External Evaluator
Dr. roojwan