



## UNIVERSITI TEKNOLOGI MARA

### CSC742: STORAGE AND RETRIEVAL ALGORITHMS

<b>Course Name (English)</b>	STORAGE AND RETRIEVAL ALGORITHMS <b>APPROVED</b>
<b>Course Code</b>	CSC742
<b>MQF Credit</b>	3
<b>Course Description</b>	This course intends to prepare students a strong background concerning the automated storage and retrieval of information. Text is the primary way that human knowledge is stored, and after speech, it is the primary way of transmitted to one and another. This course will mainly focus on the techniques for storing, searching and retrieving for textual documents. Both the theoretical aspects of information retrieval design and evaluation and the practical aspects of how these theories have been implemented in actual systems will be covered. Other types of IR, such as parallel and distributed IR, Multimedia IR will be introduces as an exposure to different types of IR besides text IR.
<b>Transferable Skills</b>	1) Demonstrate ability to identify and articulate self skills, knowledge and understanding confidently and in a variety of contexts 2) Demonstrate ability to manage personal performance to meet expectations and demonstrate drive, determination, and accountability. 3) Demonstrate ability to communicate clearly and confidently, and listen critically
<b>Teaching Methodologies</b>	Lectures, Blended Learning, Discussion, Presentation
<b>CLO</b>	CLO1 Apply appropriate Information Retrieval (IR) models, query formations, indexing & searching algorithms in evaluating IR systems CLO2 Analyze the various platforms and approaches in executing the information retrieval systems. CLO3 Demonstrate communication skills in developing IR systems CLO4 Demonstrate teamwork skills in developing IR systems
<b>Pre-Requisite Courses</b>	No course recommendations
<b>Topics</b>	
<b>1. Introduction to Information Retrieval</b> 1.1) State of Art of IR 1.2) IR Problem 1.3) IR System 1.4) User Interface for Search	
<b>2. Modelling</b> 2.1) Classic IR Model 2.2) Set Theoretic Model 2.3) Algebraic Model 2.4) Probabilistic Model	
<b>3. Documents: Languages and properties</b> 3.1) Metadata 3.2) Document formats and markup languages 3.3) Document preprocessing 3.4) Text properties 3.5) Organizing documents 3.6) Text Compression	
<b>4. Query Languages and Operations</b> 4.1) Query Languages 4.2) Query Properties 4.3) Relevance feedback 4.4) Query Expansion	

**5. Indexing and Searching Algorithms**

- 5.1) Inverted indices
- 5.2) Signature Files
- 5.3) Suffix trees and arrays
- 5.4) Sequential searching
- 5.5) Multidimensional indexing

**6. Retrieval Evaluation**

- 6.1) Retrieval Performance Evaluation
- 6.2) Evaluation using Reference Collections
- 6.3) Interactive Systems Evaluation
- 6.4) Search Log Analysis using click-through data

**7. Advanced IR**

- 7.1) Parallel and Distributed IR
- 7.2) Multimedia IR
- 7.3) Web searching
- 7.4) Open-Source Search Engine

Assessment Breakdown		%	
Continuous Assessment		100.00%	

  

Details of Continuous Assessment	Assessment Type	Assessment Description	% of Total Mark	CLO
	Assignment	Assignment 1	10%	CLO1
	Assignment	Assignment 2	10%	CLO1
	Assignment	Assignment 3	10%	CLO2
	Assignment	Assignment 4	10%	CLO2
	Presentation	Presentation	10%	CLO3
	Test	TEST 1	10%	CLO1
	Test	TEST 2	10%	CLO1
	Written Report	Written Report	30%	CLO4

  

Reading List	Reference Book Resources
	<ul style="list-style-type: none"> <li>• Baeza-Yates, R. &amp; Ribeiro-Neto, B. 2011, <i>Modern Information Retrieval: The Concepts and Technology behind Search</i>, 2nd Edition Ed., Addison-Wesley Professional [ISBN: 978-032141691]</li> <li>• Stefan Buettcher, Charles L. A. Clarke, Gordon V. Cormack 2016, <i>Information Retrieval: Implementing and Evaluating Search Engines</i>, The MIT Press [ISBN: 978-026202651]</li> <li>• Grant S. Ingersoll 2013, <i>Taming Text: How to Find, Organize, and Manipulate It</i>, Manning Publications [ISBN: 978-193398838]</li> <li>• Stefano Ceri, Alessandro Bozzon, Marco Brambilla, Emanuele Della Valle, Piero Fraternali, Silvia Quarteroni 2013, <i>Web Information Retrieval</i>, Springer [ISBN: 9783642393136]</li> <li>• Marcia J. Bates 2011, <i>Understanding Information Retrieval Systems</i>, CRC Press [ISBN: 9781439891964]</li> </ul>
Article/Paper List	This Course does not have any article/paper resources
Other References	This Course does not have any other resources