

## Subject Description Form

<b>Subject Code</b>	COMP4133
<b>Subject Title</b>	Information Retrieval
<b>Credit Value</b>	3
<b>Level</b>	4
<b>Pre-requisite / Co-requisite / Exclusion</b>	<b>Pre-requisite:</b> COMP2011, COMP2411
<b>Objectives</b>	<p>The objectives of this subject are to:</p> <ul style="list-style-type: none"><li>• provide the foundation knowledge in information retrieval;</li><li>• equip students with sound skills to solve computational search problems;</li><li>• appreciate how to evaluate search engines;</li><li>• appreciate the different applications of information retrieval techniques in the Internet or Web environment; and</li><li>• provide hands-on experience in building search engines and/or hands-on experience in evaluating search engines.</li></ul>
<b>Intended Learning Outcomes</b>	<p>Upon completion of the subject, students will be able to:</p> <p><i>Professional/academic knowledge and skills</i></p> <ul style="list-style-type: none"><li>(a) understand and apply the basic concepts of information retrieval;</li><li>(b) appreciate the limitations of different information retrieval techniques;</li><li>(c) write programs to implement search engines; and</li><li>(d) evaluate search engines.</li></ul> <p><i>Attributes for all-roundedness</i></p> <ul style="list-style-type: none"><li>(e) develop skills in problem solving using systematic approaches; and</li><li>(f) solve complex problems in groups and develop group work.</li></ul>

<b>Subject Synopsis/ Indicative Syllabus</b>	<b>Topic</b>
	<b>1. Fundamentals of Information Retrieval</b> System architecture; limitations of information retrieval systems; evaluation methodology; performance measures; benchmarking.
	<b>2. Models of Information Retrieval</b> Boolean retrieval models; fuzzy Boolean retrieval models; vector space models; inner product similarities; cosine similarities; term weighting schemes.
	<b>3. Query Processing</b> Query languages; basic query processing for Boolean retrieval models; query processing for vector space models; query expansion; relevance feedback.
	<b>4. Indexing Strategies</b> Inverted file construction; efficient dictionary management; indexing to support phrasal search and proximity.
	<b>5. Text Clustering</b> Hierarchical clustering algorithms; adaptive clustering algorithms; <i>k</i> -means clustering algorithms.
	<b>6. Web Retrieval</b> Characteristics of the web; spidering; weighting schemes for web documents; web link analysis.
	Laboratory Experiment and Tutorial:
	<b>Topic</b>
	1. Evaluation techniques. 2. Ranking techniques. 3. Text processing techniques. 4. Web processing techniques.
<b>Teaching/ Learning Methodology</b>	Teaching is based on lectures which include solving technical problems in information retrieval (aligned to Programme Outcome 6). Tutorials are used to provide examples of problems and to show how solutions are developed (aligned to Programme Outcomes 4, 5, 6). The mid-term is administered to students to strengthen their technical problem solving ability (aligned to Programme Outcome 5). There is a project/assignment that students need to write their report (aligned to Programme Outcomes 1,4). This project/assignment is typically group work (aligned to Programme Outcome 7).

Assessment Methods in Alignment with Intended Learning Outcomes	Specific assessment methods/tasks	% weighting	Intended subject learning outcomes to be assessed (Please tick as appropriate)					
			a	b	c	d	e	f
	<b>Continuous Assessment</b>	<b>55%</b>						
1. Assignments, Lab Exercises and Project	30%	✓	✓	✓	✓	✓	✓	
2. Mid-Term	25%	✓	✓		✓			
<b>Examination</b>	<b>45%</b>	✓	✓		✓			
Total	100%							
<p>The project/assignment is suitable to assess all the intended learning outcomes as it involves all of them. The mid-term and examination will tests the basic concepts learnt by the students as well as to see if the students are capable to use retrieval techniques and perform search engine evaluation.</p>								
<b>Student Study Effort Expected</b>	Class contact:							
	▪ Lecture					39 Hrs.		
	▪ Tutorial/Lab					0 Hrs.		
	Other student study effort:							
	▪ Project/ Assignment					28 Hrs.		
	▪ Self-Study					38 Hrs.		
Total student study effort					105 Hrs.			
<b>Reading List and References</b>	<b>Reference Books:</b>							
	1. Chowdhury, G.G., <i>Introduction to Modern Information Retrieval</i> , London, Facet, 2010.							
	2. Grossman, D.A. and Freider, O., <i>Information Retrieval: Algorithms and Heuristics</i> , Kluwer Academic Publishers, 2004.							
	3. Baeza-Yates, R.A. and Riberio-Neto, B., <i>Modern Information Retrieval</i> , ACM Press, 1999.							
	4. Grefenstette, G. (ed.), <i>Cross-language Information Retrieval</i> , Dorhrecht, The Netherlands: Kluwer Academic Publishers, 1998.							
	5. Witten, I.H., Moffat, A. and Bell, T.C., <i>Managing Gigabytes: Compressing and Indexing Documents and Images</i> , San Francisco, California: Morgan Kaufmann Publishers, 1999.							