

Subject Description Form

Subject Code	COMP4133					
Subject Title	Information Retrieval					
Credit Value	3					
Level	4					
Pre-requisite / Co-requisite/ Exclusion	Pre-requisite: COMP2011, COMP2411					
Objectives	<ul style="list-style-type: none"> • To provide the foundation knowledge in information retrieval. • To equip students with sound skills to solve computational search problems. • To appreciate how to evaluate search engines. • To appreciate the different applications of information retrieval techniques in the Internet or Web environment. • To provide hands-on experience in building search engines and/or hands-on experience in evaluating search engines. 					
Intended Learning Outcomes	<p>Upon completion of the subject, students will be able to:</p> <p><i>Professional/academic knowledge and skills</i></p> <p>(a) understand and apply the basic concepts of information retrieval;</p> <p>(b) appreciate the limitations of different information retrieval techniques;</p> <p>(c) write programs to implement search engines;</p> <p>(d) evaluate search engines;</p> <p><i>Attributes for all-roundedness</i></p> <p>(e) develop skills in problem solving using systematic approaches;</p> <p>(f) solve complex problems in groups and develop group work.</p>					
Subject Synopsis/ Indicative Syllabus	<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: left; padding: 5px;">Topic</th> </tr> </thead> <tbody> <tr> <td style="padding: 5px;">1. Fundamentals of information retrieval System architecture; limitations of information retrieval systems; evaluation methodology; performance measures; benchmarking.</td> </tr> <tr> <td style="padding: 5px;">2. Models of information retrieval Boolean retrieval models; fuzzy Boolean retrieval models; vector space models; inner product similarities; cosine similarities; term weighting schemes.</td> </tr> <tr> <td style="padding: 5px;">3. Query processing Query languages; basic query processing for Boolean retrieval models;</td> </tr> </tbody> </table>		Topic	1. Fundamentals of information retrieval System architecture; limitations of information retrieval systems; evaluation methodology; performance measures; benchmarking.	2. Models of information retrieval Boolean retrieval models; fuzzy Boolean retrieval models; vector space models; inner product similarities; cosine similarities; term weighting schemes.	3. Query processing Query languages; basic query processing for Boolean retrieval models;
Topic						
1. Fundamentals of information retrieval System architecture; limitations of information retrieval systems; evaluation methodology; performance measures; benchmarking.						
2. Models of information retrieval Boolean retrieval models; fuzzy Boolean retrieval models; vector space models; inner product similarities; cosine similarities; term weighting schemes.						
3. Query processing Query languages; basic query processing for Boolean retrieval models;						

	<p>query processing for vector space models; query expansion; relevance feedback.</p> <p>4. Indexing strategies Inverted file construction; efficient dictionary management; indexing to support phrasal search and proximity.</p> <p>5. Text properties and information extraction Zipf law; Heap's law; HTML markup; XML markup; Document Object Models; (DOM); string searching; filler pattern matching.</p> <p>6. Text clustering Hierarchical clustering algorithms; adaptive clustering algorithms; <i>k</i>-means clustering algorithms.</p> <p>7. Web retrieval Characteristics of the web; spidering; weighting schemes for web documents; web link analysis.</p> <p>8. Text categorization Rocchio classifier; <i>k</i>-NN classifier; naïve Bayesian classifier and evaluation methodology.</p>																																															
	<p>Laboratory Experiment and Tutorial:</p> <table border="1" data-bbox="440 880 1445 1099"> <tr> <td data-bbox="440 880 1445 947">Topic</td> </tr> <tr> <td data-bbox="440 947 1445 1099"> <ol style="list-style-type: none"> Evaluation techniques. Ranking techniques. Text processing techniques. Web processing techniques. </td> </tr> </table>	Topic	<ol style="list-style-type: none"> Evaluation techniques. Ranking techniques. Text processing techniques. Web processing techniques. 																																													
Topic																																																
<ol style="list-style-type: none"> Evaluation techniques. Ranking techniques. Text processing techniques. Web processing techniques. 																																																
<p>Teaching/Learning Methodology</p>	<p>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).</p>																																															
<p>Assessment Methods in Alignment with Intended Learning Outcomes</p>	<table border="1" data-bbox="432 1413 1457 1895"> <thead> <tr> <th rowspan="2">Specific assessment methods/tasks</th> <th rowspan="2">% weighting</th> <th colspan="6">Intended subject learning outcomes to be assessed (Please tick as appropriate)</th> </tr> <tr> <th>a</th> <th>b</th> <th>c</th> <th>d</th> <th>e</th> <th>f</th> </tr> </thead> <tbody> <tr> <td>1. Assignments, Lab Exercises and Project</td> <td>30%</td> <td>✓</td> <td>✓</td> <td>✓</td> <td>✓</td> <td>✓</td> <td>✓</td> </tr> <tr> <td>2. Mid-term</td> <td>25%</td> <td>✓</td> <td>✓</td> <td></td> <td>✓</td> <td></td> <td></td> </tr> <tr> <td>3. Examination</td> <td>45%</td> <td>✓</td> <td>✓</td> <td></td> <td>✓</td> <td></td> <td></td> </tr> <tr> <td>Total</td> <td>100 %</td> <td colspan="6"></td> </tr> </tbody> </table> <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>	Specific assessment methods/tasks	% weighting	Intended subject learning outcomes to be assessed (Please tick as appropriate)						a	b	c	d	e	f	1. Assignments, Lab Exercises and Project	30%	✓	✓	✓	✓	✓	✓	2. Mid-term	25%	✓	✓		✓			3. Examination	45%	✓	✓		✓			Total	100 %							
Specific assessment methods/tasks	% weighting			Intended subject learning outcomes to be assessed (Please tick as appropriate)																																												
		a	b	c	d	e	f																																									
1. Assignments, Lab Exercises and Project	30%	✓	✓	✓	✓	✓	✓																																									
2. Mid-term	25%	✓	✓		✓																																											
3. Examination	45%	✓	✓		✓																																											
Total	100 %																																															

Student Study Effort Expected	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.
Reading List and References	<p>Reference Books:</p> <ol style="list-style-type: none"> 1. G.G. Chowdhury. Introduction to Modern Information Retrieval, London, Facet, 2010. 2. D.A. Grossman and O. Freider. Information Retrieval: Algorithms and Heuristics, Kluwer Academic Publishers, 2004. 3. R.A. Baeza-Yates, B. Riberio-Neto. Modern Information Retrieval, ACM Press, 1999. 4. G. Grefenstette (ed.) Cross-language Information Retrieval, Dorhrecht, The Netherlands: Kluwer Academic Publishers, 1998. 5. I.H. Witten, A. Moffat and T.C. Bell. Managing Gigabytes: Compressing and Indexing Documents and Images, San Francisco, California: Morgan Kaufmann Publishers, 1999. 	