

Subject Description Form

Subject Code	COMP433
Subject Title	Information Retrieval
Credit Value	3
Level	4
Pre-requisite / Co-requisite/ Exclusion	Pre-requisite: COMP311 (not applicable for 61025), COMP305 Co-requisite/Exclusion: Nil
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><u>Professional/academic knowledge and skills</u></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><u>Attributes for all-roundedness</u></p> <p>(e) develop skills in problem solving using systematic approaches;</p> <p>(f) solve complex problems in groups and develop group work.</p> <p>Alignment of Programme Outcomes:</p> <p>Programme Outcome 1: This subject contributes to having students to practice their writing skill with project document and report writing.</p> <p>Programme Outcome 4: This subject contributes to developing student critical thinking through tutorials on solving problems. They will practice more in doing their project.</p>

	<p>Programme Outcome 5: This subject contributes to measuring the technical problem solving ability by administering quiz or classwork.</p> <p>Programme Outcome 6: This subject contributes to informing students about the advancement of information retrieval via lectures or tutorials.</p> <p>Programme Outcome 7: This subject contributes to team work with group-based project for students to practice team spirit.</p>											
<p>Subject Synopsis/ Indicative Syllabus</p>	<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: center; padding: 5px;">Topic</th> </tr> </thead> <tbody> <tr> <td style="padding: 5px;"> <p>1. Fundamentals of information retrieval System architecture; limitations of information retrieval systems; evaluation methodology; performance measures; benchmarking.</p> </td> </tr> <tr> <td style="padding: 5px;"> <p>2. Models of information retrieval Boolean retrieval models; fuzzy Boolean retrieval models; vector space models; inner product similarities; cosine similarities; term weighting schemes.</p> </td> </tr> <tr> <td style="padding: 5px;"> <p>3. Query processing Query languages; basic query processing for Boolean retrieval models; query processing for vector space models; query expansion; relevance feedback.</p> </td> </tr> <tr> <td style="padding: 5px;"> <p>4. Indexing strategies Inverted file construction; efficient dictionary management; indexing to support phrasal search and proximity.</p> </td> </tr> <tr> <td style="padding: 5px;"> <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> </td> </tr> <tr> <td style="padding: 5px;"> <p>6. Text clustering Hierarchical clustering algorithms; adaptive clustering algorithms; <i>k</i>-means clustering algorithms.</p> </td> </tr> <tr> <td style="padding: 5px;"> <p>7. Web retrieval Characteristics of the web; spidering; weighting schemes for web documents; web link analysis.</p> </td> </tr> <tr> <td style="padding: 5px;"> <p>8. Text categorization Rocchio classifier; <i>k</i>-NN classifier; naïve Bayesian classifier and evaluation methodology.</p> </td> </tr> </tbody> </table> <p style="margin-top: 20px;">Laboratory Experiment and Tutorial:</p> <table border="1" style="width: 100%; border-collapse: collapse; margin-top: 10px;"> <thead> <tr> <th style="text-align: center; padding: 5px;">Topic</th> </tr> </thead> <tbody> <tr> <td style="padding: 5px;"> <p>1. Evaluation techniques.</p> <p>2. Ranking techniques.</p> <p>3. Text processing techniques.</p> <p>4. Web processing techniques.</p> </td> </tr> </tbody> </table>	Topic	<p>1. Fundamentals of information retrieval System architecture; limitations of information retrieval systems; evaluation methodology; performance measures; benchmarking.</p>	<p>2. Models of information retrieval Boolean retrieval models; fuzzy Boolean retrieval models; vector space models; inner product similarities; cosine similarities; term weighting schemes.</p>	<p>3. Query processing Query languages; basic query processing for Boolean retrieval models; 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>	Topic	<p>1. Evaluation techniques.</p> <p>2. Ranking techniques.</p> <p>3. Text processing techniques.</p> <p>4. Web processing techniques.</p>
Topic												
<p>1. Fundamentals of information retrieval System architecture; limitations of information retrieval systems; evaluation methodology; performance measures; benchmarking.</p>												
<p>2. Models of information retrieval Boolean retrieval models; fuzzy Boolean retrieval models; vector space models; inner product similarities; cosine similarities; term weighting schemes.</p>												
<p>3. Query processing Query languages; basic query processing for Boolean retrieval models; 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>												
Topic												
<p>1. Evaluation techniques.</p> <p>2. Ranking techniques.</p> <p>3. Text processing techniques.</p> <p>4. Web processing techniques.</p>												
<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). Quizzes and/or classworks are administered to students to strength their technical problem solving ability</p>											

	(aligned to Programme Outcome 5). There is a project that students need to write their report (aligned to Programme Outcomes 1, 4). This project is typically a group project (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
	1. Assignments							
	2. Lab exercises							
	3. Project		30%	✓	✓	✓	✓	✓
	4. Mid-term		25%	✓	✓		✓	
	5. Examination		45%	✓	✓		✓	
	Total		100 %					
<p>Explanation of the appropriateness of the assessment methods in assessing the intended learning outcomes:</p> <p>The project 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>								
Student Study Effort Expected	Class contact:							
	▪ Lecture							39 Hrs.
	▪ Tutorial/Lab							0 Hrs.
	Other student study effort:							
	▪ Project							14 Hrs.
	▪ Self Study							17 Hrs.
	Total student study effort							70 Hrs.
Reading List and References	<p>Reference Books:</p> <p>1. G.G. Chowdhury. An Introduction to Modern Information Retrieval, London, Facet, 2004.</p> <p>2. R.A. Baeza-Yates, B. Riberio-Neto. Modern Information Retrieval, ACM Press, 1999.</p> <p>3. D.A. Grossman and O. Freider. Information Retrieval: Algorithms and Heuristics, Kluwer Academic Publishers, 1998.</p> <p>4. G. Grefenstette (ed.) Cross-language Information Retrieval, Dorhrecht, The</p>							

Netherlands: Kluwer Academic Publishers, 1998.

5. I.H. Witten, A. Moffat and T.C. Bell. Managing Gigabytes: Compressing and Indexing Documents and Images, New York: Van Nostrand Reinhold, 1994.