Subject Description Form

Subject Code	COMP4133
Subject Title	Information Retrieval
Credit Value	3
Level	4
Pre-requisite / Co-requisite / Exclusion	**Pre-requisite**: COMP2011/COMP2013 & COMP2411
Objectives	The objectives of this subject are to: 1. provide the foundation knowledge in information retrieval; 2. equip students with sound skills to solve computational search problems; 3. appreciate how to evaluate search engines; 4. appreciate the different applications of information retrieval techniques in the Internet or Web environment; and 5. provide hands-on experience in building search engines and/or hands-on experience in evaluating search engines.
Intended Learning Outcomes	Upon completion of the subject, students will be able to: *Professional/academic knowledge and skills* (a) understand the basic concepts of information retrieval; (b) explain the limitations of different information retrieval techniques; (c) design and develop information retrieval systems; (d) perform the evaluation and critical analysis of the performance of the retrieval system; *Attributes for all-roundedness* (e) develop skills in problem-solving using systematic approaches; and (f) solve complex problems in groups and develop group work.

<table>
<tr>
<td>Subject Synopsis/ Indicative Syllabus</td>
<td>
<table>
<tr><th>Topic</th></tr>
<tr><td>1. Fundamentals of Information Retrieval

System architecture; limitations of information retrieval systems; evaluation methodology; performance measures; benchmarking.</td></tr>
<tr><td>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>3. Query Processing

Query languages; basic query processing for Boolean retrieval models; query processing for vector space models; query expansion; relevance feedback.</td></tr>
<tr><td>4. Indexing Strategies

Inverted file construction; efficient dictionary management; indexing to support phrasal search and proximity.</td></tr>
<tr><td>5. Text Clustering

Hierarchical clustering algorithms; adaptive clustering algorithms; k-means clustering algorithms.</td></tr>
<tr><td>6. Web Retrieval

Characteristics of the web; spidering; weighting schemes for web documents; web link analysis.</td></tr>
</table>
Laboratory Experiment and Tutorial:
<table>
<tr><th>Topic</th></tr>
<tr><td>1. Evaluation techniques.
2. Ranking techniques.
3. Text processing techniques.
4. Web processing techniques.</td></tr>
</table>
</td>
</tr>
<tr>
<td>Teaching/ Learning Methodology</td>
<td>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).</td>
</tr>
</table>

Assessment Methods in Alignment with Intended Learning Outcomes

Specific assessment methods/tasks	% weighting	Intended subject learning outcomes to be assessed					
		a	b	c	d	e	f
Continuous Assessment	**55%**						
1. Assignments, Lab Exercises and Project	30%	✓	✓	✓	✓	✓	✓
2. Mid-Term	25%	✓	✓		✓		
Examination	**45%**	✓	✓		✓		
Total	100%						

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.

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

Reference Books:

1. Chowdhury, G.G., *Introduction to Modern Information Retrieval*, London, Facet, 2010.
2. Grossman, D.A. and Freider, O., *Information Retrieval: Algorithms and Heuristics*, Kluwer Academic Publishers, 2004.
3. Baeza-Yates, R.A. and Riberio-Neto, B., *Modern Information Retrieval*, ACM Press, 1999.
4. Grefenstette, G. (ed.), *Cross-language Information Retrieval*, Dorhrecht, The Netherlands: Kluwer Academic Publishers, 1998.
5. Witten, I.H., Moffat, A. and Bell, T.C., *Managing Gigabytes: Compressing and Indexing Documents and Images*, San Francisco, California: Morgan Kaufmann Publishers, 1999.