

## Subject Description Form

<b>Subject Code</b>	COMP3133
<b>Subject Title</b>	Chinese Language Processing
<b>Credit Value</b>	3
<b>Level</b>	3
<b>Pre-requisite / Co-requisite / Exclusion</b>	<b>Pre-requisite:</b> COMP1011
<b>Objectives</b>	<p>The objectives of this subject are to:</p> <ul style="list-style-type: none"> <li>• provide essential knowledge of computer representation of natural language text and practical techniques to handle and process natural language text, with a specific emphasis on the Chinese language and its co-processing with other languages such as English; and</li> <li>• provide fundamental skills to design and develop innovative software systems with natural language processing techniques and tools for different communities.</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 the technical difficulties in representing and processing text written or encoded in Chinese;</li> <li>(b) understand the system support for entering and displaying Chinese in computers;</li> <li>(c) understand fundamental Chinese language (and natural language in general) processing concepts, processes and major problems;</li> <li>(d) understand the applications of Chinese (and natural language in general) processing;</li> <li>(e) apply fundamental knowledge and advanced techniques of natural language processing to develop textual information processing systems; and</li> <li>(f) handle multi-lingual text representation and processing issues.</li> </ul> <p><i>Attributes for all-roundedness</i></p> <ul style="list-style-type: none"> <li>(g) solve problems using systematic approaches; and</li> <li>(h) learn independently and be able to search for the information required in solving problems.</li> </ul>

<b>Subject Synopsis/ Indicative Syllabus</b>	<p><b>Topic</b></p> <p><b>1. Representation, Input and Output of Chinese Characters</b> Chinese and Universal Character Set Standards (such as GB, Big5 and Unicode), Character Encoding Schemes (such as ISO2022, EUC and UTF), Code Set Compatibility Issues, Chinese Character Input and Output Processing.</p> <p><b>2. Fundamental of (Chinese) Natural Language Processing</b> Morphological Analysis, Word Stemming and Segmentation, Syntactic Analysis and Sentence Parsing, Semantic Analysis and Sense Disambiguation, Discourse Analysis and Co-Reference Resolution.</p> <p><b>3. Selected Topics in (Chinese) Natural Language Processing Applications</b> Language Modeling, Opinion Mining and Sentiment Classification, Dialogue and Conversation System, Question Answering and Summarization, Information Retrieval and Extraction, Machine Translation, etc.</p>																																																																																												
<b>Teaching/ Learning Methodology</b>	<p>Lectures teach students on the main concepts of the course, together with comprehensive examples, class exercises or questions and answers for easy understanding.</p> <p>Tutorials and lab sessions offer the opportunity for students to review the lecture materials and master the practical techniques and necessary tools for effective system and application development.</p> <p>Project and/or assignments give students the opportunity to develop analytical and problem solving skills through system implementation and interpersonal communication.</p>																																																																																												
<b>Assessment Methods in Alignment with Intended Learning Outcomes</b>	<table border="1"> <thead> <tr> <th rowspan="2">Specific assessment methods/tasks</th> <th rowspan="2">% weighting</th> <th colspan="8">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> <th>g</th> <th>h</th> </tr> </thead> <tbody> <tr> <td><b>Continuous Assessment</b></td> <td rowspan="5"><b>55%</b></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>1. Assignments</td> <td>✓</td> <td></td> <td>✓</td> <td>✓</td> <td></td> <td>✓</td> <td></td> <td>✓</td> </tr> <tr> <td>2. Lab Exercises</td> <td></td> <td>✓</td> <td>✓</td> <td>✓</td> <td>✓</td> <td>✓</td> <td>✓</td> <td>✓</td> </tr> <tr> <td>3. Project/ Presentation</td> <td></td> <td>✓</td> <td></td> <td>✓</td> <td>✓</td> <td>✓</td> <td>✓</td> <td></td> </tr> <tr> <td>4. Mid-Term</td> <td></td> <td>✓</td> <td>✓</td> <td>✓</td> <td></td> <td>✓</td> <td>✓</td> <td></td> </tr> <tr> <td><b>Examination</b></td> <td><b>45%</b></td> <td></td> <td>✓</td> <td>✓</td> <td>✓</td> <td></td> <td>✓</td> <td>✓</td> <td></td> </tr> <tr> <td><b>Total</b></td> <td><b>100 %</b></td> <td colspan="8"></td> </tr> </tbody> </table>									Specific assessment methods/tasks	% weighting	Intended subject learning outcomes to be assessed (Please tick as appropriate)								a	b	c	d	e	f	g	h	<b>Continuous Assessment</b>	<b>55%</b>									1. Assignments	✓		✓	✓		✓		✓	2. Lab Exercises		✓	✓	✓	✓	✓	✓	✓	3. Project/ Presentation		✓		✓	✓	✓	✓		4. Mid-Term		✓	✓	✓		✓	✓		<b>Examination</b>	<b>45%</b>		✓	✓	✓		✓	✓		<b>Total</b>	<b>100 %</b>								
Specific assessment methods/tasks	% weighting	Intended subject learning outcomes to be assessed (Please tick as appropriate)																																																																																											
		a	b	c	d	e	f	g	h																																																																																				
<b>Continuous Assessment</b>	<b>55%</b>																																																																																												
1. Assignments		✓		✓	✓		✓		✓																																																																																				
2. Lab Exercises			✓	✓	✓	✓	✓	✓	✓																																																																																				
3. Project/ Presentation			✓		✓	✓	✓	✓																																																																																					
4. Mid-Term			✓	✓	✓		✓	✓																																																																																					
<b>Examination</b>	<b>45%</b>		✓	✓	✓		✓	✓																																																																																					
<b>Total</b>	<b>100 %</b>																																																																																												

<b>Student Study Effort Expected</b>	Class contact:	
	▪ Lecture	26 Hrs.
	▪ Laboratory	13 Hrs.
	Other student study effort:	
	▪ Self-study	40 Hrs.
	▪ Homework, Quizzes and Assignments	40 Hrs.
	Total student study effort	119 Hrs.
<b>Reading List and References</b>	<b>Reference Books:</b> <ol style="list-style-type: none"> <li>1. Lunde, Ken, <i>CJKV Information Processing: Chinese, Japanese, Korean and Vietnamese Information Processing</i>, O'Reilly &amp; Associates, 2008.</li> <li>2. Huang, J.K.T. and Huang, T.D., <i>An Introduction to Chinese, Japanese and Korean Computing</i>, Singapore: World Scientific, 1989.</li> <li>3. Jurafsky, Dan and Martin, James H., <i>Speech and Language Processing, An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition</i>, 2008.</li> <li>4. Manning, Christopher D. and Schütze, Hinrich, <i>Foundations of Statistical Natural Language Processing</i>, The MIT Press, 1999.</li> <li>5. Bird, Steven, Klein, Ewan and Loper, Edward, <i>Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit</i>, O'Reilly Media, 2009.</li> </ol>	