Subject Description Form

Subject Code	COMP3133
Subject Title	Chinese Language Processing
Credit Value	3
Level	3
Pre-requisite / Co-requisite / Exclusion	**Pre-requisite**: COMP1011/COMP1012/ENG2002
Objectives	The objectives of this subject are to: 1. 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 2. provide fundamental skills to design and develop innovative software systems with natural language processing techniques and tools for different communities.
Intended Learning Outcomes	Upon completion of the subject, students will be able to: *Professional/academic knowledge and skills* (a) grasp fundamental concepts, processes and major problems in Chinese language (and natural language in general) processing; (b) well understand the applications of Chinese (and natural language in general) processing; (c) apply fundamental knowledge and advanced techniques of natural language processing to develop textual information processing systems; (d) handle multi-lingual text representation and processing issues; *Attributes for all-roundedness* (e) solve problems using systematic approaches; and (f) be able to search for the information required in solving problems.

Subject Synopsis/ Indicative Syllabus

Topic
1. Representation, Input and Output of Chinese Characters 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.
2. Fundamental of (Chinese) Natural Language Processing Morphological Analysis, Word Stemming and Segmentation, Syntactic Analysis and Sentence Parsing, Semantic Analysis and Sense Disambiguation, Discourse Analysis and Co-Reference Resolution.
3. Selected Topics in (Chinese) Natural Language Processing Applications Language Modeling, Opinion Mining and Sentiment Classification, Dialogue and Conversation System, Question Answering and Summarisation, Information Retrieval and Extraction, Machine Translation, etc.

Teaching/ Learning Methodology

Lectures teach students on the main concepts of the course, together with comprehensive examples, class exercises or questions and answers for easy understanding.

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.

Project and/or assignments give students the opportunity to develop analytical and problem-solving skills through system implementation and interpersonal communication.

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							
1. Assignments		✓	✓		✓		✓
2. Lab Exercises	**55%**	✓	✓	✓	✓	✓	✓
3. Project/ Presentation			✓	✓	✓	✓	✓
4. Mid-Term		✓	✓		✓	✓	
Examination	**45%**	✓	✓		✓	✓	
Total	100%						

Student Study Effort Expected	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.
Reading List and References	**Reference Books:** 1. Lunde, Ken, *CJKV Information Processing: Chinese, Japanese, Korean and Vietnamese Information Processing*, O'Reilly & Associates, 2008. 2. Huang, J.K.T. and Huang, T.D., *An Introduction to Chinese, Japanese and Korean Computing*, Singapore: World Scientific, 1989. 3. Jurafsky, Dan and Martin, James H., *Speech and Language Processing, An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*, 2008. 4. Manning, Christopher D. and Schütze, Hinrich, *Foundations of Statistical Natural Language Processing,* The MIT Press, 1999. 5. Bird, Steven, Klein, Ewan and Loper, Edward, *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*, O'Reilly Media, 2009.	