

Subject Description Form

Subject Code	COMP3133
Subject Title	Chinese Language Computing
Credit Value	3
Level	3
Pre-requisite / Co-requisite/ Exclusion	Pre-requisite: COMP1011
Objectives	<ul style="list-style-type: none">• To provide the students with a basic understanding of the foundation in system coding and design for text processing with a specific emphasis on the Chinese language and its co-processing with other languages such as English.• To provide training in software design methodology for the recognition of human/machine interface.
Intended Learning Outcomes	<p>Upon completion of the subject, students will be able to:</p> <p><u>Professional/academic knowledge and skills</u></p> <ul style="list-style-type: none">(a) understand the technical difficulties in representing and processing text written or encoded in Chinese;(b) understand the (operating) system support for displaying and entering Chinese in computers;(c) understand some Chinese information processing applications (e.g. Chinese information retrieval, Chinese abstraction and Machine Translation) that demonstrates the integrated use of various techniques;(d) handle Chinese text data encoded in various standards or format (e.g. GB and Big5);(e) handle the software development of Chinese enabled software applications in the context of developing internationalized software and in adhering to international software development practices;(f) handle some basic algorithmic problems and some basic computational Chinese linguistic techniques to enable efficient and intelligent Chinese enabled software applications. <p><u>Attributes for all-roundedness</u></p> <ul style="list-style-type: none">(g) solve problems using systematic approaches;(h) learn independently and be able to search for the information required in solving problems.

Subject Synopsis/ Indicative Syllabus	Topic
	1. Characteristics of Chinese language Historical development; geographical variations (dialects); linguistic descriptions (character, morph, word, phrase and sentence) and quantitative analysis (Zip Law, coverage curve).
	2. Representation of Chinese character sets Mathematical description of representation; character set organization (e.g., GB and Big5); encoding schemes (ISO2022 and UTF); discussion of (de facto) standard character sets (for PC, workstation and network); character decoding techniques and character conversion problems; unbounded alphabet representation and processing.
	3. Output processing of Chinese Font technology; typesetting terminology and text rendering process; bitmap fonts (representation, compression and scaling problems); vector / outline fonts (Limn algorithm); X-Windows fonts (BDF and Postscript) and font related operations (installation, specification, extraction and user-defined glyph addition); automatic glyph construction.
	4. Input processing of Chinese Introduction to Chinese input processing by pen, image, speech and keystrokes; design and evaluation issues for different keystroke input methods: shape-based (e.g. Q9), phonetic-based and shape-phonetic based input methods; input method architecture (e.g. for Microsoft Windows and X-Windows).
	5. Software development for Chinese computing Open systems, internationalization, localization, ANSI-C model, Windows programming for Chinese computing (Microsoft- and X-Windows).
	6. Selected topics in Chinese computing Character set selection (NP-complete); hashing functions for Chinese character sets; string searching (KMP, BM and Quick); dictionary lookup; string-set searching; word segmentation; Chinese information retrieval; abstraction or machine translation.
Laboratory Experiment:	
	Topic
	1. Locale Setting for operating systems and Locale detection
	2. checking of character encoding using binary reading tool
	3. Foundations of Multilingual Website design and web page language setting
	4. Installation of Input Method Engine (IME) to operating system
	5. System font set detection
	6. Character string data handling and processing in programming environment, codeset announcement and conversion
	7. Programming using Internationalization methodology

Teaching/Learning Methodology	<p>Lectures teach students on the main concepts of the course, together with comprehensive examples, and class questions and answers for easy understanding.</p> <p>Tutorials and lab sessions offer the opportunity for students to review the lecture materials through online exercises and also the use of programming tools to learn to program.</p> <p>Programming assignments will give students the opportunity to solve problems through implementation where they understand and practice on how programs can be written and compiled to run to complete certain tasks.</p> <p>Homework assignments help students to develop analytical and problem solving skills.</p>																																																																												
Assessment Methods in Alignment with Intended Learning Outcomes	<table border="1" data-bbox="459 594 1421 1140"> <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>1. Assignments</td> <td rowspan="4">55%</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</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>5. Examination</td> <td>45%</td> <td></td> <td>✓</td> <td>✓</td> <td>✓</td> <td></td> <td>✓</td> <td>✓</td> <td></td> </tr> <tr> <td>Total</td> <td>100 %</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	1. Assignments	55%	✓		✓	✓		✓		✓	2. Lab exercises		✓	✓	✓	✓	✓	✓	✓	3. Project		✓		✓	✓	✓	✓		4. Mid-term		✓	✓	✓		✓	✓		5. 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	g	h																																																																				
1. Assignments	55%	✓		✓	✓		✓		✓																																																																				
2. Lab exercises			✓	✓	✓	✓	✓	✓	✓																																																																				
3. Project			✓		✓	✓	✓	✓																																																																					
4. Mid-term			✓	✓	✓		✓	✓																																																																					
5. Examination	45%		✓	✓	✓		✓	✓																																																																					
Total	100 %																																																																												
Student Study Effort Expected	<table border="1" data-bbox="459 1140 1448 1585"> <tr> <td colspan="2">Class contact:</td> <td></td> </tr> <tr> <td>▪ Lecture</td> <td></td> <td>26 Hrs.</td> </tr> <tr> <td>▪ Laboratory</td> <td></td> <td>13 Hrs.</td> </tr> <tr> <td colspan="2">Other student study effort:</td> <td></td> </tr> <tr> <td>▪ Self-study</td> <td></td> <td>40 Hrs.</td> </tr> <tr> <td>▪ Homework, quizzes and assignments</td> <td></td> <td>40 Hrs.</td> </tr> <tr> <td colspan="2">Total student study effort</td> <td>119 Hrs.</td> </tr> </table>		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.																																																						
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	<p>Textbook:</p> <ol style="list-style-type: none"> Ken Lunde, Chinese, Japanese, Korean and Vietnamese Information Processing, 2nd Edition, O'Reilly & Associates, 2008. <p>Reference Books:</p> <ol style="list-style-type: none"> Hopcroft, J.E. and Ullman, J.D., Introduction to Automata, Theory and Languages, Addison-Wesley, 1979. McGilton, H. and Campione, M., Postscript by Example, Addison-Wesley, 1993. 																																																																												

3. Dan Jurafsky and James H. Martin, *Speech and Language processing*, 2nd Edition, Prentice Hall, 2008
4. Berry, K. and Hargraves, K.A., *GNU Font utilities (Limn algorithm)*,
5. Kano, N. *Developing International Software*, Microsoft Press, 1995.
6. Nutt, G.J., *Open Systems*, Prentice Hall, 1992.
7. Huang, J.K.T. and Huang, T.D., *An Introduction to Chinese, Japanese and Korean Computing*, Singapore: World Scientific, 1989.
8. The Unicode Consortium, *The Unicode Standard, Version 5.0*, Addison-Wesley Developer Press, November 3, 2006.
9. 胡裕樹，*現代漢語*，三聯書局，1992.
10. *ACM Transactions on Asian Language Information Processing*.
11. *International Journal on Computer Processing of Oriental Languages*.