SUBJECT DESCRIPTION FORM

Subject Title: Fundamentals of Chinese Computing

Subject Code: COMP5412

Credit Value : 3

Pre-requisite: (Subject title and code no, if any) Nil

Recommended background knowledge:

Data Structures and Algorithms (COMP305) or equivalent

Mutual Exclusions: Fundamentals of Chinese Computing (COMP553)

Learning Approach:

Characteristics of Chinese language are taught before computer representation and processing of Chinese. For tutorials, exercises and computer software demonstrations will be given to consolidate understanding in class. At least one assignment will be given to develop a realistic Chinese computing application.

42 hours of class activities including - lecture, tutorial, lab, workshop seminar where applicable

Assessment:		
	Continuous Assessment	45%
	Test, and Examination	55%

Objectives:

To provide software engineers and managers with:

1. the basic understanding of Chinese computing;
2. the basic skill in operating within a Chinese computing environment;
3. the basic vocabulary for effective communication when computer representation and processing of Chinese is involved;
4. the basic programming skill for Chinese computing;
5. the basic skill to deal with algorithmic problems posed by the large Chinese alphabet;
6. the basic knowledge for the strategic design of software for the different language communities.

Learning Outcomes:

After completing this subject, students should be able to:

1. be aware of the different Chinese character sets and their encoding schemes;
2. identify and convert the character codes generated by one encoding scheme to another;

The Department reserves the right to update the syllabus contents. Please note that the learning approach for the same subject could vary slightly due to different delivery modes.

3. demonstrate the ability to design and implement internationalized software and localization components; and
4. design and implement effective input and output systems.

Keyword Syllabus:

Characteristics of Chinese Language
Historical Development, Geographical Variations (Dialects) and Linguistic Descriptions (Character, Morph, Word, Phrase and Sentence).

Representation of Chinese Character Sets
Mathematical description of representation, character set organization, encoding schemes (ISO2022 and UTF), discussion of (*defacto*) standard character sets (for PC, workstation and network), character decoding techniques and character conversion problems.

Output Processing of Chinese
Typesetting terminology and text rendering process, Bitmap fonts (representation, compression and scaling problems), Vector/Outline fonts (Limn algorithm), X-Window Fonts (BDF and Postscript) and font related operations (Installation, Specification, Extraction and User-Defined Glyph addition), Automatic Glyph Construction.

Input Processing of Chinese
Introduction to Chinese input processing by pen, image, speech and keystrokes, Shape-based keystroke input method, Phonetic-based input method, Shape-phonetic based input method, Evaluation metrics for keystroke input method, Phrase-based input method, Sentence-based input method, Microsoft Windows Input Method Architecture, X-Windows Input Method Architecture.

Software Development for Chinese Computing
Internationalization, Localization, ANSI-C model, Windows programming for Chinese Computing (Microsoft- and X-Windows).

Algorithmic Problems in Chinese Computing
Character set selection (NP-complete), Hashing functions for Chinese character sets, String searching (KMP, BM or Sunday), Dictionary lookup for Chinese computing, String-set searching (Aho and Corasick Algorithm).

Indicative reading list and references:

胡裕樹，現代漢語，三聯書局 1992

Berry, K. and Hargraves, K.A. *GNU Font utilities (Limn algorithm), on WWW*.

Hopcroft, J.E. and Ullman, J.D., 1979, *Introduction to Automata, Theory and Languages*, Addison-Wesley.

Huang, J.K.T. and Huang, T.D., 1989, *An Introduction to Chinese, Japanese and Korean Computing*, Singapore, World Scientific.

Jain, K.A., 1989, *Fundamentals of Digital Image Processing*, PHI.

Kano, N., 1995, *Developing International Software*, Microsoft Press.

Lunde, K., 1993, *Understanding Japanese Information Processing*, O Reilly & Associates.

McGilton, H. and Campione, M., 1993, *Postscript by Example*, Addison-Wesley.

The Unicode Consortium 1991-96. *The Unicode Standard, Version 2.0*, Addison-Wesley Developer Press.

Communications of COLIPS

Journal and Conference Proceedings of Computer Processing of Oriental Languages

Journal of Chinese Information Processing