

MASTER'S THESIS

Chinese character synthesis: towards universal Chinese information exchange

Yiu, Lai Kuen Candy

Date of Award:
2003

[Link to publication](#)

General rights

Copyright and intellectual property rights for the publications made accessible in HKBU Scholars are retained by the authors and/or other copyright owners. In addition to the restrictions prescribed by the Copyright Ordinance of Hong Kong, all users and readers must also observe the following terms of use:

- Users may download and print one copy of any publication from HKBU Scholars for the purpose of private study or research
- Users cannot further distribute the material or use it for any profit-making activity or commercial gain
- To share publications in HKBU Scholars with others, users are welcome to freely distribute the permanent URL assigned to the publication

Chinese Character Synthesis:
Towards universal Chinese information exchange

YIU Lai Kuen, Candy

A thesis submitted in partial fulfillment of the requirements
for the degree of
Master of Philosophy

Principal Supervisor: Dr. WONG Wai

Hong Kong Baptist University

October 2003

Abstract

A serious problem for Chinese information exchange in the Internet era is the sheer number of Chinese characters. Commonly used character encoding systems cannot include all characters, since they are not all catalogued; as a result, fonts cannot possibly contain all the characters either. In professional and scholarly documents, unencoded characters arise often. This situation hinders the development of Chinese information exchange: special care must be taken to handle unencoded characters using techniques such as embedding characters as images.

This thesis presents a systematic attempt to solve this problem. The approach is based on the intrinsic characteristic of Chinese characters that each character is formed by combining strokes and radicals. A Chinese character description language named *HanGlyph* has been defined, in order to capture the topological relation of the strokes in a character. *HanGlyph* descriptions are converted into graphical representations through the Chinese Character Synthesis System (CCSS), implemented in METAPOST. Preliminary results show that the approach is feasible and can be applied to many areas other than information exchange.

Contents

Declaration	i
Abstract	ii
Acknowledgments	iii
Table of Contents	iv
List of Tables	viii
List of Figures	ix
1 Introduction	1
I Chinese Character Description Language	
2 Properties of Chinese characters	5
2.1 History of Chinese characters	5
2.2 Composition of Chinese characters	6
3 Chinese character description language — <i>HanGlyph</i>	8
3.1 The Syntax of <i>HanGlyph</i>	11
3.2 Strokes	14
3.3 Glyph Expressions	23
3.3.1 Operators	23
3.3.2 Relations	25
3.4 The Operation: top-bottom	26

3.4.1	Dimension	27
3.4.2	Alignment	27
3.4.3	Touching	28
3.4.4	Scaling	29
3.5	The Operation: left-right	29
3.5.1	Dimension	29
3.5.2	Alignment	30
3.5.3	Touching	31
3.5.4	Scaling	31
3.6	The Operation: fully-enclosed	31
3.6.1	Dimension	32
3.6.2	Touching	32
3.6.3	Scaling	33
3.7	The Operation: partially-enclosed	33
3.7.1	Dimension	33
3.7.2	Touching	33
3.7.3	Scaling	34
3.8	The Operation: crossing	34
3.8.1	Dimension	35
3.8.2	Alignment	35
3.8.3	Scaling	36
3.9	Macro Definitions	38
3.10	Character Definitions	38

II Chinese Character Synthesis System

4	The Structure of CCSS	41
5	Translator	43
5.1	Translation strategy	44
5.2	Parser	45
5.3	METAPOST output generator	45
5.4	Pre-defined macro handler	47

5.5	Files handler and current counter	47
6	Chinese character synthesis library	49
6.1	Parallel operations	50
6.1.1	Spacing	50
6.1.2	Dimensions	53
6.1.3	Alignments	54
6.1.4	Touch	56
6.1.5	Scale	58
6.2	Enclosed operation	59
6.2.1	Estimate the enclosed area	59
6.2.2	Scale the inner element	61
6.2.3	Dimensions, alignments and touch	61
6.2.4	Scale	65
6.3	Crossing operation	65
6.3.1	Definition of intersections	67
6.3.2	Dimensions	68
6.3.3	Alignments	68
6.3.4	Crossing operation	69
6.3.5	Scale	71
7	Stroke Library	72
7.1	Points	72
7.2	Skeleton	74
7.3	Outline	76
7.4	Serif	77
7.5	Stroke family	77

III Results and Conclusion

8	Results	80
8.1	Current state of CCSS	80
8.1.1	Using CCSS in interactive mode	80
8.1.2	Using CCSS in batch mode	81

8.1.3	Getting help from CCSS	81
8.1.4	Output from CCSS and METAPOST	82
8.1.5	Running script automatically	82
8.2	Skeleton characters rendering	82
8.3	CCSS with L ^A T _E X	83
8.3.1	The implementation	86
8.3.2	Discussion	88
8.4	Different styles of rendering	88
8.4.1	Changing pen parameters	88
8.4.2	Changing the outline properties	90
8.5	Changing outline parameters	92
8.5.1	Examples	92
8.6	Discussion	92
9	Conclusion	95
A	Pre-defined <i>HanGlyph</i> Macros	96
	Curriculum Vitæ	101