A Chinese Character System Modeling and Development

by

Chen Xiaojian

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

Master of Science in Software Engineering

Faculty of Science and Technology
University of Macau

2013

Approved by _____

Supervisor

_____

_____

_____

Date _____

In presenting this thesis in partial fulfillment of the requirements for a Master's degree at the University of Macau, I agree that the Library and the Faculty of Science and Technology shall make its copies freely available for inspection. However, reproduction of this thesis for any purposes or by any means shall not be allowed without my written permission. Authorization is sought by contacting the author at

Address: PUTIAN, FUJIAN, CHINA

Telephone: (86) 13532291732
Fax:
E-mail: kevincxj@hotmail.com

Signature _____

Date _____

UNIVERSIDADE DE MACAU
仁義禮知信
澳門大學

University of Macau

Abstract

A CHINESE CHARACTER SYSTEM MODELING AND DEVELOPMENT

by Chen Xiaojian

Thesis Supervisor:
Associate Professor LI Xiaoshan
Master of Science in Software Engineering

Chinese character system is the most complicated character system in the world. There are over ten thousand general Chinese characters. The system has a large amount of information and knowledge, which includes Pinyin, structure, component, words, calligraphy and so forth. Therefore, fast searching from the massive information is a meaningful and challenging work.

In this thesis, we first analyze and design the system using UML, and then implement it in Java. A variety of information of Chinese characters can be searched in the system efficiently with multiple conditions. A standard XML schema is designed for documenting the whole information of 6763 general Chinese characters (GB2312-80 of the People's Republic China) by focusing on structural information. Based on the established character information database, a group of searching and browsing functions are implemented in the software system. In addition, a character synthesis algorithm is designed and implemented, so that people can compose a new Chinese character interactively by the components from a given set of Chinese characters.

Compared with the existing systems, our system is particularly suitable for searching and browsing the structural information of Chinese characters conveniently. The developed system would be helpful for people to analyze and study Chinese characters.

TABLE OF CONTENTS

LIST OF FIGURES iii

LIST OF TABLES v

LIST OF ABBREVIATIONS vi

CHAPTER 1: INTRODUCTION 1

1.1 BACKGROUND KNOWLEDGE 1

1.2 MOTIVATIONS 2

1.2.1 INTEGRATE THE WHOLE INFORMATION OF CHINESE CHARACTERS 2

1.2.2 HELP PEOPLE TO LEARN CHINESE CHARACTER 2

1.2.3 PROVIDE STATISTICAL ANALYSIS OF CHINESE CHARACTER 3

1.2.4 CHINESE CHARACTER SYNTHESIS BY EXISTING CHINESE CHARACTERS 3

1.3 OBJECTIVES 4

1.3.1 DEFINE CHINESE CHARACTERS 4

1.3.2 CREATE A DICTIONARY OF CHINESE CHARACTER 4

1.3.3 BUILD A PLATFORM FOR CHINESE CHARACTER SYNTHESIS 4

1.4 BASIC CONCEPTS OF CHINESE CHARACTER 4

1.5 SCOPE OF THE PROJECT 9

1.6 RELATED WORK 9

CHAPTER 2: SYSTEM ANALYSIS 11

2.1 FUNCTIONAL REQUIREMENTS 11

2.2 USE CASES 11

2.2.1 IDENTIFY ACTORS 11

2.2.2 IDENTIFY USE CASES 12

2.2.3 USE CASE DIAGRAM 12

2.3 CONCEPTUAL CLASS DIAGRAM 14

CHAPTER 3: SYSTEM DESIGN 17

3.1 THE XML-DEFINED CHINESE CHARACTER 17
3.2 CHINESE CHARACTER INFORMATION MANAGEMENT 21
3.2.1 ADDING NEW CHINESE CHARACTER INFORMATION 21
3.2.2 MODIFICATION OF CHINESE CHARACTER INFORMATION 23
3.2.3 REMOVEMENT OF CHINESE CHARACTER INFORMATION 29
3.3 BROWSING CHINESE CHARACTER INFORMATION 30
3.4 SYNTHESIS OF CHINESE CHARACTER 34
CHAPTER 4: IMPLEMENTATION 43
4.1 DETAILED TECHNOLOGY 43
4.1.1 SEARCHING CHINESE CHARACTER BY COMPONENT 43
4.1.2 QUERYING CHINESE CHARACTER WITH SIMILAR CHARACTERISTICS 46
4.1.3 THE TECHNOLOGIES FOR CHINESE CHARACTER SYNTHESIS 51
4.2 DEVELOPING TOOLS 63
4.2.1 ECLIPSE V3.2 63
4.2.2 MYSQL 64
CHAPTER 5: CONCLUSIONS 65
5.1 CONTRIBUTION 65
5.2 FUTURE WORK 66
BIBLIOGRAPHY 67

UNIVERSIDADE DE MACAU
仁義禮知信
澳門大學

LIST OF FIGURES

Number *Page*

Figure 1: Similarity among some Chinese characters....3
Figure 3: The calligraphic order of Chinese character “明”....9
Figure 2: An example of Chinese character “明”....9
Figure 4: System actors....12
Figure 5: Use case diagram for the system....13
Figure 6: Elaboration for searching Chinese character....13
Figure 7: Elaboration for Chinese character information management....14
Figure 8: Conceptual class diagram....15
Figure 9: Partial object diagram for Chinese character “强”....16
Figure 10: The shape of a Chinese character....19
Figure 11: Sequence diagram for adding new Chinese character....22
Figure 12: Chinese character information management through the toolbar....23
Figure 13: The dialog box for “Add New Character”....23
Figure 14: Sequence diagram for modifying Chinese character (1)....24
Figure 15: The dialog box for “Search Chinese Character”....25
Figure 16: Sequence diagram for modifying Chinese character (2)....26
Figure 17: The list of Chinese characters....26
Figure 18: The dialog box for “Modify Character”....27
Figure 19: Sequence diagram for modifying Chinese character (3)....28
Figure 20: Sequence diagram for removing Chinese character....29
Figure 21: The dialog box for “Remove Character”....30
Figure 22: Sequence diagram for searching Chinese character....31
Figure 23: The interface for searching Chinese character....31
Figure 24: The interface for listing the found Chinese characters....31

Figure 25: Sequence diagram for browsing Chinese character 32
Figure 26: The interface for browsing Chinese character.....33
Figure 27: The main interface for browsing Chinese character information.....34
Figure 28: Sequence diagram for Chinese character synthesis.....35
Figure 29: The initial interface for Chinese character synthesis.....36
Figure 30: An Input field for the set of Chinese characters.....36
Figure 31: The interface is displayed after the button is clicked 37
Figure 32: The picture of a Chinese character.....38
Figure 33: Extracting the component from the corresponding character 39
Figure 34: The whole process for extracting a component “亻” 39
Figure 35: The whole process for extracting a component “口” 40
Figure 36: The whole process for extracting a component “木” 40
Figure 37: The final version of the synthetic Chinese character 41
Figure 38: The interface after a node is double-clicked from the tree.....42
Figure 39: Searching Chinese character by component “子” 45
Figure 40: Search tree of component “子” 45
Figure 41: The location of each component in Chinese character “保” 47
Figure 42: The tree chart of Chinese character “呆” 50
Figure 43: The example of querying character with similar characteristics.....51
Figure 44: Comparison with a character “呆” and a synthetic character “保” 60
Figure 45: Comparison with a character “藻” and a synthetic character “保” 60
Figure 46: Comparison with a character “朵” and a synthetic character “保” 61
Figure 47： The set of Chinese character templates 61
Figure 48: The synthetic Chinese character “保” 62
Figure 49: The optimal solution for synthesis of Chinese character 62
Figure 50: lists the Chinese characters which contain “木” 63
Figure 51: The main interface of Eclipse V3.2.....64

LIST OF TABLES

Number *Page*

Table 1: The various kinds of radicals 6

Table 2: The various kinds of strokes 7

Table 3: The various kinds of structures 8

UNIVERSIDADE DE MACAU
仁義禮知信
澳門大學

LIST OF ABBREVIATIONS

XML. e**X**tensible **M**arkup **L**anguage

UML. **U**nified **M**odeling **L**anguage

AD. **A**nno **D**omini

GB. **G**uojia **B**iaozhun (国家标准)

IDE. **I**ntegrated **D**evelopment **E**nvironment

UNIVERSIDADE DE MACAU
仁義禮知信
澳門大學

ACKNOWLEDGMENTS

In my thesis research work, I have a lot to be thankful for. First of all, I especially would like to thank my supervisor Prof LI Xiaoshan for giving me so many helpful ideas and kindest care, I feel very lucky to be his student. Secondly, thanks to my partners Miss Yang Lijie and Mr. Li Wei who are always trying to help me and encourage me. Thirdly, I also would like to thank all the people at N105 who I met and worked with. Finally, I would like to thank the University of Macau, and thank all my teachers. The university provides me a good platform for development, and the teachers give me a careful cultivation.

Thank you, University of Macau. Thank you, my dear teachers. Thank you, my dear classmates.