

学校编码: 10384

分类号_____密级_____

学号: X2008221049

UDC_____

厦 门 大 学

工 程 硕 士 学 位 论 文

“毛体”字库的研发

"Mao Ti" Character Warehouse linkstone Is Sent Out

张 晋 晖

校内指导教师: 陈毅东 副教授

校外指导教师: 谢孟荣 教授

专 业 名 称: 计算机应用技术

论文提交日期: 2012 年 月

论文答辩时间: 2012 年 月

学位授予日期: 2012 年 月

答辩委员会主席: _____

评 阅 人: _____

2012 年 月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下，独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果，均在文中以适当方式明确标明，并符合法律规范和《厦门大学研究生学术活动规范（试行）》。

另外，该学位论文为（）课题（组）的研究成果，获得（）课题（组）经费或实验室的资助，在（）实验室完成。（请在以上括号内填写课题或课题组负责人或实验室名称，未有此项声明内容的，可以不作特别声明。）

声明人（签名）：

年 月 日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

1. 经厦门大学保密委员会审查核定的保密学位论文，于 年 月 日解密，解密后适用上述授权。

2. 不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年 月 日

摘要

字库，就是我们使用计算机是显示汉字的图像源。计算机调用字库显示汉字，不是直接调用相同的图像，而是调用这个汉字的内码，程序通过这个内码，再到相应的图像源（字库）当中寻找相应的图像信息，并画到屏幕上或者打印到纸上。所以字库的制作是中文信息处理的基础工作。

毛泽东同志是中国当代的书法巨匠,被公认为二十世纪最杰出的书法家之一。现在毛泽东的书法在很多地方用到，没有毛体字库，那么软件中就无法显示、打印“毛体”字库。“毛体”字库的研制，可以填补这一空白。

本论文中首先介绍了汉字的发展历程，目前汉字字库的现状、“毛体”字库研究的意义等。

其次，字库的制作过程也是对毛泽东书法深入了解的过程。在论文中对毛泽东书法的发展历程，形成的原因、毛泽东书法特点作了详细介绍。并说明“毛体”草书自成一体的原因。在整个说明过程中，配以了各种图片，以达到更好的效果。

字库的制作使用了 Font Creator，它是一个可以编辑只做 TrueType 字体的软件，在论文中对它的使用作了简要的介绍。

最后对毛泽东书法字体的收集、整理过程，对每个“毛体”的挑选和使用，以及“毛体”字库的设计过程做了说明，并说明了遇到的困难和字库存在的不足。

关键词：字库；毛泽东书法；信息处理

Abstract

Font, that is, we use the computer display of Chinese characters is the source of the image. Call computer font display Chinese characters, rather than a direct call to the same image, but the Chinese call the code, the procedure code, corresponding to the image source (Font), which look for the image information and pictures to the screen Or to print on paper. So the character is the production of Chinese information processing on the basis of the work.

Comrade Mao Zedong of China's contemporary masters of calligraphy, acknowledged as the twentieth century one of the most prominent calligrapher. Mao's calligraphy is now in use in many places, there is no body hair character, "Body hair" of the development of character, you can fill this gap.

First of all, this paper introduces the character of the development of Chinese character font of the current status quo, "body hair" character study of the meaning, and so on.

Second, the character of the production process is also an in-depth understanding of the calligraphy of Mao Zedong. In the paper on Mao's calligraphy course of development, the causes, characteristics of the calligraphy of Mao Zedong made a detailed presentation. Throughout that process, with a variety of images in order to achieve better results.

The font used in the production of Font Creator, it is a production editor of the TrueType font software, in its use of paper in a brief introduction.

Finally, Mao's calligraphy font collection, collation process, each of the "body hair" in the selection and use, as well as the "body hair" character of the design process described and illustrated the difficulties and shortcomings font.

Keywords: character; calligraphy of Mao Zedong; information processing

目 录

摘要	I
ABSTRACT	II
第一章 绪论	1
1.1 立项背景及研究意义	1
1.2 国内外研究发展状况及存在的问题	2
1.2.1 1980 年国标(GB)字库 6763 字	2
1.2.2 1983 年台湾 BIG5 字库 13461 字	4
1.2.3 1993 年 UCS 统一的中日韩汉字 20902 字	4
1.2.4 1995 年 GBK 1.0 字库 21886 字	5
1.2.5 2000 年 GBK2.0 字库 27533 字	6
1.3 本文的主要研究工作	8
1.4 本文组织结构	9
第二章 毛泽东的书法发展	10
2.1 发展阶段	10
2.1.1 第一阶段——青年时代	10
2.1.2 第二阶段——1921 年到 1949 年	11
2.1.3 第三阶段——1949 年进北京以后	12
2.2 “毛体”草书自成一体	13
2.3 “毛体”形成的原因	16
2.3.1 酷爱书法，有天分	16
2.3.2 勤学苦练，持之以恒	16
2.3.3 虚心学碑帖，博采众长	17
2.3.4 胸襟宽大，见识广博	17
2.3.5 善创造	17
2.4 毛泽东的书法艺术探微	19
2.4.1 行书	19
2.4.2 行草	22
2.4.3 草书	23
2.5 毛泽东书法的历史地位	25
2.5.1 书风	25
2.5.2 与古代书法的联系	26
2.5.3 书风形成的原因	27
2.6 小结	30

第三章 字体软件 FONT CREATOR 使用介绍.....	31
3.1 字体基本知识.....	31
3.1.1 光栅字体 (.FON)	31
3.1.2 矢量字体 (.FON)	32
3.1.3 Postscript 字体(.PFM)	32
3.1.4 TrueType 字体 (.TTF)	32
3.2 FONT CREATOR PROGRAM 的使用	32
3.2.1 准备工作.....	33
3.2.2 启动工作.....	33
3.2.3 制作过程.....	34
3.2.4 字体测试.....	35
3.2.5 后期处理.....	35
3.3 小结	36
第四章“毛体”字库的设计与实现.....	37
4.1 “毛体”字的搜集、整理	37
4.1.1 书法的收集.....	37
4.1.2 对字体的整理.....	37
4.2 “毛体”字的特点和选取	38
4.3 手写字与信息字的关系与区别	38
4.3.1 结构保持不变.....	38
4.3.2 结构比例的确定.....	38
4.3.3 比画保持原有风格.....	39
4.4 “毛体”制作过程	39
4.4.1 挑选单字.....	39
4.4.2 修字.....	39
4.4.3 字库制作.....	39
4.4.4 字体测试.....	40
4.5 “毛体”书法的特点	40
4.5.1 用笔.....	40
4.5.2 结构.....	40
4.5.3 章法.....	41
4.5.4 笔画.....	41
4.5.5 落款.....	41
4.6 字库的结构和特点	42
4.7 程序调用字库.....	42
4.8 结论.....	44
第五章 总结与展望	45

5.1 工作总结	45
5.2 对将来研究工作的展望	46
参考文献	47
致谢	48

厦门大学博硕士论文摘要库

TABLE OF CONTENTS

ABSTRACT IN CHINESE	I
ABSTRACT IN ENGLISH	II
CHAPTER 1 INTRODUCTION	1
1.1 BACKGROUND AND RESEARCH SIGNIFICANCE OF A PROJECT	1
1.2 RESEARCH STATUS AND PROBLEM	2
1.2.1 1980 NATIONAL STANDARD (GB) 3 CHARACTER 6763 WORDS.....	3
1.2.2 IN 1983 THE WORD TAIWAN BIG5 FONT 1346 1	4
1.2.3 UCS 1993, A UNIFIED CJK CHARACTERS 20902 WORDS	4
1.2.4 FONT 1995 GBK 1.0 6 21 886 WORDS	5
1.2.5 2000 GBK2.0 27,533 WORD FONT	6
1.3 THE MAIN RESEARCH PAPER	8
1.4 THIS ORGANIZATIONAL STRUCTURE	9
CHAPTER 2 ELOPMENT OF MAO ZEDONG'S	
CALLIGRAPHY	10
2.1 STAGE OF DEVELOPMENT	10
2.1.1 Phase I - 10 youth	10
2.1.2 Phase II - 1921 to 1949	11
2.1.3 Phase III - into Beijing in 1949 afte	12
2.2 "BODY HAIR" CURSIVE SELF-CONTAINED	13
2.3 "BODY HAIR" CURSIVE SELF-CONTAINED	16
2.3.1 passion for calligraphy and talented.....	16
2.3.2 hard work and practice, erseverance.....	17
2.3.3 humbly learn rubbings, absorbing.....	17
2.3.4 Large-minded, knowledgeable	17
2.3.5 create good.....	17
2.4 EXPLORATION MAO'S CALLIGRAPHY	19
2.4.1 running scrip.....	19
2.4.2 Cursive.....	22

2.4.3 cursiv	23
2.5 ALLIGRAPHY OF MAO ZEDONG'S HISTORICAL STATUS	25
2.5.1 Wind books	25
2.5.2 contact with ancient calligraphy	26
2.5.3 Wind causes the formation of the book.....	27
2.6 SUMMARY	30
CHAPTER 3 DESCRIBES THE FONT USED SOFTWARE FONT CREATOR	31
3.1 FONT BASICS	31
3.1.1 Raster font (. FON)	31
3.1.2 Vector fonts (. FON).....	32
3.1.3 Postscript fonts (. PFM)	32
3.1.4 TrueType 字体 (.TTF)	32
3.2 FONT CREATOR PROGRAM	32
3.2.1 TrueType fonts (. TTF).....	33
3.2.2 Start of work.....	33
3.2.3 Production process	34
3.2.4 font test.....	35
3.2.5 Post-processing	35
3.3 SUMMARY	36
CHAPTER 4 "BODY HAIR" CHARACTER DESIGN AND IMPLEMENTATION	37
4.1 "BODY HAIR" CHARACTERS TO COLLECT, COLLATE	37
4.1.1 calligraphy collection.....	37
4.1.2 font finishing.....	37
4.2"BODY HAIR" AND SELECT THE CHARACTERISTICS OF WORDS	38
4.3 handwriting and word relationships and differences between the inforMa -TION.....	38
4.3.1 The structure remains unchanged	38
4.3.2 Structure determination of the proportion of	38
4.3.3 gesture to keep the original style	39

4.4 "BODY HAIR" PRODUCTION PROCESS	39
4.4.1 The selection of words	39
4.4.2 repair word	39
4.4.3 Production font	39
4.4.4 font test	40
4.5 "'BODY HAIR" FEATURES CALLIGRAPHY	40
4.5.1 pen	40
4.5.2 Structure	40
4.5.3 tricks	41
4.5.4 strokes	41
4.5.5 inscription	41
4.6 THE STRUCTURE AND CHARACTERISTICS OF THE FONTS	42
4.7 PROGRAM CALLED FONT	42
4.8 CONCLUSION	44
CHAPTER 5 SUMMARY AND OUTLOOK.....	45
5.1 SUMMARY OF WORK	45
5.2 PAIRS IN THE FUTURE PROSPECTS FOR RESEARCH	46
REFERENCES.....	47
ACKNOWLEDGEMENTS	48

第一章 绪论

1.1 立项背景及研究意义

汉字是汉语书写的最基本单元，汉字从商周时代的甲骨文演变到当代的方块形，已经历了数千年的漫长进程，是世界上使用历史最悠久、使用人口最多的文字。汉字的数量是随着时间的推移而不断增加的。西汉的《仓颉篇》仅有 3300 字；东汉的《说文解字》收录了 9353 字，将汉字构造规律概括为“六书”：象形、指事、会意、形声、转注、假借。其中，象形、指事、会意、形声四项为造字原理，是“造字法”；而转注、假借则为用字规律，是“用字法”。；清朝的《康熙字典》突破了 47000 字；现在的《汉语大字典》所收汉字高达 54678 个。各地小范围流通使用的地名、人名用字还有许多，如果全部收集起来，总数大约有 9 万。

尽管汉字的数量如此庞大，但常用汉字的数量是有限的。孙中山先生写《三民主义》，总字数约 16 万，仅用了 2134 个不同的字；《毛泽东选集》一至四卷，总字数超过 66 万，也只用了 2981 个不同的字；小说《骆驼祥子》，总字数达 10 万多，用到的汉字数为 2413。郭沫若先生曾说过，汉字的数目大体上有五万多字的光景，这五万多字中绝大多数已经不使用了^[1]，目前一般知识分子日常所使用的大概有五、六千^[2]。也就是说绝大多数汉字仅在古籍资料中才能见到。

汉字是承载文化的重要工具，目前留有大量用汉字书写的典籍。不同的方言都使用汉字作为共同书写体系，因而汉字在历史上对中华文明的传播起到了重要作用，并成为东南亚文化圈形成的内在纽带。在汉字发展过程中，留下了大量诗词、对联等文化，并形成了独特的汉字书法艺术。

毛泽东书法艺术是毛泽东思想的一个重要组成部分，他的书法已经成为一个独特的书体，世人尊称为“毛体”。“毛体”的形成与毛泽东同志领导的中国革命是分不开的。他是一个伟大革命家的宽广胸怀、豪放性格、独特气质、渊博知识融入到自己的书法之中。“毛体”的诞生，充实了中华文明的精神宝库，在中国书法史上具有划时代的意义。毛泽东的书法作品深受广大群众的喜爱。半个多世纪以来，越来越多的爱好者开始研习毛体。广大毛体书法爱好者充分认识到毛

泽东是中国书法史上独树一帜的书法艺术大师。

但是随着毛体书法在国内的盛行，随着毛泽东书法的不断披露面世，随着国内拍卖行业的兴起，对毛泽东书法的作伪会越来越组织化、公开化，形势很严峻，作为毛体书法家和毛泽东书法研究的工作者，识别伪作，揭露假冒伪劣，维护毛泽东书法的权威与尊严，责任很大，担子很重。

所以利用计算机信息化技术，处理毛泽东的书法艺术，更好地识别伪作，揭露假冒伪劣，维护毛泽东书法的权威与尊严！

1. 在众多的汉字字库中，毛泽东毛笔书法字库相对比较少，而在日常工作中，“毛体”字用的非常多，没有毛体字库，那么软件中就无法显示、打印“毛体”字库。“毛体”字库的研制，可以填补这一空白。这对于以后研究毛泽东的各个方面，尤其是其书法艺术的研究有着重要的历史意义。

2. 毛泽东主席的书法个性突出，书法作品从点画造型、格式建构、笔法运用到气势态韵等散发着独特的艺术魅力。毛泽东的文学和书法举世公认，很多文章已经编入中、高等学校教科书。书法造诣已被评为近百年十大书法家之一。字库还是一个让书法家“存储”自己某一时期书艺成就的最好形式。如果能够把在书法上的艺术成就以字库的形式保存下来，不仅能够流传千古，也是对自己所取得的书法艺术成就的最好保护和交代。

3. 字库的制作过程也是对“毛体”书法的一个深入了解的过程。利用现代技术对“毛体”书法文字应用、发展和研究起到巨大的推动作用。对于研究毛泽东书法的结构和笔顺起到有益的作用，可以为后人研究毛泽东书法与计算机技术的结合提供借鉴。

1.2 国内外研究发展状况及存在的问题

近二十多年来，我国的中文信息交换和处理技术取得了突飞猛进的发展，从而使计算机的汉字库也变得越来越庞大，与此同时，也存在了不少问题。

1.2.1 1980年国标(GB)字库 6763字

从1975年开始，我国为了研究汉字的使用频度，进行了大规模的字频统计工作，在数以亿计的浩瀚文献资料中，统计出不同的汉字数为6335个，其中3000

多字的累计使用频度达到 99.9%，而另外的 3000 多字累计频度不到 0.1%，说明了常用汉字与次常用汉字的数量合起来不足 7000 个，这就为国家制定汉字库标准提供了依据。

1980 年，国家标准总局颁布了国家汉字信息交换用编码，全称《信息交换用汉字编码字符集——基本集》，国家标准号为：GB2312-80，自 1981 年 5 月 1 日起实施^[4]，通行于大陆。新加坡等地也使用此编码。

GB2312 收录 7445 个图形字符，其中汉字占 6763 个，还选入了 682 个非汉字字符，包含有数字、一般符号、拉丁字母、日本假名、希腊字母、俄文字母、拼音符号、注音字母等。

GB2312 规定“对任意一个图形字符都采用两个字节表示，每个字节均采用七位编码表示”，我们习惯上称第一个字节为“高字节”，第二个字节为“低字节”。

GB2312 将代码表分为 94 个区，对应第一字节；每个区 94 个位，对应第二字节。94 个区中，01-09 区为符号、数字区；16-87 区为汉字区；10-15 区、88-94 区是有待进一步标准化的空白区。

GB2312 将收录的汉字分成两级：第一级是常用汉字，计 3755 个，置于 16-55 区，按汉语拼音字母/笔形顺序排列；第二级汉字是次常用汉字，计 3008 个，置于 56-87 区，按部首/笔画顺序排列。

GB-2312 的实施，保证和满足了一般汉字处理技术的基本要求，从而奠定了中文信息处理的基础。它对于初创阶段电脑中文的应用和发展，功不可没。但经过实践，也存在了不少问题：

1. 收字量过小。目前学术界认为汉字总数在六万左右，而它仅收了十分之一。该标准制定的主要依据是 1956 年公布的《汉字简化方案》和当时报刊的字频统计，它显然不符合 21 世纪今天的需要，也不能全面表达三千年文明史留下的大量文献，更不会为汉字文化圈其他国家和民族所认同。
2. 收字不合理。基本集中有不少稀见字，如“厓、茺、廋”等，在本已狭小的编码区间中白白占据了宝贵的位置。
3. 两级汉字库各收汉字 3000 有余，但其排列次序一级为拼音，二级为偏旁，并未采取一以贯之的排列方式，造成了人为的困难和混乱。

4. 在技术处理上, 现已发现的差错有 80 处之多。当时, 我国大陆各种中文 DOS 版本、Windows 3.1/3.2 版本, 装入的字库都是国标一二级字库。正由于当时对汉字的实际需要量认识不足, 在使用 GB-2312 的过程中常带来不少尴尬和混乱。如遇到“镨、啰、瞭、祔、祔、罌、贇、贇、蠹、麤、犇……”等汉字, 既无法输入, 又不能打印。特别是自从实行个人存款账户实名制后, 便经常碰到打不出来的“失名户”。在高校招生实行计算机录取后, 有相当一部分学生的名字在计算机里成了“黑三角”, 影响了录取工作的进程。同样的现象也出现在户籍管理和测绘、古籍、辞书整理等方面。产生这些现象的根本原因是计算机中所装的 6 千多个汉字不够用。

1.2.2 1983 年台湾 BIG5 字库 13461 字

1983 年 10 月, 台湾国家科学委员会、教育部国语推行委员会、中央标准局、行政院共同制定了《通用汉字标准交换码》, 后经修订于 1992 年 5 月公布, 更名为《中文标准交换码》, BIG5 是台湾资讯工业策进会根据以上标准制定的编码方案, 它通行于台湾、香港地区, 是一个繁体字编码方案, 虽存有一些瑕疵, 但广泛应用于电脑行业, 尤其是互联网中, 从而成为一种事实上的行业标准。

BIG5 码是双字节编码方案, 共收录 13461 个汉字和符号, 包括:

△符号 408 个。

△常用字 5401 个, 包括台湾教育部颁布的《常用国字标准字体表》的全部汉字 4808 个, 台湾教科书常用字 587 个, 异体字 6 个。

△次常用字 7652 个, 包括台湾教育部颁布的《次常用国字标准字体表》的全部汉字 6341 个, 《罕用国字标准字体表》中使用频率较高的字 1311 个。

1.2.3 1993 年 UCS 统一的中日韩汉字 20902 字

随着汉字全面向信息化社会转型, 随着计算机在各个领域的广泛应用, 随着国际间的交流与合作的扩大, 信息处理应用对字符集提出了多文种、大字量、多用途的要求。1993 年, 国际标准化组织 ISO 发布了一个编码标准, 即 Universal Multiple-Octet Coded Character Set (简称 UCS), 大陆译为《通用多八位编码字符集》, 台湾译为《广用多八位元编码字元集》, 它与 Unicode 编码完全兼容。ISO/IEC 10646.1-1993 是该标准的第一部分《体系结构与基本多文种平面》。

嗣后, 国家技术监督局发布了 GB13000.1-93 国家标准(1993 年 12 月 24 日),

对 UCS 以国家标准的形式予以认可，即 GB13000.1 等同于 ISO 10646.1。该标准采用了全新的多文种编码体系，收录了 20902 个汉字，使编码空间扩大。同时还编制了《统一的中日韩汉字》，又称“CJK 统一汉字”（C 指中国，J 指日本，K 指韩国）。而其中的中国部分，包括了源自中国大陆的 GB2312、GB12345、《现代汉语通用字表》等法定标准的汉字和符号，以及源自台湾的 CNS11643 标准中第 1、2 字面（基本等同于 BIG-5 编码）和第 14 字面的汉字和符号。

新标准的诞生，是电脑中文取得突破性进展的里程碑。它不仅对我国计算机的发展产生深远的影响，也对汉字文化圈各国、各地区的文化交流和科技进步起到重大的作用，尤其在对汉字的视听识别和人工智能方面做出了贡献。它促进了汉字文化圈计算机信息的直接交换，也促进了软件的更新以及数据量的扩大，特别是沟通海峡两岸的计算机技术，有益于祖国的统一。

但是，新标准的建立，也使人们引起某些误解和疑虑，如：

1. 由于收入了许多《汉字简化方案》所规定的“废弃”字，产生了标准和方案相互矛盾的假象，往往被人们误以为是在变相地恢复繁体字。我们说，标准只是为了使用、交换和生产的便捷而约定的规则，它允许某些灵活性、宽容性、人为的约定性和使用的变通性，所以它不涉及对《汉字简化方案》的否定和对繁体字的肯定。

2. 近十年来，我国在编码输入法的开发上可以说是百花争艳、万“码”奔腾，但绝大多数都是建立在 GB-2312 基础上的。新标准字库建立后，自然需要新输入法的研究。众所周知，字库的增大，意味着编制输入码的难度要有平方和立方乘积的增长。

1.2.4 1995 年 GBK 1.0 字库 21886 字

为了配合 UNICODE 的实施，全国信息技术标准化技术委员会于 1995 年 12 月 1 日制订了《汉字内码扩展规范》(GBK)，英文名称 Chinese Internal Code Specification。国家技术监督局标准化司、电子工业部科技与质量监督司于 1995 年 12 月 15 日联合以技监标[1995]229 号文件的形式，将它确定为技术规范指导性文件，并予以发布和实施。这一版的 GBK 规范为 1.0 版。GB 即“国标”，K 是“扩展”的汉语拼音第一个字母。GBK 向下与 GB2312 编码兼容，向上支持 ISO 10646.1 国际标准，是前者向后者过渡过程中的一个承上启下的标准。

Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.

厦门大学博硕士论文摘要库