中文信息处理基础教程

余锦凤　萧志青　编著

北京大学出版社

中文信息处理基础教程

余锦凤　萧志春　　编著

北京大学出版社
北　京

内 容 简 介

《中文信息处理基础教程》一书阐明了中文信息处理的意义和范围、汉字的各种属性、汉字的特色、汉字的嵌套结构理论、各种汉字编码方案的设计原则和设计目标、"汉语拼音双拼方案"、"汉语全拼和双拼统一方案"的设计方法、优化"形码设计"步骤和一种形音结合的编码方案;给出了一种易学快速的形码——完备三码;给出了可处理国际标准要求的2万多个汉字的新的数字编码输入技术——佳法码;归纳出书写汉字的笔顺原则;阐明了一部分汉字编码基础理论,这些理论有利于人们识别各种编码方案的优劣;介绍了汉字输出的必需设备——汉字库(点阵汉字库、方正矢量字库、TrueType曲线字库)。其所包含的教学内容——嵌套字素编码输入法、佳法码输入法和嵌套微型曲线字库等三项技术于2000年5月18日获第三届清华创业计划大赛的"技术创新"、"最新创意"两项优胜奖,即是对本书先进性、实用性的一个新的实际证明。此中,嵌套字素编码法已获中国发明专利,1987年参加"中华杯中文电脑公开赛",以"后期速率第一名"获金杯奖,1997年通过国家教委主持的专家认定。本书从理论上循序渐进地引导读者深入了解并探求汉字计算机键盘输入的最佳方法。可以预料,在本世纪逐渐升温的国际性"汉语热"中,本书将对国内外学生学习利用计算机处理汉字,有效地提高汉语、汉字方面的素质,产生积极作用。本书附有光盘,适用作中文信息处理、中文信息管理、文秘等有关专业的教材,也适用于需要了解、应用、研究中文输入、输出技术的人士。

图书在版编目(CIP)数据

中文信息处理基础教程/余锦凤等编著.-北京:北京大学出版社,2002.7

ISBN 7-301-05665-6

Ⅰ.中… Ⅱ.余… Ⅲ.汉字信息处理-教材 Ⅳ.TP391

中国版本图书馆CIP数据核字(2002)第041750号

书　　名:中文信息处理基础教程

著作责任者:余锦凤等

责任编辑:沈承凤

标准书号:ISBN 7-301-05665-6/TP·0660

出 版 者:北京大学出版社

地　　址:北京海淀区中关村北京大学校内　100871

网　　址:http://cbs.pku.edu.cn

电　　话:出版部62752015　发行部62754140　编辑部62752038

电子信箱:zpup@pup.pku.edu.cn

排 版 者:北京因温特有限公司

印 刷 者:北京大学印刷厂

发 行 者:北京大学出版社

经 销 者:新华书店

787毫米×1092毫米　16开本　15印张　374千字

2002年6月第1版　2002年6月第1次印刷

定　　价:20.00元

前　言

当此信息化时代，世界各国日趋广泛地需要在电子计算机及各种电子产品、通信设备上处理汉字信息，中文信息处理技术已经居于不容忽视的地位。2000年3月份，国家颁布强制执行含有两万多字的ISO-IEC 10646国际标准，拼音、音形输入方法及社会上流行的各种汉字输入法都受到了新的挑战，人们希望有一种易学好用高效的汉字输入法来适应这个国际标准字符集。

美国的微软公司、IBM公司分别在中国成立了研究院和研究中心，其主要研究方向就是中文信息处理。移动电话、手持机(包括PDA、汽车电脑)、可佩戴式计算机和一些掌上型电子设备的键盘仅有数字键，因而又产生了以数字为基础的汉字输入方法这一新的需求热点。美国、日本及我国的香港特区正在着力于发展一种具有输入功能的双向寻呼机，必然面临中文信息处理问题。应该如何高效率地利用电子字典查找汉字和单词，也是中文信息处理亟待解决的问题。在计算机网络方面，有关高效中文搜索引擎、中文电子邮件、中文电子商务等与中文信息处理相关的技术有很好的前瞻性。

在此高新科学技术大潮的推动下，中文的计算机处理研究正需要开辟一条熔铸古今、贯通中西的道路。它既不能囫囵吞枣泥古不化，也不能邯郸学步崇洋不悟，更不可趋时媚俗沦为商品经济的附庸或拜金主义的奴隶，而应该立足于优化新世纪新时代学人的人品、学风、知识结构和思维品格。《中文信息处理基础教程》一书的宗旨就是希望以跨学科、跨国界的新视野，采用恰当的教学手段、启迪青年学子的思路，逐渐创制出全新的中文计算机文化。汉字的计算机处理是其基础，汉字的键盘输入处理则是基础之基础。虽然，市面上已经出现了一些语音输入产品，但从客观情况看来，目前都还处于实验室研究阶段，离真正实用还有一段差距。在手写汉字识别领域，也已有了一些产品，然而，还有漫长的攻关之路要走。我们大家热切地期盼各种具有真正实用性的处理中文信息技术的成功，但在今后数十年内，汉字键盘输入方法仍会是处于主导地位的输入技术。

20世纪70年代以来，出现过的微机键盘输入方案约达数百种，它们在时间的长河中经过大浪淘沙的筛选，继而能适应操作系统的更换(由早期的CCDOS直至当前流行的WINDOWS)，能为人们提供处理GB 2312-80所含6000多汉字的输入方法只剩下十来种；能用以大致处理ISO-IEC 10646国际标准字符集第一期工程所含20902个汉字的输入方法只有三四种；而真正做到每字只需三码就能快速、准确地处理ISO-IEC 10646国际标准字符集所规定的全部27484个汉字的输入方法就只有《嵌套字素编码输入法》了。嵌套字素编码输入法具有独到的符合科学原理的汉字结构理论。其制定的设计要求之一是“不用切换控制键也能贯通输入汉字、数字、外文、标点及科技符号”。这不仅使嵌套字素编码法成为久经磨砺、愈用愈好的一种字形结构输入方法，而且又派生出一种数字编码法——佳法码和一种形声码。佳者，一方面言其学、用俱佳，另一方面，寓意于谐音“加”，“佳法码”，亦即“加法码”是也。在移动电话成为世界发展的一种流行趋势，拉丁语系和其它拼音语系文字也都面临着要用数字编码的时代，采用佳法码输入竟然显示出汉字比其它文字更为简捷、快速、准确，而且不用附加键盘的突出的优越性，表现出汉字具有坚韧、灵活、适应性极强的生命力。其制定的另一设计思想是“所选用的字素经简单变换能直接生成汉字字形发生器”，它造就了一种矢量字库的专利技术——《嵌套字素编

码的笔画、点阵混合式汉字发生器》和另一种嵌套微型曲线汉字库。嵌套微型曲线汉字库以不到2MB的存储单元能容纳宋、仿宋、楷、黑四种字体，共8万多汉字，所用存储单元仅为WINDOWS系统中提供的字库的1/16。

中文信息处理就是上述技术发展背景下应运而生的一门新兴的多边缘学科，一些大学正在开设“中文信息处理专业”，从而产生的需要就是相应的教材。北京大学教务部和教材科果断地将《中文信息处理基础教程》定为2000年的立项教材，这一举措正好适应了教育领域即将出现的教材需要热点。本书具有文理交融的特点。第一章简单介绍了中文信息处理的意义和范围；第二章阐明了汉字的各种属性；第三章摘要介绍了汉字的特色，；第四章提出了汉字的嵌套结构理论；第五章归纳出汉字的笔画及其使用情况和汉字的笔顺原则；第六章简介了人们必须了解的二进制运算及其与十进制之间相互转换原理、信息交换用汉字编码集ISO-IEC 10646中汉字编码与十六进制的关系；第七章介绍了汉字编码中使用的一些术语并且阐明了一部分汉字编码基础理论，这些理论有利于人们识别各编码方案的优劣；第八章综合了各种汉字编码方案的设计原则、设计目标，提出了设计“汉语拼音双拼方案”、“汉语全拼和双拼统一方案”的方法，并给出了优化“形码设计”步骤和一种形音结合的编码方案；第九、第十两章阐明了“嵌套字素编码输入法”是如何设计的，以及如何使用它；第十一章给出了一种易学快速的形码——完备三码；第十二章给出了能处理国际标准要求的2万多个汉字的新颖、易学、好用的数字编码输入技术——佳法码（它适用于台式电脑、PDA、移动电话、双向BP机、电子辞典、电子翻译器、电子记事本、信息家电、其它手持终端等微型设备）；第十三章介绍了汉字输出的必需设备——汉字库，重点介绍了点阵汉字库、方正矢量字库，顺便介绍了TrueType曲线字库。以上内容只是中文信息处理的基础部分，其所包含的教学内容——嵌套字素编码输入法、佳法码输入法和嵌套码微型曲线字库等三项技术于2000年5月18日获第三届清华创业计划大赛的“技术创新”、“最新创意”两项优胜奖。其中，嵌套字素编码法已获中国的发明专利，1987年参加“中华杯中文电脑公开赛”，以“后期速率第一名”获金杯奖，1997年通过国家教委主持的专家认定并获得与会全体专家的一致推荐。

《中文信息处理基础教程》是我和萧忠义、萧志春通力合作的结果，它凝聚了我们三位编著者在中文信息处理领域廿多年的基础研究经验及本人在北京大学信息管理系开设此门课程十多年的教学经验。一方面从理论上引导读者循序渐进地深入了解并思索寻求汉字计算机键盘输入的最佳方法。同时，也指导学习者快速掌握两种值得学习的非常实用的键盘输入方法。可以预料，在本世纪渐趋升温的国际性“汉语热”中，本书将对国内外学生学习利用计算机处理汉字，有效地提高汉语、汉字方面的素质，产生积极作用，可作为中文信息处理、中文信息管理、文秘等范畴有关专业的教材，也适用于需要了解、应用、研究中文输入、输出技术的人士。

本书作为此类教材的“引玉”之砖，错误、疏漏、不足之处在所难免，敬请各方人士赐教、斧正。

本书附有光盘，光盘内容有：(1)嵌套字素编码法输入系统；(2)嵌套字素编码法教学幻灯片；(3)本书第十四章全部内容。若需购买光盘，请与北京大学出版社电子出版部联系。

联系方式：

电话：(010)62757513，62757146　　传真：(010)62757513　　邮编：100871

地址：北京 中关村成府路205号　　E-mail：wy@pup.pku.edu.cn

余锦凤

2001年10月于北京大学

目　录

第一章　中文信息处理简介

1.1　中文信息处理的意义

信息一词，其内涵宽泛，外延无限，至今尚未有一公认的统一的定义。就广义而言，举凡来自人类的生存环境和与生存环境有关的天体宇宙、地球上的山川河流、矿物等各方面的自然现象、一切生物（包括人）的生态（包括死亡后存留的遗骸）现象、人类的社会现象（政治、经济、军事、文化、商业、科学技术及工农业生产、生活活动）所产生的各种状态和消息都含有信息。所以，信息的含意丰富，而且可以有数据、文字、声音、图形、图像等多种多样的表现形式，称之为信息的多元化表示。用计算机处理多元化信息，属于信息处理技术的范畴。

信息是客观存在，若根据需要去正确地利用信息，信息就会产生相应的价值。要使信息产生广泛的社会价值就需要传递。传递和保存信息都需要处理技术。在电信技术发明以前，人们只能用人工通信，或者用其他简单的表示方式或各种约定来传递信息。电子通信技术的发展，从电话电报开始，直到传真、电视，从有线通信发展到无线通信，直到微波、光纤通信、卫星通信，信息的传递速率大大提高。20世纪40年代发明了电子计算机，用于处理数值运算。由于信息之多、信息之复杂，且要求处理信息快而准确，所以对信息进行的加工处理必然离不开计算机技术，信息处理这一术语就自然而然地隐含了计算机技术。随着软件技术的发展进步，“数据”逐渐用以表示广义的信息，从而发展了数据信息处理的应用技术。利用计算机处理数据信息，除了作信息传输外，主要是对信息按某种规律或作某种意义的加工，使它适应某种特定目标的需要。例如，气象预报的信息处理就是结合信息传感技术，对采集到的原始数据按照所设计的数学模型进行处理，得出的结果用作气象预报的资料。因此，用计算机处理或加工信息扩大了信息的利用范围，使信息的利用价值大大提高。计算机信息技术日益成为现代社会的科技进步、经济发展、人类文明进程所不可缺少的东西。它和物质、能源一起被视为现代人类社会生存和发展的三大要素，形成了蒸蒸日上的信息产业。

应用计算机处理多元化信息，属信息处理技术范畴。一方面，微型机及其相关产品的普及应用为信息处理技术的实用化提供了基础。另一方面，软件技术飞速发展，不仅使数据和文字信息处理技术更加完善，而且开拓了信息处理技术的更新的应用领域，比如模式识别、语音识别及语音合成、自然语言处理、语言的翻译等技术领域。计算机指纹识别技术在刑侦破案、取代锁和钥匙、作为存取财物的有效凭证等方面得到了应用。

计算机还具有利用数据通信技术实现的计算机网络通信功能。传统的信息处理是局限于信息的存储与检索，是狭义的信息处理；传统的通信技术是以传输模拟信号为主，只须完成信息的传输或转移。经计算机存储和处理的信息可以在两台或多台计算机或数据处理设备之间、两地或多地之间互相传输，更加增强了信息处理技术的效能，扩展了信息处理技术和通信技术的内容，使信息处理技术和通信技术结合起来，形成了广义的信息处理技术，即兼有信息处理与信息传输功能的计算机通信技术。

在多元化的信息中，文字信息是一种最普遍的表示形式。例如，文件、信函、报表、记录、印刷品等基本上采用文字表达的形式。

“中文信息处理”一词是从20世纪70年代流行起来的，实际上，自古以来，中文信息处理工作源远流长。可以说，自从有了中文(汉字)，即相应地出现了中文信息处理的工作。从开始编制第一部汉字字典和编写第一篇文摘起就开始了中文信息的分析与综合处理的研究。然而，现代人们言及的“中文信息处理”包括了有关中文信息的采集、存储、传输和利用，是指利用电子计算机和现代通信、照相、排版等自动化技术对汉字信息进行输入输出整理、加工、转换、传输、复制等各种处理的一项新兴的科学技术。其交叉性使之成为“信息科学”的分支；其综合性应用使之成为“系统工程”的一个实例。它涉及到语言文字学、计算机科学、信息科学、工程心理学、数理统计学、声学、自动识别技术、人工智能、网络技术、文献检索学等等。故可以说它是一门新兴的多边缘科学。中国要实施先进的信息处理技术手段，中文信息化是一项重要的资源开发工作。中文信息网已逐渐成为我国现代化社会的神经系统，它将促进人民文化和社会生产效率迅速提高。中文信息处理工程已建立起现代化中文语言文字信息系统，使凝聚在语言文字中的知识信息发挥更大效能，使汉语汉字得到最佳利用。

计算机中文信息处理技术从70年代至今，历经20多年，完成了由初级阶段向比较成熟阶段的过渡，这是微电子技术和IT技术高速发展以及迫切的应用需求所促成的。

现在，许多移动电话都已具备中文菜单和显示中文短信息功能，但都有缺陷，还不是真正意义上的“全中文”。只有当它既能显示中文又能输入和处理中文，也就是说，能直接利用手机进行中文输入时，才可以说是“全中文”。然而，一般移动电话仅有数字键，这无疑对汉字数字输入法(简称数字码)提出了很迫切也是很高的要求。顺便说一下，在WAP技术成为新的热点之时，连英文也面临着需要编码输入的严峻事实。

当前，美国、日本及我国香港特区都在大力发展一种双向寻呼机，它同时具有输入功能，即，它同样也面临着中文处理问题。还有电子字典，如何高效、规范化地利用电子字典查找汉字和单词，也是中文信息处理应该解决的问题。

信息家电是一个热门话题，它也面临着中文信息处理的问题。另外，从计算机本身的发展来看，手持机(包括PDA和汽车电脑)和可佩戴式计算机的中文信息处理尚有诸多问题需要解决。可佩戴式计算机还处于发展初期，其应用领域广泛，尤其在军事上有很大的用途，面临新军事革命的挑战，我国在研究其相应设备时，首先遇到的就是中文信息处理问题。

微软和IBM公司分别在中国成立了研究院和研究中心，广揽人才，其主要研究方向就是中文信息处理。

在计算机网络方面，中文信息处理将具有更加广阔的前景。高效的中文搜索引擎、电子邮件、中文电子商务等技术均与中文信息处理密切相关。移动电话、信息终端等电子设备对以数字为基础的计算机汉字输入方法的需求又成为研究领域的新热点。在语音识别汉字输入方面，硬件的进一步微型化、连续语音识别、噪声背景下的语音识别以及汉语口语理解等都是亟待解决的难点。手写汉字识别技术方面，联机状态下的笔写入方式，通常的麻烦就是字与字间书写的停顿时间不易控制，写得慢了，多部首的组合汉字被分了家，造成错字；写得快了，或字与字间的停顿太短，会将两个单字拼凑成一个字，又成了错字。

尽管有调整改变手写速度“快速、中速、慢速”等技术措施，实用中却使人感到频繁换用鼠标时的不便乃至产生厌烦情绪而不愿使用了。非特定的脱机手写汉字识别的困难则更多。

目前仍处于实验室研究阶段，尚未进入真正实用状态，还有许多棘手难题需要逐步解决。因此，在今后数十年内，中文键盘输入方法仍会是处于主导地位的输入技术。

2000年3月份，国家颁布强制执行两万多字的ISO-IEC 10646国际标准后，纯拼音或音形混合输入方法以及社会上流行的一些汉字输入法都受到了严峻挑战，人们期待着新的易学好用高效的输入方法早日产生，本书就是要为中文信息处理技术往纵深发展打下一定的基础。

1.2 中文信息处理涉及的范围

信息的表示形式是多元化的。文字信息是大多数信息表示形式的基础，而文字信息处理则是基础的基础。中文信息处理包含中文文字信息处理、中文文献信息处理以及中文的各种管理系统和服务性系统。

利用计算机解决汉字的信息处理问题是20世纪中期以来的事，它包含有输入、存储、处理、传送、输出等环节。下面着重介绍输入和输出两个环节。

1.2.1 汉字的输入技术

1. 单字、词汇和语句的键盘输入

(1)专用型的中键盘或大键盘整字输入方式

①大键盘：一键一字输入方式。

②中键盘：一键多字输入方式。

(2)通用小键盘

①拼音方式：利用字音编码输入。

·汉语拼音方式：全拼音方式，例创 chuàng

双拼方式，例创 uug(ch→u，àng→g)

·注音字母方式：创ㄔㄨㄤ

注音字母：

ㄅㄆㄇㄈㄉㄊㄋㄌㄍㄎㄏㄐㄑㄒㄓㄔㄕㄖㄗㄘㄙㄧㄨㄩㄚㄛㄜㄝㄞㄟㄠㄡㄢㄣㄤㄥㄦ

②拼形方式：利用字形特征编码输入。

·笔画笔形式

·偏旁部首式

·字形结构式

③混合式。

·音形混合：以音为主，以形为辅

·形音混合：以形为主，以音为辅

2. 手写输入方式

3. 语音输入方式

4. 扫描方式

5. 传真方式

1.2.2 汉字的输出技术

1. 汉字的输出有多种方式

(1)屏幕显示:显像管显示器、液晶显示器;

(2)打印机:针打式、喷墨式、激光打印;

(3)语音输出;

(4)绘图仪;

(5)传真机。

2. 汉字输入输出所必需的汉字库

计算机系统中存储汉字字形信息的字库,字库分为三种类型:

(1)点阵字库;

(2)矢量字库;

(3)曲线字库:整字轮廓字库、压缩字库。

1.2.3 中文信息处理基础理论方面的研究内容

(1)汉字识别(包括印刷体字、限制性手写体字及一般手写体字);

(2)汉语语音识别(包括语音波形编码和解码、语音的分解与合成);

(3)汉语自然语言的理解和处理;

(4)汉语的机器翻译;

(5)中文文献的自动勘误、自动标引和自动编文摘;

(6)汉字的单字、词汇使用频度的研究;

(7)汉语的语词、语法、语料库研究;

(8)中文信息处理应用平台研究;

(9)汉字编码理论研究;

(10)汉字编码方法研究;

(11)汉字编码方案评测标准研究。

1.2.4 中文文献信息处理工作内容

(1)利用各种编辑软件进行编辑排版。

(2)利用制表软件编制各种表格。

(3)利用数据库软件建立各种各样的文献信息数据库及其他各种应用软件系统,例如:

①研制各种类型图书馆或文献服务中心的集成式管理系统、检索系统;

②档案部门的集成式管理系统、检索系统;

③出版社、书店的集成式管理系统、检索系统;

④各种书刊文献、档案的自动分类系统、自动编文摘系统或其他的智能式文献处理系统。

1.2.5 应用中文的各种管理系统和服务性系统

国家各部门、厂矿企业、银行、医院、酒店的管理系统,专家系统,信息咨询检索系统,电化教学系统,远程教育系统,电子印刷排版系统,办公自动化系统,翻译系统,通信系统,财会系

统,售票系统,咨询服务系统,电话系统等等,多不胜数。随着计算机信息处理应用范围的扩大,中文信息处理技术还将逐步深入和提高。

思 考 题

1. 中文信息处理的意义。
2. 汉字的输入技术有几种方式。
3. 中文文献信息处理工作内容。

第二章　汉字属性

所谓属性是指事物本身所固有的性质、特点，包括状态、动作、关系等。如，人是地球上的主宰力量，也是人类社会必不可少的组成部分，因此，人是地球上人类社会的主要属性。又如，大千世界都处于运动之中，因此，运动就是物质的根本属性。对汉字而言，其属性当指汉字本身所具有的特性。它包括有形、音、义、量、体、频、序等方面。前五个方面表示汉字的个体特征，后二个方面表示汉字间的相互关系状况。

2.1　汉字字形

字形，是指字的外形轮廓和内部结构组合而成的集合体。汉字的形状特征就其外部轮廓而言，是方块状。细察汉字的内部结构，可以看到其令人惊叹的万千变化，正是这种丰富生动的结构变化而形成了各个不同的单字，致使人们能够见到的正式出版的汉字字典所收纳的单字量达数万种之多。

就外表看，大体上可以说篆书是横短竖长的长方形；隶书是横长竖短的扁方形；楷体则是横竖相等的正方形。当然，如果仔细分辨，即使是在楷体中，也并非所有的字全是正方形，还有一些汉字的形状是别的状态，如，外部轮廓是圆形的字有：○、  、 等；整字为复合图形的字有：  、  、  等等；呈菱形的字就有：十、个、小、令、子、卡等；呈横长竖短的矩形的字有：一、二、曰、四等；呈横短竖长的矩形的字有：日、目、月等；呈正三角形的字有：人、厶、众、品等；呈倒三角形的字有：丅、丁、了、丫等。据此，应该指出，要区分每个字的形状，必须要区分字的内外结构才行，不能只考虑外部轮廓形状，还要考虑其内部结构才能确定其形状。例如以“口”为轮廓的字就有数百个之多，只有细分其内部结构才能真正把这数百个汉字的形状区分出来。

2.2　汉字字体

汉字字体是指汉字的各种不同形体或汉字书法中各种流派和风格。如，印刷用的字体就有：宋体、仿宋体、黑体、楷体、圆体、魏碑、行书、草书、篆书和隶书等。又如汉字书法流派中有：钟(繇)体、王(羲之)体、欧(阳询)体、颜(真卿)体、柳(公权)体等。

汉字是由各种笔画组成的，而笔画又是由点、横、直、撇、折、捺、提、钩等组成的。除“点”外，其他笔画仅由无粗细变化的线条组成时，则称在先秦使用的这种字体叫篆书。篆书分为大篆和小篆，大篆泛指殷商时甲骨文、金文，通行于春秋战国时秦国的籀文和通行六国的文字。而小篆是在秦始皇统一中国后，采用李斯的建议，推行统一文字政策，以籀文为基础淘汰通行于六国的异体字形成的，它形体匀圆齐整。秦朝对汉字进行了规范化，使得汉字笔画发展成既有粗细变化，又有笔锋、笔头和方向的变化，使汉字字体进入到开始使用隶书的阶段。也就是说，隶书始于秦代，真正普遍使用则在汉魏时期。秦代程邈将隶人(指胥吏)在抄写小篆时应用的一种简易书写体加以搜集整理，故后世便有程邈创隶书之说。在南北朝以后，楷体便普遍流

行使用了。行书和草书也在同一时期得到相应的发展。在北魏时期产生了一种碑志造像等刻石文字，即魏碑体。魏碑体笔势舒畅流利而又苍劲坚实、奇峰险峻，结构跌宕起伏，笔意朴拙且存隶书遗风，。到宋代，产生了用于印刷字的宋体。直到1916年左右，钱塘的丁辅之、丁善之等集宋代刻本字样仿刻了一种活字字体，称之为仿宋体。到20世纪中叶，又出现了黑体、圆体、综艺体等印刷用字体。其字例及其特征列于表2.1中。

表 2.1

字体	字　例	特　征	应用范围
报宋	文化的结晶	笔画纤细，字体清秀工整，印刷效果清晰明快	适用于报纸、杂志的正文
书宋	文化的结晶	字体端正清秀，结构均匀，笔法严谨，美观实用	适用于书刊、杂志的正文
宋三	文化的结晶	横轻竖重，笔画均匀，字体严谨，排印清晰	适用于书刊、杂志、宣传品的正文
小标宋	**文化的结晶**	字型端正，结构匀称，笔画横细竖粗，布局严谨、稳重。	书、报、杂志的大小标题字及说明用字
大标宋	**文化的结晶**	字体端庄严谨，笔画粗细分明，布局凝重沉稳	适用于书、报、杂志的各类标题字
宋黑	**文化的结晶**	宋体字的结构中，溶入黑体粗壮稳重的特点，活泼美观，适用性广	适用于书、报、杂志的各类标题字
仿宋	文化的结晶	笔画粗细均匀，字型俊秀挺拔，布局严谨，错落有致	书、报、杂志及古籍、诗词等的正文和标题
楷体	文化的结晶	字体朴实端正，笔法舒展有力，流畅自然，结构匀称	适用于书、报、杂志和各级教材的中小标题及正文
细等线	文化的结晶	笔画纤细，字型方正，结构均匀，排列整齐	适用于书刊、杂志的正文及地图、广告用字
中等线	文化的结晶	笔画粗细均匀，字型端正典雅，构体清晰	适用于书、报、杂志的中小标题及绘图、制表、广告用字
黑体	**文化的结晶**	笔画均匀，字型端正古朴，构体平稳充实	适用于书、报、杂志的各类标题

续表

字体	字　例	特　征	应用范围
大黑	**文化的结晶**	字体粗重平稳，横竖比例一致，构体庄重，引人注目	适用于报纸大标题及书籍、画报的美术装帧和广告用字
细圆	文化的结晶	笔画圆顺舒展，结构秀逸婉转，字体匀称美观	适用于书、报、杂志的正文及装帧、广告等宣传用字
准圆	**文化的结晶**	字型圆润舒展，笔法婉转柔和，结构古朴美观	适用于书、报、杂志的各类标题字及装饰用字
粗圆	**文化的结晶**	字型圆润饱满，笔法舒展柔和，美观大方	适用于书、报、杂志的标题及装饰用字
琥珀	**文化的结晶**	字型圆润饱满，新颖活泼，结构错落有致，粗而不重，肥而不臃	适用于书、报、杂志和各类印刷品的标题及装饰用字
综艺	**文化的结晶**	字型见方，结构饱满，笔法新颖雅致，具有独特艺术效果	适用于书、报、杂志及宣传印刷的标题或装饰用字
水柱	**文化的结晶**	意在笔先，笔断意连，造型独特，具有流动感	书、报、杂志的标题及正文字，或装饰用字
姚体	文化的结晶	字体隽秀工整，结构狭长，字间距较大，排印清晰，整齐划一	适用于书籍、报刊、杂志的标题字
隶变	文化的结晶	字型微扁，笔画舒展，字体秀逸平和，古朴庄重	适用于书、报、杂志的标题及正文，也可作装饰用字
隶书	文化的结晶	字体浑厚饱满，婉转流畅，笔法古朴典雅，韵味深长	适用于书、报、杂志的各类标题字和装饰、宣传用字
魏碑	文化的结晶	字体苍劲坚实，结构跌宕起伏，笔意朴拙，不避锋芒	适用于书籍、报刊、、杂志的标题字
行楷	文化的结晶	字体飘逸洒脱，行笔流畅自然，刚劲舒展，书法韵味浓	适用于书籍、报刊、杂志的标题字和宣传、装饰用字

续表

字体	字　例	特　征	应用范围
中楷	文化的結晶	字体端正严谨，布局均匀，笔法流畅自然，排印效果好	适于书、报、杂志和各类教材的标题及于文字
幼线	文化的結晶	笔画纤细均匀，字形方整，构体自然，张驰有序	适用于书、报、杂志的正文
黑一	文化的结晶	字体工整，结构匀称，排印效果清晰明快	适用于书、报、杂志的中小标题及标记、说明、广告用字
小黑	文化的结晶	横平竖直，字型端正，笔画均匀，结构沉稳充实	适用于书、报、杂志的中小标题字
超粗黑	**文化的结晶**	字型方正饱满，笔画粗重坚实，庄重醒目，号召力强	适用于报纸及书籍、画报的标题及宣传用字
彩云	文化的結晶	字体活泼，珠圆玉润，结构错落，交叠有序，新颖美观	适用于书、报、杂志和各类印刷品的标题及装饰用字
新秀丽	文化的結晶	字体清秀美观，结构端正均匀，印刷效果好	适于报刊、杂志的正文

2.3　汉字字量

汉字字量就是汉字的数量。汉字到底有多少个？到现在为止，恐怕谁也说不清楚，道不明白，无人能给出一个确定值。汉字在历史的长河中不断地产生，由于历朝历代都没有设立管理文字使用的机构，当人们使用已有文字表达不出自己的意思时，便会造出一些新的字来。不同行业的人都可根据本行业的需要造字，造出来的字又可用不同手段进行加工改造，为记录同一事件造出不同字形的字是常有的事，从而产生了很多同音同义异形的异体字。还有的人为了减少汉字认读、书写之苦而造出了不少简化字，其中一部分简化字因考虑不周引起语义问题而被政府部门废弃。可是这些字却仍然被一些编字典的人搜集起来放进字典中去了，于是出现了汉字的总数量在某一历史阶段陡然增加的现象。

汉字字量在我国各个朝代所编纂的字典中基本上能反映当时的情况，但并不可能把所有的字都收进去。西安有一种小吃叫“  面”。其中的“  ”字迄今为止未被任何一本正式出版的字典所收录。表 2.2 列出我国历代有代表性的字典所收集的汉字字量。