

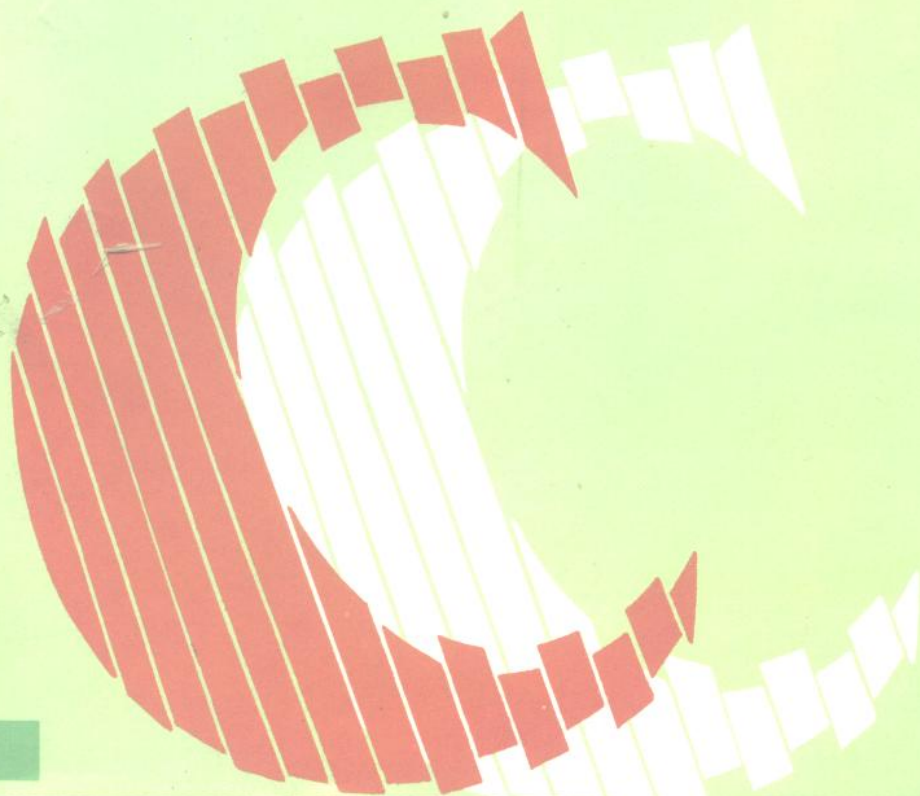
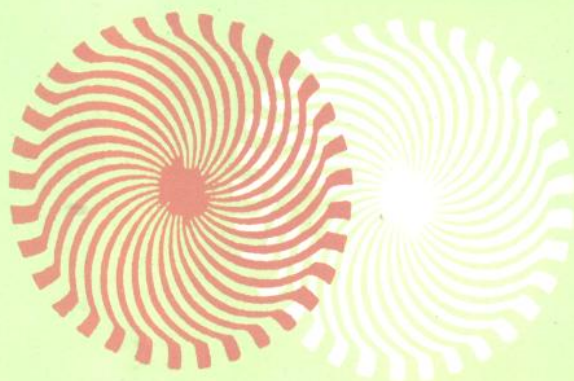
职业学校电子类教材(计算机技术专业)

# 汉字录入 与编辑技术

●(第二版)

●王式杰 邢小茹 编著

●电子工业出版社



Computer

中国计算机教育出版集团 中国计算机教育出版中心

# 汉字录入 与编辑技术

第二版

李人俊 王中强 主编

中国计算机教育出版中心



职业学校电子类教材(计算机技术专业)

# 汉字录入与编辑技术(第二版)

王式杰 邢小茹 编著

電子工業出版社  
Publishing House of Electronics Industry

## 内 容 简 介

本书内容建立在学生已经学习和掌握了西文 DOS 操作系统的基础之上。全书共分五章,第一章介绍了汉字信息处理的基本概念;第二章简单介绍了汉字操作系统的知识;第三章对最基本最流行的汉字输入方法做了讲解;第四章是汉字编辑软件 WPS 的使用方法;第五章是汉字编辑软件 CCED 的使用方法。书中所依据的软件都选用了目前最新的版本。为了方便读者学习和使用,本书在每一章后增加了部分习题,供教师和学生选用。习题多来自作者多年教学实践,也参照了社会各种计算机考试的试卷。本书编写时,在考虑教材体系的前提下,融汇了劳动部、国家教委相应级别考试的要求和内容,使读者在完成本课程学习的同时,也能顺利通过相关的考试。

本书适用于各级各类中等职业教育学校的学生在计算机基础课学习中使用。

JS416/18

丛 书 名:职业学校电子类教材(计算机技术专业)

书 名:汉字录入与编辑技术(第二版)

编 著 者:王式杰 邢小茹

责任编辑:吕 迈

排版制作:电子工业出版社计算机排版室排版

印 刷 者:中国科学院印刷厂

出版发行:电子工业出版社出版、发行 URL:<http://www.phei.com.cn>

北京市海淀区万寿路 173 信箱 邮编 100036 教材发行部电话:66708597

经 销:各地新华书店经销

开 本:787×1092 1/16 印张:10 字数:256 千字

版 次:1997 年 8 月第一版 1998 年 9 月第 8 次印刷

书 号:ISBN 7-5053-4020-4  
G·315

定 价:11.00 元

凡购买电子工业出版社的图书,如有缺页、倒页、脱页者,本社发行部负责调换

版权所有·翻印必究

# 出版说明

职业技术教育是提高全民文化技术素质的重要手段，是经济发展的重要支柱。在现有教育体系下，为适应社会生产力发展的水平，必须大力发展职业技术教育事业，有效地对社会新成长的劳动力和全体职工进行职业技能培训，以培养跨世纪的职业技术人才。

受国家教委的委托，电子工业出版社负责规划、组织出版全国中等职业技术学校实用电子技术和计算机技术两个专业的教材。电子工业出版社以电子工业为背景，以本行业的科技力量为依托，与教学第一线的优秀教师相结合，充分反映我国电子工业发展的实际情况，到目前已规划并组织出版两个专业的教材共60余个品种。这些教材深受广大职业技术学校师生的好评，为促进和大力发展我国的职业教育做出了积极的努力。

市场经济的发展要求职业学校进一步打破陈规，培养有创造性的、能适应新技术发展的、有实际技能的中级技术人才。根据《国务院关于大力发展职业技术教育的决定》、《职业教育法》及劳动部颁发的《职业技能鉴定规范》，电子工业出版社会同电子行业的专家学者、全国部分重点职业技术学校的领导和教师，在深入调查研究的基础上，制定了两个专业的指导性教学计划。教学计划强调了技能培养，并充分考虑到各个学校课程设置，师资力量，教学条件的差异，突出了“宽基础多模块，大菜单小模块”灵活办学的精神，为两个专业教材的配套与完善奠定了基础。

新一轮职业技术学校教材的出版有三个特点。

一、针对原有教材在内容、实用性等方面的不足，对原有教材进行了修编，一方面补充了新知识，一方面完善了内容和体系的衔接。

二、适应电子技术、特别是计算机技术日新月异的发展，与职业技术学校开设的新专业课相配套，推出了一批新教材。电子技术专业突出数字化、集成化技术；计算机技术专业内容涉及多种流行软件及实用技术。

三、教材的编写一改过去又深又厚的模式，突出“小模块”薄而精的特点，为不同学校依据自己的师资力量和办学条件灵活地选择不同的专业模块组合提供方便，同时也减轻了学生负担。

党中央、国务院提出实施科教兴国战略和可持续发展战略，教育体制改革必将进一步深化，加之科学技术的迅猛发展，编写职业技术学校教材始终是一个新课题。希望全国各地职业技术学校的广大师生多提宝贵意见，帮助我们紧跟职业技术教育和科学技术的发展不断地提高教材的编写质量，以便更好地为广大师生服务。

## 全国职业高中电子类教材工作领导小组

### 组长:

姚志清 (中国电子工业总公司教育局副局长)

### 副组长: (以下按姓氏笔划为序)

孙金兰 (北京市教育局职教办副主任)

李 群 (黑龙江省教委职教处处长)

李步斗 (江苏省教委职教处处长)

赵家鹏 (机电部电子类专业教材办主任)

褚家蒙 (四川省教委职教处副处长)

### 成员:

王仲伦 (甘肃省教委职教处副处长)

刘志平 (北京市职教中心教研员)

苏 丹 (新疆维吾尔自治区教委副主任)

张兆松 (山东省教委职教处副处长)

李宏栋 (天津市教育教研室职教室主任)

李启源 (广西壮族自治区教委职教处副处长)

何肃波 (吉林省教委中职处副处长)

张荫生 (上海市中等职业教育中心副校长)

何雪涛 (浙江省教委职教处主任科员)

杨玉民 (北京市教育局副局长)

林春赞 (湖北省教委职教处处长)

费爱伦 (上海市教育局中职处副处长)

梁 义 (辽宁省教委中职处副处长)

葛玉刚 (河北省教委职教处处长)

韩学理 (陕西省教育科学研究所副所长)

翟汝直 (河南省教委职教研究室主任)

### 秘书长:

邓又强 (电子工业出版社副总编)

### 副秘书长:

王玉国 (电子工业出版社编辑)

# 全国职业高中电子类教材编审委员会

## 主任委员:

杨玉民 (北京市教育局副局长)

## 副主任委员: (以下按姓氏笔划为序)

刘志平 (北京市职教中心教研员)

张萌生 (上海市中等职业教育中心副校长)

## [实用电子技术编审组]

### 组长:

刘志平 (北京市职教中心教研员)

### 副组长:

李蕴强 (天津市教育教研室教研员)

陈其纯 (江苏省苏州市电子职业中学教研组长)

张晓明 (黑龙江省教育学院职教部教研员)

### 组员:

白春章 (辽宁省教育学院职教部教研员)

朱大海 (河北省教科所研究室主任)

孙介福 (四川省教科所职教室主任)

刘洪志 (河南省新乡市机电部22所职高教师)

沈大林 (北京市宣武职教中心副校长)

陈先铭 (广西壮族自治区柳州市一职高教研组长)

吴恒丰 (湖北省武汉市第一职教中心教务主任)

杜德昌 (山东省教学研究室教研员)

张志强 (甘肃省武威职业学校校长)

周金波 (河南省郑州市教委职业教研室副主任)

金国砥 (浙江省杭州市红星职业中学教研组长)

杨荫彪 (河北省河北机电学校电子系主任)

俞兰浦 (上海市静安职业学校校长)

徐洪吉 (吉林省吉林大学机关学校教师)

## [计算机编审组]

### 组长:

张萌生 (上海市中等职业教育中心副校长)

### 副组长:

王森 (河北省军械工程学院计算所副教授)

王道生 (辽宁省沈阳工学院计算机系副教授)

史建军 (山东省青岛市教育局教研员)

**组员：**

王世学 （黑龙江省哈尔滨市职业学校教师）  
刘永振 （吉林省吉林大学计算中心副教授）  
刘逢勤 （河南省郑州市第三职业中专教研组长）  
肖金立 （天津市电子计算机职业中专教师）  
陈文华 （浙江省温州市职业技术学校教研组长）  
严振国 （江苏省无锡电子职业中学教务副主任）  
吴清萍 （北京市财经学校副校长）  
钟 葆 （上海市中等职业教育中心 OA 教研组长）  
戚文正 （湖北省武汉市第一职教中心教师）

## 前 言

1993年电子工业出版社出版了《汉字录入与编辑技术》，由于计算机技术的迅猛发展，原书内容需要更新，因此我们修编写了该书，以适应形势的发展。

考虑到目前计算机实际应用现状，我们侧重讲解了UCDOS系统，而对其它汉字系统只做了简单的介绍。由于本课程是建立在学生已经学习和掌握了西文DOS操作系统的基础之上，所以书中对DOS系统中常用的命令不再涉及，需要时就直接引用有关的知识。现在流行的汉字输入方法较多，原书用了四章的篇幅进行阐述，本书将其精简为一章，只对最基本最流行的方法做了讲解。本书删去了原书中汉字WORDSTAR，保留了原书中的WPS软件，并加以充实，另外增加了对CCED软件的介绍。为了方便课堂教学，本书在每一章后增加了部分习题，供教师和学生选用。

目前一个学习计算机的热潮正在社会中兴起，出现了多种计算机的考试和鉴定，本书编写时，在考虑教材体系的前提下，融汇了劳动部、国家教委相应级别考试的要求和内容，使学生在完成本课程学习的同时，也能顺利通过相关的考试。

本书的内容都具有极强的实践性，无论是对计算机专业，还是对非计算机专业的学生，本书所教授的知识在未来的工作岗位上都是最基本的必备知识；本书所要求的技能都是最起码的技能，掌握了这些技能就可以胜任相当多的工作，每一个学习计算机的学生都应该十分熟练的掌握它们。因此，适当加大机上的操作量是必要的手段，不如此不足以达到熟练的地步。

建议本课程至少使用108课时，按每学期18周计算，周课时为6课时；也可以按周课时3课时安排，进行两个学期。一般说来，相对集中使用课时，教学效果会好一些，其中机上的操作练习不得少于一半时间。

本书一至四章由北京宣武职业教育中心学校王式杰老师编写，第五章由石家庄市第二职业中专邢小茹老师编写。由于时间的关系，加之作者水平有限，书中错误之处在所难免，敬请广大读者在使用过程中提出批评指正。

编 者

1997年3月

# 目 录

|                      |      |
|----------------------|------|
| <b>第一章 计算机汉字处理概述</b> | (1)  |
| 第一节 汉字的特点            | (1)  |
| 第二节 计算机汉字处理的概念       | (2)  |
| 第三节 计算机汉字处理的发展概况     | (2)  |
| 第四节 汉字字库             | (3)  |
| 第五节 计算机的汉字外部码        | (5)  |
| 第六节 计算机的汉字内部码        | (6)  |
| 第七节 计算机的汉字交换码        | (6)  |
| 第八节 计算机的汉字输出码        | (6)  |
| 习题                   | (7)  |
| <b>第二章 汉字操作系统</b>    | (9)  |
| 第一节 汉字操作系统简述         | (9)  |
| 第二节 UCDOS 系统         | (10) |
| 一、UCDOS 简介           | (10) |
| 二、UCDOS 的系统组成        | (11) |
| 三、UCDOS 的系统配置        | (12) |
| 四、UCDOS 的启动和退出       | (13) |
| 五、UCDOS 的功能键         | (14) |
| 习题                   | (16) |
| <b>第三章 汉字输入方法</b>    | (18) |
| 第一节 汉字输入方法的分类        | (18) |
| 第二节 区位法汉字输入          | (18) |
| 一、区位与汉字区位输入法         | (19) |
| 二、区位法的特点             | (19) |
| 第三节 拼音法汉字输入          | (19) |
| 一、智能全拼法              | (20) |
| 二、双拼法                | (24) |
| 第四节 五笔字型法汉字输入        | (26) |
| 一、汉字的三个层次            | (27) |
| 二、基本字根及其键位           | (27) |
| 三、汉字的拆分原则            | (31) |
| 四、单字的输入              | (32) |
| 五、词语的输入              | (34) |
| 六、简码                 | (35) |
| 七、学习键                | (36) |
| 第五节 自然码法汉字输入         | (36) |

|                             |      |
|-----------------------------|------|
| 一、自然码输入法的启动 .....           | (37) |
| 二、自然码的编码规则 .....            | (39) |
| 三、单字的输入 .....               | (40) |
| 四、词语的输入 .....               | (43) |
| 习题 .....                    | (45) |
| <b>第四章 汉字编辑软件 WPS</b> ..... | (53) |
| 第一节 WPS 系统简介 .....          | (53) |
| 一、WPS 的系统文件 .....           | (53) |
| 二、WPS 系统的安装 .....           | (54) |
| 三、WPS 的启动 .....             | (54) |
| 四、WPS 的主菜单 .....            | (55) |
| 五、WPS 的命令菜单 .....           | (56) |
| 第二节 WPS 的基本编辑操作 .....       | (58) |
| 一、光标的移动 .....               | (58) |
| 二、字符的插入与删除 .....            | (58) |
| 三、段落的划分与合并 .....            | (61) |
| 四、段落的重排 .....               | (61) |
| 五、存盘与退出 .....               | (62) |
| 第三节 WPS 的字块操作 .....         | (62) |
| 一、字块的建立 .....               | (62) |
| 二、字块的移动 .....               | (63) |
| 三、字块的复制 .....               | (64) |
| 四、字块的删除 .....               | (64) |
| 五、字块的取消 .....               | (64) |
| 六、字块的存盘 .....               | (64) |
| 七、字块的读取 .....               | (65) |
| 第四节 文件的查找与替换 .....          | (65) |
| 一、查找字符串 .....               | (65) |
| 二、替换操作 .....                | (66) |
| 第五节 WPS 的版面控制 .....         | (67) |
| 一、边界的设定 .....               | (67) |
| 二、行距和字距 .....               | (68) |
| 三、字符的升高 .....               | (69) |
| 四、字符的后退 .....               | (69) |
| 五、分栏 .....                  | (69) |
| 第六节 WPS 的制表方法 .....         | (70) |
| 一、规则表与不规则表 .....            | (70) |
| 二、自动制表 .....                | (71) |
| 三、手动制表 .....                | (72) |
| 四、取消连线 .....                | (73) |

|                                |      |
|--------------------------------|------|
| 五、制表连线 .....                   | (73) |
| 六、标尺行 .....                    | (74) |
| 第七节 WPS 的打印控制 .....            | (74) |
| 一、汉字的字体 .....                  | (74) |
| 二、汉字的字号 .....                  | (76) |
| 三、西文字体和字号 .....                | (77) |
| 四、修饰 .....                     | (77) |
| 五、划线 .....                     | (79) |
| 六、前景与背景 .....                  | (79) |
| 七、阴影 .....                     | (80) |
| 八、模拟显示与打印输出 .....              | (80) |
| 第八节 WPS 的其它操作 .....            | (82) |
| 一、窗口的操作 .....                  | (82) |
| 二、计算器操作 .....                  | (83) |
| 第九节 WPS 主菜单中的其它功能 .....        | (84) |
| 一、编辑非文书文件 .....                | (84) |
| 二、打印文书文件 .....                 | (84) |
| 三、文件服务功能 .....                 | (84) |
| 四、帮助信息 .....                   | (84) |
| 习题 .....                       | (85) |
| <b>第五章 汉字字表处理软件 CCED</b> ..... | (89) |
| 第一节 CCED 简介 .....              | (89) |
| 一、CCED 的安装和启动 .....            | (89) |
| 二、CCED 的编辑屏幕 .....             | (90) |
| 三、CCED 的菜单 .....               | (91) |
| 四、CCED 磁盘目录的操作 .....           | (91) |
| 五、CCED 的存盘、退出 .....            | (93) |
| 第二节 CCED 的基本编辑 .....           | (94) |
| 一、光标的移动 .....                  | (94) |
| 二、文件中的文字修改 .....               | (95) |
| 三、文件中格式的编辑和整理 .....            | (95) |
| 第三节 查找与替换 .....                | (96) |
| 一、查询操作 .....                   | (97) |
| 二、替换操作 .....                   | (97) |
| 三、文件的特殊查询 .....                | (98) |
| 第四节 块操作 .....                  | (98) |
| 一、块的分类与定义 .....                | (98) |
| 二、块的移动 .....                   | (99) |
| 三、块的复制 .....                   | (99) |
| 四、块的删除 .....                   | (99) |

|                               |       |
|-------------------------------|-------|
| 五、块的打印 .....                  | (100) |
| 第五节 文件的特殊编辑 .....             | (100) |
| 一、新闻格式的文件 .....               | (100) |
| 二、窗口操作 .....                  | (102) |
| 三、文件的分页功能 .....               | (103) |
| 四、一些特殊的编辑操作 .....             | (103) |
| 第六节 CCED 的表格处理 .....          | (105) |
| 一、规则表格的生成 .....               | (105) |
| 二、手动制表 .....                  | (105) |
| 三、表格的修改与调整 .....              | (106) |
| 四、表内数据处理 .....                | (107) |
| 五、将已有的报表生成数据库 .....           | (108) |
| 第七节 CCED 的数值计算 .....          | (109) |
| 一、编辑屏幕上的数值计算 .....            | (109) |
| 二、表格内的数据计算 .....              | (109) |
| 三、多表间的数据计算 .....              | (112) |
| 第八节 模拟显示与打印输出 .....           | (113) |
| 一、打印控制主菜单 .....               | (113) |
| 二、打印字体与字号 .....               | (114) |
| 三、打印字符的修饰 .....               | (116) |
| 四、划线、前景、背景、阴影与灰度 .....        | (116) |
| 五、打印版面的控制 .....               | (118) |
| 六、打印预览 .....                  | (119) |
| 七、打印输出 .....                  | (120) |
| 第九节 CCED 对数据库报表的输出 .....      | (120) |
| 一、报表的分类 .....                 | (120) |
| 二、几个基本文件 .....                | (121) |
| 三、报表文件的生成 .....               | (121) |
| 四、报表的输出 .....                 | (124) |
| 习题 .....                      | (124) |
| 附录 A 汉语拼音方案 .....             | (131) |
| 附录 B 五笔字型二级简码表 .....          | (132) |
| 附录 C 自然码简码词表 .....            | (133) |
| 附录 D WPS 菜单操作及命令表 .....       | (135) |
| 附录 E WPS 命令索引表 .....          | (136) |
| 附录 F CCED 5.03 菜单操作及命令表 ..... | (140) |
| 附录 G CCED 5.03 编辑命令表 .....    | (141) |
| 附录 H CCED 5.03 打印控制命令表 .....  | (146) |

# 第一章 计算机汉字处理概述

自1946年世界上第一台电子计算机ENIAC问世以来,至今已经过去了半个世纪。这五十年中,计算机工业及计算机技术以无以伦比的速度,得到了迅猛的发展,成为现代化社会的一个重要支柱。能够熟练地掌握和使用计算机,也成了现代人的一个重要标志。

中国要进行经济建设,要把我国建设成为一个现代化的社会主义强国,就必须在各行各业中使用计算机这一强有力的工具。普及计算机知识、掌握计算机技术,是发展的需要,也是社会的现实。

但是毕竟计算机是外国人发明的,他们使用的是拼音文字,而我们的母语是汉语,在历史上、文化上都有巨大的反差。要使计算机在我们的社会中发挥作用,就必须通过“计算机的汉字处理”这一关口。“学习国外的先进经验,走自己的路。”经过中国计算机工作者的多年努力,这一难题已经得到了较好的解决。许多行之有效的方法已经在实践中得到检验,得到推广。随着人们对信息认识的日益深化和计算机应用领域的不断扩大,汉字信息处理已经成为计算机学科的一个重要分支,目前在我国的生、产、科、研、管、理、生、活等各个方面,正在大面积地普及和应用计算机。学习和使用计算机,尤其是用计算机进行汉字信息处理,已经成为许多人的迫切要求,这个热潮正以燎原之势燃遍全国。

目前,各行业的计算机应用,大量表现为文字处理,其中又以汉字处理为主要工作内容。这一章中我们主要讲解在汉字计算机处理中有关的一些基本概念,以使读者对这一工作的范畴及过程有个总体的了解。

## 第一节 汉字的特点

我国的汉字,自殷商时代的甲骨文开始,大约已有三千多年的历史。在漫长的时间里发展成为今天的样子。汉字的文字量很大,最新的字典中有汉字六万多个,常用字大约也有几千个,这是一般外文不能相比的,因此,汉字的这一特点给汉字的计算机处理带来了极大的困难,这是第一个特点。

第二,汉字的字形复杂,它是一种二维的方块字,每个汉字又可以分为笔画、字根、整字等几个层次,与西文拼音文字截然不同,而计算机使用的恰是英文键盘,这又是一个难以解决的矛盾。更何况汉字中还有许多异体字存在,不但使得汉字的字量增大许多,而且使汉字的书写更加复杂化。

第三,汉字有多种字体,为了适应不同的需求,就要备有多种不同字形的汉字库,这就造成了计算机花销过大的情况。

第四,汉字的发音呈多样化形态,一字多音与多字同音的现象屡见不鲜,加之为数众多的方言和民族语言,造成汉字输入编码不易解决重码的问题。

第五,汉字的字义复杂,同一汉字在不同情况下使用,可以代表多种含义,因此,单独使用汉字的字义作为输入方法是不可行的。

第六,汉字排序难,英文只有26个字母,排序很容易,而汉字数量大,结构复杂,同音字多,

不能很好解决排序问题,就会造成计算机检索的困难。

上述这些情况,都给在我国大面积普及计算机的工作造成了几乎难以逾越的困难。但是我们的文字仍是世界上优秀的文字,是世界上用最少的符号表达多含义的文字。经过计算机工作者多年的努力,上述问题基本上得到了解决。形形色色的汉字操作系统、汉字输入方法、汉字处理软件相继问世,为计算机在我国的推广应用打下了基础。

## 第二节 计算机汉字处理的概念

在现代社会中,信息量在急剧地增加,而信息的表现形式则是多样的,文字信息就是一种最通用、最普遍的信息表现形式。无论是公文、文件、信函、报表,还是各种印刷出版物,大多数都使用文字的形式来记录信息。文字信息处理的应用范围非常广泛,从编辑文稿、建立文件档案资料、排版印刷,到行政管理、办公自动化,以至翻译、情报检索、通讯等等。可以说,凡需用文字表达信息的场所,都可以使用文字信息处理技术。文字信息处理的实质就是把文字信息数字化。就是用一个固定的数码代表一个字母或文字,这一数码就叫做“代码”。在计算机内部处理文字信息时,就可以象处理纯数字一样来进行。需要输出时,再把替代的数码还原成相应的字母或文字。也就是说,汉字信息处理就是计算机对汉字代码的数据处理,使得在计算机上使用汉字和使用英文一样的方便。

我们所讲的计算机文字处理的内容包括以下三个方面:

文字信息的输入(主要指各种输入方法)

文字信息的处理(主要指各种编辑方法)

文字信息的输出(主要指显示和打印的方法)

目前,计算机的技术水平、生产规模和应用程度已经成为衡量一个国家现代化水平的显著标志。能否有效地对自己本民族的文字进行计算机处理,关系到一个国家和民族的兴亡。

汉字信息处理就是对文字信息中的汉字信息进行处理的过程。现在在世界上使用汉字的人数越来越多,而我国是汉字的发源地,研究和开发汉字信息处理技术,是我国计算机工作者无可推卸的责任,学习计算机汉字处理技术也是我国学生义不容辞的义务。

## 第三节 计算机汉字处理的发展概况

在七十年代中期,我国开始系统地研究和开发汉字信息处理技术,当时使用的中文信息处理方法还比较简单,数字处理和汉字信息处理是分别进行的。编制一个“汉字驱动程序”,需要输入或输出汉字时,临时调用这个程序,执行输入、输出汉字信息的任务,用户程序与汉字处理程序未经操作系统统一调度运行,效率低,使用繁琐。八十年代初期,微机大量推广使用,西文的操作系统扩充了汉字功能,改变为中西文兼容的操作系统,产生了CCDOS系统,得到了广泛的应用,从此汉字操作系统在我国的计算机发展中起到了不可磨灭的作用,一些著名的数据库系统也得到了汉化。

在汉字操作系统发展的同时,大批计算机工作者对汉字输入方法的研究也产生了大约五百余种的设计方案,其中几十种有实用价值的方案很快得到推广,现在已经成为汉字输入方法的主流。

与此同时,汉字文字识别、汉字语音处理、汉语计算机辅助教学等领域的研究也取得了许

多可喜的成果。利用计算机对汉字进行的基础研究也取得了不少重要的成果,比如在字频统计、人口姓氏频度统计等方面,都有了重要的研究进展。

八十年代以来,汉字设备的研制和生产发展也很快,这里包括汉字键盘、汉字存储芯片和汉卡、汉字印字机、汉字终端等硬件的开发和研制。

目前在我国的各级各部门中计算机汉字信息处理技术得到了广泛的应用,如在各行政机关中的办公自动化系统、情报检索系统、各企业的管理信息系统、电子出版系统、通信系统、专家系统、翻译系统等领域。

但是,在微机汉字处理工作中,还存在着方法、代码不够规范,字库缺乏适用的统一标准,词库还不具备广泛的适应性,还不能做到自然语言处理等缺陷,还有许多工作要做。

## 第四节 汉字字库

要想让计算机能处理汉字,就需要在计算机中存有汉字的字模(字形),这就象铅字印刷时要有各种铅字字模一样。产生汉字字形的方法有模拟式和数字式两种,前者如字模板、全息照相等,文字质量较高,价格较便宜;后者是在计算机中使用的主要方法,它输出文字一致性好、速度快。

我们把每个汉字分别写在一个划分为 M 行, N 列的网格方块内,方块内的每个小方格是一个点,有笔画的方格涂成黑点,用二进制中的数字 1 代表;没有笔画的方格内部空白,用二进制中的数字 0 代表,这样一个汉字,就可以用若干个二进制数字来表示了。这个方块就叫做一个  $M \times N$  的点阵,如果 M 等于 N,我们就简称这个点阵为 M 点阵,比如 16 点阵、24 点阵、32 点阵等。不论多大的点阵,字形信息一般都是以字节的形式存储在存储器中的。点阵数越大,所表现的字形越精确、越逼真,所占用的存储量也越大。一个 16 点阵的汉字字形就要用 16 乘 16 等于 256 个二进制位表示,也就是说要占 32 个字节的存储量,而这仅仅是实用中最小的点阵。图 1-1 是一个 16 点阵的汉字字形。

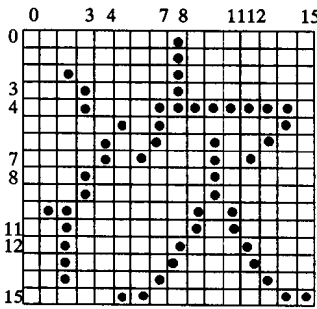


图 1-1

汉字字形按其存储方式可分为整字存储和压缩存储两大类。把汉字字形点阵信息按字节全部存入存储器中,需要时直接读出,就是整字存储,它使用方便、原理简单、速度快,但需要较多的存储量;而压缩存储不将字形信息直接存储,它先将点阵信息用压缩技术进行压缩,再将压缩后的信息存入存储器,使用时先将压缩信息还原成点阵信息再用。压缩的方法很多,应用较广泛的矢量法就是其中之一。

汉字字模除了字模尺寸和点阵大小的规格外,还要求有多种字体。存储这些信息的存储器叫做汉字库(或称汉字字模库)。一般采用磁盘或只读存储器做存储介质,前者存储成本较

低,但读取速度也较慢,这种用磁盘做存储介质的汉字库叫做“软字库”。用只读存储器芯片(ROM)做存储介质的汉字库叫做“硬字库”,它读取速度快,价格稍高于软字库。现在常用的还有一种叫做“汉卡”的设备,它是一块装有存储器芯片和逻辑电路芯片的电路板,它能提供根据有关汉字编码方案得出的汉字键盘码到汉字机内码变换的译码程序,在上面还装有汉字字模芯片,提供汉字的输出字模信息,也就是说,汉卡中包含有汉字库。汉卡的种类很多,它们分别适用于不同的机型和不同的汉字输入方法,也分别具有不同的功能。

由于汉字字数多,为了在全国范围内有一个统一的标准,我国自七十年代末就开始编制国家标准字符集,目前在计算机汉字处理中最常用的是《中华人民共和国国家标准 GB2312-80 信息交换用汉字编码字符集-基本集》。它根据汉字频度统计的结果选择了 6763 个常用汉字并收入了 682 个非汉字图形符号,包括 202 个一般符号、60 个序号、22 个数字、外文(包括英文、日文、希腊文、俄文)字母 335 个、26 个汉语拼音符号、37 个汉语注音字母等总共 7445 个图形符号,并为它们分配了标准代码,供汉字信息交换使用。其中汉字的字符被分为两类,一类是一级汉字,共 3755 个,按汉语拼音的字母顺序排列,同音字以笔形顺序排列;另一类是二级汉字,共 3008 个,按部首排列,部首次序及同部首字按笔画数排列。基本集中每个符号用两个字节表示。对于我国绝大多数汉字信息处理系统来说,只要具有这个基本集中的汉字就能满足各种使用要求,这些汉字的使用覆盖率可达 99% 左右。实际上只要具备了一级汉字,就大体上能满足工作的需求,因为最常用的汉字不过在三、四千个左右。整个基本集可以被看成是一个 94×94 格的表格,表中每一行叫做一个“区”,每一列叫做一个“位”,区和位的编号都是从 1 到 94。表中 1 区至 3 区是各种专用的符号,3 区中间开始到 8 区,顺序是英文、日文、希腊文、俄文和汉语拼音等各种文字的字母,第 9 区是制表用的符号。10 至 15 区目前是空白,留待进一步补充新字符使用。一级汉字排列在第 16 区至第 55 区;二级汉字排列在第 56 区至第 87 区,自 88 区开始,直到最后都是空白。这个字符集规定了一个国家标准,为未来进一步在全国规范汉字信息处理创造了条件。图 1-2 是国标基本集的示意图。

|             |   |   |   |     |
|-------------|---|---|---|-----|
|             | 、 | 。 | · | ... |
| !           |   |   |   | ... |
| ·<br>·<br>· |   |   |   | ... |
| 啊           |   |   |   | ... |
| ·<br>·<br>· |   |   |   | ... |
| 于           |   |   |   | ... |
| ·<br>·<br>· |   |   |   | ... |

图 1-2