

3 NEW

3 N 电脑自修·培训丛书

翁瑞琪 主编

文字录入与 处理

新视野、新技术、新方法

李俊旺 刘学民
柏风雷 翁瑞琪
编著

14
1



TP391.14
LJW/1

3N 电脑自修·培训丛书

翁 瑞 琪 主 编

文字录入与处理

李俊旺 刘学民 编著
柏风雷 翁瑞琪

天津科学技术出版社

责任编辑:李树云

3N 电脑自修·培训丛书

翁瑞琪 主编

文字录入与处理

李俊旺 刘学民 编著
柏风雷 翁瑞琪

*

天津科学技术出版社出版

天津市张自忠路 189 号 邮编 300020

河北省迁安县印刷厂印刷

新华书店天津发行所发行

*

开本 787×1092 毫米 1/16 印张 21.5 字数 514 000

1997 年 3 月第 1 版

1997 年 3 月第 1 次印刷

印数:1—3 500

ISBN 7-5308-2064-8

TP·93 定价:31.50 元

《3N 电脑自修·培训丛书》

编委会名单

主 编	翁瑞琪		
编 委	李宗耀	金朝崇	袁忠良
	王艺梅	王德然	杨建基
	李俊旺	于长云	张 炜
	黄国胜	王定一	徐 彤
	刘 彤		

JS361/29

丛书总序

计算机的发展与普及,使计算机成为人类不可缺少的重要助手。现在,越来越多的人看到计算机对当今社会各个领域所产生的巨大影响,看到计算机对科技进步、经济发展以至人类生活所起的重要作用。在当今社会,计算机知识和计算机应用能力是人们必须掌握的基本知识和基本技能。不懂计算机、不会使用计算机就难以适应今天的工作和生活。要适应当今的计算机时代,必须掌握计算机文化。

为加速计算机应用的普及,尽快让更多的人掌握计算机文化、掌握应用计算机的基本技能,我们组织编写了这套新视野、新技术、新方法电脑丛书,即3N电脑自修·培训丛书,包括:《微机操作与使用》《文字录入与处理》《微机故障诊断与维修》《FoxPro 原理与应用》《C 语言与程序设计》《Windows 应用与编程》《微机绘图与动画技术》《多媒体速成与应用》八种。

本丛书提供了轻松快捷地学习计算机文化的途径,并从使用角度出发,引导读者较快掌握计算机的应用。

本丛书可供计算机初学者自学,也可用作初学者的培训教材和作为计算机用户和爱好者操作与应用计算机的参考用书。

本丛书由国内长期从事计算机教育与培训工作的教师编写。书中总结了他们多年的教学与实践经验,他们的丰富经验将使广大读者得到启迪。在此对他们的无私奉献表示谢意。

期望本丛书的出版能为我国计算机事业的发展与计算机应用的普及起到积极作用。

翁瑞琪

1996年1月

前 言

计算机在“文字录入与处理”方面的应用是计算机应用中最为广泛的方面之一。

本书共分1章。

第1章介绍汉字操作系统Super —CCDOS、UCDOS和CCBIOS 2.13H的使用与操作方法。

第2章的内容为文字录入法,介绍了键盘的使用及指法训练,介绍了拼音录入、五笔字型录入、表形码录入和区位码录入的方法。

第3章介绍文字处理系统WPS的使用与操作方法。

第4章介绍中文字表编辑软件CCED的使用与操作方法。

为便于读者理解和掌握,本书各章均提供了丰富的实例和练习,章末都附有习题。

本书的附录部分还给出了汉字编码。

本书可供计算机初学者自学,也可用作计算机初学者学习文字录入与处理方面的培训班的培训教材,并可供计算机用户和爱好者参考使用。

参加本书编写的有李俊旺、刘学民、柏风雷和翁瑞琪。第1章、第3章和第4章由李俊旺执笔,第2章的第1~3和5节由柏风雷执笔,第2章的第4节和附录由刘学民执笔,全书由翁瑞琪统一定稿和审定。

限于水平,错误在所难免,欢迎广大读者和专家们批评指正。

目 录

第1章 汉字操作系统	(1)
1.1 汉字信息处理	(1)
1.1.1 汉字信息处理的意义及特点	(1)
1.1.2 汉字信息处理	(3)
1.2 Super—CCDOS 汉字操作系统的使用	(7)
1.2.1 Super—CCDOS 的特点、功能、运行环境	(7)
1.2.2 Super—CCDOS 的模块功能介绍	(9)
1.2.3 Super—CCDOS 的启动	(12)
1.2.4 Super—CCDOS 的菜单使用	(14)
1.2.5 Super—CCDOS 的功能键	(19)
1.3 UC DOS 汉字操作系统的使用	(20)
1.3.1 UC DOS 的特点及功能、运行环境及安装	(20)
1.3.2 UC DOS 系统组成	(23)
1.3.3 UC DOS 系统的优化	(24)
1.3.4 UC DOS 的启动与退出	(26)
1.3.5 UC DOS 的功能键	(27)
1.3.6 UC DOS 的特殊显示	(28)
1.3.7 UC DOS 的汉字打印	(32)
1.4 CCBIOS 2.13H 汉字操作系统的使用	(33)
1.4.1 CCBIOS 2.13H 的特点、功能、运行环境	(33)
1.4.2 CCBIOS 2.13H 的组成	(34)
1.4.3 CCBIOS 2.13H 的安装与启动	(34)
1.4.4 CCBIOS 2.13H 的功能键	(35)
1.4.5 CCBIOS 2.13H 的汉字输入法	(36)
1.4.6 CCBIOS 2.13H 的汉字库与汉字的显示和打印	(37)
1.4.7 CCBIOS 2.13H 提供的几个实用程序简介	(39)
习题1	(40)
第2章 文字录入法	(41)
2.1 键盘的使用及指法训练	(41)
2.1.1 101 键盘介绍	(41)
2.1.2 击键的正确姿势	(43)

2.1.3	击键的正确指法及要领	(44)
2.1.4	指法练习	(44)
2.1.5	汉字输入法简介	(51)
2.2	拼音输入法	(52)
2.2.1	直接拼音法	(52)
2.2.2	紧缩拼音法	(52)
2.2.3	拼音双音输入法	(55)
2.3	《五笔字型》输入法	(63)
2.3.1	对汉字进行剖析	(63)
2.3.2	《五笔字型》字根总表	(64)
2.3.3	《五笔字型》字根键盘及助记词	(66)
2.3.4	《五笔字型》编码规则	(69)
2.3.5	键内汉字编码	(69)
2.3.6	键外汉字编码	(71)
2.3.7	简码、重码、容错码	(75)
2.3.8	词语	(78)
2.3.9	Z 学习键	(78)
2.3.10	练习	(79)
2.4	表形码输入法	(82)
2.4.1	汉字编码及汉字部件	(82)
2.4.2	单笔画部件	(84)
2.4.3	多笔画部件之一:离聚型部件	(85)
2.4.4	多笔画部件之二:交叉型部件	(86)
2.4.5	多笔画部件之三:包围型部件	(87)
2.4.6	多笔画部件之四:粘连型部件	(91)
2.4.7	多笔画部件之五:字架型部件	(92)
2.4.8	组合部件	(95)
2.4.9	拆字规则总结	(96)
2.4.10	汉字部件总表	(96)
2.4.11	词组、简码、重码	(98)
2.4.12	《普及版》及“万能”版简介	(100)
2.5	区位码输入法	(100)
习题2		(101)
第3章	WPS 文字处理系统	(102)
3.1	WPS 文字处理系统概述	(102)
3.1.1	运行环境及系统组成	(102)
3.1.2	系统安装及启动	(103)
3.1.3	主菜单介绍	(105)

3.1.4	命令菜单的使用	(108)
3.1.5	基本概念	(109)
3.2	文本编辑	(112)
3.2.1	编辑方式	(113)
3.2.2	光标移动命令	(113)
3.2.3	插入及改写命令	(114)
3.2.4	删除及恢复命令	(115)
3.2.5	分行与分页	(116)
3.3	文件操作	(117)
3.3.1	文件的打开及关闭	(117)
3.3.2	保存文件及读取文件	(118)
3.3.3	文件密码的设置与修改	(118)
3.3.4	举例	(119)
3.4	块操作	(119)
3.4.1	块的标记和取消	(119)
3.4.2	块的移动、复制与删除	(121)
3.4.3	块的列方式	(121)
3.4.4	块的磁盘操作	(122)
3.4.5	大规模块的操作以及复制CCDOS块	(123)
3.4.6	举例	(124)
3.5	查找与替换	(125)
3.5.1	查找和替换命令	(125)
3.5.2	方式选择及通配符的使用	(127)
3.6	打印控制符的设置	(128)
3.6.1	打印字样控制符	(129)
3.6.2	打印格式控制符	(133)
3.6.3	打印控制符的特性及有效范围	(135)
3.6.4	打印控制命令汇总表	(135)
3.6.5	举例	(137)
3.7	排版	(138)
3.7.1	左、右边界设置命令	(138)
3.7.2	段落重排(^B)	(139)
3.7.3	设置左边界点数命令(^OE)	(139)
3.7.4	标尺及控制符显示的开与关	(140)
3.7.5	设定制表站的命令(^OI)	(140)
3.7.6	设置Tab宽度命令(^OK)	(140)
3.7.7	Tab制表键的使用	(141)
3.8	制表	(141)
3.8.1	自动制表命令(^OA)	(141)

3.8.2	制表连线与取消制表线命令	(142)
3.8.3	手动制表	(143)
3.8.4	举例	(143)
3.9	窗口及其他功能	(144)
3.9.1	窗口操作	(144)
3.9.2	WPS 的其他功能	(146)
3.10	模拟显示与打印输出	(148)
3.10.1	模拟显示	(148)
3.10.2	打印输出	(150)
3.10.3	改变当前打印参数	(152)
3.10.4	安装新的24针打印机参数	(153)
3.11	文件服务与帮助功能	(155)
3.11.1	文件服务功能	(155)
3.11.2	帮助功能	(156)
3.11.3	WPS 错误信息及其含义	(157)
3.11.4	WPS 返回码	(158)
3.13	图文编排系统 SPT 的使用	(159)
3.13.1	SPT 的运行环境	(159)
3.13.2	SPT 的安装与启动	(159)
3.13.3	SPT 的操作风格与工作流程	(160)
3.13.4	SPT 的功能详解	(162)
3.13.5	参数区的操作及其他	(170)
习题3		(171)
第4章	CCED 4.0 中文字表编辑软件	(176)
4.1	CCED 软件概述	(176)
4.1.1	CCED 软件的功能特点及运行环境	(176)
4.1.2	CCED 的版本改进及文件组成	(177)
4.1.3	CCED 软件的安装及参数调试	(179)
4.1.4	CCED 的启动	(186)
4.2	基本编辑	(188)
4.2.1	编辑状态及下拉菜单	(188)
4.2.2	光标移动	(190)
4.2.3	文书及非文书内容的输入	(191)
4.2.4	行的插入、连接与复制	(193)
4.2.5	字符及行的删除与恢复	(194)
4.2.6	字符串的搜索与替换	(196)
4.2.7	存盘、退出及文件加密	(197)
4.3	块操作	(199)

4.3.1	块的定义与撤销	(199)
4.3.2	块的标志(显示)与寻找	(200)
4.3.3	块的移动、复制、删除、打印等操作	(201)
4.3.4	文件间的块操作	(204)
4.3.5	块操作一览表	(204)
4.3.6	举例	(205)
4.4	排版	(205)
4.4.1	自动排版	(205)
4.4.2	段落重排	(206)
4.4.3	行居中与左右对齐	(206)
4.4.4	多栏目文书编辑	(207)
4.5	多窗口功能及其他	(208)
4.5.1	多窗口与多文件编辑	(208)
4.5.2	中西文切换	(210)
4.5.3	在(CED)内部执行DOS命令	(210)
4.5.4	数据演示功能	(211)
4.6	制表格	(212)
4.6.1	表格生成	(212)
4.6.2	表格的修改与调整	(213)
4.6.3	表格中的复制	(215)
4.6.4	设置线保护	(216)
4.6.5	表内数据的整理	(217)
4.7	数据计算	(219)
4.7.1	文本编辑版面上的计算	(219)
4.7.2	表格内数据的计算	(220)
4.8	打印	(224)
4.8.1	打印命令及分页控制	(224)
4.8.2	运用原始控制码控制打印效果	(228)
4.8.3	集约控制符的引入及使用	(231)
4.8.4	打印标签、信件	(237)
4.9	利用DBST 程序实现dBASE 数据报表输出	(238)
4.9.1	样本表格文件	(238)
4.9.2	输出报表	(241)
4.9.3	报表打印控制	(243)
4.10	利用CCEDLT 程序进行文件转换与表格叠加	(244)
4.10.1	文件转换	(244)
4.10.2	表格叠加	(246)
4.11	CCED 4.0 编辑命令一览表	(247)
习题4		(250)

附录 I 国家标准《信息交换用汉字编码字符集(基本集)》(GB2312—80)	(253)
附录 I 国标GB—2312(80)第16区~第87区 五笔字型与表形码汉字编码码本	(257)

第1章 汉字操作系统

1.1 汉字信息处理

1.1.1 汉字信息处理的意义及特点

1. 汉字信息处理的意义

信息科学已经成为一门新兴的学科。它研究信息的来源、产生、获取、识别、转换、组织、存储、处理、检索、表达、评价以及提供与信息有关的理论和方法论。现代社会是充满信息的社会,并且社会信息量日趋庞大和复杂,处理形式也多样化,从数据到文字、图形、图像、声音和自然语言。所以,对信息进行处理和管理是社会的需要。由于社会对共享信息资源的需要,就要求大容量和高速传递信息(如美国现正在搞“信息高速公路”,而我国也正在搞“三金”工程——金桥、金关、金卡工程。),要求既经济又高度可靠的信息系统。本章主要介绍汉字信息以及汉字信息的处理。我们用计算机存储和控制信息,加工和处理信息时,常常涉及到汉字信息处理,随着计算机系统功能的不断提高,应用领域也迅速扩展,汉字信息处理的概念、涵义、作用及应用范围也进一步扩大,内容更加丰富。其应用已从一般的制表、文档、文书发展到书刊和报纸的自动编辑和排版,日常事务的处理,办公室自动化,文字翻译,公共咨询服务,数据通讯等。实际上,汉字信息处理技术已逐渐渗透到社会活动的所有方面。并正在促使社会经济、科学技术和家庭生活日新月异的变革。如现代的电子词典、英汉字典以及家庭用的电脑记事簿等。

我们的祖先创造了汉字,使中华民族具有今天灿烂的文化和文明。目前,世界上约有四分之一的人口使用汉字,汉字也是联合国规定的通用文字之一。我国的汉字是一种象形文字、字量多、字型复杂。在构成的汉字信息处理系统中,除了通用硬件设备之外,还需配备汉字输入输出设备以及相应软件处理系统,使软件处理既能适合汉字处理,又能适合西文处理。目前,我国的科技工作者已研制出十多种适合不同应用范围的汉字处理系统,使汉字信息处理发生了巨大的变化。

2. 汉字信息处理涉及的范围

汉字信息处理技术是一项综合性的技术,其核心技术是计算机技术,所涉及的技术范围包括以下几个方面:

(1) 汉字字量 在汉字信息处理技术中,需要使用多少个汉字是重要问题,我国汉字多达5~6万个,国标中收集一级常用汉字3755个,使用频度占95%,二级常用汉字3008个,共6763个。因此,一个汉字处理系统应收集多少个汉字容量,视实际的使用频度来确定。

(2) 汉字字体 在印刷排版的汉字处理系统中,对字体各类的要求较高。一般来说,汉字字体可分为宋体、仿宋体、楷体、黑体等。而每一种字体又有方体、长体、扁体的区别。

(3) 汉字索引 可以从不同的角度检查汉字,例如以笔画、偏旁或部首来检索,也可以用汉字发音音序来检索,或以国际交换码区/位号检索,还可能有其它检索方法,其目的都是以简

捷的法则准确地查得某个汉字或它的标准编码。

(4)使用频度 对于不同的汉字,其使用频度的差别是很大的。同一汉字在不同专业领域中使用时,其频度也有差异。根据汉字的频度,我们可以把汉字分为常用字、次常用字、稀常用字、罕用字等几个等级,在建立不同种类的汉字处理系统中,必须根据使用频度来选用字库中所收容的汉字。

(5)字型分解 字型分解就是把汉字分解为部首、字首、笔画、位点。分解字型是为了找出汉字结构的规律。通过对字型的分析研究,可以选取最少量的字根,合理地组成所需的汉字,在以字型特征为基础的汉字编码方法中,特别要注重字型分解的研究工作。

(6)汉字词组 在汉字信息处理技术中,对词的研究很重要。在汉字输入方案中,对于使用频度很高的词,可以用软件方法设定,用一个键位代表一个词,也可以根据需要使用某个键位所代表的词,也可以把词组组成词典,存入系统的存储器中,以便建立索引或进行查找。

(7)计算机硬件及软件 信息的处理离不开计算机的硬件设备,例如汉字键盘、汉字显示终端、汉字输出设备等。根据汉字系统规模的大小和应用特点,还可配置其他外设或外部设备。

汉字信息处理还需要软件支持,诸如操作系统、高级语言编译程序,还需配上汉字操作系统、汉字处理(管理)程序以及汉字字库等。

3. 汉字信息处理的特点

世界上的文字可以分为三种:表意文字、音节文字和音素文字。表意文字单独成为一个体系,音节文字和音素文字合并成为拼音文字体系。

西文(如英文、俄文等)是拼音文字,用计算机处理西文信息是较容易的。汉字是表意文字(或象形文字),每个汉字都有特定的形体、一定的读音和确定的意义,称之为汉字的形、音、意三要素。用计算机处理汉字信息,有以下特点:

(1)汉字数量多 在几千年的历史演变中,汉字的数量随语言的发展而增多。据资料记载,殷商代只有2000个字左右,东汉收录汉字9353个,清朝康熙字典收集汉字42174个,1915年出版的中华大字典收入汉字达44908个。在这众多的汉字中,高频字约占100个,常用字约占3000个,次常用字约占4000个,罕见字约占8000个,死字约占29808个。可见,常用的字还是少数。目前,微机上实现的汉字系统所用汉字一般为国标一级汉字3755个或国标一、二级汉字6763个。70年代初有关科研单位对报纸、期刊、教材、专著等读物进行抽样调查统计并由计算机处理,得出经常使用的汉字约7000~8000个左右。

(2)汉字字形复杂 汉字笔画繁简不一,少至一笔画一字,多至30画一字,笔画方向和形状变化多。汉字的构字元素(点、横、竖、撇、捺、折、叉、勾)以平面结构方式非线性地构成文字单位,不像拼音文字那样由字母的线性排列方式构成文字单位。另外,汉字的异体字多,一个汉字往往有几种不同的写法,认识一种写法不行,还必须记住另外的写法。

(3)汉字同音字和多音字数量大 同音字是指音同字不同,即一个字音信息对应多个字型信息。据统计,几乎100%的汉字都有同音字,少则几个,多则几十个以至一百多个,平均计算,每个汉字一般有18~20个同音字。

多音字是指同一字形读音不同,即一个字形信息对应多个字音信息。据统计,多音字至少占汉字总数的10%。

(4)汉字字体美观,汉语精练准确 汉字字形结构工整美观,形象表意。世界上只有汉字有书法,成为一门艺术;汉语精练准确,在联合国各种文字书写的文书中,最薄的必定是中文

本。利用计算机的大容量、高速度和图形处理能力处理汉字,恰到好处地发挥了计算机的优势。

4. 汉字信息处理技术的发展

迄今为止,虽各种实用性汉字信息处理系统已推广应用,但汉字信息处理技术基础理论方面的研究还在不断深入,标准化工作也在不断进行,各种应用软件不断涌现,为适应经济建设需要,汉字信息处理技术正向系统集成化和智能化方向发展。

(1)系统集成化 将孤立的汉字信息系统集成为一个系统,可以在更大范围内共享信息资源,可以进行全局的优化或管理,从而提高系统的效率、效能和效益,这是汉字信息系统的发展方向。

汉字信息系统的集成一般是与运行的、异构系统的集成。实现异构系统的集成,要解决从硬件到软件一系列的问题。但对汉字信息系统来说,网络的集成、数据的集成以及应用程序的集成是最重要的。

1)网络的集成 网络的集成是要解决异构系统间的数据传送和互访问题,是系统集成化的基础。

2)数据集成 数据的集成是解决异构系统间数据共享问题。要数据共享,仅仅相互交换数据是不够的,还要解决各子系统间的数据互操作,不同数据模型的转换和数据语义的差异等问题。

3)应用程序的集成 要求应用程序可以在不同的子系统上运行,也就是各个子系统应对应用程序提供统一的界面。标准化有利于汉字信息系统的集成,但集成不能完全依赖于标准化。

目前,相当部分的企、事业单位、大专院校利用NOVELL网络原理把本单位原以单台机运行的汉字信息处理系统统一联网,使各处、室的汉字信息实现了共享和集中管理,极大地提高了系统效率。

(2)功能智能化 传统的汉字信息系统智能化程度低,主要存在的一些弱点是:系统是被动的,只能被动执行用户要求完成的任务;系统缺少推理功能,只能向用户提供系统中的存储信息,而不能从存储的信息推理出其他有用的信息。

基本上以事务性操作为主。只有少量的汉字信息系统具有决策功能,或附有适用于特定领域的专家系统,仅限于局部的应用。

为了提高汉字信息处理系统的效能,充分发挥系统的效益,须要克服上述缺点。因此,未来的汉字信息处理系统应是数据密集型的,知识密集型的,可提供诸如:提示和报警、自动跟踪记录或统计、专人或专题情报服务、预测或规划、决策和咨询等功能,而且随着应用的发展、智能化程度越来越高。

1.1.2 汉字信息处理

西文计算机只能识别计算机字符集中的字母、数字和特定符号,不能直接处理汉字。

要实现汉字信息处理,一种方法是设计专用的汉字信息处理系统;另一种方法是利用现有的西文计算机的软、硬件资源,加以改造,使之实现汉字与西文兼容。无论采用哪种方法,汉字信息处理系统均应包括汉字输入、汉字信息处理、汉字字库和汉字输出(包括显示和打印)四个部分。其组成如图1-1所示。

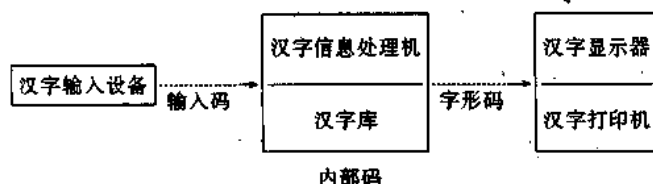


图1-1 汉字信息处理系统框图

1. 汉字的编码输入

输入设备的作用是按各种编码规则将汉字和ASCII字符输入到计算机内部,并将输入码映射成内部码。

汉字是二维空间的图形结构,要计算机处理,首先必须解决汉字的输入问题,即必须使用某一规则将数万个汉字变为有限个元素,自左至右,线性排列成一维空间的等价信息。换句话说,必须将汉字编码,至少要把5000~8000个常用汉字的字形与汉字代码一一对应。

据统计,国内外已有400多种汉字输入编码方案。这些汉字编码方案,就其汉字输入计算机的方式,可归纳为以下四种:

(1)数字编码方式 这种方式是将所有被编码的汉字,按一定的音序或形序排列编号,然后,以数字形式表示汉字。属于这类编码方式的有国标码、电报码等。

国标码,即1981年国家标准局制定并颁布的“信息交换用汉字编码字符集基本集图形字符分区表”规定了汉字信息交换用的基本图形字符及其二进制编码,是一种用于计算机汉字处理和汉字通信系统的标准交换代码。

国标码又称区位码,把国标分为94个区,每个区94个位,这就构成了一张汉字交换用区位代码表。所有汉字及符号的94个区划分为以下四个组:

- 1) 1~15区:图形符号区,其中1~9区为标准区,10~15区为自定义符号区;
- 2) 16~55区:一级常用汉字区共3755个汉字,该区的汉字按汉语拼音排序;
- 3) 56~87区:二级非常用汉字区共3008个汉字,该区的汉字按部首排序;
- 4) 88~94区:自定义汉字区。

例如汉字“啊”,它的区位码为十进制1601,位于16区01位。

这类编码的特点是整齐,容易实现。但代码的取值与汉字的形、音、意没有直接的联系,故用户难以掌握。

(2)字形编码方式 这种方式取码于汉字字型特征。按取码特征部位的大小可分为整字编码输入法、字元(构件)编码输入法和笔画编码输入法。

整字编码输入通常采用大键盘,将常用的汉字直接做在键盘的盘面上,一键一码,代表一个汉字。大键盘一般具有3000个左右汉字。这种方法的优点是直观,无重码,缺点是效率低。

字元编码输入法即把汉字分解为一个个基本构件,然后按一定规则进行组合输入。例如:按字的书写顺序、从左到右、自上而下,先外后内进行字元拆分编码等,这种优点是形象、直观。缺点是组字规则不易掌握。

笔画编码输入法是先将字形结构化为用笔画表示,然后,再按一定规则进行组码。这种方法规则简单,但码长不一,输入效率低。

(3)字音编码方式 这种方式取决于汉字的读音特征。根据汉语拼音进行编码。只要有一定的汉语拼音基础,掌握起来并不困难。缺点是重码较多,要解决同音字的区分问题。

(4) 形音结合编码方式 字形编码有形象、直观的优点,但字元与键面标示之间存在矛盾。字音编码虽然实施简单,但受同音字、生僻字、方言、口语等影响,难于掌握。形音结合编码方式有两种:一种是以字形分解为主,辅以字音;另一种是以字音为主,辅以字形。优点是取两者之长,缺点是对形码和音码两种规则都要比较熟悉。

2. 汉字内码

上述所讲编码对计算机来说,都称为“外部码”,各种外部码,长短不同,格式各异,差别很大,要在计算机内进行各种处理,是很困难的。因此,把每一个汉字和计算机内的一个代码唯一地对应起来,不管是哪一种编码方案,输入的外部码最后都转换为这种机内代码,就能方便地处理了。这种机内代码,叫内码。西文也有内码,这就是ASCII码,一个ASCII码,用最高位为0的一个字节8位二进制内码表示,例如:字符A的内码是01000001B。汉字的内码形式,在CC-DOS中,一个汉字的国标码的二个字节的最高位都设置为“1”而实现的。例如:汉字“大”的国标码是3473D,即二字节是00110100B和01110011B,将二个字节的最高位置“1”,得到10110100B和11110011B,即B4F3H,这就是汉字“大”的内码。

注:B—Bin表示二进制;

D—Dec表示十进制;

H—Hex表示十六进制。

3. 汉字字形库

(1) 汉字字形码 通常,把每个汉字离散成网点,每点以一个二进制位表示,组成汉字的点阵字形(字模),称为汉字字形码。它是汉字点阵字形的数据表示形式。目前,16×16、24×24、36×36的汉字点阵字形码已有统一的国家标准。汉字“大”的16×16汉字点阵字形码如图1-2所示。

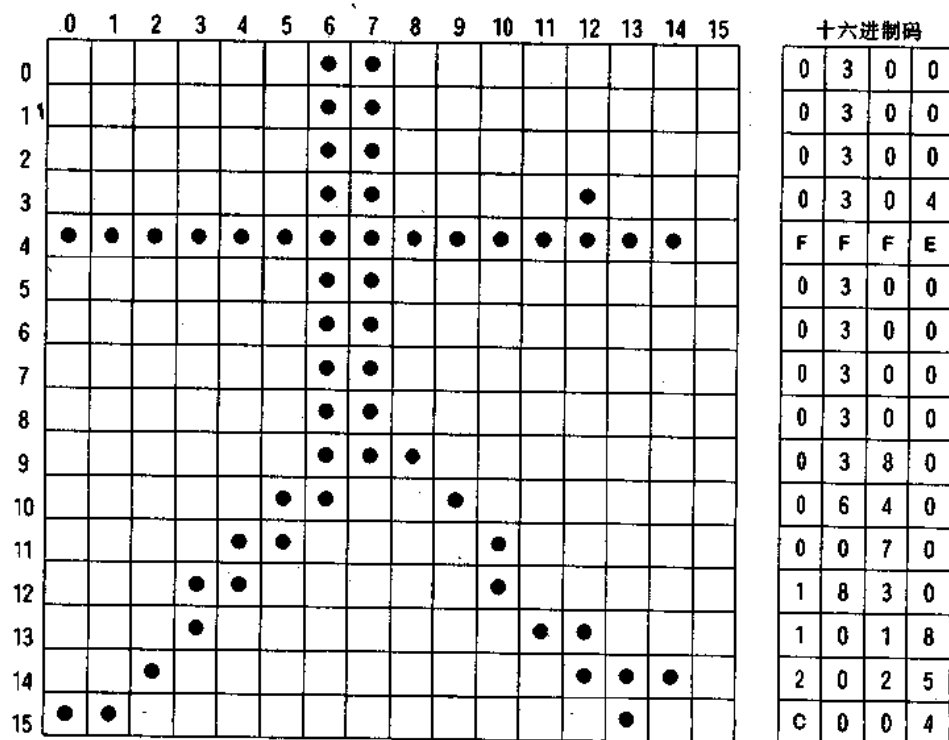


图1-2 16×16汉字点阵字形及代码