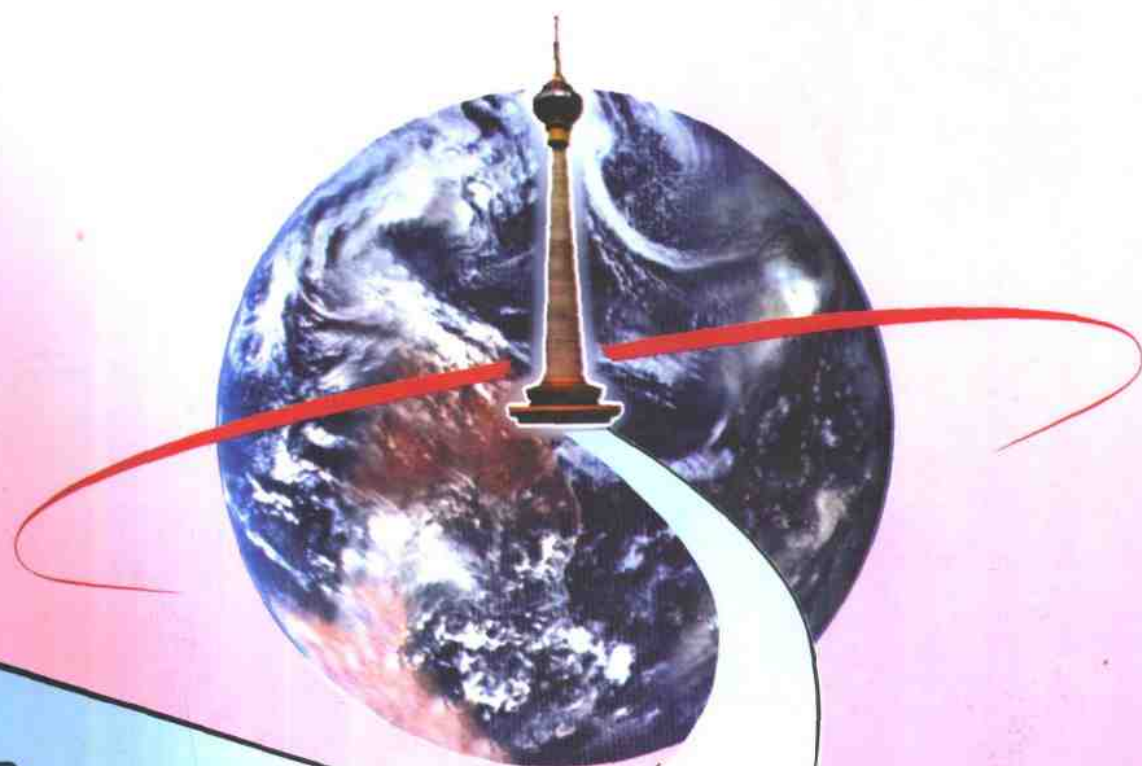




全国非计算机专业等级考试、
自学考试辅导丛书

新编文字处理技术 自学辅导

本丛书编委会



电子工业出版社

PUBLISHING HOUSE OF ELECTRONICS INDUSTRY

新编文字处理技术 自学辅导

本丛书编委会

电子工业出版社
Publishing House of Electronics Industry

内 容 提 要

本书是全国非计算机专业等级考试、自学考试辅导丛书的第六册,全书以通俗、浅显的文字介绍了微机文字处理的基本技术。全书突出基本技能的培养,并配有作业及答案。书中收集了初学者易出的错误,并给予答疑辅导。本书适合于参加各类计算机等级考试的读者自学使用,亦可作为计算机基础教育的入门教材。

全国非计算机专业等级考试、自学考试辅导丛书

新编文字处理技术自学辅导

本丛书编委会

责任编辑:应月燕

*

电子工业出版社出版

北京市海淀区万寿路 173 信箱(100036)

电子工业出版社发行 各地新华书店经销

北京市顺义县李史山胶印厂印刷

*

开本 787×1092 毫米 1/16 印张 17.5 字数:435 千字

1997 年 5 月第一版 1997 年 5 月第一次印刷

ISBN 7-5053-3842-0/TP·1655

定价:20.00 元

全国非计算机专业等级考试、自学考试辅导丛书编委会

顾 问 刘乃琦

策 划 王明君

编 委 许 远 何成彦 冷 潜
冯晓琳 黄 坤 陈周坤

补充说明

在编写本书的过程中,编者感到难度相当大,因为本书是写给参加计算机等级考试的读者的,他们都具有高等学历,如果仅从汉字系统和汉字编辑软件的使用,以及如何录入汉字的角度来讲,显然不适宜。在编写过程中,编者参考了许多同类书籍,并根据自己对大纲的理解拟定了编写总体纲要,为了便于广大读者学习,并与我们相互交流,特对编写时的几个问题进行如下说明:

(1) 讲文字处理的知识不能只限于用户级,因为大学生不同于打字录入人员,对文字处理系统的使用不仅要知其然,而且要知其所以然,所以本书打破常规,简单明了地从文字处理系统的历史发展角度来谈学习它的必要性;接着又简单地介绍了文字处理系统的组成、一般原理等稍具理论性的知识,从而使读者了解到计算机为什么能处理汉字和计算机是怎样处理汉字的。这就突破了仅仅讲述“如何用计算机处理汉字”的传统教学方式。

(2) 参加等级考试的读者们不仅需要知道当前常用的各种汉字输入方法的使用法,如何利用汉字编辑软件进行文字处理等必备知识,而且更重要的是要了解各种字处理软件的文件组成,如何安装这些系统,出现一般性的故障如何排除等有关系统维护的知识。读者更需要培养系统维护方面的能力,所以在现行的考试大纲的基础上,我们着重拓展这方面的基础知识。

(3) 参加等级考试的大学生们,一般不会从事专业录入工作,因此对于汉字输入方法的讲解,本书编者认为,全拼双音与双拼双音已达到考试要求,若需达到更快的输入速度,再选学五笔字型或自然码即可。

全国非计算机专业等级考试、自学考试辅导丛书编委会

前 言

随着科学技术的迅猛发展,计算机已成为各个学科领域不可缺少的应用工具,计算机知识和应用能力已成为当代大学生知识和能力结构的一个重要组成部分,也是我国教育培养跨世纪人才最突出的需要加强的环节之一。目前在高校中普遍开展的计算机知识和应用能力等级考试正有效地推动这一目标的实现。1993年12月国家教委考试中心颁布了在全国进行计算机应用能力认证考试文件,这必将进一步推动全社会学习计算机、使用计算机的热潮。与此有关的教材和参考资料的需求与日俱增。

到目前为止,各省、自治区、直辖市都举办了计算机等级考试,此类的书为数不少,本书的出版正是在充分吸收先期出版的同类图书的优点、克服存在的弊病基础上,推陈出新,更上一层楼。

为了达到这个目的,我们在编撰过程中特别注意了以下三个问题。

(1) 适当放低起点,但不降低总的要求,充分贯彻《国家教育委员会全国计算机等级考试大纲》的要求,兼顾目前存在的多种等级考试的要求,循序渐进,深入浅出,对基本内容讲透、讲够,对于易出的错误,给予明确指出,同时对《大纲》进行适当扩充,以保证该书具有一定的实用性与超前性。

(2) 注意考试科目的基本知识讲授,在掌握基础知识的同时,适当地进行基本技能的训练,而不以“习题”和“模拟试题”为主。我们认为扎扎实实的掌握基本技能,完全可以达到有关考试的要求。

(3) 本书适合于读者自学,也适合于有关专业进行课堂教学,每道习题均有答案,习题本着精辟、典型的原则进行收录。

(4) 本书本着实用、广谱的原则,兼容各级各类考试的要求,适合于以下几种考试:

- 各省、市、自治区组织的非计算机专业计算机等级考试
- 国家教委考试中心的非计算机专业计算机能力考试
- 国家教委考试中心的计算机专业的计算机水平和资格考试

全国非计算机专业等级考试、自学考试辅导丛书编委会

本书阅读方法

◁必学/了解

这部分内容是基础性的和稍具理论性的,了解它们是为了学习后续内容的需要。

◁必学/重点

这部分内容是实质性的,要求准确地理解,熟练地掌握,读者阅读时可多加推敲。

◁自学答疑

这部分的内容收集了初学者易产生的一些概念错误、操作错误,它是本书的画龙点睛之处,对于缩短学习周期,提高学习效率有重要指导意义,应认真对待。

◁选 学

这部分的内容超出了考试大纲的要求,若读者在实际应用中遇到这些问题或有空余的时间,不妨一看。

◁实用技巧

这部分是一些极具实用性的技巧,当你“山穷水尽疑无路”时,它也许能让你有“柳暗花明又一村”的感觉。

◁作 业

这部分收集了一些典型的习题,要求认真完成。

◁答 案

本部分是“作业”部分的答案,也许你的答案和书中的答案不一样,这很正常,也许你的答案更先进(向你祝贺),也许我们的答案有误(欢迎指正)。

目 录

第一章 微机文字处理概述	(1)
§ 1.1 开篇	(1)
§ 1.2 汉字的信息化表示	(2)
§ 1.3 汉字文字处理系统的基本组成	(9)
第二章 常见汉字操作系统(I)	(18)
§ 2.1 CCDOS 汉字系统——第一个成熟的汉字系统	(18)
§ 2.2 Super-CCDOS 的使用(I) 系统的组成与启动	(23)
§ 2.3 Super-CCDOS 的使用(II) 输入法的选择与系统功能的设置	(28)
§ 2.4 Super CCDOS 的使用(III)——若干问题的解答	(38)
第三章 常见汉字操作系统(II)	(44)
§ 3.1 UCDOS 汉字系统(I)——特点概述	(44)
§ 3.2 UCDOS 汉字系统(II)——组成·安装·启动	(47)
§ 3.3 UCDOS 汉字系统(III)·系统功能详解	(53)
§ 3.4 汉字系统的讨论	(58)
第四章 常用汉字输入法(I) 拼音输入	(63)
§ 4.1 Super-CCDOS 的拼音输入法(I)——全拼输入法	(63)
§ 4.2 Super-CCDOS 的拼音输入法(II)——双拼输入法	(66)
§ 4.3 Super-CCDOS 的拼音输入法(III)——拼音输入技巧	(68)
§ 4.4 其它系统的拼音输入法	(71)
第五章 常用汉字输入法(II)——五笔字型输入	(77)
§ 5.1 汉字的结构层次——学习五笔字型的基础	(77)
§ 5.2 记识基本字根——学习五笔字型的关键	(81)
§ 5.3 字根组字与汉字拆分	(91)
§ 5.4 五笔字型汉字编码	(95)
§ 5.5 五笔字型输入的若干问题	(105)
§ 5.6 五笔字型学习必备资料	(112)
第六章 常用汉字输入法(III)——自然码、区位码、电报码	(122)
§ 6.1 自然码输入法(I)——编码概述	(122)
§ 6.2 自然码输入法(II)——自然码软件的使用	(127)
§ 6.3 自然码输入法(III)——汉字输入	(129)

§ 6.4 流水码输入——国标区位码与电报码	(132)
§ 6.5 五笔画输入法	(134)
第七章 WPS 文字处理系统	(137)
§ 7.1 WPS 系统的启动	(137)
§ 7.2 WPS 的菜单操作	(140)
§ 7.3 WPS 系统的文本编辑(I)——基本概念	(147)
§ 7.4 WPS 系统的文本编辑(II)——基本操作	(149)
§ 7.5 WPS 系统的文本编辑(III)——设置打印控制字符	(153)
§ 7.6 WPS 系统的文本编辑(IV)——版面设计	(168)
§ 7.7 WPS 系统的文本编辑(V)——块操作·标记	(172)
§ 7.8 WPS 系统的文本编辑(VI)——查找与替换	(179)
§ 7.9 WPS 系统的文本编辑(VII)——制表·多窗口操作	(183)
§ 7.10 模拟显示与打印	(189)
§ 7.11 WPS 系统的文件操作	(195)
§ 7.12 WPS 系统的其它辅助操作	(197)
第八章 CCED 字表处理系统	(202)
§ 8.1 CCED 软件概述	(202)
§ 8.2 CCED 的编辑操作(I)——基本编辑操作	(211)
§ 8.3 CCED 的编辑操作(II)——块操作	(218)
§ 8.4 CCED 的编辑操作(III)——文稿排版打印	(225)
§ 8.5 CCED 的表格处理(I)——简单报表的处理	(233)
§ 8.6 CCED 的表格处理(II)——数据库报表处理	(252)
附录	(262)
附录 A WPS 命令速查表(WPS NT 1.0)	(262)
附录 B CCED 编辑命令速查表	(266)
附录 C CCED 打印控制命令对照表	(270)

第一章 微机文字处理概述

【导读提要】微机的应用在当今社会中已经无处不在,也许有人会问:在我国,微机应用得最多的领域是什么?那就是本书的主题——微机文字处理。那么这方面的知识总体结构如何?包含哪些内容?作为一个计算机的应用人员,对于文字处理这方面的知识应该了解多少?了解到什么程度?如何学习这方面的知识等一系列的问题将在本章中给予简单明确的回答。

本章中要求了解的内容有:

- 文字处理的概念
- 汉字的编码
- 区位码与国标码的区别及互换
- 汉字输入码(外码)的特点
- 文字处理系统的基本组成
- 汉字系统的概念
- 字库的概念与功能

§ 1.1 开 篇

文字处理技术是一门对文字信息进行加工的技术,包含文字的编辑、排版、印刷等。文字处理自古有之,并不是计算机时代的特有产物。自从有文字记载开始,人类就不断地对文字信息进行加工处理,中国古代的大发明家毕升的活字印刷术就是古代文字处理的例证。发明计算机的初衷并不是为了进行文字处理,当时主要是为了满足科学计算上的需求,后来,工程技术人员发现计算机的“连续、自动地工作”、“不知疲倦”的特征特别适合用于文字处理领域。开始是利用计算机编辑一些简单的英文文本,由于当时硬件条件的限制(例如没有足够的内存存放字体、字型信息,打印机还是采用电传打字机或菊花轮的打字机)和软件算法的局限(汉字在计算机中如何表示还未获解决,汉字的输入方法还未形成体系)。这种“文字处理”仅限于英文文本,而且不能达到印刷工艺的要求。随着计算机硬件技术的发展,大容量的存储设备出现了,计算机的运算速度显著提高,能精密地输出字符的激光打印机出现后,印刷业开始大规模地应用计算机,外文图书的编辑、出版率先走向电子化,美国在60年代末就研制成功了英文的精密照排系统。当时我国的汉字出版物仍然是采用铅排的落后手段。汉字真正进入计算机应该是在1958年,邮电部门研制成功了汉字电报译码机,该机从通信电路接收到由发报局发来的汉字代码后,便自动将其转换为该代码所对应的汉字点阵,然后在纸上打印出来。进入70年代,汉字的计算机处理技术已经十分迫切,1974年8月,“748”工程上马,该工程的主要目的是研制精密汉字编辑排版系统、汉字情报检查系统、汉字通信系统和汉字终端设备。“748”工程投入资金约一亿元,“华光方正”电子出版系统就是748工程的直接成果之一,“华光方正”以及后来的“北大方正”是目前中国最有影响的电子出版系统。70年代末期,以8位微处理机(如Z80)为代

表的微型机开始大规模推广,针对它的 CP/M 操作系统汉化的成功,开创了微机操作系统汉化的先河。

本书所讲的文字处理即是指利用计算机进行文字处理,其中的文字主要指汉字。

一、本书主要讲述什么

◁学习指导

微机文字处理的内容相当广泛,既有许多理论性很强的内容,又有很多实际操作性很强的内容。本书主要从应用角度来介绍“文字处理”技术,主要包括:

- (1)基本的文字处理理论知识
- (2)汉字操作系统的结构、基本原理与应用
- (3)汉字文字处理系统的组成
- (4)常用汉字系统的安装、使用维护
- (5)常用汉字编辑软件的使用
- (5)常用的汉字输入法

对于各部分的内容的要求如下:

- (1)了解有关的理论,掌握常用的概念
- (2)熟练地安装、使用常见的汉字系统
- (3)能够排除汉字系统的一些软故障,能进行简单维护
- (4)熟练掌握一种汉字输入方法
- (5)熟练掌握常用的汉字编辑等软件和字表处理软件

二、文字处理的步骤

◁必学/了解

尽管目前文字处理系统很多,但从总的来说,用计算机进行文字处理的步骤与方法总是大同小异的,可以总结为以下三点:

- (1)将文字信息输入计算机,并且在计算机中以一定的方式表示文字信息。
- (2)对输入的文本进行修改、美化、排版等操作,这是主要的工作。
- (3)将处理好的文本保存于磁盘或输出到打印机上,或做成制版胶片供印刷用。

在用计算机进行文字处理的过程中,需要用到以下的软件:

- (1)首先要启动计算机,要用到 MS-DOS 或 Windows 等操作系统。
- (2)其次,要启动汉字系统,可以用 CCDOS、Super-CCDOS、UCDOS 等汉字系统(Windows 系统下的汉字软件应在 Windows 3.x 中文版下使用),启动汉字系统的目的在于使得原先只能处理西文信息的计算机能够处理汉字,使得计算机能在汉字系统的控制下完成汉字的输入、显示、排版、输出等工作。

(3)启动相应的汉字编辑软件,选取相应的汉字输入法进行汉字的录入,而后利用编辑软件提供的编辑排版功能对文本进行加工。

- (4)启动汉字打印驱动程序,使得打印机能从汉字字库中读出字型信息,能够打印汉字。

§ 1.2 汉字的信息化表示

汉字的信息化表示就是根据计算机处理汉字的要求,对汉字进行编码。

我们知道,在计算机中,只能表示 0、1 两种状态,因此在计算机内部,一切信息都只能用二进制数字串的形式来表示。如果我们想把汉字变成计算机能够识别的信息,就必须把它变为二进制代码的形式,这就需要进行编码。所以,编码的过程就是按照某种规则把汉字转换成二进制数字串的过程。

在汉字系统中,存在着多种汉字编码形式。一般来说,在系统的不同部位可根据环境使用不同的汉字代码,这些代码就组成了一个汉字代码体系。为了适应各模块对汉字信息进行处理的不同要求,汉字系统应提供相应的程序来完成这些代码的转换和处理。下面我们来讨论一下汉字操作系统中的各种代码。

一、ASCII 码

<必学/重点

ASCII 码是 American Standard Code For Information Interchange(美国标准信息交换码)的缩写。在 IBM-PC 系列微机中,字母和符号用国际化标准组织规定的 ASCII 码来表示。

表 1-1

ASCII 值	字符	控制字符	ASCII 值	字符	ASCII 值	字符	ASCII 值	字符
000	(null)	NUL	032	(space)	064	@	096	'
001	☐	SOH	033	!	065	A	097	"
002	●	STX	034	"	066	B	098	b
003	♥	ETX	035	#	067	C	099	c
004	♦	EOT	036	\$	068	D	100	d
005	♣	END	037	%	069	E	101	e
006	♠	ACK	038	&	070	F	102	f
007	(beep)	BEL	039	*	071	G	103	g
008		BS	040	(	072	H	104	h
009	(tab)	HT	041	)	073	I	105	i
010	(line feed)	LF	042	*	074	J	106	j
011	(home)	VT	043	+	075	K	107	k
012	(form feed)	FF	044	,	076	L	108	l
013	(carriage return)	CR	045	-	077	M	109	m
014		SO	046	.	078	N	110	n
015		SI	047	/	079	O	111	o
016	▶	DI	048	0	080	P	112	p
017	◀	DC1	049	1	081	Q	113	q
018		DC2	050	2	082	R	114	r
019	!!	DC3	051	3	083	S	115	s
020		DC4	052	4	084	T	116	t
021		NAK	053	5	085	U	117	u
022		SYN	054	6	086	V	118	v
023		ETB	055	7	087	W	119	w
024	↑	CAN	056	8	088	X	120	x
025	↓	EM	057	9	089	Y	121	y
026	→	SUB	058	:	090	Z	122	z
027	←	ESC	059	;	091	[	123	{
028		FS	060	<	092	\	124	
029	◆	GS	061	=	093	]	125	}
030	▲	RS	062	>	094	^	126	~
031	▼	US	063	?	095	_	127	-

在计算机中,每个 ASCII 码用一个字节来表示,ASCII 码制是七位编码,其范围为 0000000~1111111,共有 $2^7=128$ 种状态,每种状态对应表示一个数字、字母、字符或控制代码。由于 ASCII 码存储时占用一个字节(即 8 个二进制位),而编码只有七位,故最高位作为奇偶校验位。在一般情况下,高位闲置不用,置为零,这样,一个 ASCII 码用八位二进制代码来表示(见表 1-1)。

从 ASCII 码表中可以看到,在它表示的 128 个字符中,0~31 为控制代码,它们可用来进行数据通信时的传输控制、打印或显示的格式控制、对外部设备的操作控制等特殊功能,这些字符是不能直接显示或打印出来的。还有 127 为删除符,也是不可显示、打印的。第 32~126 为可显示打印字符,共 95 个,对应表示字母、数字及通用符号等。这些字符通常都可在计算机键盘上找到,按键后就可将该字符的编码输入。例如,若用户键入“ABC”三个字母,则存入计算机存储器内,其代码是 01000001、01000010、01000011 3 个二进制数字串。

二、汉字信息交换码

< 必学/重点

汉字信息交换码是国家颁布的系统间通讯或交换信息用的汉字代码,是为不同的汉字系统之间交换汉字信息而设定的。我国的《信息交换用汉字编码字符集基本集》(即 GB2312-80)于 1981 年公布,该字符集以 94 个可显示的 ASCII 码字符为基集,由其中任意两个代码组成一个汉字交换码(两字节),简称国标码。所以,国标码与 ASCII 码属同一制式,可以认为是扩展的 ASCII 码(即从字节扩展为字)。

1. 区位码

在 GB2312-80 中,由行、列号可唯一地确定一个汉字,分别用两位无符号十进制数来表示,我们称行号为区,列号为位,将其合并在一起组成四位无符号十进制数,称为区位码,也即汉字输入码之一。

区和位分别可用 94 个 ASCII 码字符集中的可显示符号来表示,故共有 94 个区,每个区可分为 94 个位(其区位 1~94 对应 ASCII 码的 33~126),可表示 $94 \times 94 = 8836$ 个表格单元,每个单元为一个汉字,其实,表中对应汉字只有 6763 个(其余未用的区域为空白)。表中汉字按使用频率分为两级:第一级(第 16~55 区)汉字 3755 个,属常用字,按汉语拼音顺序排列,同音字按笔画排列;第二级(第 56~87 区)汉字 3008 个,属非常用字,按偏旁部首排序。

2. 国标码

上述区位码的区与位的取值范围分别为 01~94,显然,其值仍属 ASCII 代码范围,且出现了小于 32 的控制代码部分,为了使其代码均为可打印的字符代码,在 GB2312-80 标准中规定分别用十六进制数 21 到 7E 的代码来表示行与列(即将每个区码或位码分别加上十进制数 32),即国标码用四位十六进制数来表示。按照 GB2312-80 中的顺序表示:①各种非汉字图形字符;②一级汉字;③二级汉字;④空白区(以待扩展)。其分布示意图如图 1-1 所示。

【例 1】“啊”字在表中的 16 区第 1 位,求它的区位码与国标码。

显然,“啊”字的区位码为 1601(十进制),对应的国标码计算方法如下:

第一字节 = $16 + 32 = 48d = 30H$; 第二字节 = $01 + 32 = 33d = 21H$;

故国标码为 3021H,对应的二进制代码为 00110000,00100001。

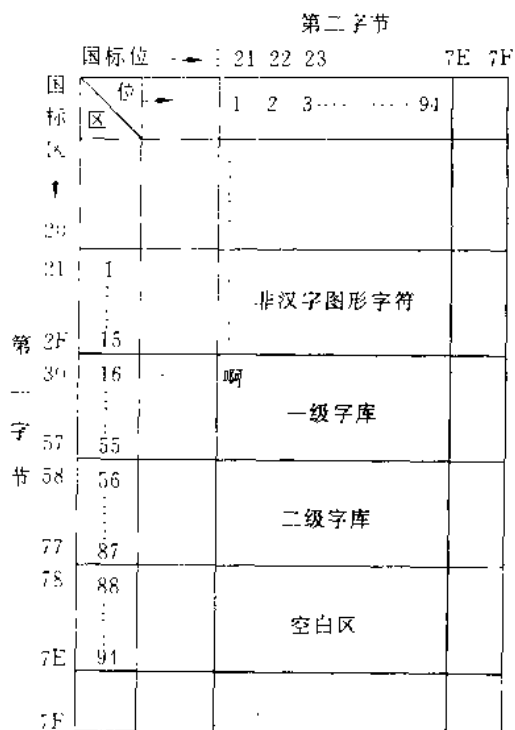


图 1-1

三、汉字输入码

◁必学/重点

汉字输入码又称外码。它是指从键盘上输入的代表一个汉字的全部或部分信息,是人与计算机进行信息交换时所采用的代码,也就是通常所指的汉字编码。对于同一个汉字,采用的编码方式不同,就会得到不同的输入码。例如:汉字“啊”,区位码输入方式下外码为“1601”,而在拼音方式下输入码为“a”,五笔字型输入方式下为“KBSK”等。

汉字输入码位于人机界面上,所以它应是易为用户接受、熟悉和掌握的,一般要求一种好的汉字输入编码应该具有以下特点:

- (1)易于记忆,甚至不需记忆;
- (2)编码的字符串(一般均使用字母或数字)尽量短,短的代码可获得较快的输入速度;
- (3)编码与汉字的对应性好,无重码或重码少。

所谓“重码”,就是一个代码和几个汉字相对应的现象。重码的存在,并不能看作是输入编码的“禁区”,可采用选择的方法确定实际的汉字,但是重码的存在无疑会降低键盘输入速度。

输入码的编码方案,人们已研究了不下数百种,并从中优选出十余种,但大多数还不完全理想。直到现在,汉字输入的问题仍然是影响汉字系统普及应用的关键问题之一。对汉字编码方案的研究仍在继续之中。在 IBM-PC 机上实现的汉字输入编码有多种,例如:

- (1)以 GB2312-80 为基准的国标码、区位码

国家标准信息交换用汉字编码(GB2312-80)所规定的机器内部编码同样可以用作输入码。每个汉字对应用 4 个十六进制的数字来表示,因此在输入时,只要键入四个数字字符即可

输入一个汉字,其优点是无重码,缺点是难以记忆。

例如“大”字的国标码是 3474,键入这四个数字键就可输入“大”字。

(2)以发音为基础的拼音码

这是以文字改革委员会公布的汉语拼音方案为基础的输入编码,只要懂汉语拼音的人都可掌握这种汉字输入方式。汉语拼音的编码方式,对于懂拼音的人无需专门进修,而且一旦掌握也无须记忆,因此人们乐于使用。但是,由于汉字的同音字为数众多,在键入一个拼音码的全部字母后,还必须进行同音字的选择(利用显示屏底部的专门的提示行来进行),故输入速度较低。

汉字系统对拼音码的输入采用了两种方法。第一种是完全按照汉字的拼音码,即所谓“全拼音”方式。比如汉字“王”,键入时只须用 WANG,然后在提示行中发音为 WANG 的众多汉字中选出“王”的编号即可。另一种则采用了简化的键入方法,对常用的复韵母和双字母的声母按规定作了简化。比如 WANG 中的 ANG 被简化为 H,于是“王”字的输入码为 WH。必须注意的是:一般情况下,全拼音或简化系统的输入方法不是两者兼可的,所以使用简化拼音的系统中将不允许全拼中的大部分复合韵母介入。因此,使用者必须记忆简化拼音码对应表。

(3)以字形为基础的码

这类编码的典型代表是五笔字型编码,它将汉字笔画划分为点、横、竖、撇、捺、折五种,并且规定了如何将一个汉字拆分成基本字根的一整套方案,它的特点是见字识码,重码率低,输入速度快,但难记。

(4)以发音和字形结合为前提的声韵部形码

这类编码以自然码为典型代表,自然码的编码方案由两部分组成,一是声韵双拼编码,另外一部分是形义编码,将一些常用的编码部首作为自然码的一部分,既可用拼音输入,又可用拼音加字形的输入,充分体现了音、形、义相结合的特征,成功地解决了“易学”与“高效”之间的矛盾。

(5)电讯中通用的电报码

这种编码是把邮电系统中普遍使用的电报码的编码方案直接作为汉字输入的方式,每个汉字用 4 位数字表示。邮电部门的专业人员可以方便地掌握这种汉字的编码方式。

(6)汉字的其它输入法

除了使用标准键盘,采用汉字编码的输入方式以外,还有一些其它的输入方式。

大键盘输入方式。在计算机系统中加配一个专门用于汉字输入的整字键盘(也称“汉字大键盘”),在各键上标有对应的汉字,按盘面的汉字字数分,有 2K 键盘(2000 字)和 4K 键盘(4000 字)。它借助键盘上的矩阵式开关产生信息,由一个键盘控制器通过不断地扫描查询而感知,并检测出按键的位置,然后转换成该汉字的内码。

这种操作是一键一字,操作方便、直观;但因盘中字数很多,非专职操作人员查找比较困难。然而对于汉字使用面不太广,涉及汉字种类不太多的使用环境,这种大键盘的工作方法是比较方便的。

另外,随着科技的进步,产生了一些更为理想的汉字输入设备,即利用汉字光学阅读器,直接阅读汉字。印刷体汉字的阅读设备已趋成熟,投入配套使用;手写体汉字的阅读设备也已开发,但由于这类设备的价格昂贵,一般不在微型机上使用。

四、汉字机内码

◁必学/重点

汉字机内码又称汉字内码,它是计算机系统内部进行存储、传递和运算所使用的代码,是汉字系统体系结构设计的基础。对于一个汉字系统来说,输入码可以是多种多样的,但机内码却是唯一的,同一汉字的不同输入码对应一个相同的机内码。

在计算机系统内部,西文字符的机内码一般都采用一个字节的 ASCII 码来表示。ASCII 码用七位二进制数进行编码,最高位置零。由 GB2312-80 可知,一个汉字要用两个字节(两个可显示的 ASCII 码字符)来表示,每个字节也只用了七位,其最高位未作定义。按理说,国标码可以直接作为计算机系统的机内码,但是,为了保证中西文兼容,这就意味着机内码中必须保持西文机内码(ASCII)的使用,同时又要允许国标码的使用,这就难免会出现二义性,无法区分是汉字机内码还是 ASCII 码。例如,机内有两字节,内容为 30H 和 21H,它们既可表示国标码“啊”,又可表示西文字符“0”和“1”的 ASCII 码。为了保证原西文系统功能,又能对汉字信息进行处理,人们对汉字机内码进行了设计,采用了把国标码的每个字节的最高位置 1 的方法形成汉字机内码,以示与 ASCII 码的区别,这种代码称为变形国标码。

这样,当两个字节的最高位都是 1 时表示汉字机内码,当两个字节的最高位都是 0 时,表示两个 ASCII 码字符,当两个字节中有一个最高位为 1 时,作未定义处理。这样,即解决了西文机内码与汉字机内码的二义性,又保证了汉字机内码与国标码之间有简单的对应关系,所以,汉字系统中一般都采用变形国际码作为机内码。

五、汉字字形码(字模)

◁必学/重点

汉字字形就是指确定一个汉字字形点阵的代码。由于汉字信息是以图形方式显示的,所以在进行汉字输出时,必须为汉字输出设备(打印机和显示器)提供汉字的字形数据——汉字字形码。

目前汉字信息处理系统中产生汉字字形的方式,大多是数字式的,即以点阵的方式形成汉字。这种方法不依赖于汉字本身的结构,技术简单,速度较快,并可与其它文字符号混用。

点阵法是将一个汉字平行投影到一个二维平面网格上,笔画所经过的网格单元用点来表示(见图 1-2),其余为空格,这样就得到了一个汉字字形的图形表示。在进行存储时每个计算机位对应一个表格单元,其值为 1 表示点,值为 0 表示空格,这样就把一个汉字的图形变成了二进制信息。对于 16×16 点阵来说,由于一行有 16 格,对应 16 个计算机位共占 2 个字节,一个汉字有 16 行,共需 $2 \times 16 = 32$ 字节的存储空间。这样的点阵即为一个汉字字形码或称一个字模。

我们以 16×16 的汉字字形点阵来说明一个汉字的字形码在计算机中是怎样表示的。

【例 2】我们来看一下“旦”字的字形码,如图 1-2 所示。

16×16 的点阵表示用 16 行、16 列的一个方阵来表示一个汉字。

六、汉字地址码

◁必学/难点

字形是以点阵形式存入计算机的,故需要占用较大的存储空间而不利于处理,所以在计算机内进行存储、传递、处理都采用机内码,字形码只有在输出过程中才有意义。因此,字形码被存放在一个名为汉字库的文件中,当系统需要输出时,再从汉字库中找出对应的汉字字模,进行输出。汉字地址码就是指汉字字模库中存储汉字字形码的逻辑地址码。

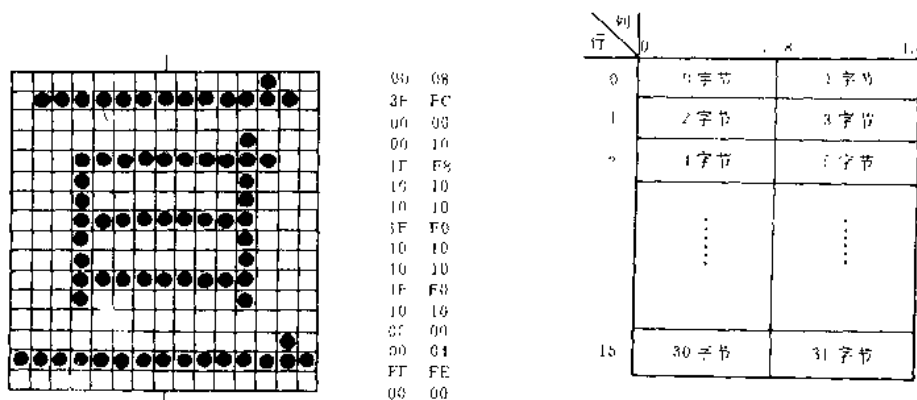


图 1-2

在进行输出时,系统按照输出汉字的机内码找到对应的地址码,再由地址码找出汉字库中对应的字形码。由于机内码与字形码之间有严格的一一对应关系,不允许有重码,所以机内码与地址码之间也是一一对应的。在汉字库中,汉字字形码一般是按国标码的顺序排列的,所以汉字地址码与汉字机内码之间有着很简单的换算关系,以便于查找汉字字形码。

如何计算汉字的地址码

自学答疑

例如,CCLIB 是存放 16×16 点阵汉字字模的字库。汉字点阵由 32 个字节组成。将 CCLIB 定义成记录长度为 32 字节的随机文件,那么每个汉字字模就是随机文件中的一个记录,而汉字的记录号与该汉字的区位码有关,对于 CCDOS2.10 的汉字库,可用下面的式子求出:

$$QC = \begin{cases} 6 & : \text{当 } QM = 9 \text{ 时} \\ QM & : \text{当 } QM > 1 \text{ 及 } QM \leq 15 \text{ 时} \\ QM - 8 & : \text{当 } QM \geq 16 \text{ 时} \end{cases}$$

$$JL = (QC - 1) \times 94 + WM$$

式中:

QM —— 某汉字的区码; WM —— 某汉字的位码;
QC —— CCLIB 的区码; JL —— 某汉字在 CCLIB 中的记录号。

例如,“啊”字的区位码是“1601”,根据上式可得:

$$QC = 16 - 8 = 8$$

$$JL = (8 - 1) \times 94 + 01 = 659$$

即“啊”字在 CCLIB 中的记录号是 659,也就是说“啊”字在 CCLIB 字库中对应第 659 号字形信息,由于每个字形占 32 个字节,所以“啊”字在字库中存储的物理地址应该是从 $659 \times 32 = 21408$ 字节开始存放的,到 21439 字节为止,故“啊”字的地址码为 21408。

说明:对于不同的汉字库文件,地址码的计算方法不完全一样,但所遵循的原理是相同的。

七、代码转换

必学/了解

由前讨论可知,在汉字系统中存在着多种汉字编码,在系统的不同部位采用不同的编码形