



中等职业学校
21世纪计算机规划教材



汉字录入 与编辑技术

□ 陈 潇 黄恺昕 编著

冶金工业出版社

中等职业学校 21 世纪计算机规划教材

汉字录入与编辑技术

陈 潇 黄恺昕 编著

北 京

冶金工业出版社

内 容 简 介

本书介绍了五笔字型、智能 ABC、微软拼音、98 王码、中文二笔输入法及郑码输入法等常见的汉字输入法，其中重点介绍了五笔字型汉字输入法的编码规则与具体输入法。本书从实际应用出发，讲述了文字编辑软件 Word 2003 的基础知识及在 Word 2003 环境下的文字、表格、图形的基本操作方法，并且在最后还有大量的上机指导与练习。通过对本书的学习，读者能够快速地从汉字输入到文件输出的所有技术。

本书既可作为中等职业学校计算机基础课程的教材，也可作为培训学校的培训教材或计算机初学者的自学参考书以及办公人员的学习书籍。

图书在版编目 (CIP) 数据

汉字录入与编辑技术 / 陈潇等编著. —北京: 冶金工业出版社, 2005.12

中等职业学校 21 世纪计算机规划教材

ISBN 7-5024-3869-6

I. 汉... II. 陈... III. 汉字信息处理—专业学校—教材 IV. TP391.12

中国版本图书馆 CIP 数据核字 (2005) 第 139392 号

出版人 曹胜利 (北京沙滩嵩祝院北巷 39 号, 邮编 100009)

责任编辑 戈兰

佛山市新粤中印刷有限公司印刷; 冶金工业出版社发行; 各地新华书店经销

2006 年 1 月第 1 版第 1 次印刷

787mm×1092mm 1/16; 12.5 印张; 282 千字; 190 页

20.00 元

冶金工业出版社发行部 电话: (010) 64044283 传真: (010) 64027893

冶金书店 地址: 北京东四西大街 46 号 (100711) 电话: (010) 65289081

(本社图书如有印装质量问题, 本社发行部负责退换)

前 言

一、关于本书

随着计算机的广泛应用,掌握文档的处理技术已成为办公人员必备的技能之一。本书精选了几种最常用的文字输入法予以介绍。汉字录入与编辑技术是中等职业技术学校的一门操作性很强的课程。

本书的主要目的是让学生在学习基本理论的基础上进行大量的实践操作,最后能够掌握 Word 软件进行文档处理。

二、本书结构

本书共分 10 章,其具体结构如下:

第 1 章:主要介绍了计算机汉字编码方法,汉字字库。

第 2 章:主要介绍了打字要求与指法以及 Windows XP 系统操作等内容。

第 3 章:主要介绍了汉字录入方法的基本知识。

第 4 章:主要介绍了五笔字型输入法的字根、编码及汉字输入方法。

第 5 章:主要介绍了智能 ABC 输入法和微软拼音输入法两种拼音输入法。

第 6 章:主要介绍了三种其他常用的输入方法:98 王码、中文二笔输入法、郑码输入法。

第 7 章:主要介绍了运用 Word 编辑软件进行文字、表格、图形的处理。

第 8 章:主要介绍了文档的排版方法、样式和模板的使用。

第 9 章:主要是各主要知识点的上机指导和练习。

第 10 章:综合练习。

建议本书教学至少使用 90 课时。在教学中,应重点突出汉字录入的教学,尽量在课程结束时,使学生的汉字录入速度能提高到较高的水平。在所有课时中,需要至少一半课时作为上机操作时间,学校条件允许的情况下可在讲授的同时进行操作。

三、本书特点

本书在内容编排上不但注重录入知识的启蒙作用,更注重内容的实用性和易掌握性,以培养读者的实际操作和应用能力为最终目的。

全书内容全面,语言简练,通俗易懂,并且在每章中插入了大量的实例和练习,方便读者在学习过程中巩固和提高。通过对本书的学习,读者能够快速地了解汉字录入与编辑技术,达到理想的学习目标。

四、本书适用对象

本书由暨南大学的陈潇、黄恺昕编写。

本书的作者是计算机教学第一线的教师,在长期的教学实践过程中积累了丰富的教学

实战经验。他们根据中等职业技术学校学生的认知规律和特点，总结归纳出一套理论联系实际的教学方法，在讲解各知识结构和技巧的同时，将重要的知识点融于实例中，这样，既便于教学又利于自学，为培养读者的灵活应用能力和创造能力奠定了基础。因此，本书不仅可作为中等职业学校计算机技术专业的教材，也可作为培训学校的培训教材或计算机初学者的自学参考书以及办公人员的学习书籍。

由于编者水平有限，编写时间仓促，书中疏漏之处在所难免，敬请读者朋友批评指正，联系方法如下：

电子邮箱：**service@cnbook.net**

网址：**www.cnbook.net**

本书的电子教案及习题参考答案可从该网站免费下载，此外，该网站还有一些其他相关书籍的介绍，可以方便读者选购参考。

编 者

2005 年 9 月

目 录

第 1 章 计算机汉字编码基础知识	1
1.1 编码及汉字处理	1
1.1.1 英文字符与 ASCII 码	1
1.1.2 汉字及其特点	2
1.1.3 计算机汉字处理	3
1.2 汉字编码	3
1.2.1 汉字字符集	3
1.2.2 汉字内码	4
1.2.3 汉字输入码	5
1.2.4 汉字字形码	5
1.2.5 汉字字库	6
小结	6
习题一	6
一、选择题	6
二、思考题	7
第 2 章 打字操作基础及系统使用	8
2.1 打字操作基础	8
2.1.1 打字姿势与要求	8
2.1.2 鼠标的操作	8
2.1.3 键盘的操作	9
2.2 Windows XP 操作基础	11
2.2.1 Windows XP 简介	11
2.2.2 登录和退出系统	12
2.2.3 Windows XP 的基本操作	14
2.2.4 中文输入法的设置	16
小结	16
习题二	17
一、选择题	17
二、思考题	17
第 3 章 汉字录入介绍	18
3.1 汉字录入概述	18
3.1.1 汉字录入的现状	18
3.1.2 汉字录入技术简介	18

3.2 键盘汉字录入技术简介	19
3.2.1 概述	19
3.2.2 汉字录入的编码方法	19
小结	20
习题三	21
一、选择题	21
二、思考题	21
第4章 五笔字型输入法	22
4.1 五笔字型简介	22
4.2 五笔字型输入法	22
4.2.1 汉字的结构	22
4.2.2 基本字根及其键位	25
4.2.3 汉字的编码规则	29
4.2.4 末笔字型交叉识别码	32
4.2.5 五笔字型汉字编码流程图	34
4.2.6 五笔字型字根表编码实例	34
4.2.7 简码	37
4.2.8 词组输入	38
4.2.9 容错码	39
4.2.10 Z键的作用	39
小结	40
习题四	40
一、选择题	40
二、思考题	40
第5章 拼音输入法	41
5.1 智能 ABC 输入法	41
5.1.1 智能 ABC 输入法的特点	41
5.1.2 输入法界面及操作	41
5.1.3 输入规则	43
5.1.4 使用技巧	45
5.2 微软拼音输入法	46
5.2.1 微软拼音输入法的特点	46
5.2.2 输入法界面及操作	48
5.2.3 输入规则	50
5.2.4 使用技巧	52
小结	53
习题五	54

一、选择题.....	54
二、思考题.....	55
第 6 章 其他常用汉字输入法	56
6.1 98 王码	56
6.1.1 98 王码简介	56
6.1.2 98 王码与 86 版五笔字型的区别	56
6.1.3 98 王码码元键盘	56
6.1.4 98 王码的输入方法.....	58
6.2 中文二笔输入法	59
6.2.1 二笔输入法简介	59
6.2.2 二笔输入法的特点.....	59
6.2.3 输入法界面及操作.....	59
6.2.4 编码规则	61
6.2.5 高级技巧	64
6.3 郑码输入法.....	65
6.3.1 郑码输入法简介	65
6.3.2 郑码输入法的特点.....	66
6.3.3 郑码输入法的键盘布局	66
6.3.4 输入基本过程	68
小结	70
习题六	70
一、选择题.....	70
二、思考题.....	71
第 7 章 Word 2003 文字处理	72
7.1 Word 2003 操作简介	72
7.1.1 启动/退出 Word 2003.....	72
7.1.2 Word 2003 的工作界面	72
7.2 输入和编辑文本	74
7.2.1 创建新文档	74
7.2.2 文本输入	74
7.2.3 选定文本	76
7.2.4 修改文本	77
7.2.5 查找和替换	78
7.2.6 拼写和语法检查	79
7.2.7 错误操作处理	80
7.2.8 保存文档.....	80
7.2.9 打开文档和插入文件.....	81

7.3 表格处理.....	81
7.3.1 创建表格.....	81
7.3.2 编辑表格.....	83
7.3.3 设置表格格式.....	84
7.3.4 表格数据处理.....	86
7.4 图形图片处理.....	88
7.4.1 绘制与编辑图形.....	88
7.4.2 插入图片与图片处理.....	90
7.4.3 插入艺术字.....	92
7.4.4 插入文本框.....	93
7.4.5 编辑数学公式.....	94
小结.....	96
习题七.....	96
一、选择题.....	96
二、思考题.....	98
第8章 文档排版.....	99
8.1 字符格式.....	99
8.1.1 字号、字体、效果.....	99
8.1.2 字符间距.....	101
8.1.3 文字效果.....	102
8.2 格式化段落.....	102
8.2.1 段落缩进.....	103
8.2.2 更改标尺单位.....	103
8.2.3 行间距与段间距.....	103
8.2.4 对齐方式.....	104
8.3 项目符号和编号.....	105
8.4 边框和底纹.....	106
8.4.1 为选定段落添加边框.....	106
8.4.2 为选定段落添加底纹.....	106
8.4.3 页面边框.....	107
8.5 页面设置.....	107
8.5.1 页边距.....	107
8.5.2 纸张类型.....	108
8.5.3 文档网格.....	108
8.5.4 页眉页脚.....	109
8.6 混合排版图片、文字.....	110
8.7 打印及打印预览.....	111
8.7.1 打印预览.....	111

8.7.2 打印文档	112
8.8 样式	113
8.8.1 建立新样式	113
8.8.2 修改样式	114
8.8.3 复制样式	114
8.8.4 删除样式	115
8.8.5 应用样式	115
8.9 模板	115
8.9.1 使用模板创建新文档	115
8.9.2 创建模板	116
8.9.3 应用模板	117
小结	118
习题八	118
一、选择题	118
二、思考题	119
第9章 上机练习	120
9.1 指法练习	120
9.1.1 键盘指法分工	120
9.1.2 数字键盘的指法练习	122
9.1.3 英文打字训练软件	123
9.2 汉字上机练习	126
9.2.1 汉字上机练习素材	126
9.2.2 86 版五笔字型输入法练习	137
9.2.3 98 版五笔字型输入法练习	147
9.2.4 智能 ABC 输入法练习	160
9.2.5 微软拼音输入法练习	161
9.2.6 二笔输入法练习	161
9.2.7 郑码输入法练习	171
9.3 输入和编辑文本上机练习	173
9.3.1 Word 2003 基本操作	173
9.3.2 Word 2003 文本处理	174
9.3.3 表格制作	175
9.3.4 图形处理	176
9.4 文档排版上机练习	180
第10章 综合操作	183
10.1 综合操作一	183
10.2 综合操作二	184

10.3	综合操作三.....	185
10.4	综合操作四.....	185
10.5	综合操作五.....	186
10.6	综合操作六.....	187
10.7	综合操作七.....	188
10.8	综合操作八.....	189
10.9	综合操作九.....	190

第 1 章 计算机汉字编码基础知识

教学目标

本章主要介绍计算机汉字输入的预备知识，要求读者了解计算机汉字输入处理的一些基本概念，共有如下两方面的内容：

- (1) 编码及汉字处理。
- (2) 汉字编码。

1.1 编码及汉字处理

计算机处理的不仅仅是一些数值信息，还需要处理大量的非数值符号信息，如英文字母、汉字等。例如，当用计算机写文章时，就需要先将各种符号、英文字母、汉字等基本信息输入计算机，然后再利用计算机进行编辑排版处理。

1.1.1 英文字符与 ASCII 码

1. 字符编码

字符编码是指用计算机能够识别的二进制代码表示各种数字、字母、专用符号和控制符号等。所有的字符编码都是有国家标准和国际标准之分的，因为它涉及世界范围内信息的表示、交换、处理、存储等基本问题。

在计算机系统中，主要的通用字符集编码有两种：

- (1) 主要应用于 IBM 大型机的 EBCDIC 码。
- (2) 主要应用于微型计算机和小型机的 ASCII 码。

2. ASCII 码

目前，计算机处理文字信息的编码标准绝大多数是使用的国际上通用的字符集编码——美国标准信息交换码 (American Standard Code for Information Interchange)，简称 ASCII 码。用 ASCII 码表示的字符称为 ASCII 字符。一般 ASCII 码有 7 位版本和 8 位版本之分。7 位版本的 ASCII 码采用 7 位二进制代码表示每一个字符，它能表示的字符数量为 $2^7 = 128$ 个。它们主要包括：10 个十进制数字符，26 个大写字母和 26 个小写字母，32 个运算符和标点符号，34 个通用控制字符。如表 1-1 所示为 7 位版本的 ASCII 码字符表。

表 1-1 ASCII 码字符表

$b_7b_6b_5$ $b_4b_3b_2b_1$	000	001	010	011	100	101	110	111
0000	NUL	DLE	SP	0	@	P	,	p
0001	SOH	DC1	!	1	A	Q	a	q
0010	STX	DC2	"	2	B	R	b	r
0011	ETX	DC3	#	3	C	S	c	s
0100	EOT	DC4	\$	4	D	T	d	t
0101	ENQ	NAK	%	5	E	U	e	u
0110	ACK	SYN	&	6	F	V	f	v

续表 1-1

$b_7b_6b_5$ $b_4b_3b_2b_1$	000	001	010	011	100	101	110	111
0111	BEL	ETB	'	7	G	W	g	w
1000	BS	CAN	(	8	H	X	h	x
1001	HT	EM	)	9	I	Y	i	y
1010	LF	SUB	*	:	J	Z	j	z
1011	VT	ESC	+	;	K	[	k	{
1100	FF	FS	,	<	L	\	l	
1101	CR	GS	-	=	M	]	m	}
1110	SO	RS	.	>	N	^	n	~
1111	SI	US	/	?	O		o	DEL

如果需要确定某个字符的 ASCII 码,首先从上表中找到该字符所在位置,然后再找出其对应的列 ($b_7b_6b_5$) 和行 ($b_4b_3b_2b_1$) 的编码,将编码按列、行 ($b_7b_6b_5b_4b_3b_2b_1$) 组合便是该字符的 ASCII 码。

需要特别指出的是,十进制数字字符的 ASCII 码与它们的二进制数表达法是有区别的。例如,十进制数 3 的七位二进制表示为 (0000011),而十进制数字字符“3”的 ASCII 码为: $(0110011)_2 = (33)_{16} = (51)_{10}$

由此可以看出,数 3 与数字字符“3”在计算机中的表示是不一样的。数 3 表示的是数的大小,并可以参与数学运算;而数字字符“3”只是一个符号,它不能参与数学运算。

8 位版本的 ASCII 码使用 8 位二进制代码来编码,当最高位为 0 时,表示基本 ASCII 码;当最高位为 1 时,表示扩充 ASCII 码。扩充的 ASCII 码通常被各国用于自己语言文字的代码。

1.1.2 汉字及其特点

各国现在使用的文字大体上可分为两种类型:标音文字和表意文字。标音文字也叫拼音文字,是用一套简单的表意字母来拼写复杂的表意文字;表意文字是直接以大量的复杂的符号系统表达的文字。

汉字是记录汉语的书面符号系统,属于表意文字,也是世界上历史最悠久的文字之一。世界上使用汉语的人口最多,约占世界人口的五分之一,因此汉语是联合国指定的七种工作语言之一。

与世界上的其他语言相比,汉语具有自己的特点:

(1) 汉字属于表意文字,一个字记录一个音节,同时表示一定的意义,从而用汉字记录汉语很方便。但是,这种表意文字不可避免地会出现众多的文字,最新的汉语字典中有六万多个汉字,即使国家标准局公布的标准汉字编码 B2312-80 中也收录了 6763 个汉字。众多的汉字给计算机信息处理带来了困难。

(2) 汉字适应汉语同音语素多的特点。汉字用不同形体的方块字记录同音语素,使它们在书面表达上有了区别,例如“巴、八、爸、扒”等。但是,同音字加上多音字,会造成计算机拼音编码重码率高的缺点。

(3) 汉字具有“超方言”的魔力,适应汉语方言复杂的特点。汉语方言复杂,同一

个语素，用方音来念，可以迥然不同。两千年来，汉字形体上没有发生大的变化。统一而稳定的汉字，使得不同方言区的人可以看懂彼此写的书面语。形体复杂的汉字可能对稳定汉字字形有一定的促进作用，但是，却给汉字的计算机键盘输入带来了难题。

(4) 汉字独具艺术魅力，它们给华夏文化带来丰富的内容。历史悠久的华夏文化创造出多种多样的漂亮的汉字字体，例如宋体、隶书、楷书等。富于变化的汉字形体，为其成为审美对象提供了足够的条件。要用计算机展示出各种各样的汉字字体，势必需要各种形体的汉字库，从而需要占用大量的计算机存储空间。

1.1.3 计算机汉字处理

电子计算机内部能够处理的信息，只能是数字化的。对拼音文字来说，一个字符就对应一个编码，因而输入非常方便，一种拼音文字只有一种基本的输入方法，会打字母就会输入文字。

汉字输入计算机，遇到三大难题：数量大；字形复杂；存在大量同音字和多音字。因此，设计出简便而高效的编码方案，是有困难的。不过，中国的计算机专家经过 10 多年的悉心研究，成功地开发出了若干种成熟的汉字信息处理系统。

目前，汉字编码方案有 500 多种，分属音码、字形码、音形码等。这些方案各有优略。好的方案必须是输入方便、处理容易、存储节省、重码率低、输入和处理效率高、传输可靠、设备经济实用等等。

汉字处理的流程中，需要分别采用不同的编码。例如，输入汉字时要用输入码；计算机处理汉字时需要采用机内码；显示汉字时要用字型显示码；交换汉字时又要用交换码。

1.2 汉字编码

1.2.1 汉字字符集

汉字字符集是计算机或系统间进行通信或交换信息的标准代码，它是中文信息处理的基础。

1. 汉字国标码

国标码是国家标准汉字编码的简称。GB2312-80 是国家标准局公布的标准汉字编码，在 GB2312-80 中共收录了汉字 6763 个，非汉字图形符号 682 个。

在这个标准中，每个字符用双字节二进制编码表示，每个字节的最高位都置 0，只用低 7 位表示，其中有 34 个编码起控制作用，所以每个字节只能有 94 个编码用于表示字符，两个字节总共可以表示的字符数就是 94×94 ，即 8836 种不同的字符，也就组成一个 94×94 的矩阵。

在这个矩阵中，每一行称为一个“区”，每一列称为一个“位”。每一个汉字的国标码是由一个区码和一个位码所组成，所以一个汉字的区位码用十六进制数表示就是四位数，其中前两位是区码，后两位是位码，如：“林”字的区位码为“3354H”，“字”字的位码为“5178H”。

在区位码矩阵中，94 个区可以分为两部分：

(1) 1~15 区为图形符号区，其中，1~9 区为标准符号区，10~15 区为自定义符号区。

(2) 16~94 区为汉字区, 其中汉字又可以分为三部分:

- ① 16~55 区为一级常用汉字区, 共 3755 个汉字, 这个区是按汉字的拼音排序的。
- ② 56~87 区为二级非常用汉字区, 共 3008 个汉字, 这个区是按汉字的部首排序的。
- ③ 88~94 区为用户自定义汉字区。

通过对常用汉字和非常用汉字进行区分, 可以提高系统在字库中搜索汉字的效率。

2. GBK 字符集

我国虽然颁布了汉字国标码, 但是, 只有大陆执行该标准, 而在我国台湾, 港澳地区多用 BIG-5 码, 世界其他地区的汉字文化圈中也存在一些其他的汉字编码方案。汉字编码的不统一, 造成了各种汉字处理系统之间无法通用的麻烦。为使世界上包括汉字在内的各种文字的编码走上标准化、规范化的道路, 1992 年 5 月国际标准化组织 ISO 通过了 ISO/IEC10640 号标准, 即《通用多八位编码集 (UCS)》, 同时我国也制定了新的国家标准 GB13000-1993 (简称 CJK 字符集)。全国信息标准化技术委员会在此基础上发布了《汉字扩展内规范》, 其中收集了中国、日本、韩国三国汉字共 20902 个 (大字符集字库, 简称 GBK 字符集), 可以在很大程度上满足汉字处理的要求。

3. 汉字扩充编码

国家标准 GB18030-2000《信息交换用汉字编码字符集基本集的扩充》是我国继 GB2312-1980 和 GB13000-1993 之后最重要的汉字编码标准, 是未来我国计算机系统必须遵循的基础性标准之一。

GB 18030 是 GB 2312 的扩展, 收录了 27484 个汉字, 总编码空间超过 150 万个码位, 为解决人名、地名等用字问题提供了解决方案, 为汉字研究、古籍整理等领域提供了统一的信息平台。

GB 18030 采用单/双/四字节混合编码。双字节部分收录内容主要包括 GB13000.1 全部 CJK 汉字 (20902 个)、有关标点符号、表意文字描述符 (13 个)、增补的汉字和部首/构件 (80 个)、双字节编码的欧元符号等; 四字节部分收录了上述双字节字符之外的, 包括 CJK 统一汉字扩充 A 在内的 GB 13000.1 中的全部字符。

GB18030-2000 在字汇上与 GB 13000.1-1993 兼容, 并包含了其中收录的所有汉字、藏、蒙、维等少数民族文字, 以及世界上几乎所有的语言文字, 为中文信息在因特网上的传输和交换提供了保障。

1.2.2 汉字内码

汉字内码也称机内码, 是汉字在计算机内部进行存储、传输的代码, 是由区位码演变而来的。实际上它是将每个汉字的区码和位码的最高位置“1”而形成的, 所以汉字内码也由两个字节表示。汉字内码的最高位置“1”, 其实是为了汉字在计算机内部进行传输时, 能够与 ASCII 代码相区分的标志位, 也就是说每一个字节的代码的最高位是区分 ASCII 码和汉字代码的标志。如:

“林”的国标码: 00110011 01010100

“林”的内码: 10110011 11010100

区码和位码的范围都在 01~94 内, 如果直接用它作为内码就会与基本 ASCII 码发生冲突, 因此汉字的内码采用如下的运算规定:

高位内码 = 区码 + 20H + 80H

低位内码 = 位码 + 20H + 80H

在上述运算规则中加 20H 应理解为基本 ASCII 的控制码；加 80H 意在把二进制最高位置“1”，以与基本 ASCII 码相区别，或者说是识别是否是汉字的标志位。

例：将汉字“玻”的区位码转换成机内码：

$$\begin{aligned}\text{高位内码} &= (18)_{10} + (20)_{16} + (80)_{16} \\ &= (00\ 010\ 010)_2 + (00\ 100\ 000)_2 + (10\ 000\ 000)_2 \\ &= (10\ 110\ 010)_2 \\ &= (B2)_{16} = B2H\end{aligned}$$

$$\begin{aligned}\text{低位内码} &= (3)_{10} + (20)_{16} + (80)_{16} \\ &= (00\ 000\ 011)_2 + (00\ 100\ 000)_2 + (10\ 000\ 000)_2 \\ &= (10\ 100\ 011)_2 \\ &= (A3)_{16} = A3H\end{aligned}$$

$$\begin{aligned}\text{内码} &= (\text{区码} + 20H + 80H)(\text{位码} + 20H + 80H) \\ &= (1\ 011\ 001\ 010\ 100\ 011)_2 = B2A3H\end{aligned}$$

1.2.3 汉字输入码

目前汉字的输入法主要有键盘输入、文字识别和语音识别。键盘输入是用汉字的键盘编码（又称汉字外码），通过键盘输入汉字。目前较常用的汉字键盘输入码的编码方案有 500 多种，不同的输入法的同一个汉字编码各不相同，其编码的长度也不统一，不过，大多数采用 4 个键。

国内常用的汉字输入码有下面几种类型：

- (1) 以国标 GB2312-80，GBK 字符集为基准的国标码，也有人称数字码。
- (2) 以汉字拼音方案为基础的音码。
- (3) 以汉字构形为基础的形码。
- (4) 以音码加形码的混合码，也称为音形码。

输入码到机内码的转换依赖于输入码表。每一种输入码都有其各自的转换码表。

1.2.4 汉字字形码

汉字字形码也称字模，是供计算机输出（显示和打印）汉字用的二进制信息。目前汉字信息处理系统中产生汉字字形的方式大多数是数字式的，即以点阵的方式形成汉字。例如，如果一个汉字是由 16×16 个点阵组成的，那么就需要有一个 16×16 的点阵数据与之对应，则一个汉字占 16 行，每行上有 16 个点，每一个点在存储器中用一个二进制的位来存储，值为“0”表示暗，值为“1”表示亮。

根据不同字体的不同要求，点阵的大小也有不同，通常使用的点阵有： 16×16 、 24×24 、 32×32 、 48×48 、 64×64 、 128×128 点阵等。由于计算机存储器的每个字节由 8 位二进制组成，如 32 个点则需要 4 个字节来存放。

如果使用的点阵越大，那么所占用的存储空间也越大，当然，其输出的汉字字模质量就越好。例如， 32×32 的字模质量就会比 16×16 的好，但其所使用的存储空间就会相差一

倍, 32×32 点阵图如图 1-1 所示。

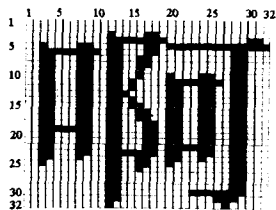


图 1-1

$$16 \times 16 \div 8 = 32 \text{ 字节}$$

$$32 \times 32 \div 8 = 128 \text{ 字节}$$

1.2.5 汉字字库

1. 什么是汉字字库

全部汉字字形码的集合叫汉字字库, 也称汉字字模库。一个 32×32 点阵的汉字需要 128 个字节的存储空间, 对于 GB2312-80 (6763 个汉字) 则需要 $6763 \times 128 = 865664$ 个字节的存储空间。

2. 硬字库和软字库

硬字库将字库固化在一个单独的存储芯片中, 再与其他必要的器件组成接口卡 (通常称为汉卡), 插接在计算机上, 早期多采用此方式。

软字库是以文件的形式存放在硬盘上, 是目前计算机使用的方式。

3. 显示字库和打印字库

显示字库用于在屏幕上显示文字之用, 所以, 当系统启动时, 显示字库就会被调入内存备用。为了取得快速的显示速度, 显示字库的字模点阵比较小, 通常为 16×16 点阵、24×24 点阵或 48×48 点阵。

打印字库用于打印输出之用, 因此, 只有输出时才调用打印字库。打印字库的字模点阵要比显示字库的大, 从而取得更好的输出质量。

小结

本章主要介绍了如下内容:

(1) 文字编码。介绍了字符编码的用途, 以及字符编码所使用的标准: EBCDIC 码和 ASCII 码。

(2) 汉字处理。汉字属于表意文字, 具有数量大, 字形复杂, 存在大量同音字和多音字的特点。因此, 针对以上特点, 介绍了汉字编码方案必须达到的要求。

(3) 汉字编码。介绍计算机汉字处理的标准 (汉字字符集), 计算机处理过程所使用的汉字内码、汉字输入码、汉字字形码和汉字字库。

习题一

一、选择题

1. 在计算机中汉字信息的存储和运算都是采用 ()。