

计算机中文信息处理

下册

赵珀璋

徐

力

编著

جوڭگو يېزى ئىقلىرى ئېنگۇر ماتىسىيە لىرىنى

ھىسا بىلاش ماشىنىسىدا بىر تە رەپ قىلىش
جوڭگو جازۇلارى ئېنگۇر ماتىسىياسىن
سە يىتە ۋە ماشىناسىندا ئېز جايلى ۋە تە

전자제생기중국 언어문자 정보처리

计算机中文信息处理

计算机中文信息处理

计算机
中文信息处理
下 册

宇航出版社

内 容 简 介

本书全面系统地论述了计算机的中文信息处理和中文信息的计算机处理有关基础知识、基本原理、设计要领和使用方法。全书共分十四章,下册从第九章开始,书后的附录一、二分别为思考题和综合试题。附录三至八分别给出了 GB1988—80“信息处理交换用的七位编码字符集”、GB2311—80“信息处理交换用的七位编码字符集的扩充方法”、GB2312—80“信息交换用汉字编码字符集 基本集”、GB5007.1~5007.2—85“信息交换用汉字 24×24 点阵字模集及数据集”、GB5199.1~5199.2~85“信息交换用汉字 15×16 点阵字模集及数据集”、GB6345.1~6345.2—86“信息交换用汉字 32×32 点阵字模集及数据集”的有关节录。

这是第一本作为我国研究生教材的中文信息处理技术的书籍,可作为高校教材,也可供从事计算机科学、计算机工程、计算机应用和中文信息处理系统研制、生产、开发和使用部门的科技、管理人员参考,对提高和普及我国计算机的广泛应用,对文字改革及语言研究将会起到促进作用,对世界各国研究中文信息处理技术,也将起到一定的推动作用。

计算机中文信息处理(Jìsuáncì Zhōngwén Xīnxi Chùlǐ)

下 册(Xiàcè)

编著:赵珀璋(Zhào pò zhang)

徐 力(Xú Lì)

责任编辑:廖寿琪(Liào shòu-qí)

*

宇航出版社出版 新华书店北京发行所发行

国防工业出版社激光照排室用华光型机照排

各地新华书店经销

天津市静一胶印厂印刷

*

开本: 787×1092 1/16 印张: 33.875 字数: 845 千字

1989年6月第1版第1次印刷 印数:1 - 4000册

ISBN 7-80034-005-8/TP·002 定价: 15.50 元

目 录

第九章 中文信息处理系统	1
9.1 汉字终端	2
9.1.1 汉字终端系列的构成	3
9.1.2 汉字终端系列的技术条件	4
9.2 汉字微型机系统	6
9.2.1 硬件系统	6
9.2.2 软件系统	7
9.2.3 中西文兼容微型机	8
9.3 汉字输入键盘	8
9.3.1 全键式整字大键盘	8
9.3.2 笔触式整字大键盘	9
9.3.3 整字盘盘面标准设计	9
9.3.4 整字键盘结构设计	11
9.4 CKBD—III 型汉字智能键盘	17
9.4.1 键盘的构成	18
9.4.2 工作原理	19
9.5 汉字显示终端	19
9.5.1 字符式汉字终端	20
9.5.2 图形式汉字终端	20
9.5.3 集群式汉字显示终端	21
9.5.4 汉字显示终端与计算机的连接方式	25
9.6 独立式汉字显示终端举例	25
9.6.1 CCRT—III 汉字智能显示器	25
9.6.2 CCRT—III 型显示控制器	26
9.6.3 TRN—III 型中西文智能终端	32
9.6.4 软件配置	34
9.6.5 功能码解释及终端代码	38

9.7 图形终端.....	44
9.7.1 迭代法.....	44
9.7.2 插补法.....	48
9.8 汉字印字技术.....	51
9.8.1 印字机的分类.....	51
9.8.2 点阵冲击式印字机.....	52
9.8.3 热敏印字机.....	54
9.8.4 喷墨式印字机.....	55
9.8.5 其它种类印字机.....	56
9.8.6 汉字印字方式选择.....	57
9.8.7 汉字印字机.....	57
9.8.8 汉字印字机系统举例.....	58
9.8.9 汉字尺寸、点阵与分辨率的关系.....	64
9.9 汉字制表.....	64
9.9.1 名词术语.....	64
9.9.2 汉字制表举例.....	68
9.9.3 格式描述语言.....	70
9.9.4 汉字报表语言的设计.....	77
9.9.5 中文版式设计语言.....	81
9.10 中西文兼容计算机系统.....	85
9.10.1 “化合式”中西文兼容计算机系统.....	85
9.10.2 “混合式”中西文兼容计算机系统.....	86
9.10.3 终端仿真举例.....	88
9.11 通用汉字报表处理应用程序.....	109
9.11.1 文件结构及名词术语.....	110
9.11.2 应用程序的功能介绍.....	112
9.12 计算机-激光汉字编辑排版系统.....	123
9.12.1 激光照排机光学系统.....	124
9.12.2 激光编排系统软件.....	124
9.12.3 报版编译系统.....	127
9.12.4 精密字模自动制作与补字系统.....	128
9.13 通用汉字打印模块设计.....	132
9.13.1 通用汉字打印模块的提出.....	132
9.13.2 通用汉字打印模块设计.....	132
9.13.3 打印机参数表生成程序举例.....	134
9.14 小结.....	137
参考文献.....	138
第十章 中文信息通信	140
10.1 数据通信基础知识.....	142
10.1.1 数据传输线路分类.....	142
10.1.2 数据通信方式分类.....	143
10.1.3 数据交换方式分类.....	144
10.2 传输控制规程.....	145
10.2.1 面向字符的传输控制规程.....	145

10.2.2 面向比特的传输控制规程	146
10.3 汉字通信有关问题	146
10.4 汉字联机系统	149
10.5 通信规程转换器	151
10.5.1 通信规程转换器的组成	151
10.5.2 通信规程转换器的基本功能	153
10.5.3 通信规程转换器的软件结构	153
10.5.4 汉字智能多路规程仿真器	154
10.5.5 通信规程监测器	158
10.6 国家标准 GB3453-82“数据通信基本型控制规程”及高级数据链路控制规程(HDLC)简介	161
10.6.1 字符结构	161
10.6.2 传输控制字符	162
10.6.3 电文格式	163
10.6.4 数据通信阶段	165
10.6.5 差错控制	166
10.6.6 恢复规程	166
10.6.7 会话型传输控制规程	166
10.6.8 编码独立的信息传输控制规程	167
10.6.9 多站选择	167
10.6.10 全双向传输控制规程	167
10.6.11 高级数据链路控制规程 HDLC 简介	167
10.7 中文计算机网络	171
10.7.1 微型机局域网概述	172
10.7.2 OMNINET 局域网	179
10.7.3 以太 (ETHERNET) 局域网	189
10.7.4 中西文兼容局域网络举例	193
10.7.5 网络层次设计若干问题	198
10.7.6 中西文兼容电文传输网简介	199
参考文献	202
第十一章 汉字和汉语语音识别处理	204
11.1 汉字的特征与表示	206
11.2 汉字识别的预处理	208
11.2.1 汉字的平滑处理	210
11.2.2 汉字的细化处理	211
11.3 汉字笔画有向特征抽取方法	215
11.4 印刷体汉字识别法	219
11.4.1 结构分析法	219
11.4.2 模板匹配法	222
11.4.3 几种判别函数方法	225
11.4.4 象限端点转动惯量识别法	226
11.5 汉字识别中的分类法	228
11.5.1 复杂指数分类法	228
11.5.2 四边代码分类法	229

11.6	印刷体汉字识别的粗分类方法	230
11.6.1	汉字R变换码粗分类方法	230
11.6.2	汉字外轮廓四角八码粗分类方法	231
11.6.3	汉字内轮廓四角八码粗分类方法	233
11.6.4	汉字脉冲码粗分类方法	234
11.6.5	汉字长短的粗分类方法	236
11.7	轮廓投影识别法	236
11.8	手写体汉字识别	238
11.8.1	联机识别系统	239
11.8.2	汉字直接写入法	246
11.8.3	脱机识别系统	247
11.9	手写体汉字数据库	251
11.9.1	正规化二值图案的制作	251
11.9.2	矩形外接文字框的检出	252
11.9.3	正规化	253
11.9.4	手写体汉字数据库形式	253
11.10	模糊数学简介	253
11.10.1	模糊集合的定义	255
11.10.2	模糊集合的运算	255
11.10.3	模糊关系	266
11.10.4	模糊聚类分析	268
11.10.5	模糊聚类分析步骤	272
11.11	汉字模糊笔画特征抽取方法	275
11.11.1	汉字笔画特征	276
11.11.2	特征判别方法	276
11.12	汉字模糊笔画笔顺分析	278
11.12.1	汉字的模糊结构	278
11.12.2	笔画之间的关系	280
11.12.3	汉字的二维构形向一维笔序的变换	281
11.13	汉字树形识别法	282
11.13.1	预处理	282
11.13.2	汉字字根分离	283
11.13.3	汉字字根识别	284
11.13.4	字根集的模糊聚类分析	285
11.13.5	字根识别字典的自动生成	286
11.13.6	字根识别算法	288
11.13.7	汉字语法树的生成	288
11.13.8	汉字组装	288
11.13.9	汉字字典的查找	288
11.13.10	识别系统流程图	289
11.14	语音信号处理基础	289
11.14.1	语音信号	289
11.14.2	语音信号处理	290
11.14.3	数字信号处理	291

11·14·4	数字语音处理	291
11·14·5	语音处理的时域方法	293
11·14·6	语音波形的数字表示	295
11·14·7	短时傅里叶分析	295
11·14·8	同态语音处理	296
11·14·9	语音的线性预测编码	296
11·14·10	人机语声通信的数字语音处理	296
11·15	汉语的语音特点	300
11·16	微型机语音识别接口	302
11·16·1	语音参数表达式	303
11·16·2	实时端点测定	304
11·16·3	特征提取和规正	306
11·16·4	训练	307
11·16·5	识别	308
11·16·6	软件结构及调用	310
11·16·7	试验考核情况	312
11·17	口呼识别系统	317
11·17·1	语音图样量化方法	317
11·17·2	差距的计量法	319
11·17·3	100个中国地名的口呼识别实验	320
11·18	汉语普通话四声识别	322
11·19	计算语音学	329
11·19·1	CSIPS 系统功能	330
11·19·2	CSIPS 系统结构	331
11·19·3	汉字语音输出系统	332
11·19·4	CSIPS 1000 系统	335
11·20	数学形态学原理及其基本运算	340
11·20·1	蚀与胀运算	340
11·20·2	开与闭运算	343
11·20·3	形态学特点及并行算法	344
11·21	利用形态方法识别国家标准 GB2312—80 基本集印刷体汉字	345
11·21·1	边框特征和第一级分类	345
11·21·2	局部特征和第二级分类	346
11·21·3	端点、结点和识别	346
11·21·4	结构单元及其提取	346
11·21·5	边框特征提取	347
11·21·6	端点和结点特征提取	348
11·21·7	局部特征提取	348
11·21·8	匹配运算	349
	参考文献	349
	第十二章 汉语自然语言处理	351
12·1	语言工程	352
12·2	汉语语言理解系统	355

12.3	汉语拼音输入系统	359
12.3.1	汉语拼音输入系统 HP-2	359
12.3.2	自动切分系统	362
12.3.3	同音自动处理系统	362
12.4	机器翻译概述	363
12.5	机器翻译的两个主要方面	365
12.5.1	机器翻译中汉语有关问题	365
12.5.2	机器翻译专用软件包	368
12.6	机器翻译系统设计	370
12.6.1	机器翻译的分类法	370
12.6.2	机器翻译程序设计方法	371
12.7	机器翻译中目标语的生成	373
12.7.1	短语(成份)词组的转换与生成	373
12.7.2	以动词为核心的句法结构转换与生成	374
12.8	自然对话智能机	376
12.8.1	自然对话基本方法	377
12.8.2	语义框架及其在语言理解过程中的操作	377
12.9	中文信息五维模型	379
12.10	自然语言信息处理的心理学过程	383
12.11	汉语词汇自动统计	384
12.11.1	什么是词	384
12.11.2	词汇切分理论	384
12.11.3	词汇切分方法	386
12.11.4	切词词典建立	388
12.11.5	系统结构	391
12.12	汉语的结构分析	393
12.12.1	单词的分类	393
12.12.2	短语的分类	393
12.12.3	句子的分类	393
12.12.4	汉语的里层结构	394
12.12.5	汉语的表层结构	395
12.12.6	计算机汉语理解	396
12.12.7	计算机汉语生成	396
12.13	计算语言学	397
12.14	中文信息的知识处理	398
	参考文献	399
	第十三章 少数民族文字处理	401
13.1	蒙古文信息处理	402
13.1.1	蒙古文信息处理交换用编码字符	404
13.1.2	蒙古文键盘设计	405
13.1.3	蒙古文显示终端设计	405
13.1.4	蒙古文打印机设计思想	408
13.2	满文信息处理	408
13.3	朝鲜文信息处理	409

13.3.1 朝鲜文编码字符集	410
13.3.2 朝鲜文键盘设计	411
13.4 藏文信息处理	412
13.4.1 藏文的元音	412
13.4.2 藏文的结构	413
13.4.3 藏文编码字符集	414
13.4.4 TCES 藏文信息处理系统	415
13.5 维吾尔文信息处理	415
13.5.1 维吾尔文特殊的排列格式	418
13.5.2 维吾尔文编码字符集	421
13.5.3 维吾尔文键盘	425
13.5.4 维吾尔文的屏幕编辑	428
13.5.5 维吾尔文印字技术	434
13.5.6 维吾尔文字形设计	435
13.6 彝文信息处理	435
13.7 小结	437
参考文献	440
第十四章 中文信息处理有关标准	442
14.1 国家已颁布的有关标准	442
14.2 国家正在制订或将要制订的有关标准	443
14.3 国家标准“文字和符号成形设备的增补控制功能”编制说明	444
14.3.1 编制原则	444
14.3.2 几个具体问题	444
14.4 国家标准“信息交换用汉字编码字符集 辅助集”设计思想	450
14.5 汉字编码字符集的使用	450
附录一 思考题	455
附录二 综合试题	460
附录三 GB1988—80《信息处理交换用的七位编码字符集》	465
附录四 GB2311—80《信息处理交换用七位编码字符集的扩充方法》	473
附录五 GB2312—80《信息交换用汉字编码字符集 基本集》	484
附录六 GB5007.1~5007.2—85《信息交换用汉字 24×24 点阵字模集及数据集》	516
附录七 GB 5199.1~5199.2—85 《信息交换用汉字 15×16 点阵字模集及数据》	521
附录八 GB6345.1~6345.2—86《信息交换用汉字 32×32 点阵字模集及数据集》	527

本章首先论述中文信息处理系统的构成及其性能指标,然后举几个例子,目的是使读者能够得到中文信息处理的完整概念。

中文信息处理系统如何划档分类,目前还没有统一的看法。这里介绍的是笔者的个人意见。根据就是中文信息处理与西文没有本质的区别。甚至可以说,从信息处理角度来看,中西文信息处理是一致的。显然在系统结构及其分类上,也应该是一致的,而不要人为地把中西文信息处理分开,致使中西文信息软硬件系统不能兼容、资源不能共享,这对中文信息的处理也是不利的。

中文信息处理系统结构框图如图 9—1 所示。这是从系统功能配置角度来划分的。共分三级:

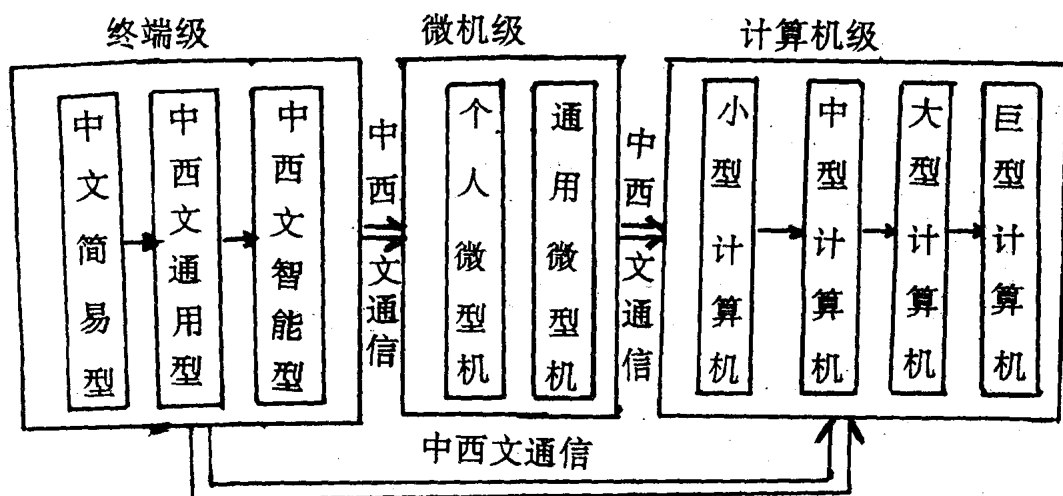


图 9—1 中文信息处理系统结构框图

1. 终端级

终端级由中文简易型终端、中西文通用型终端和中西文智能型终端三种组成。它是中文信息处理系统的主要部件,也是关键部件之所在,是人机接口界面。因其需要量大,故要求成本低。在某种意义上说,终端级的性能价格比直接影响着系统的性能价格比。

2. 微型机级

中文微型机和西文微型机应向上兼容,不必搞出一种只能处理中文信息而不能处理西文信息的纯中文微型机。它与终端的差别在于:

- (1)系统结构上:终端没有标准总线,而微型机具有标准总线结构。
- (2)软件系统上:终端只有监控程序,而微型机配有操作系统。
- (3)应用功能上:终端主要作联机用,而微型机主要是单机工作或联网工作。

3. 计算机级

计算机级分为小型机、中型机、大型机和巨型机四大类。相应地构成小、中、大、巨四个系统。它具有数据库、计算机网络等丰富的软硬件资源。中文信息处理系统研制生产的应用开发都必须充分注意现有计算机系统软硬件资源。

9 • 1 汉字终端

1. 汉字终端是人机接口的界面。换句话说,人和计算机打交道是通过终端进行的,计算机有无处理中文信息的能力是通过终端表现出来的。因此说,中文信息处理系统的最低配置,其终端必须具备处理中文信息的能力。

2. 汉字终端是中文信息处理系统的基础设备之一。在主要使用汉语进行交流信息的我国国情下,只有配置了汉字终端才有可能充分利用计算机系统的软硬件资源,把计算机的功能和应用范围扩充到中文信息处理中。

3. 汉字终端是一种综合性强的基本设备。它具有输入输出功能,具有很强的联机处理能力和一定的脱机处理能力。

因此可以说,终端设备是一种承上启下的设备。对其性能价格比要求较高,因为任何一个系统都需要大量的终端设备。

中文信息处理目前已进入新的阶段,主要表现在:

1. 配置方面:

由简单终端到系统,由单用户到多用户,由单机到计算机网络。

2. 输入方面

由单一方案到多种方案兼容;由键盘输入到图形报表、语音识别输入和文字识别输入。

3. 字库方面

由软字库到硬字库,由低分辨率到高分辨率,由单种字型字号到多种字型字号,由集中式字库到分布式字库。

4. 显示方面

由单色到彩色,由字符式到图形式,由低分辨率到高分辨率,由单屏到多屏显示。

5. 输出方面

由针式到非击打式印字系统,由字符到图形字符混合输出,由单种字型字号到多种字型字号,由文字输出到语音输出。

6. 软件方面

由应用级处理中文发展到系统级支持中文,进而应用软件支持中西文兼容,由终端、单机到通信网络技术和数据库技术。

7. 系统方面

由应用级发展到系统级,进而发展到生成级系统。从而能够实现较高的中西文兼容能力。

汉字终端是一种能够通过通信信道发送和接收汉字信息的设备。它主要以联机方式工作,通常由一个汉字输入设备和某种类型的汉字显示设备组成。

9.1.1 汉字终端系列的构成

1. 简易汉字终端

此类汉字终端不提供执行用户程序的功能,其处理部分由主机完成。此类终端具有下述功能:

1)能输入输出汉字、字符和图形符号。

2)保持西文哑吧型终端的显示属性。

3)具有基本屏幕编辑功能。

4)能接受二种以上汉字输入编码方法。

5)在主机控制下,依靠主机资源,实现某些汉字信息处理功能,并保持同类西文终端的全部功能。

6)适用于工业控制领域的简易汉字终端,还能实现工业过程显示,如趋势显示、极值显示、误差分析和分数比较等。

7)具有自检功能

此类终端主要用作电报终端和电传机。可选配汉字印字机或其它输出设备,并配有国家标准汉字交换码和国家标准 15×16 或 24×24 字模点阵汉字字形库。

2. 通用汉字终端

这类汉字终端可不提供执行用户程序的能力。由主机控制或自身管理程序支持实现初步的汉字和字符信息处理能力。

通用汉字终端应具有下述功能:

1)能输入输出汉字、字符、图形符号和图形。

2)具有屏幕编辑功能,满屏显示汉字不少于 12×40。

3)保持西文灵巧型终端的显示属性。

4)能接受三种以上汉字输入编码方法。

5)在主机控制下,利用主机资源或在监控程序支持下,具有汉字信息处理功能,并保持西文终端各种功能。

6)配有可选的汉字打印接口,供硬拷贝输出。

7)对主机具有中西文兼容联机通信能力,选用调制解调器等通信设备,可适应远程终端等应用场合。

8)具有自检功能

9)可通过更换功能模块来扩充功能。

通用汉字终端选配汉字印字机或其它输出设备、终端通信部件和国家标准 15×16 和 24×24 字模点阵汉字字形库。

3. 智能汉字终端

这类汉字终端可提供执行用户程序的能力。在主机控制或自身管理程序支持下,实现较

强的汉字和字符信息加工能力。

智能汉字终端应具有下述功能：

- 1) 具有较强的汉字、字符处理能力
- 2) 保持西文智能型终端的显示属性。
- 3) 具有全屏幕编辑功能。
- 4) 能接受多种汉字输入编码方法。
- 5) 有一定的脱机汉字信息处理能力。
- 6) 配有可选的标准串行和并行接口。

7) 对主机有较强的中西文兼容脱机通信能力, 选用调制解调器等通信设备, 可适应于远程终端等多种应用场合。对于同一通道上的几条输入输出数据流进行多路复用与识别。

8) 配有二种程序设计语言和丰富应用程序。

9) 具有自检功能。

10) 通过更换功能模块(固化软件模块或硬件模块)来扩充功能, 变为各种专用汉字终端。

11) 配有盒式磁带, 提供辅助存储器能力。

以上主要从功能上给汉字终端系列分为三种类型, 根据标准化、系列化和模块化原则, 各种类型终端应注意功能模块扩充和相互兼容。

9.1.2 汉字终端系列的技术条件

这些技术条件对各类终端有所差别, 这种差别随具体应用各不相同, 这里不作具体规定。留给系统设计和用户去选配, 但基本条件必须具备。

1. 硬件系统

1) 系统组成

A. 基本配置

- 处理器、汉字库、ROM、RAM、显示器、标准键盘和标准接口。

B. 扩充选件

汉字印字机、盒式磁带机、穿孔输出机、纸带读入机、标准大、中键盘、图形输入板、文字或语音传感器等。

2) 处理机

包含 CPU 和 RAM 内存储器。具有输入、打印、传输等主要部件串并行接口。

3) 汉字字形库

包含符合国家标准信息交换用汉字编码字符要求的汉字、图形符号以及用户专用图形字符。汉字字型必须选用国家标准信息交换用汉字 15×16 , 24×24 或 32×32 点阵字模集及数据集。随着超大规模集成电路技术发展, 还可选用标准 48×48 或 64×64 汉字点阵宋体、仿宋体、黑体和楷书的字模集及数据集。汉字库用 MASK ROM 组件构成。

4) 汉字显示器

CRT 或液晶等显示设备, 根据不同使用要求, 形成系列产品:

A. 低档 CRT 显示器

30 厘米单色, 无闪烁, 光栅扫描体制。分辨率不低于 512×256 , 满屏汉字不少于 480 个。

B. 中档 CRT 显示器

30. 48 厘米以上单色或彩色(至少七种颜色),无闪烁,光栅扫描体制。分辨率不低于 640×400 ,满屏汉字不少于 24×40 个。可有强辉度,加闪烁,反视频局部保护区等显示功能,有汉字图形处理和一定的窗口处理能力。

C. 高档 CRT 显示器

分辨率在 1024×1024 以上。满屏汉字在 24×40 以上。具有汉字图形或汉字图象处理能力,16 种以上彩色和一定的窗口处理功能。

5) 联机接口

采用标准软硬接口。传输波特率在 $50 \sim 19200$ BPS 范围内并能按标准值选择。

可有多多个视频接口,可扩充多个显示屏。采用标准转换电平。

6) 输入键盘

A. 大键盘

在键盘盘面结构由文字输入区、词组输入区和功能键输入区组成。键位总数在 2K 或 4K 键左右,并有盘外字输入能力。

B. 中键盘

中键盘键位在 200 个左右,能支持几种输入编码方案,并且和小键盘向上兼容。

C. 小键盘

符合国家标准通用字符键盘要求,还有供自定义的功能键。

D. 微键盘

微键盘键位都在 50 个以下。常由 $0 \sim 9$ 个数字键,或其它字符和功能键。可接便携式汉字终端。

7) 汉字印字机

可接需要选配不同型号的汉字印字机。同一行中可输出多种字型。

A. 16 针或 18 针点阵式图形打印机、双向打印、输出速度不低于每秒 20 个汉字。打印纸幅面可变。

B. 24 针点阵式图形打印机、双向打印、输出速度不低于每秒 35 个汉字。纸幅面可变。

C. 可与激光打印机等其它输出设备的标准接口对接。

2. 软件部分

汉字终端应配置齐全的软件,保证与主机联接,并能独立完成一定的汉字信息处理功能。

汉字终端软件应按模块化、标准化和通用化的要求开发和设置。

汉字终端软件应有功能较强的监控程序、联机程序、汉字输入输出程序、汉字库管理程序、打印程序及扩充功能的软件模块等。

其通信终端有加密、解密、自动启动收发(呼叫应答)及通信规程控制、格式处理等功能。

(1) 监控程序

应能解释各种命令(包括通过功能键输入的命令和主机发来的命令)。并执行各种操作。

应包括初始化程序和自身维护程序。

(2) 联机程序

应能适应主机操作系统的各项要求,不能因接汉字终端而要求主机操作系统或其它系统软件进行重大修改使输入信息转化为主机能够接收的信息。

从主机的角度看,汉字终端在功能上等价于原系统的西文终端。

(3) 汉字输入程序

实现从键盘输入到内部码的转换,完成汉字、词组装入等功能。

应能适应多种输入手段并支持多种汉字输入编码方案。

(4) 汉字库管理程序

字库调用程序保证实时提供显示或打印所需要的字形信息。

应有较方便的造字及修改功能。

(5) 打印程序

具有多种格式的打印功能,至少有大、中、小三种字型尺寸。

(6) 屏幕编辑程序

按不同用途提供相应的屏幕编辑功能,包括字编辑、行编辑、段编辑、光标移动和复位,屏幕上下左右滚动等功能。

应具有汉字/西文混合编辑和混合输出的功能。

(7) 表格处理程序

应具有对各种表格进行定义、处理、存取、修改等便于操作的功能。

(8) 图形处理程序

至少能以线段为单位制作平面图形,图中能够出现文字(中文和西文)。

(9) 诊断服务程序

至少开机后通过诊断服务程序可以检查终端的软硬件工作状态。进而要求对各种存储器进行奇偶校验、差错定位,对内部传输差错和超时进行检出、再运行和差错定位,并备有诊断读、诊断写等维护指令。

9.2 汉字微型机系统

一般来说,简易终端可作为电传机或电报机,因此可以设计为全中文的或者是以中文为主,西文为辅(此处西文亦可作为中文来处理)中西文兼容。实际上这已是作为独立脱机工作。而通用型和智能型,必须要和主机相结合才能很好发挥作用,因此,终端性能与主机紧密相关。

汉字微型机就是具有汉字信息处理功能的微型计算机。那么一台微型机怎样才能具有汉字处理功能呢?具有汉字信息处理功能的微型机,其软硬件功能模块需作哪些扩充?

9.2.1 硬件系统

从体系结构角度看,汉字微型机与西文微型机没有什么差别。只是增加了汉字信息处理功能,配置上有所扩充:

1. 汉字输入键盘

汉字输入键盘有大、中、小、微型四种,对汉字微型机来说,应有多种键盘接口(软接口和硬接口)。

2. 显示器

无论字符式或图形式显示器,都具有汉字显示能力,其汉字显示格式至少为 12×40 个汉字。对于中档显示器,应有 24×40 个汉字满屏显示。具有汉字图形处理能力和较强的窗口功能。而对于高档显示器,应具有很强的汉字图形或汉字图象处理能力和窗口功能。其彩色应

有 4096 种以上可调节。满屏汉字应有 1000 个以上并能显示 24×24 字模点阵宋体字形。

3. 印字机

汉字印字机 24 针以上点阵打印机或激光打印机。打印字形至少为 24×24 或 32×32 字模点阵的宋体。还应能打印 48×48 或 64×64 字模点阵的宋体、仿宋体、黑体和楷书。激光打印机应能打出多种字号。打印速度为 40 字/秒以上。有汉字图形打印能力。

配有与喷墨、静电印字机等输出设备标准接口能力。

4. 汉字库

至少配有国家标准信息交换用汉字编码字符集所包含的汉字及图形符号。其汉字字模点阵应为国家标准 15×16 , 24×24 两种 MASK ROM 硬字库。

还可酌情应配有 32×32 点阵宋体、 48×48 (或 64×64) 点阵宋体、仿宋体、黑体和楷书四种软字库或硬字库。亦可配有一种 48×48 (或 64×64) 点阵字模通过软件换成所需的字型字号。此外,还可配有相应的繁体字库等。

还需有一定量的造字、造图形或符号。

随着集成电路技术发展,尽量采用硬字库,以便提高效率。

5. 汉字识别装置

可选择配上印刷体汉字识别或手写体识别传感器,如摄象扫描器、图形输入板等设备。

6. 语音信息处理装置

可选择配上汉语语音信息处理装置,如话筒、A/D、D/A 板,语音分析合成器语音板、扬声器等设备。

7. 汉卡

一种带有汉字库和一些固化汉字处理程序的专用插板,插入标准总线(也有装在设备上,如汉字印字机等)。随着集成电路技术发展,汉卡应用会越来越多种多样、越来越广泛。

9.2.2 软件系统

为增加汉字信息处理功能,除了必须配置中西文兼容操作系统。还需扩充下列应用软件:

1. 通用汉字输入模块

一台汉字微型机,应能支持各种汉字输入编码方案,故必须配上通用汉字输入模块。而不是只配固定几种输入编码方法。

2. 通用汉字输出模块

一台汉字微型机,亦应能支持多种型号汉字印字机,故必须配通用汉字输出模块。

3. 汉字库管理程序

此软件包对汉字库进行维护,并有增、删改等功能,也能对各种字型、字号变换和调用。

4. 汉字编辑程序

能对汉字,汉字和西文混合正确实施文本编辑和屏幕编辑功能。亦要有汉字图形或汉字图象混合编辑能力。并具有开窗,窗口移动和缩放功能。

5. 表格处理程序

汉字表格程序有其与西文表格处理程序不同的地方,要特别仔细设计。

6. 汉字程序设计语言

7. 汉字数据库技术