

微机汉字操作系统 实用开发技术

● 钱培德 董庭辉 主编

WEIJIHANZICAOZUOXITONG
SHIYONGKAIFAJISHU



北京师范学院出版社

微机汉字操作系统 实用开发技术

钱培德 董庭辉 杨季文 编著
朱巧明 吕 强 何卓敏

北京师范学院出版社

1991年·北京

内 容 简 介

汉字操作系统是汉字信息处理系统的核心部分,汉字操作系统技术是汉字信息处理技术的关键。本书以 IBM—PC 系列机为硬件环境,系统和深入地阐述了汉字操作系统,并以 CC—DOS 的典型版本为例,重点介绍汉字操作系统的开发技术。

全书共分九章,其内容可分为三个部分:汉字操作系统基础、汉字操作系统结构分析与设计,汉字操作系统优化与二次开发技术。本书为力图理论联系实际,并附上了必要的实例程序,以突出实用性和技术性。本书还尽可能力求反映国内微机汉字操作系统的水平。本书既可作为计算机专业人员和计算机用户的参考书,也可作为高等院校计算机专业的教材。

微机汉字操作系统实用开发技术

钱培德 董庭辉 杨季文 编著
朱巧明 吕 强 何卓敏

*

北京师范学院出版社出版发行
(北京阜成门外花园村)

全国新华书店经销
郑州市耿河印刷厂印刷

开本:787×1092 1/16 印张:30.75 字数:715 千字
1991 年 10 月北京第一版 1991 年 10 月郑州第一次印刷
印数:0,001—3,000 册
ISBN7—81014—559—2/G·454
定价 18.8 元

前 言

现代社会是充满信息的社会，对信息进行处理和管理是社会的需要。由于信息量日趋庞大、信息结构日趋复杂，用计算机对信息进行处理与管理已势在必行。汉字是我国的通用文字，所以我国社会的绝大部分信息是汉字信息。如果不解决计算机的汉字信息处理问题，那末计算机的使用在我国会受到极大的限制。因此，研究与开发汉字信息处理技术具有重大的现实意义。

微型计算机的问世大大拓宽了计算机的应用范围，目前 IBM-PC 和长城系列机已成为我国使用的主流机种。电子工业部第六研究所开发的 CC-DOS 汉字操作系统，对我国的计算机应用具有巨大的推动作用。CC-DOS 的诞生，意味着我国的汉字信息处理技术达到了一个新的水平，同时也表明我国研制的汉字操作系统已进入实用阶段。

我们曾经开发了 UniFLEX 多用户操作系统的汉化模块，CH-DOS 汉字操作系统，PG 型汉字信息处理系统和组合式通用型汉字操作系统 GAOK。我们还对汉字操作系统作了深入的研究，在国内首先分析了 CC-DOS。我们在汉字信息处理技术方面积累了一定的经验，并具有相当的基础。为适应当前国内汉字信息处理技术飞速发展的形势，为了满足广大计算机科研、工程技术人员和计算机用户学习与开发汉字操作系统的需要，我们合作编写了这本书，集中介绍微机汉字操作系统的实用开发技术。为了突出技术性和实用性，并顾及先进性，本书在阐述系统结构和开发设计技术时，将以 CC-DOS V4.0 为蓝本。我们力图本书能对有关人员有较大的参考价值，对推广汉字操作系统能起一定的作用。

全书内容共分三部分。第一部分(第一章至第三章)为基础篇，介绍汉字操作系统的基本知识和使用方法，并介绍了与汉字信息处理有关的 DOS 技术；第二部分(第四章至第七章)为系统分析与设计篇，介绍汉字操作系统的结构，并逐个模块地介绍设计方法和实现方法；第三部分(第八章和第九章)，为开发优化篇，介绍对汉字操作系统进行系统优化和再开发的技术。

在这本书的形成过程中，得到了许多同志的支持和帮助，特别是为本书进行审稿的教授和专家，对书稿提出了不少有益的意见。我们谨在此向这些热心的同志深表谢意。

本书由钱培德、董庭辉、杨季文、朱巧明、吕强、何卓敏合作完成，并由钱培德主持编写并最后定稿。由于汉字操作系统的复杂性和我们水平的限制，所以尽管我们在各方面均作了较大的努力，但是书中一定仍会有不足之处，还望读者不吝指教。

作 者

1991 年 1 月

目 录

第一章 系统概述	(1)
第一节 汉字系统的发展与特点	(1)
一、引 言	(1)
二、汉字信息处理技术的发展	(1)
三、CC-DOS V4.0 操作系统的特点	(2)
第二节 系统设计与实现	(3)
一、设计思想	(3)
二、汉字机内码的设计	(4)
三、实现方法	(5)
第三节 代码体系和字库结构	(6)
一、代码体系	(6)
二、字库与字模	(7)
第二章 CC-DOS V4.0 的操作与使用	(10)
第一节 引 言	(10)
一、CC-DOS V4.0 系统文件介绍	(10)
二、硬件支撑环境	(12)
三、软件支撑环境	(12)
第二节 系统的启动	(12)
一、从软盘启动系统	(12)
二、从硬盘启动系统	(14)
三、在带汉卡的机器上启动系统	(14)
四、装入出错	(15)
第三节 命令功能及汉字输入方法	(15)
一、系统功能	(15)
二、汉字输入方法	(16)
第四节 汉字的输入	(19)
一、区位码输入方式	(19)
二、首尾码输入方式	(21)
三、拼音码输入方式	(25)
四、快速码输入方式	(28)
五、汉字输入注意事项	(30)
第五节 词组的使用及词库的建立	(30)

一、词组输入方法	(30)
二、词库的定义	(31)
第六节 系统功能键的使用	(36)
一、制表功能键	(36)
二、字典功能键	(38)
三、修改码表功能键	(40)
四、设置图形符功能键	(42)
五、高频字统计功能键	(44)
六、退出汉字系统功能键	(45)
七、改变字符颜色功能键	(45)
八、字符图形方式转换功能键	(46)
九、建立自动光标功能键	(46)
十、建立纯中文方式功能键	(47)
十一、打印控制功能键	(48)
十二、自选输入方式功能键	(48)
十三、模糊输入功能键	(49)
第七节 造字工具的使用	(50)
一、造字程序的功能	(50)
二、造字程序的装入	(50)
三、改 码	(52)
四、造 字	(53)
五、删 除	(56)
六、部首功能	(57)
七、显 示	(58)
八、退 出	(58)
第八节 汉字的打印输出	(58)
一、24×24 点阵的 24 针打印输出	(58)
二、16×16 点阵的 24 针打印输出	(64)
三、16×16 点阵的 9 针打印输出	(64)
第三章 汉字信息处理有关的 DOS 技术	(65)
第一节 文件系统的数据结构和算法	(65)
一、文件目录表 FDT	(65)
二、文件分配表 FAT	(69)
三、文件分配表的使用	(73)
第二节 访问磁盘的方法	(77)
一、磁盘的逻辑结构	(77)
二、磁盘 I/O 参数表	(79)
三、由簇号到逻辑扇区号的转换	(82)

四、DOS 提供的磁盘扇区读写功能	(83)
第三节 磁盘 I/O 驱动程序	(90)
一、逻辑扇区号到扇区物理地址的转换	(90)
二、13H 号中断处理程序	(91)
第四节 可执行文件的结构	(92)
一、*.COM 文件的结构	(93)
二、*.EXE 文件的结构	(93)
三、实例分析	(98)
 第四章 系统自举	(101)
第一节 系统文件及其结构	(101)
一、内核文件	(101)
二、BIOS 文件	(102)
第二节 内核的自举	(104)
一、内核自举总流程	(104)
二、引进记录的执行	(105)
三、IBMBIO.COM 的自举	(106)
四、IBMDOS.COM 的自举	(106)
五、COMMAND.COM 的自举	(106)
第三节 BIOS 的自举	(107)
一、CCCC.COM 的执行	(107)
二、CCCC.OVR 的执行	(110)
三、CCCC.OV1 的执行	(114)
四、M2024E.COM 的执行	(117)
 第五章 显示器控制模块	(122)
第一节 模块总述	(122)
一、工作原理	(122)
二、模块结构和调用方法	(123)
三、字库组织	(125)
四、视频存贮区	(127)
五、模块的总体流程	(129)
第二节 视频初始化	(130)
一、概 述	(130)
二、视频初始化的实现	(131)
第三节 光标功能的实现	(133)
一、光标位置的读取	(133)
二、光标的产生与定位	(133)
三、RRAM 映象区的浮动	(137)

第四节 字符的读出与显示	(143)
一、字符的读出	(143)
二、字符的显示	(143)
三、字符的 TTY 方式显示	(154)
第五节 提示行管理与其它	(157)
一、提示行管理	(157)
二、汉字库字模的读写	(162)
三、光标的建立与删除	(164)
四、屏幕的滚动	(164)
 第六章 键盘管理模块	 (167)
第一节 模块总述	(176)
一、工作原理	(167)
二、模块结构和调用方法	(168)
三、键盘中断处理程序	(168)
第二节 重要的工作区	(169)
一、键盘缓冲区及其指针	(169)
二、输入码对照表	(172)
三、主要工作区	(174)
第三节 主体流程	(175)
一、键盘管理模块的总体流程	(175)
二、各功能的流程	(176)
第四节 字符输入的实现	(179)
一、0 号功能块为流程	(179)
二、字符处理子程序	(182)
第五节 功能符的处理	(186)
一、功能符处理子程序	(186)
二、CTRL+Fi 符的处理	(187)
三、ALT+Fi 符的处理	(192)
四、输入制表处理	(194)
第六节 输入码的处理	(196)
一、区位码处理程序	(196)
二、区位码转换程序	(200)
三、拼音一首尾码处理程序	(204)
四、拼音一首尾码转换程序	(207)
第七节 高频字统计和字典功能	(214)
一、高频字统计处理	(214)
二、字典功能的实现	(216)
第八节 词组输入处理	(221)

一、词组输入原理	(221)
二、词组输入码对照表	(222)
三、主要工作区	(222)
四、词组处理程序	(223)
五、词组选择处理程序	(228)
六、词组恢复符处理程序	(229)
第七章 打印机驱动模块	(230)
第一节 模块总述	(230)
一、工作原理	(230)
二、打印机及其驱动程序	(231)
三、汉字字型的变换	(231)
四、模块确定的结构	(232)
第二节 模块主体和工作区	(232)
一、打印机驱动模块的主体	(232)
二、主要的工作区	(233)
第三节 打印参数的定义	(238)
一、总 述	(238)
二、打印字型的定义	(239)
三、字间距的定义	(242)
四、行间距的定义	(244)
五、行宽和工作方式的定义	(248)
第四节 字符的打印输出	(249)
一、主体流程	(249)
二、西文方式下字符的打印输出	(250)
三、汉字方式下字符的打印输出	(251)
第五节 行信息的打印输出	(260)
一、缓冲区内容的输出	(260)
二、字模读取与变换	(265)
三、行纵扩处理	(277)
第六节 屏幕硬拷贝	(282)
一、总体流程	(282)
二、图形方式下的屏幕硬拷贝	(282)
三、一列信息的硬拷贝	(286)
第八章 高分辨率视频适配器及其程序设计	(292)
第一节 引 言	(292)
第二节 EGA 视频适配器及其程序设计	(293)
一、概 述	(293)

二、EGA 视频 RAM 结构	(293)
三、象元的写入	(293)
四、象元的读出	(301)
第三节 MCGA 视频适配器及其程序设计	(304)
一、概 述	(304)
二、MCGA 新性能和视频 BIOS	(306)
三、字母数字模式的程序设计	(307)
四、图形模式的程序设计	(308)
五、视频 DAC 的程序设计	(310)
第四节 VGA 视频适配器及其程序设计	(311)
一、概 述	(311)
二、图形模式的程序设计	(312)
三、视频 DAC 的程序设计	(316)
第五节 CGE400 视频适配器及其程序设计	(318)
一、概 述	(318)
二、图形模式的程序设计	(320)
三、视频 BIOS	(323)
第六节 汉字显示技术的一般方法	(324)
一、概 述	(324)
二、图形文本特性	(325)
三、汉字显示问题的解决	(327)
四、汉字显示的其它问题	(329)
五、中文窗口的讨论	(330)
第九章 汉字系统优化技术	(332)
第一节 “透明法”扩充技术	(332)
一、CC-DOS 的实现思想	(332)
二、“透明法”原理	(333)
三、“透明法”应用和举例	(335)
第二节 外码连接技术	(337)
一、引 言	(337)
二、外码连接模块的流程	(338)
三、自定义编码外部模块的程序设计	(340)
第三节 词组输入技术	(352)
一、词组输入技术	(352)
二、充分发挥计算机的潜力	(353)
三、编码的词方向	(354)
四、高级动态造词功能的设计	(355)
五、重码的智能处理	(364)

六、词的主动意义	(375)
第四节 联想输入技术	(376)
一、概 述	(376)
二、单字联想输入技术的数据结构	(376)
三、词组联想输入技术的数据结构	(377)
第五节 磁盘碎片集成技术	(379)
一、问题的提出	(379)
二、与 DFC 有关的文件管理机制	(380)
三、一个 DFC 的实现	(381)
四、进一步的讨论	(384)
第六节 扩展内存的使用方法	(385)
一、何为扩展内存	(385)
二、80286 的存储器管理	(386)
三、虚拟盘 VDISK	(388)
四、15H 号中断程序的 87H 号子功能	(389)
五、虚拟汉卡	(391)
六、打印驱动程序对扩展内存的使用	(398)
第七节 进出 80286 保护方式	(398)
一、进出 80286 保护方式概述	(399)
二、进出保护方式的准备	(400)
三、进入保护方式	(401)
四、退出保护方式	(402)
五、自定义词库的检索	(402)
附录 A 信息交换用汉字编码字符集基本集	(411)
附录 B 信息处理用现代汉语五千词表	(442)
参考资料	(469)

第一章 系统概述

第一节 汉字系统的发展与特点

一、引言

计算机的应用由以数值计算为主,发展到以非数值性的数据处理为主,这是计算机应用领域中的一个重大突破,它大大拓宽了计算机的应用范围,促使计算机应用的发展。然而,计算机应用的发展,必然向计算机科学技术提出许多问题和要求,这无疑将促进计算机科学技术和计算机产业的发展。

在现代社会里,有大量的信息需要进行存储、传递和处理,如果用人工来完成这些工作,则要花费大量的劳动,而且由于信息量的庞大和人脑工作的特点,往往不能达到非常满意的结果。用计算机实现对这些信息的处理,已经势在必行。

我国的通用文字是汉字,所以,我国社会中进行处理的信息主要是汉字信息。然而,汉字信息的输入、输出与处理,均要比西文信息的相应处理困难得多,至今世界上各计算机厂(公司)研制的计算机多为西文处理系统,与其配套的软件系统(系统软件、实用程序和数据库等)也是建立在西文信息处理系统的基础上的,一般不能实现汉字的输入与输出。显然,不能很好地解决在计算机上进行汉字信息处理这个问题,就不可能在我国推广计算机的应用,各行各业也就不可能实现现代化的管理。从我国计算机应用的现状来看,其使用效率不高的重要原因也在于没有很好地解决汉字信息的处理问题,这已经成了阻碍我国计算机事业发展的重要因素。我国是汉字的发源地,对于汉字结构的特性及使用情况最为熟悉,对于发展汉字信息处理技术的要求最为迫切,受益也最大。因此,我国理应在汉字信息处理领域中走在世界的最前列。

二、汉字信息处理技术的发展

我国在六十年代后期就开始对汉字信息处理技术进行探索和研究,在七十年代中期开始系统地研究和开发这项技术,并获得了一些技术条件和研制经验。从七十年代末开始,由于大规模集成电路存储器和成套的微处理机芯片进入我国应用领域,因而在很大程度上促进了汉字信息处理技术的发展,不仅使原有的一些技术得到更新,而且研制成了一些新型的汉字输入输出设备,在技术指标、可靠性和实用性方面,均有很大的提高。并且能用国内研制的汉字设备和计算机配置成多种应用系统,特别是以微处理器为基础的汉字信息处理系统发展更为迅速。目前,国内在继续进行汉字基础理论研究的同时,已制订了汉字信息处理设备与系统的研制和生产规划。汉字信息处理系统的推广应用工作愈益得到政府部门和各类业务部门的重视。可以预期,在今后几年内,我国的汉字信息处理技术将会以更快的速度向前发展。

我国的汉字信息处理系统已由试验阶段发展到了成熟阶段。在汉字信息处理系统的配置中,除了提供必要的汉字设备与接口外,最重要的是软件的配置,而其中以汉字操作系统为最主要。我国在汉字操作系统的研制和开发方面作了不少工作,一般均是对已有的西文操作系统进行扩充和改造,使其成为能处理汉字信息的汉字操作系统。例如,在微型机 CP/M 操作系统中加入了汉字输入输出的管理模块,形成了能支持高级语言 and 应用程序处理汉字信息的新系统。对于 VAX 系列的小型计算机系统,也完成了把它的 VMS 操作系统扩充为具有汉字处理功能的 CCVMS 操作系统。对于著名的 UNIX 操作系统,国内也已经完成了对其多种版本及变种的汉化工作。

随着 IBM-PC 微型计算机的引入和发展,使得这种计算机已成为我国的主流机种。为了使它具有处理汉字信息的功能,著名的汉字操作系统 CC-DOS 问世了,CC-DOS V4.0 为其最新版本。我们将以此系统为例,来讨论汉字操作系统。

三、CC-DOS V4.0 操作系统的特点

为了增加对 CC-DOS V4.0 的认识,故有必要在这儿讨论一下这个系统的特点。

1、字库的分区选择驻留

CC-DOS V4.0 向用户提供了选择汉字库驻留内存的最灵活手段,用户可以任意选择驻留内存的汉字库区数。其两个极端是:汉字库全部驻留内存和汉字库全部驻留外存,其折衷方案是一级汉字驻留内存,二级汉字驻留外存。这无疑会给用户带来极大的方便。在汉字库全部驻留外存时,系统只占用内存 40K 左右字节。

2、采用了反拷贝和保密措施

CC-DOS V4.0 在系统盘上采用了激光反拷贝技术,以保证其版权不受侵犯。为了保护其设计技术,CC-DOS V4.0 的系统文件进行了保密处理。这些文件中的内容不是真正的可执行代码,故无法对它们进行反汇编。这些文件装入内存后,由系统对其作恢复解密处理后,才能正常运行。另外,系统文件中还采用了反跟踪措施,有效地实现了 CC-DOS V4.0 的反拷贝与保密。显然,这些方法的采用,会给 CC-DOS V4.0 的分析带来很大的困难。

3、具有附加输入码软接口

CC-DOS V4.0 中设计了一个附加输入码的软接口。通过这个软接口,很容易把一种附加的汉字输入方式连到系统中去。CC-DOS V4.0 除了提供“区位”、“首尾”、“拼音”和“快速”四种基本的汉字输入方式外,还提供一种附加的汉字输入方式。附加的汉字输入方式可以是“大众”、“五笔”、“仓颉”、“广东拼音”、“电报”、“拼形”、“声韵”、“声声”和“笔形”方式中的任意一种,由用户自由选择。这样就大大方便了用户。

4、具有字典和高频字统计功能

字典功能是指在某一种输入方式下输入一个汉字后,提示行中会同时显示出在另外几种输入方式下输入该汉字时应输入之编码。这有利于用户在学习某种输入方式时熟悉汉字的输入编码。有了高频字统计功能后,系统在工作时会动态地、自动地统计出用户最常用的 10 个一码常用字和 10 个两码常用字。当输入相应编码时,这些常用字会首先显示在提示行中,作到“常用字先见”,大大减少了翻页几率,提高了输入速度。由于这种统计是动态的,故可以适合不同用户的使用特点。

15、可进行彻底的中西文方式间的切换

汉字操作系统具有中西文方式间的切换功能是必要的，一般汉字操作系统均具有此功能。中文方式要比西文方式占用较多的内存资源，一般系统从中文方式切换到西文方式时，其多占的内存资源是不释放的。CC-DOS V4.0 在这方面作了改进，它从中文方式切换到西文方式时，会把多占的内存资源全部释放出来，作到了彻底切换。

6、具有通用性较强的显示器控制模块

CC-DOS V4.0 的显示器控制模块具有通用性，系统能够在自举时自动识别系统配置的显示卡，把显示卡分为彩色卡、单色卡和长城 CEGA 卡三种，然后调相应的控制模块进行适配。因此 CC-DOS V4.0 的适用范围较广，不象以前的 CC-DOS 版本那样，只能适用于一种显示卡(CGA 卡)。

7、具有功能完备的打印机驱动程序

CC-DOS V4.0 为常用的打印机均配备了打印机驱动程序。每种打印机有两种驱动程序，分别用于打印输出 16×16 点阵汉字和 24×24 点阵汉字。CC-DOS V4.0 的打印机驱动程序开放打印机的全部固有控制命令。用户可以用键盘命令或在程序中用特定的命令序列来定义输出字型、行间距、行宽和字间距，为打印各种表格提供了方便。另外，打印机的屏幕硬拷贝程序也得到加强，用户可以变换硬拷贝输出的字型。

8、实现汉卡方案与软方案一体化

一般来说，汉字系统只能采用汉卡方案和软方案中的一种，而 CC-DOS V4.0 可以适用于这两种方案。也就是说，如果系统中配置有汉卡，就采用汉卡方案，这时不再需要把汉字库装入内存。系统对系统中是否配置汉卡的识别是自动进行的。

9、对系统内核的动态汉化

CC-DOS V4.0 的内核采用 PC-DOS 的内核，为了使其能与汉字 I/O 较好地适配，必须对它作相应改动(即汉化)。CC-DOS V4.0 对内核的汉化是在系统自举时进行的。这样，系统盘上就不必带上内核文件，可使用用户自己的内核文件。

10、与长城系列机有较好的兼容性

长城系列机是国产微型机，它具有大量的用户。这种机器的操作系统为 GW-BIOS V3.0，但是这种操作系统与以前的 CC-DOS 版本不完全兼容，使得有些汉化过的软件不能在这两种系统上通用。CC-DOS V4.0，考虑到了这一点，采取了一系列的措施，实现了它与 GW-BIOS V3.0 间的兼容。

第二节 系统设计与实现

一、设计思想

CC-DOS 是一个汉字操作系统，设计 CC-DOS 的目的是使 IBM-PC 成为一个汉字信息处理系统。汉字信息处理系统的关键是在计算机对汉字的数据处理上，让计算机使用汉字和使用西文一样的方便和容易。这样才能使古老的汉字真正地焕发青春，才能使汉字真正永远地自立于世界民族文字之林。我们应从这方面来提出 CC-DOS 的设计思想。

PC-DOS 是 IBM-PC 的主操作系统, 它已配有丰富的软件, 并且已为广大用户所接受。CC-DOS 应该与 PC-DOS 兼容, 才能借用他人开发出来的、经长期的实践检验的丰富软件财富, 这有利于该系统的推广应用, 也有利于程序的交流。所以, CC-DOS 应以 PC-DOS 为基础, 对其有关部分进行改造而成。它必须具有 PC-DOS 的全部功能和使用方法, 它还应有 PC-DOS 所不具有的汉字信息处理功能, 包括汉字 I/O 和汉字信息的处理等, 并且使用中文与使用西文一样方便。要使用户感到, 在计算机系统内, 西文和中文都是符号, 只是用不同的形式表示而已。

从信息处理的观点看, 汉字信息处理与西文信息处理没有本质的区别, 两者均是非数值处理, 都要求计算机系统具有资源、信息共享、高级语言、数据库和数据通讯等功能的支持, 因而, 只要用某种方式在 CC-DOS 内部表示汉字信息, 增加汉字 I/O 功能, CC-DOS 就可以成为中西文兼容操作系统。

汉字操作系统必须具有汉字信息处理软件, 它包括汉字信息加工软件和支持汉字处理的软件工具。

汉字信息加工软件直接对汉字信息进行加工处理, 它包括键盘汉字输入处理程序, 显示器汉字显示处理程序和打印机汉字打印处理程序等。

键盘汉字输入程序主要接收给定的某种或几种汉字输入方案, 并进行重码处理, 使每输入一个汉字编码就转换成唯一的汉字机内码。

显示器汉字显示处理程序主要是根据汉字机内码发生相应的汉字字形, 并在显示器上显示出来, 另外还要解决汉字字形点阵的检索问题。

打印机汉字打印处理程序主要是根据汉字机内码产生相应的汉字字形, 并在打印机上打印出来, 还要解决各种规格字型的产生, 汉字点阵的旋转和字库点阵的检索。

CC-DOS 的运行环境是 IBM-PC, 我们希望原来在此环境下运行的软件均具有汉字功能, 所以上述三个处理程序应嵌入相应的外部设备驱动程序中。

支持汉字处理的软件工具包括汉字库的建立与维护程序、汉字输入码对照表生成程序和中文编辑程序等, 这些均是支持汉字处理的服务程序。这些程序一般处在用户级, 用户可以根据自己的需要来建立汉字处理软件工具。为了系统的可靠性、实用性和方便性, 系统应该多带这类软件, CC-DOS 也应以实用程序的形式向用户提供这方面的软件工具。

二、汉字机内码的设计

汉字机内码(亦称汉字内码)是系统内部处理和存储汉字而使用的代码。众所周知, 西文字符的机内码多采用一个字节来表示的 ASCII 码, 有的系统则采用 EBCDIC 码。一般只使用七位来表示 128 个字符, 而把高位用作奇偶校验(或者不用)。我国的国标 GB2312-80 规定, 一个汉字用两个字节来表示, 目前规定每个字节也只用七位, 其高位尚未作定义。

为了保证系统的中西文兼容, 这就意味着系统的机内码中必须保持 ASCII 码(IBM-PC 采用该码作西文字符的机内码)的使用, 同时又要允许汉字机内码的使用, 并且使两者之间没有冲突。如果用 GB2312-80 中的国标码作为汉字机内码, 则在系统中同时存在 ASCII 码和国标码时, 将会产生二义性。例如, 机内有两个字节的内容分别为 30H 和

21H, 它们既可表示汉字“啊”的国标码, 又可表示字符“0”和“!”的 ASCII 码。所以原原本本地采用国标码作为汉字机内码是不行的, 必须要加以适当的变换, 同时应使变换的结果符合以下准则:

- (1)信息的位数要少, 尽量减小信息的冗余度;
- (2)与汉字国标码有尽可能简单的对应关系;
- (3)排除汉字机内码与西文机内码之间的二义性;
- (4)便于汉字信息处理中的基本操作和运算;
- (5)便于纳入到各种程序设计语言的字符类型中;
- (6)与原西文系统的兼容性要好;
- (7)便于扩充汉字字符集。

根据上述准则, CC-DOS 将国标码的每个字节的高位置成 1, 作为汉字机内码, 这种编码称作变形国标码。这样作既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极简单的对应关系。

三、实现方法

IBM-PC 系统的 BIOS(基本输入输出系统)存放在系统板(主板)上的 ROM 中, 所以它亦被称为 ROM-BIOS。它由若干个独立的外部设备驱动模块组成, 每个模块对应于一种外部设备, 为这种设备的驱动和控制程序。从而, ROM-BIOS 在汇编语言级上向用户系统程序提供外部设备的设备层控制功能, 包括盒带机操作、上电自检、显示器, 以及通讯口、键盘和打印机的字符传送与图形发生等, 这些操作均无需用户考虑设备的细节。ROM-BIOS 是用户与硬件间的第一界面, 外设的更改或增减均对用户透明。

但是, ROM-BIOS 的各个模块的入口地址被存放在中断向量表中, 该表的地址是 0000:0000H-03FFH。其它软件对外部设备的使用, 均通过直接或间接调用 ROM-BIOS 中的有关模块来实现。对这些模块的调用方法是采用软中断来完成的, 其中断号为 10H-1AH。CPU 响应软中断后, 就根据中断号所对应的中断向量, 转向相应的模块执行。

根据上述情况, CC-DOS 扩充或修改了 ROM-BIOS 中与汉字 I/O 有关的模块, 它们是: 显示器控制模块(10H 号中断处理程序)、键盘管理模块(16H 号中断处理程序)、打印机驱动程序(17H 号中断处理程序)和屏幕硬拷贝驱动程序(5H 号中断处理程序)。在系统自举时把这些新的模块引入内存, 并使它们驻留下来, 然后把中断向量表中相应模块的入口地址改为指向内存中新模块。经过这样修改后, 当用相应的软中断来调用 BIOS 时, 就不再执行 ROM-BIOS 中的模块, 而转到内存中的相应模块执行。这就是系统实现的基本方法。

第三节 代码体系和字库结构

一、代码体系

在汉字信息处理系统中存在着多种汉字编码。一般来说,在系统的不同部位,可根据其环境给汉字定以相应的编码。从而在汉字信息处理系统中存在着数种汉字编码,这些编码构成了一个汉字代码体系。为了适应汉字信息处理系统各组成部分对汉字信息进行处理的不同要求,CC-DOS 提供各种程序模块来完成这些代码的转换和处理。下面对 CC-DOS 中存在的汉字代码作一介绍。

1、汉字输入码

目前,将汉字输入计算机的方法大体上有三大类:

(1)用拼音字母、数字或符号将汉字编成代码的间接输入,通常称为汉字编码输入;

(2)采用字根或字元从键盘拼形输入,或以整字从大键盘上直接输入,通常称为汉字键盘输入;

(3)采用语言识别和文字识别手段的高级输入,通常称为自然语言输入。

当前,汉字主要由人工采用第一种方法通过键盘输入,汉字输入码就是为这种输入方法而编制的汉字代码。国内研制的汉字输入码方案大约有四、五百种,这些汉字输入码大体可归纳为四种类型,即流水码、纯音码、音形码和纯形码。

汉字输入码位于人机界面上,它应该是面向用户的,它应易为用户熟悉、掌握和接受,所以常常要求它的规则简单、易于记忆、操作方便、编码容量大、码长短、输入速度高和重码率低。一般来说,由于用户不同,用途有别,故汉字系统中常配有数种汉字输入方式,即可以支持数种汉字输入码。

2、汉字机内码

汉字机内码是系统内部处理和存贮汉字统一使用的汉字代码。不同的系统可使用不同的机内码。CC-DOS 采用变形国标码作为其汉字机内码,详细情况已在本章第二节中作了介绍。

3、汉字地址码

汉字地址码是指汉字字模库(这里主要指整字形的点阵式字模库)中存储各汉字字形信息的逻辑地址码。CC-DOS 中的汉字字模库有两种,它们分别存放在 RAM(或 EPROM)中和磁盘上。故 CC-DOS 中的汉字地址码有两种,一种采用内存地址数来表示(对应于 RAM 和 EPROM 中的字库),另一种采用盘地址来表示(对应于磁盘中的字库)。由于 CC-DOS 的汉字字模库中汉字字形信息排列的规则(在下面给出),使得 CC-DOS 的汉字地址码与汉字机内码之间存在着一个简单的函数关系,很容易实现两者间的转换。

4、汉字交换码

汉字交换码是一种用于汉字信息处理系统之间,或者与通信系统之间进行信息交换的汉字代码。汉字交换代码位于一台机器的出口和另一台机器(包括输出设备与记录设备)的入口之间。为了要达到系统设备之间或记录介质之间信息交换的目的,汉字交换码必须采