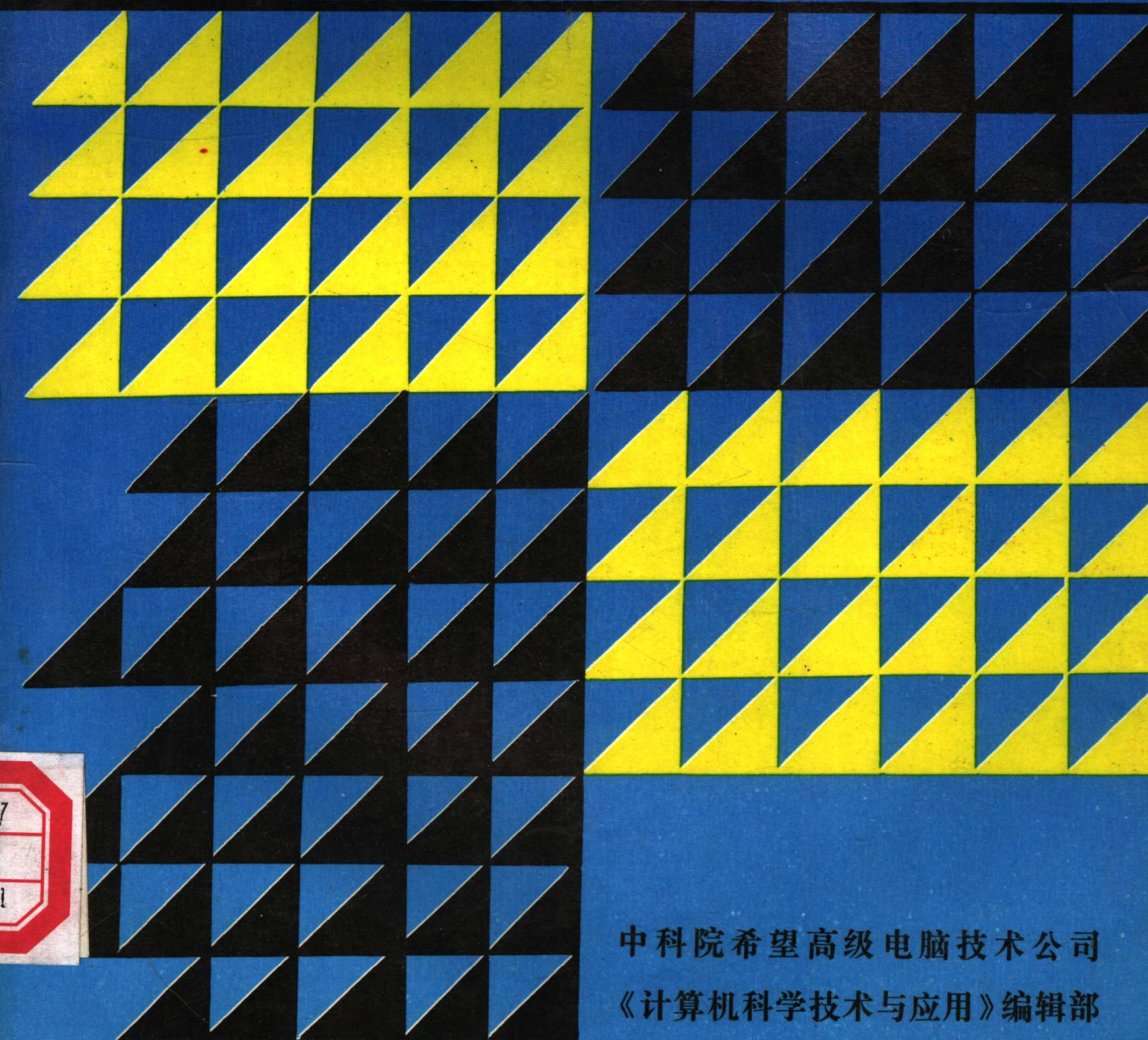


CC-DOS V 4.0

# 汉字操作系统实用技术



中科院希望高级电脑技术公司  
《计算机科学技术与应用》编辑部

73.87221

151

# 汉字操作系统实用技术

钱培德 杨季文 编著  
朱巧明 吕 强

# 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,用计算机实现对信息的处理与管理,已势在必行。我国的通用文字是汉字,所以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,计算机的使用在我国将受到极大的限制。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,IBM—PC计算机进入我国后,加快了我国计算机应用的发展。电子工业部第六研究所开发的CC—DOS汉字操作系统,则大大加剧了这种发展。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。由于CC—DOS使用的广泛性,在计算机界掀起了一股“汉字操作系统热”,追求汉字操作系统技术的人越来越多,他们的学习要求也越来越迫切。

我们——苏州大学计算机工程系操作系统研究与开发组,开发了UniFLEX多用户操作系统的汉化软件、CH—DOS汉字操作系统、PG型汉字信息处理系统和通用联想式汉字操作系统。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 V2.1,现在又首先分析了CC—DOS V4.0。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。

为了适应当前国内汉字信息处理技术飞速发展的形势,为了满足计算机科研、工程技术人员和计算机用户应用与学习操作系统的需要,我们编写了这本书,集中介绍汉字操作系统的实用技术。为了突出技术性和实用性,并顾及先进性,本书以CC—DOS V4.0为例进行阐述。CC—DOS V4.0是目前CC—DOS的最新版本,由于它采用了加密和保密措施,所以它所用的某些技术,至今仍笼罩着神秘的色彩。我们力图使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。

本书的内容共分三部分。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,阐述了CC—DOS V4.0的使用方法与技巧,并介绍了有关的DOS技术;第二部分(第四章至第七章)为系统分析篇,对CC—DOS V4.0作出了全面分析,包括对自举过程、显示器控制模块、键盘管理模块和打印机驱动模块的分析,给出了各模块的详细流程图和入口地址;第三部分(第八章至第九章)为开发优化篇,介绍对汉字操作系统进行系统优化和二次开发的技术,包括高分辨率显示适配器的程序设计、输入部分的优化、输出部分的优化与再开发等。

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,对书稿提出了许多有益的意见。我们谨在此向这些热心的同志深表谢意。

由于汉字操作系统的复杂性、写作时间的仓促和我们水平的限制,所以尽管我们在各方面均作了较大的努力,但是书中一定会有不少缺点与不妥之处,还望广大读者不吝指教。

最后,我们还要向CC—DOS V4.0的设计者表示敬意,向《计算机科学技术与应用》编辑部的全体同志表示感谢。

作 者

一九九〇年二月

于苏州大学计算机工程系

# 目 录

## 第一章 系统概述

|                |       |
|----------------|-------|
| 第一节 汉字系统的发展与特点 | ( 1 ) |
| 第二节 系统设计与实现    | ( 3 ) |
| 第三节 代码体系和字库结构  | ( 5 ) |

## 第二章 CC—DOS V4.0的操作与使用

|                 |        |
|-----------------|--------|
| 第一节 引言          | ( 9 )  |
| 第二节 系统的启动       | ( 11 ) |
| 第三节 命令功能及汉字输入方法 | ( 13 ) |
| 第四节 汉字的输入       | ( 16 ) |
| 第五节 词组的使用及词库的建立 | ( 24 ) |
| 第六节 系统功能键的使用    | ( 29 ) |
| 第七节 打印输出        | ( 35 ) |

## 第三章 汉字信息处理有关的DOS技术

|                  |        |
|------------------|--------|
| 第一节 文件系统的数据结构和算法 | ( 41 ) |
| 第二节 访问磁盘的方法      | ( 51 ) |
| 第三节 磁盘I/O驱动程序    | ( 63 ) |
| 第四节 可执行文件的结构     | ( 66 ) |

## 第四章 系统自举

|              |        |
|--------------|--------|
| 第一节 系统文件及其结构 | ( 75 ) |
| 第二节 内核的自举    | ( 78 ) |
| 第三节 BIOS的自举  | ( 80 ) |

## 第五章 显示器控制模块

|              |         |
|--------------|---------|
| 第一节 模块总述     | ( 88 )  |
| 第二节 视频初始化    | ( 94 )  |
| 第三节 光标功能的实现  | ( 96 )  |
| 第四节 字符的读出和显示 | ( 101 ) |
| 第五节 提示行管理与其它 | ( 109 ) |

## 第六章 键盘管理模块

|            |         |
|------------|---------|
| 第一节 模块总述   | ( 115 ) |
| 第二节 重要的工作区 | ( 117 ) |
| 第三节 主体流程   | ( 120 ) |

|     |            |       |
|-----|------------|-------|
| 第四节 | 字符输入的实现    | (123) |
| 第五节 | 功能符的处理     | (128) |
| 第六节 | 输入码的处理     | (132) |
| 第七节 | 高频字统计和字典功能 | (142) |
| 第八节 | 词组输入处理     | (146) |

## 第七章 打印机驱动模块

|     |          |       |
|-----|----------|-------|
| 第一节 | 模块总述     | (153) |
| 第二节 | 模块主体和工作区 | (155) |
| 第三节 | 打印参数的定义  | (159) |
| 第四节 | 字符的打印输出  | (163) |
| 第五节 | 行信息的打印输出 | (169) |
| 第六节 | 屏幕硬拷贝    | (178) |

## 第八章 高分辨率视频适配器及其程序设计

|     |                    |       |
|-----|--------------------|-------|
| 第一节 | 引言                 | (183) |
| 第二节 | EGA视频适配器及其程序设计     | (183) |
| 第三节 | MCGA视频适配器及其程序设计    | (193) |
| 第四节 | VGA视频适配器及其程序设计     | (198) |
| 第五节 | CGE 400视频适配器及其程序设计 | (202) |
| 第六节 | CGE 400的汉字显示       | (207) |

## 第九章 汉字系统优化技术

|     |              |       |
|-----|--------------|-------|
| 第一节 | “透明法”扩充技术    | (211) |
| 第二节 | 外码连接技术       | (215) |
| 第三节 | 词组输入技术       | (218) |
| 第四节 | 联想输入技术       | (225) |
| 第五节 | 磁盘碎片集成技术     | (227) |
| 第六节 | 扩展内存的使用方法    | (232) |
| 第七节 | 进出80286的保护方式 | (244) |

|    |               |       |
|----|---------------|-------|
| 附录 | 信息处理用现代汉语五千词表 | (249) |
|----|---------------|-------|

# 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贮、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,通用语言是汉语。所以,我国社会中要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,各行业也就不可能实现现代化的管理。从我国目前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贮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,在微型机的 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40多K字节。

#### 2. 采用了反贝拷和保密措施

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,指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”,大大减少了翻页机率,提高了输入速度。由于这种统计是动态的,故可以适合于不同用户的使用特点。

#### 5. 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。每种打印机均有两种驱动程序,分别用于打印输出 $16 \times 16$ 点阵汉字和 $24 \times 24$ 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,为打印各式表格提供了方便。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,可使用用户自己的核心文件。

#### 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,采取了一系统的措施,实现了它与GW-BIOS V3.0间的兼容。

## 第二节 系统设计 with 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、被长期的实践检验的丰富的软件财富,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,西文与中文都是符号,只是用不同的信息表示而已。

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的操作系统。

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。

键盘汉字输入处理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型点阵的检索问题。

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贮汉字而使用的代码。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表示，目前规定每个字节也只用七位，其高位未作定义。

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及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、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。

组成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，改为指向内存中的新模块。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,故一个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)中存贮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序列的规则(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,汉字交换码必须采取统一的形式。目前国内计算机系统所采用的标准信息处理交换码,是根据国际标准制定的,即GB1988《信息处理交换用的七位编码字符集》;还制定了相应的代码扩充标准,即GB2311《信息处理交换用七位编码字符集的扩充方法》。因此汉字交换码应与GB1988兼容,并根据GB2311所规定的方法进行编制。由于汉字数量远远大于七位编码所能表示的容量,所以根据GB2311,一个汉字必须用两个或两个以上的七位编码表示。我国首先对六千多个常用汉字制定了交换码的国家标准,即GB2312《信息交换用汉字编码字符集——基本集》,其中每个汉字用对应于GB1988的两个七位码来表示。

## 5. 汉字字形码

由于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,故汉字字形码是指确定一个汉字字形点阵的代码。随着汉字字形点阵的规格不同和汉字字形信息的格式不同,汉字字形码也就不同。CC-DOS中有两种汉字字形点阵规格(1×16点阵和24×24点阵)及两种汉字字形信息格式(显示式和打印式),所以CC-DOS中存在着三种汉字字形码。

# 二、字库与字模

汉字库的主要作用是为汉字输出设备提供汉字的字形数据。汉字库中存放着一批汉字的字形码(字形点阵信息)或字形压缩码,前者能直接表现汉字的字形,后者只能间接表现汉字字形。汉字库由两大部分构成:字形码存贮器和字形码读取程序。从汉字信息在系统内部流程的角度来看,汉字库是一个处于汉字信息输出点上的代码转换器,它吸收的是待输出的机内码,吐出的是字形码。汉字库的设计应追求三项目标:字形质量高、读取速度快和系统开销小。根据汉字库中信息的存贮可分为不压缩型和压缩型。不压缩型汉字库中存放的是其所含汉字的字形码,它的优点是字形质量高和读取速度快,它的缺点是系统开销大。压缩型汉字库中放的是字形压缩码,这样可以大大减少系统开销,一般可减少1/2—3/4的存贮容量。压缩型汉字库的主要缺点是由于对字结构的拆拼而产生字形的失真。另外,在读取速

度上亦要相对慢一些

CC-DOS中有两个汉字库，一个是常用汉字库，其中的字形点阵为 $16 \times 16$ ，用于输出一般的汉字字形。这个汉字库在系统自举时部分或全部驻留在内存中。另一个是 $24 \times 24$ 点阵汉字字形的汉字库，用于输出较高质量的汉字字形。由于这个汉字库的容量较大，故其驻留在磁盘上。CC-DOS的这两个汉字库均为不压缩型汉字库，它们包含国标一、二级汉字和绝大部分符号。它们中的汉字信息均按其所对应之国标码序列进行排列，使得CC-DOS中的汉字地址码与汉字机内码之间有简单的换算规则，加快了它们间的转换。

前面已经提到过，CC-DOS中存在着三种汉字字形码，这是因为CC-DOS中存在两种汉字字形点阵（ $16 \times 16$ 点阵和 $24 \times 24$ 点阵），并且存在两种汉字字形信息格式（点阵形式），汉字字形的信息格式由汉字输出设备决定。CC-DOS通常驱动两种汉字输出设备：显示器和打印机。对于显示器，它希望字形信息中的字节内容应表示点阵中横向点的情况，称为横向字形信息。而打印机则希望字形信息中的字节内容应表示点阵中纵向点的情况，称为纵向字形信息。显然，它们所对应的字形码的编码也不同。再加上汉字字形点阵有两种，照例CC-DOS中应有四种字形码。但是，由于受到显示器分辨率的限制，CC-DOS不支持 $24 \times 24$ 点阵汉字字形的显示，从而只有三种汉字字形码。图1—1是 $16 \times 16$ 点阵横向字形信息格式，图1—2是 $16 \times 16$ 点阵的纵向字形信息格式，图1—3是 $24 \times 24$ 点阵的纵向字形信息格式。CC-DOS的常用汉字库中存放的是汉字横向字形信息，在显示汉字时可以直接取用之。当要打印汉字时，必须把取到的内容转换成汉字的纵向字形信息。

|       |       |
|-------|-------|
| 第0字节  | 第1字节  |
| 第2字节  | 第3字节  |
| 第4字节  | 第5字节  |
| ⋮     | ⋮     |
| 第28字节 | 第29字节 |
| 第30字节 | 第31字节 |

图1—1  $16 \times 16$ 点阵的横向字形信息格式

|       |       |       |   |       |       |
|-------|-------|-------|---|-------|-------|
| 第0字节  | 第1字节  | 第2字节  | ⋮ | 第14字节 | 第15字节 |
| 第16字节 | 第17字节 | 第18字节 | ⋮ | 第30字节 | 第31字节 |

适用于9针打印机输出

|      |      |      |   |       |       |
|------|------|------|---|-------|-------|
| 第0字节 | 第2字节 | 第4字节 | ⋮ | 第28字节 | 第30字节 |
| 第1字节 | 第3字节 | 第5字节 | ⋮ | 第29字节 | 第31字节 |

适用于24针打印机输出

图1—2  $16 \times 16$ 点阵的纵向字形信息格式

由于打印机对打印汉字的实施方法不一样，故汉字的纵向字形信息格式也不尽相同。图 1—2 中给出了适合于 9 针打印机和 24 针打印机使用的两种  $16 \times 16$  点阵纵向字形信息格式，因为 CC-DOS 允许使用 9 针打印机或 24 针打印机来输出  $16 \times 16$  点阵的汉字。

|        |        |        |       |         |         |
|--------|--------|--------|-------|---------|---------|
| 第 0 字节 | 第 8 字节 | 第 6 字节 | ..... | 第 66 字节 | 第 69 字节 |
| 第 1 字节 | 第 4 字节 | 第 7 字节 | ..... | 第 67 字节 | 第 70 字节 |
| 第 2 字节 | 第 5 字节 | 第 8 字节 | ..... | 第 68 字节 | 第 71 字节 |

图 1—8  $24 \times 24$  点阵的纵向字形信息格式

对于高质量汉字 ( $24 \times 24$  点阵) 的输出，在 CC-DOS 中只存在打印输出，而不存在显示输出（其原因已在前面讲过），所以  $24 \times 24$  点阵汉字库中存放的是纵向字形信息。这样，在打印输出  $24 \times 24$  点阵汉字时只要直接取用就可以了，免去转换之劳。另外，CC-DOS 只支持用 24 针打印机打印  $24 \times 24$  点阵汉字，所以图 1—8 中只有一种  $24 \times 24$  点阵的纵向字形信息格式。

## 第二章 CC-DOS V4.0 的操作与使用

### 第一节 引言

CC-DOS V4.0是汉字操作系统CC-DOS的最新版本,它集其以前几个版本之长,具有功能强,性能好,速度快,软件质量高以及可扩充性好等特点。

#### 一、CC-DOS V4.0系统文件介绍

CC-DOS V4.0整个系统分布在5张360KB的软盘上,其中1号盘及2号盘上存放的是24×24点阵的汉字库,该汉字库由CLIB241和CLIB242拼装而成,安装后的24点阵字库必须放在硬盘上。安装的方法有两种,一种是通过盘上的LOAD·BAT批处理文件来进行,另一种方法是用户直接用COPY命令把两张盘的CLIB241和CLIB242拷贝成CLIB24文件,并存于硬盘的根目录下。

CC-DOS V4.0的系统盘为3号盘,运行CC-DOS V4.0时,首先要把这张盘插入A驱动器才能启动。该盘上共有七个文件(见表2—1),其中CCLIB是一个16×16点阵的汉字库,它供系统显示和打印16×16点的汉字用,CCCC.COM是CC-DOS V4.0的系统装入程序,通过它把CCCC.OVR从外存调入内存,并在内存进行重定位。CCCC.OVR是系统的键盘管理模块(INT16H)及其自举部分,它主要完成驻留键盘管理模块及根据当前系统的显示卡配置情况,调入相应的显示器控制模块。CCCC.OV1是彩色显示适配卡的显示器控制模块,CCCC.OV2是单色适配卡的显示器控制模块。3号盘上的另外两个文件是CH.EXE和CH.MSG,它们分别是高频字统计程序和高频字统计记录文件。

CC-DOS V4.0的4号盘是打印机驱动程序盘,它配备了适用于多种型号打印机的打印机驱动程序,例如M2024、TH—3037、MX—100、MX—80、FX—80、LQ—1500、TH—1350、TH—1351、NEC—9400等打印机。该打印驱动程序能打印出16×16点阵或24×24点阵的汉字,并具有定义各种字型、字间距和行间距的功能。

CC-DOS V4.0的5号盘是一张系统支持盘,它上共有14个文件,其中文件名以KEY开头的10个文件是系统提供的附加汉字输入方式处理程序,共有9种输入方式。它们分别是:五笔字形输入法,大众码输入法、仓颉码输入法、电报码输入法、广东话拼音输入法、拼形码输入法、笔形码输入法,声韵和声声码输入法等。另外还有4个文件分别是CH16.EXE、CH24.EXE、CZ.EXE和LOADCZ.EXE。其中CH16.EXE是系统提供的16×16点阵汉字的造字程序,CH24.EXE是系统提供的24×24点阵汉字的造字程序,CZ.EXE是词库建立程序,通过执行CZ.EXE,用户可自定义词组及其输入码到系统词库中,LOADCZ.EXE是词库装入程序,它可以把指定的词库调入内存,并与系统连接,供用户使用。

表2—1列出了整个系统的磁盘文件及其功能。

#### 二、硬件支撑环境

CC-DOS V4.0工作于IBM-PC及其兼容机上。它需要如下硬件条件:

- (1) 具有图形显示功能的单色或彩色显示器一台,以及与其相对应的显示适配卡;

表 2—1 CC-DOS V4.0系统磁盘文件表

| 盘 号 | 文 件 名        | 功 能 及 其 说 明            |
|-----|--------------|------------------------|
| 1   | LOAD.BAT     | 装配24×24点阵字库命令          |
|     | CLIB 241     | 24×24点阵字库的前半部分         |
| 2   | LOAD.BAT     | 装配24×24点阵字库命令          |
|     | CLIB 242     | 24×24点阵字库的后半部分         |
| 3   | CCLIB        | 16×16点阵汉字库             |
|     | CCCC.COM     | 汉字系统装入文件               |
|     | CCCC.OVR     | 汉字系统的键盘管理模块            |
|     | CCCC.OV1     | 彩色显示适配卡的显示模块           |
|     | CCCC.OV2     | 单色显示适配卡的显示模块           |
|     | CH.EXE       | 高频字统计文件                |
|     | CH.MSG       | 高频字统计记录文件              |
|     | TH1351E.COM  | TH-1351型打印机24点阵打印驱动程序  |
| 4   | TH3070E.COM  | TH-3070型打印机24点阵打印驱动程序  |
|     | M2024E.COM   | M2024型打印机24点阵打印驱动程序    |
|     | NECP7E.COM   | NEC-P7型打印机24点阵打印驱动程序   |
|     | NEC9400E.COM | NEC-9400型打印机24点阵打印驱动程序 |
|     | M2024P.COM   | M2040型打印机16点阵打印驱动程序    |
|     | NECP7P.COM   | NEC-P7型打印机16点阵打印驱动程序   |
|     | 1351P.COM    | TH-1351型打印机16点阵打印驱动程序  |
|     | 3070P.COM    | TH-3070型打印机16点阵打印驱动程序  |
|     | 9400P.COM    | NEC-9400型打印机16点阵打印驱动程序 |
|     | 9P.EXE       | 9针打印机打印16×16点阵汉字打印驱动程序 |
| 5   | KEYWB.COM    | 五笔字型键盘输入模块             |
|     | KEYDZ.COM    | 大众码键盘输入模块              |
|     | KEYDB.COM    | 电报码键盘输入模块              |
|     | KEYGY.COM    | 广音输入法键盘输入模块            |
|     | KEYCJ.COM    | 仓颉输入模块                 |
|     | KEYPY.COM    | 拼形码输入模块                |
|     | KEYSS.COM    | 声韵、声声输入模块装入程序          |
|     | KEYSS.OV1    | 声声输入模块                 |
|     | KEYSS.OV2    | 声韵输入模块                 |
|     | KEYBX.COM    | 笔形码输入模块                |
|     | CH24.EXE     | 24×24点阵汉字造字程序          |
|     | CH16.EXE     | 16×16点阵汉字造字程序          |
|     | CZ.EXE       | 词组造词程序                 |
|     | LOADCZ.EXE   | 词库装入程序                 |

(2) 至少一台360KB的双面双密度软盘驱动器及其适配卡,

(8) 至少有128KB的RAM。

### 三、软件支撑环境

CC-DOSV4.0要求有PC-DOS(MS-DOS)2.0以上的版本的支持,也就是说,CC-DOS V4.0的运行需要基于DOS的核心。因此,运行CC-DOS V4.0时,必须预先启动PC-DOS(或MS-DOS)。

## 第二节 系统的启动

### 一、从软盘启动系统

当用户的机器仅有软盘驱动器时，可用下列方法来启动CC-DOS V4.0。

(1) 首先使系统进入西文操作系统之下，由于CC-DOS V4.0要求在PC-DOS 2.0以上版本上运行，因此，用户必须先用PC-DOS 2.0以上版本的操作系统启动机器，

(2) 待系统出现提示符时，把CC-DOS V4.0的系统盘（8号盘）插入A驱动器，

(3) 若用户以前已有自定义过词组，并有词库在系统盘上，则用户可键入CCCC（后跟词库名），这样，系统启动时会自动把指定的词库装入；若用户以前没有定义过词库，则只要键入CCCC和回车键，启动汉字系统。若此时用户也键入CCCC（后跟词库名），而词库又不存在，则系统会提示“CZ” Error的出错信息；

(4) 当用户键入CCCC或CCCC（后跟词库名）时，系统给出如图2-1所示的提示。此时系统要求用户对汉字库驻留内存的情况作出选择。若用户选择0，则表示汉字库全部驻留于外存（原系统盘上），这时，系统占用的内存开销最小，但每次访问汉字库都要访问外存；若用户选择1，则把汉字库中一级字库（共3755个汉字）驻留内存；若用户选择2，

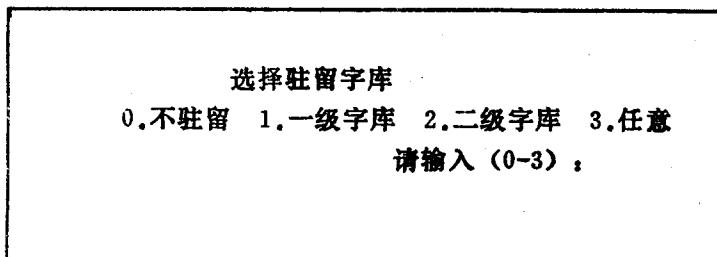


图 2—1 系统启动提示

则把整个汉字库CCLIB驻留内存，此时，访问字库的速度较快；若用户选择3，则表示用户可把汉字库分成任意两部分，一部分驻留内存，一部分驻留外存，此时系统提示出请用户输入驻留结束字，用户可用任意一种输入方法输入一个汉字，输入后系统自动把汉字库以该字为界分成两分，前半部分驻留内存，后半部分驻留外存。例如，用户用区位方式输入6565，则系统把从1区1位到65区65位以前的所有汉字库中的汉字点阵信息驻留内存，其余的驻留外存。

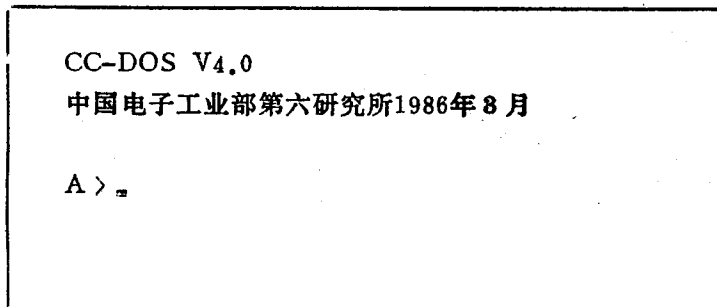


图 2—2 系统完成启动后屏幕显示内容

当用户选择完毕后，系统开始把相应的字库内容驻留于内存，经过一段时间（A驱动器红灯亮），屏幕上出现如图2—2所示的内容，这时系统的核心已全部安装完毕。

## 二、从硬盘启动系统

当用户的机器带有硬盘时，可先把系统安装到硬盘上，用硬盘来完成系统的启动。

但是，这里有一点值得注意，目前市场上流行的CC-DOS V4.0有两种版本，一种是用激光加密的，另一种是已解密的。对于加密过的系统盘即使把整个系统都装到硬盘上，启动时仍需把系统盘（3号盘）插入A驱动器中。这里我们主要介绍以不加密的系统的硬盘安装和启动方法。

（1）把1号盘插入A驱动器中，再键入命令

A: LOAD <回车>

此时，把1号盘上的CLIB241文件拷贝到C盘上，当拷贝结束后，系统提示把2号盘插入A驱动器，则用户可抽出1号盘，插入2号盘，按回车键后，系统把2号盘上的CLIB241文件与CLIB241文件拼接成一个24×24点阵的汉字库文件CLIB24，存于C盘的根目录下。

（2）把3号盘插入A驱动器，键入下列命令

COPY A: \*.\* <回车>

这样把8号盘上的7个文件全部拷贝到C盘上的根目录下。

（3）用户可根据自己系统的配置及打印机的型号，把4号盘上的相应打印驱动程序拷贝到C盘上，另外用户还可根据自己习惯用的输入方法，把5号盘的相应输入方法处理程序文件拷到C盘上。

（4）为了启动方便，可在C盘上建立一个批处理文件，方法如下：

COPY CON CCDOS4.BAT <回车>

CLS <回车>

CCCC [词库名] <回车>

M2024E <回车>

KEYWB <回车>

^Z <回车>

当用户执行这个批处理命令时，系统自动把打印机驱动程序（M2040打印机的24×24点阵汉字打印驱动程序）及五笔字型输入法与系统一起驻留内存。

值得注意的是，用户为了方便起见，也可仅把系统盘（3号盘）上的文件拷到C盘的根目录下，然后键入

CCCC <回车>

来启动系统。

## 三、在带汉卡的机器上启动系统

CC-DOS V4.0在带汉卡的长城机或者带CC-DOS V4.0汉卡的机器上启动时，只需键入CCCC [词库名]即可，此时系统不会提示选择驻留字库，因为此时字库已在汉卡上，不必驻留内存了。