

高等专科学校试用教材

计算机文字处理与自动照排系统

倪萍萍 编



上海交通大学出版社

计算机文字处理与自动照排系统

倪萍萍 编

上海交通大学出版社

内 容 提 要

本书以典型的常用的文字处理系统 C—WORDSTAR 和写研 SAPTRON APS— $\mu 5s$ 第三代自动照排系统、《科印》微机排版系统、华光第四代自动照排系统为中心,系统地介绍了计算机文字处理和自动照相排版系统的原理和使用方法。同时,对文字处理和自动照排系统有关的计算机基本知识也作了简要说明。

本书可作为大专院校出版、印刷制版专业的教材,也适用于从事计算机文字处理、出版印刷工作的工程技术人员学习参考。

沪新登字 205 号

计算机文字处理与自动照排系统

出 版: 上海交通大学出版社

(淮海中路 1984 弄 19 号)

发 行: 新华书店上海发行所

印 刷: 上海出版印刷专科学校实习工厂

开 本: 787×1092(毫米)1/16

印 张: 10.25

字 数: 254000

版 次: 1992 年 8 月第 1 版

印 次: 1992 年 9 月第 1 次

印 数: 1—5000

科 目: 264—301

ISBN7—313—00947—X/TS·8

定 价: 3.45 元

前 言

随着电子计算机的发展和逐步普及,使用电子计算机的半自动和自动照排系统正在逐步取代传统的打字、铅排方式,这场印刷排版领域内的重大变革,不仅推动了印刷厂排版工艺的技术进步,开始改变传统的出版方式,而且使办公室日常文字处理工作现代化变为现实。

编写本书的目的,是使读者对计算机文字处理和自动照排系统有系统的深入的了解。为此,本书以典型的常用的文字处理系统 C—WORDSTAR 和三种重要的自动照排系统为中心,介绍它的基本原理和使用方法。同时,也简要介绍了学习上述内容时要用到的计算机的有关基本知识。

本书第一章阐述了印刷排版技术的发展过程,说明自动照排系统的出现是印刷排版领域内的一场重大变革。

第二章介绍了与文字处理和自动照排系统有关的计算机的一些基本知识,以及中文编辑软件 C—WORDSTAR 的使用方法。

第三章介绍了自动照相排版系统原理,重点放在与用户关系密切的汉字输入方式和计算机版面编辑方法等。

第四、五、六章分别介绍了写研 SAPTRON APS— μ 5S 第三代自动照排系统、《科印》微机排版系统和华光第四代自动照排系统。

第七章作为排版语言的一个典型例子,介绍了华光系统中使用的 BD 排版语言。

复旦大学计算机科学系张然教授审阅了本书全部书稿,并提出了宝贵意见,在此表示感谢。

编 者

目 录

第一章 印刷排版技术的发展

- 第一节 铅合金活字排版工艺..... 1
- 第二节 手选式照相排字工艺..... 2
- 第三节 自动照排系统的发展..... 2

第二章 计算机与文字处理

- 第一节 计算机系统概述..... 4
- 第二节 操作系统 11
- 第三节 中文编辑软件 C—WORDSTAR 20

第三章 自动照相排版系统原理

- 第一节 输入设备及汉字输入方式 30
- 第二节 计算机版面编辑 39
- 第三节 拍摄输出设备 43
- 第四节 自动照排系统的工艺原理 47

第四章 写研 SAPTRON—APS $\mu 5s$ 自动照排系统

- 第一节 系统构成与工艺流程 50
- 第二节 排版拍摄主机 APS $\mu 5s$ 的工作原理与特点 51
- 第三节 SAPTRON 照排系统中的其它设备 58
- 第四节 排版程序 SAPCOL 63

第五章 《科印》微机排版系统

- 第一节 系统构成与操作流程 70
- 第二节 基本版式和版式定义参数 73
- 第三节 文艺理论版排版命令 78
- 第四节 科技数学版式排版命令 97

第六章 华光计算机——激光汉字编辑排版系统

- 第一节 系统构成与工艺流程..... 105
- 第二节 高倍率汉字字形信息的压缩..... 111
- 第三节 照排控制机(TC) 115

第七章 BD 排版语言

- 第一节 说明注解和组版文件..... 121
- 第二节 执行注解和小样文件..... 127

第一章 印刷排版技术的发展

排版是书刊印刷中制版、印刷、装订三个主要工序的核心。它根据原稿内容,按照出版单位提出的版面规格和设计要求,产生以页为单位的版面。排版质量的高低,直接关系到印刷品的质量,排版时间的快慢,直接影响出书的速度,排版所化的成本,又是决定书价的重要因素之一。

第一节 铅合金活字排版工艺

印刷排版经历了从手工抄写演变到木版雕刻,再从木版雕刻演变到活字排版的发展过程。

二世纪初,东汉蔡伦改进了造纸方法,但在当时,书籍还只能靠手工抄写来传播。六——七世纪,我国发明了刻版(木版雕刻)印刷,使文化传播有了更加有效的工具。但由于是用手工雕版,每印一页书就要雕刻一块印版,印一部篇幅较大的书籍往往要化几年的功夫。而且一旦一块印版刻错一个字,整块印版就不能使用。1045年前后,毕昇首先发明了胶泥活字印刷(即中国古代四大发明之一),开创了活字排版印刷的新时代,为现代印刷奠定了基础。

1445年,德国人谷腾堡(Gutenberg)使用铅合金制作活字,把活字排版印刷术又推进了一步。1886—1890年间,西欧开始出现适用于西文排版的莱诺铸排机(Linotype)和摩诺铸排机(Monotype)等机械排字的设备。

不管是手工排字,还是机械排字,都属于熔铅铸字的热式操作工艺。在铅合金活字的生产工艺中,必须先将固态的铅加热熔化成液态,再将铅液灌注入铜模内,经冷却凝成固态的铅字。这种采用热能的排版方式即称为“热排”。“热排”工艺陈旧,速度缓慢,即使是机械化的热式排版,在速度上也将受到一定的限制。如摩诺铸排机的极限速度是2.5字/秒,超越这个速度,灌注入铜模内的铅液将来不及冷凝成型。我国的汉字字数多,字形复杂,机械化的热排设备在我国未被推广应用。我国的热排工艺仍以铅合金活字手工排版工艺为主流,其工艺流程如图1-1所示。

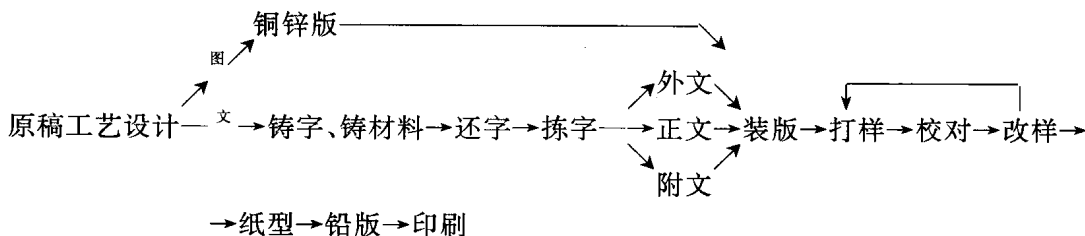


图 1-1 铅字排版工艺流程示意图

活字手工排版工艺,虽然便于改版,原材料消耗很少,但它工序繁多、生产效率低,劳动强

度大;大量的铅字不仅要占用大量有色金属、还必须占用大面积的厂房。因此,热排工艺已越来越不能满足快出书、多出书的要求,热排技术的革新势在必行。

第二节 手选式照相排字工艺

本世纪 20 年代起,国外就开始了照相排字的研制工作。由于它是利用光学摄象原理排字,所以又称为“冷排”。

照相排字的设备是照相排字机,顾名思义,就是打字机和照相机的结合。早期的照排机(即第一代照排机)是由人工根据原稿,一个字一个字地从字模板上选出经曝光拍摄(放大或缩小)而进行排版的。操作者除了要熟记字模的文字排列方法外,还要根据不同的版式要求,及时调节和控制行距、字距的推送量,为此在选字拍摄前必须先对版面作精确的计算和设计工作。

国产的手选式汉字照相排字机于六十年代初期问世,它是由操作人员根据版面计算的结果,由人工从字模板上逐字直接选出并对其拍摄,使装在暗箱内的感光材料感光而排成版面(或长条),经校对、拼版制成供照相制版用的版面(胶片或相纸)。其工艺流程如图 1-2 所示。

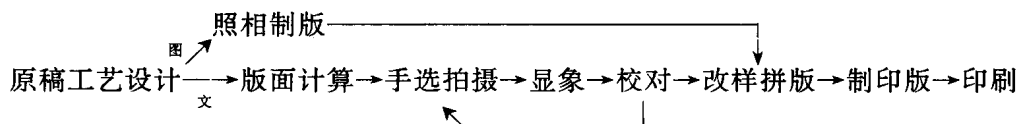


图 1-2 手选照排工艺流程示意图

手选照排机结构简单、造价低廉,便于操作和维修。手选照排工艺的优点是减少了工序,改善了工人的劳动条件,提高了劳动生产率。但是,由于它又从“活版”回到“死版”,一经成版就不便改动;同时拍摄前必须先经版面计算,而版面计算的工作量将随版式的复杂度增加而增加。手选式照排机的这两个弱点大大影响了它的推广使用。

微型电子计算机的发展给手选式照排机注入了新的生命力。它代替人工进行版面计算和某些机器部件的功能,而原有人工选字、曝光拍摄等基本特性不变,这种装有微机的新颖手选照排机又被称为一代半机。例如由日本引进的 PAVO 系列、ROBO 系列手选照排机,以及我国生产的 HDP-3 型多功能手动照排机等都是属于一代半的照相排字机。

第三节 自动照排系统的发展

利用电子计算机解决版面计算、版面修改和选字拍摄的照排系统又称为自动照排系统。

五十年代初期,美国 photon 公司应用旋转字模圆盘和氙灯频闪曝光成像原理,研制成第二代照排机,并于 1954 年发表了第一台 photon200 型自动照排系统。六十年代初,具有处理汉字功能的日本写研公司的 SAPTON、富士通公司的 FACOM 等系统相继问世。近四十年来,自动照排系统已从光机式的二代机发展到 CRT 全电子式的三代机和激光式四代机,生产工艺也日趋完善。

我国的汉字字数多、字形复杂,使发展我国的自动照排系统具有更大的难度。1970 年左右,我国上海、北京两地分别开始了对自动照排系统的研制工作,上海中华印刷厂和复旦大学等单位协作研制的 ZZPZ-701 型光机式第二代照排系统,是我国第一台投入正常生产使用的国产自动照排系统。为了迅速改变我国在文字信息处理方面的落后面貌,1974 年 8 月经国家计委批准,由当时的四机部(现电子工业部)牵头,开始了对汉字信息处理的研究工作。1975 年 5 月,北京大学参与并主持“计算机——激光汉字编辑排版系统”的研究工作。81 年 5 月通过华光 I 型机(激光式四代机)验收,85 年 8 月完成 II 型机,86 年研制了华光 III 型机,目前已推出 IV 型机。与此同时,中国印刷技术研究所于 1985 年 6 月发表了可在 IBM PC/XT 机上运行的汉字书刊排版高级文书处理系统软件包,87 年推出《科印》科技排版软件。加上近年来全国各地引进的英国摩诺公司、日本写研公司和森泽公司的照排系统,使我国自动照排技术的推广应用有了很大的发展。

以计算机为核心的自动照排系统的出现,使文字排版处理技术发展到一个崭新阶段,从根本上改变了铅活字排版和手选照排的工艺流程。推广应用自动照排系统的最终设想是:由出版社编辑部,或作者本人来完成稿件的排版工作;由照排中心进行照排并提供制版用的软件,甚至印版,然后交付印刷厂印刷、装订。随着各类照排系统的不断完善和发展,这一理想的实现已经指日可待。

第二章 计算机与文字处理

第一节 计算机系统概述

从 1946 年世界上第一台电子计算机(ENIAC)诞生以来,计算机已经经历了第一代(电子管计算机)、第二代(晶体管计算机)、第三代(集成电路计算机)和第四代(大规模集成电路计算机)的发展过程。计算机的性能和工艺取得了惊人的进步和发展,计算机的应用几乎遍及各行各业,但是计算机的结构和工作原理基本上没有改变。计算机只能执行有限数目的一些基本操作,例如算术运算中的加、减、乘、除,逻辑运算中的与、或、非,比较、移位、数据传送,以及计算机与用户或外存储器之间交换数据的输入输出等。要计算机完成任何一项指定的任务,都必须先将为完成该任务所必须执行的操作命令依次排列起来,构成一个程序(即程序设计),并把该程序和执行时所用到的一些数据预先存储到计算机的存储器中,由计算机自动取出并分析执行。除此以外,在计算机内部,程序和数据都以二进制的代码形式表示。

一、计算机硬件的基本组成

计算机系统由硬件和软件两部分组成。所谓硬件,是指构成计算机的各种器件,以及由这些器件组成的计算机部件或整个计算机。组成计算机硬件的主要部件有:

1. 主存储器

主存储器是用来存放信息(程序和数据)的装置。可以把存储器看成是由统一编号的很多房间组成,它的功能是按指定的编号(即地址)写入或读出信息。存储器能存放的信息的总量称为存储器的容量,它是衡量计算机性能的指标之一。通常取 8 个二进制位作为一个单位,叫做字节(Byte)。并以千字节(KB)或兆字节(MB)作为存储器容量的计量单位:

$$1\text{KB}=1024 \text{ 字节}$$

$$1\text{MB}=1024 \times 1024 \text{ 字节}$$

一般,主存储器由随机存储器(RAM)和只读存储器(ROM)两部分组成,它们都可以由程序指令访问。但是只读存储器的存储内容只能读出,不能写入,它常用来存放预先一次写入的固定不变的程序和常数。

2. 中央处理器(CPU)

中央处理器由运算器和控制器组成。运算器是用来完成各种运算的装置,它的功能是在控制器的控制下完成算术运算、逻辑运算和其它运算。衡量运算器性能的一项指标是运算数据的位数,又称为字长。控制器是用来控制计算机各个部件协调工作的装置,它的功能是根据指令的要求有序地、有目的地向各个部件发出控制信息。衡量控制器性能的主要指标是指令系统和运算速度。指令系统是指计算机能够执行的操作种类,不同的操作由不同的指令表示。运算速度指计算机 1 秒钟内能执行的平均指令数,通常用 MIPS(每秒钟百万条命令)为单位。

3. 输入输出设备

输入输出设备用于使用者和计算机之间的数据传输。输入设备的功能是把程序或数据转换成计算机能够识别的信息并输入到计算机内部,输出设备的功能是把计算机内部的信息(例如运算操作的结果)输出到计算机外部并变换成人们或其它设备能够识别的信息。目前,微机系统中最常用的输入设备是键盘,最常用的输出设备是显示器(CRT)和打印机。磁盘(硬盘和软盘)是既作为输入设备、又作为输出设备的辅助存储器,也称为外存储器。

4. 外存储器

中央处理器执行命令是从主存储器中按地址进行的。但是一个计算机系统的主存储器容量总是有限的,大量需要保存的程序和数据不可能全部放在主存储器中,而必须以文件形式保存在外存储器中。外存储器的特点是容量大、成本低,但存取数据的速度要比主存储器慢得多。在计算机运行过程中,可以随时把外存储器中有关文件取至主存储器,或者把主存储器中的程序或数据组织成文件后保存在外存储器中。

微机系统中最常用的外存储器是硬磁盘和软磁盘。

软磁盘(简称软盘)是一张薄塑料圆片,其表面是一层磁化物的涂层,外面装有一个纸保护套。软盘的容量取决于信息在盘面上的记录格式和记录密度。为了能正确地在软盘上读写信息,盘面上的信息是按一定的格式存储的——沿盘面半径方向划分成一个一个磁道,每个磁道又分成相同个数的扇形区域。每个扇区能记录的数据字节数称为记录密度。

在各类照排系统中所用的软盘有 8 英寸和 $5\frac{1}{4}$ 英寸两种。

8 英寸软盘分 77 个磁道,最外面的是 00 道,最里面的是 76 道。每道划分为 26 个扇区,它的记录密度有 128 个字节/区(称为单密度盘片)和 256 字节/区(称为倍密度盘片)两种。

$5\frac{1}{4}$ 英寸倍密度盘片(简称 5 英寸盘)是目前微机系统使用最多的一种软盘,它分 40 个磁道,每道 9 个扇区,常见的记录密度是 512 字节/区(如图 2-1 所示)。可见,常用的 5 英寸双面倍密度软盘片的存储容量计算公式为:

$$C = 2(\text{面}) \times 40(\text{道/面}) \times 9(\text{扇区/道}) \times 512(\text{字节/扇区}) = 360\text{KB}$$

近年来出现的 5 英寸高密度软盘片的存储容量达 1.2MB,其计算公式为:

$$C = 2(\text{面}) \times 80(\text{道/面}) \times 15(\text{扇区/道}) \times 512(\text{字节/扇区}) = 1.2\text{MB}$$

为此,5 英寸软盘驱动器也有 360KB 低密和 1.2MB 高密两种。使用时,软盘和软盘驱动器一定要兼容,即低密驱动器只能读写低密度软盘,而高密驱动器既可以读写高密软盘,也可以读写低密软盘。

盘片通过软盘驱动器读写信息时,盘片保护套下方的椭圆孔(称为磁头槽)是驱动器磁头读写信息的位置。中间的大圆孔(驱动器轴孔)用于把盘片固定在驱动器轴上高速旋转。旁边的小圆孔(索引孔)用于定位读写盘片时的初始位置,软盘片上亦有一个小圆孔,当旋转对准保护套上的小圆孔时即为读写初始定位。封套下部(8 英寸盘)或右边(5 英寸盘)的方形缺口称“写保护缺口”,所谓“写保护”是指只能从软盘片上读出信息,而不能把信息写入盘片。对于 8

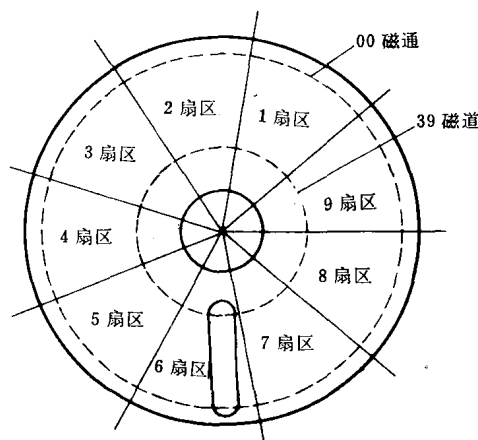


图 2-1 软磁盘磁道扇区分布图

英寸盘来说,撕掉封口实现写保护,而 5 英寸盘正好相反,封住缺口时实现写保护。

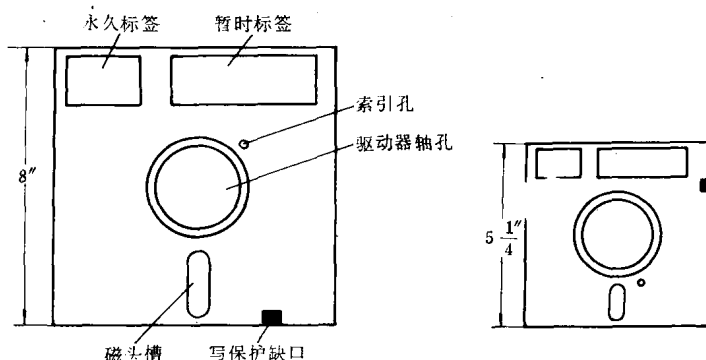


图 2-2 磁盘外形图

使用软盘时应注意如下事项:

- (1) 使用温度应在 $10^{\circ}\text{C} \sim 52^{\circ}\text{C}$ 之间。
- (2) 使用时,不要用手直接接触暴露在保护套外的盘片表面部分。盘片插入驱动器时,应将贴有标签的面朝上,磁头槽方向先插入,切勿反插或倒插。
- (3) 使用后要小心地从驱动器小门取出,并插入软盘片纸袋。
- (4) 禁止把软盘片弯曲,防止盘片被重压、划伤和弄脏。
- (5) 盘片应远离磁场,避免阳光直射。存有需要长期保存文件的盘片,最好每半年复制一次,以防信息丢失。

硬磁盘的工作原理与软磁盘相似,只是它是由若干片表面涂有磁化物的金属圆片组成。温彻斯特磁盘是微机系统中常用的一种硬磁盘,它是一种盘片不可更换的固定盘。例如,IBM PC/XT 微机系统中装有容量为 10MB 的硬盘,它的主要指标是:4 个记录面,每面 306 道,每道 17 个扇区,每扇区记录 512 个字节,可知其总容量 $C=10.4\text{MB}$ 。

二、数在计算机中的表示方法——二进制

在日常生活中,人们常遇到各种数制的数。例如,算术中常使用十进制数。计时系统中,规定 60 秒为 1 分钟,60 分为 1 小时,使用的是六十进制数。计算机内部使用的则是二进制数。

十进制数的特点是有十个数码(0,1,2,3,4,5,6,7,8,9),逢十进一。同理,二进制数只有二个数码(0 和 1),逢二进一。任何具有两个不同稳定状态的电子器件都可以用来表示二进制数中的一位数码,例如电压的高和低,电流的通和断。二进制的算术运算规则很简单,可以使计算机的处理速度十分迅速:

$$\begin{array}{lll}
 0+0=0 & 0-0=0 & 0\times 0=0 \\
 0+1=1 & 10-1=1 & 0\times 1=0 \\
 1+0=1 & 1-0=1 & 1\times 0=0 \\
 1+1=10 & 1-1=0 & 1\times 1=1
 \end{array}$$

要把一个十进制整数化为二进制整数,只需将它一次又一次地被 2 除(直到被除数等于零时为止),得到的余数(从最后一次得到的余数读起)就是用二进制表示的数。如:

$$\begin{array}{r}
 2 \overline{) 11} \quad (1 \\
 \underline{2} \\
 2 \quad (1 \\
 \underline{2} \\
 2 \quad (0 \\
 \underline{2} \\
 1 \quad (1 \\
 \underline{1} \\
 0
 \end{array}$$

得到 $11_{(10)} = 1011_{(2)}$

换言之,只要把十进制数化成以 2 为底的指数形式,其系数的顺序排列(由高次到低次)就是以二进制表示的数。例如:

$$11_{(10)} = 1 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 1 \times 2^0 = 1011_{(2)}$$

反之,要把一个二进制整数化为十进制数,只需将它的最后一位乘以 2^0 ,最后第二位乘以 2^1 ,……,依次类推,再将各项相加就得到用十进制表示的数,如:

$$\begin{aligned}
 10111_{(2)} &= 1 \times 2^5 + 0 \times 2^4 + 1 \times 2^3 + 1 \times 2^2 + 1 \times 2^1 + 1 \times 2^0 \\
 &= 32 + 8 + 4 + 2 + 1 = 47_{(10)}
 \end{aligned}$$

二进制数书写起来很不方便。例如十进制数 419 写成二进制形式是 $110100011_{(2)}$,一个五位十进制整数转换成二进制数将有 14~17 位之多。如果用十六进制数,书写起来就比二进制数短得多。

把二进制数从低位向高位,每四位数作为一组,那么每组能代表一个从 0~15 之间的十进制数(四位二进制数不会大于等于 16)。也就是说,以四位二进制数作为一组的数是逢十六进一的,它应该有 16 个数码组成——即十六进制数。

十进制数与二进制数、十六进制数的对应关系如表 2-1 所示。

表 2-1 十进制数与二进制数、十六进制数的对应关系

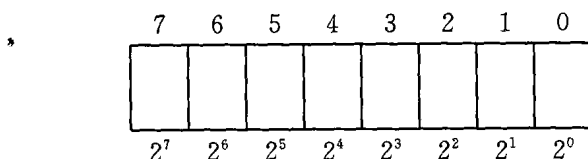
十进制数	二进制数	十六进制数
0	0000	0
1	0001	1
2	0010	2
3	0011	3
4	0100	4
5	0101	5
6	0110	6
7	0111	7
8	1000	8
9	1001	9
10	1010	A
11	1011	B
12	1100	C
13	1101	D
14	1110	E
15	1111	F

三、文字在计算机中的表示方法——机内码和字形码

需要计算机处理的文字必须存入内存储器,才能由中央处理器执行,而存储在计算机中的数据只能以0和1的组合形式存放。为此,必须把文字转换成二进制的代码形式,即文字的机内码。

1. 西文机内码

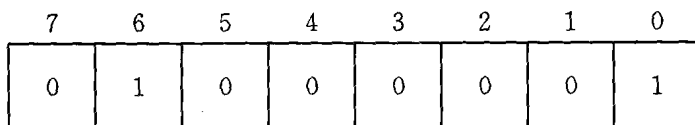
存储器的单位——字节由8位二进制位组成:



最低的第0位为1时表示 $1(2^0=1)$,第1位为1时表示 $2(2^1=2)$,第2位为1时表示 $4(2^2=4)$,……,第7位为1时表示 $128(2^7=128)$ 。可见,一个字节可以表示的最小二进制代码是 $00000000_{(2)}$,它对应于十进制数的0,或者十六进制数的 $00_{(16)}$ 。一个字节可以表示的最大的二进制代码是 $11111111_{(2)}$,它对应于十进制数的 $255(=2^0+2^1+\cdots+2^7)$,或十六进制数的 $FF_{(16)}$ 。因此,一个字节可以用来表示256种不同的代码。

如果使一个字节的內容(二进制代码)和计算机输入键盘盘面上的各种符号和控制键(统称为字符)建立一一对应的关系,那么在键盘上打入一个键就相当于向计算机输入了一个代码。什么键对应什么代码,目前已有标准规定,其中使用最多的一种标准就是ASCII码(美国信息交换标准代码——American Standard Code for Information Interchange)。

ASCII码把输入键盘上的每个字符用7位二进制组成的一个代码来表示。因此,只需一个字节就可以存储一个西文字符的代码,字节中的最高位(第7位)恒为零。例如,字母“A”的机内代码是:



用十六进制可表示为 $41_{(16)}$,用十进制表示为65。

2. 汉字的机内码

一字节的代码作为西文字符的机内码已经足够,但对汉字来说,却远远不能满足需要。但西文机内码的设计方法却可应用于汉字,即把所有可区别的汉字排成一张有序的表格,用序号作为对应汉字的机内码。

为了有利于各种汉字处理系统的信息交换,1981年我国颁布了作为国家标准的《通讯用汉字字符集(基本集)及其交换码标准》GB2312-80方案,作为信息交换用的汉字标准编码,又称为国标码。

国标码用二个7位二进制代码代表一个汉字,分列在二个字节中。每个字节的取值范围均在 $21_{(16)}\sim 7E_{(16)}$ 之内,字符集的最大容量为8838个字符。目前已收容的字符包括6763个汉字和700多个外文字母、数字、标点和符号。国标码的分布图如图2-3所示。

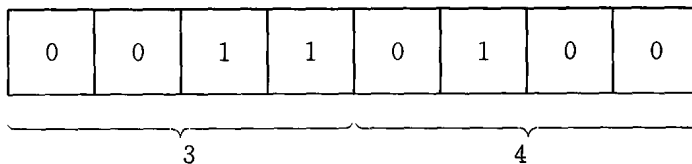
第二字节

	21 22 23.....7E	
第一 字 节	21 . . . 27 . . . 30 . . . 57 58 . . . 77 78 . . . 7E	字母、数字、图符、记号区
	空白区	
	一级汉字（3755个）（拼音字母序）	
	二级汉字（3008个）（部首笔划序）	
	空白区	

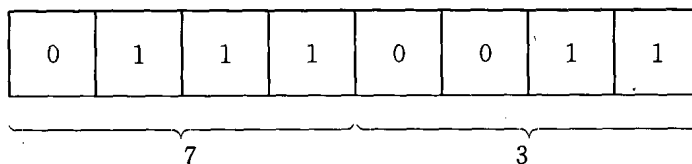
图2—3国家标准 (GB2312—80) 汉字字符集

选取汉字机内码,除了应与国标码之间有简单的对应关系外,还必须避免汉字机内码与 ASCII 码的二义性。为了区分二个字节的内容是表示一个汉字字符还是二个 ASCII 码字符,一种简单的方法是将国标码中的每个字节的最高位(第 7 位)置 1 后作为汉字字符的机内码。例如“大”的国标码是 3473₍₁₆₎:

第一字节

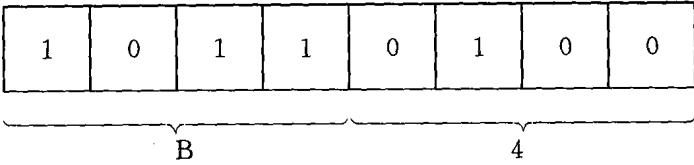


第二字节

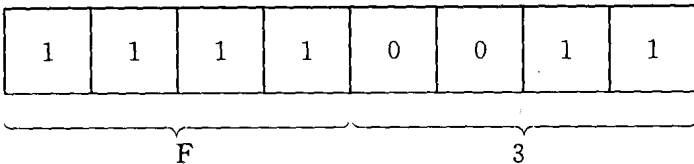


则“大”的机内码为 B4F3₍₁₆₎:

第一字节



第二字节



也就是说,把国标码的两个字节分别加上 $80_{(16)}$,就可得到该文字的机内代码。

3. 汉字的字形存储与字形码

汉字的机内代码只是汉字在计算机内的识别码,当需要输出(显示或打印)某个汉字时还必须将它的字形存储于计算机中。

计算机存储汉字字形的常用方法是把每个汉字按一定的密度离散成网点,每点以一个二进制位表示:即有笔划时为 1,没有笔划时为 0,从而就可以用一组二进制代码表示该汉字的点阵式字模,称为字形码。全部汉字点阵式字模的集合称为汉字的字模库。计算机系统中常用的汉字字模有 16×16 点阵和 24×24 点阵两种,前者一般用于屏幕的显示输出,后者主要用于针式打印机输出。图 2-4 是汉字“大”的 16×16 点阵式字模,它的字形码需要用 32 个字节来表示。用十六进制书写时,它的字形码是:

字节	代码	字节	代码	字节	代码
1	03	12	00	23	0C
2	00	13	03	24	20
3	03	14	00	25	18
4	00	15	03	26	30
5	03	16	00	27	10
6	00	17	03	28	18
7	03	18	00	29	20
8	04	19	03	30	0E
9	FF	20	80	31	C0
10	FE	21	06	32	04
11	03	22	40		

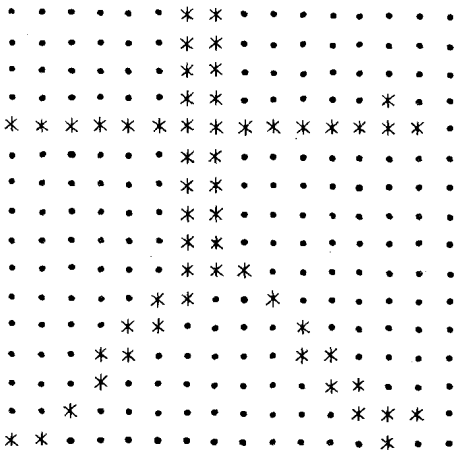


图 2-4 汉字点阵式字模

同理,对于 24×24 点阵的字模,每个汉字的字形码需占 72 个字节。但对于能应用于制版印刷的精密型字模库来说,汉字的点阵数应该随汉字字号的增大而加大,一般至少要在 25 线/毫米以上。也就是说,对于 4 毫米大小的汉字,至少需要存储 100×100 点阵的字模,即每个汉字的字形码需占用 1250 个字节。为此,精密型字库,往往需要将字形代码先经高度压缩以后再

存储于字模库,以便减少对存储器的压力。

第二节 操作系统

电子计算机发展到今天,操作系统已成为其不可缺少的组成部分。它是用户和计算机硬件之间的接口,用户需通过操作系统使用计算机。

操作系统由一类程序组成,它的作用是控制和管理计算机硬件资源(例如内存储器的分配,输入输出设备的使用等)和软件资源;合理地组织计算机工作流程;以及把复杂的人机对话简化为易学易用的操作命令供用户使用。

使用计算机,首先要学会使用操作系统的命令。但不同机种上的操作系统可能不相同,即使同一机种,也可能运行不同的操作系统。本节主要介绍 IBM PC 机上使用最多的操作系统 MS-DOS 和扩充后的中文操作系统。

一、MS-DOS 概述

MS-DOS 是美国 Microsoft 公司发行的磁盘操作系统,在 IBM PC 上使用的 MS-DOS 又叫做 PC-DOS。

MS-DOS 系统经过不断改进,已有各种版本,例如 MS-DOS2.0,MS-DOS3.0 等。版本号越高,功能越强,但不同版本间基本上是兼容的。所谓兼容,指低版本的操作命令,在高版本上同样可以使用。

1. MS-DOS 的启动

启动计算机,首先指将 MS-DOS 的有关文件从磁盘上装入内存储器,使计算机能够接受用户的操作命令。

当计算机电源未接通时,启动 MS-DOS(又称冷启动)的步骤如下:

- (1) 将 MS-DOS 软盘插入 A 驱动器,关好驱动器小门;
- (2) 打开显示器和打印机(必要时)的电源开关;
- (3) 打开主机箱上的电源开关。

这时,中央处理器对有关硬设备进行检验。检测完毕后,从 A 盘的固定位置上将 MS-DOS 的有关文件装入内存储器,然后屏幕上依次出现当前日期、时间的提示和询问。操作者应按提示的格式依次输入当前的日期和时间,如果操作者认为没有提供当前日期(或时间)的必要,也可以直接以“回车”键回答。此后,屏幕即显示出有关版本的说明性信息和等待用户输入命令的提示符:

A>

A 表示当前的工作磁盘是 A 盘,用户命令中涉及的文件名如未指明所在磁盘的盘符时,一律到当前工作盘上去查找。

当计算机电源已打开时,启动 MS-DOS(又称热启动)的步骤为:

- (1) 将 MS-DOS 软盘插入 A 驱动器,关好驱动器小门;
- (2) 同时按下〈Ctrl〉、〈Alt〉、〈Del〉三个键。

如果 MS-DOS 的有关文件已全部复制在硬磁盘 C 上,启动前应打开 A 驱动器的小门,启动时计算机即会自动地从 C 盘装入 DOS。此时,等待用户输入命令的提示符为:

C>

表示 C 盘为当前工作盘。

2. 文件

文件是一组相关信息的集合。一个程序，一篇文章都可以构成一个文件。操作系统下的所有程序和数据都是以文件的形式存储在磁盘上的。

为了便于对文件进行各种操作，磁盘上的每个文件都必须有一个唯一的标记——文件引用名。

文件引用名由盘符、文件名和扩展名三部分组成。其格式为：

[<盘符>]<文件名>[.<扩展名>]

其中盘符指文件所在盘的标记，例如“A:”“B:”或“C:”分别表示文件所在盘是 A 盘、B 盘或 C 盘，缺省时表示文件所在盘是当前盘。文件名由 1~8 个 ASCII 字符构成。扩展名一般用来说明该文件所属的类型，由 1~3 个 ASCII 字符构成。下列 ASCII 字符都可以用于文件名和扩展名中：

A~Z(大小写字母意义相同)，0~9，

\$_#&.@,!%,(),-,{},“等

文件引用名又可分成单义引用名和多义引用名两类。一个单义引用名仅和一个文件对应，而多义引用名可以对应多个文件。

当文件引用名中包含替代符号时称为多义引用名，MS-DOS 中使用的替代符有“?”和“*”共两个。“?”用来代替所在处的任一字符，“*”用来代替从所在位置到下一间隔符(指<.>或<空格>)之间的一串字符。例如 CB?.BAS 可以表示 CB1.BAS, CB2.BAS, ……等多个文件，*.BAS 则表示扩展名为.BAS 的所有文件。

3. 目录和路径

DOS2.00 以前的版本使用单一的目录结构，即所有的文件都在同一个目录的管理下。随着硬盘的使用，一个硬盘可以保存数千个文件，使用单一的目录结构就会出现检索效率太低的矛盾。自 2.00 版开始，MS-DOS 的文件系统采用树型目录结构。树中的每个结点都有一个名字以供访问，树的结点分三类：根结点表示根目录；树枝结点表示子目录；而树叶则表示普通文件。

根目录又称为系统目录，每个盘上只有一个根目录，它在盘片初始化时自动建立。子目录是包含在根目录或其它子目录中的目录，它们是通过操作员使用“建立子目录命令”而建立的。一个目录中可以建立多个子目录，每个子目录都必须具有目录名，目录名的格式与文件名相类似，也是由 1~8 个 ASCII 字符组成。

对于树状结构的文件系统，为了建立或寻找一个文件，DOS 不仅需要知道该文件的文件引用名，还必须知道包含该文件的目录名以及进入该目录的路径。DOS 检索文件的起始目录称为当前目录，DOS 初始启动时，当前目录是根目录。如果被操作的文件在当前目录中，检索该文件时只需指出文件引用名即可。如果文件不在当前目录中，则还必须指出从当前目录(或根目录)到文件所在目录的路径(分别称为相对路径和绝对路径)。说明路径的方法是依次列出每一级子目录名，每个子目录名前面加“\”斜杠。例如自根目录开始通过 FIRST, SECOND, THREE 三级子目录寻找文件 YS100 的路径为：

\FIRST\SECOND\THREE\YS100