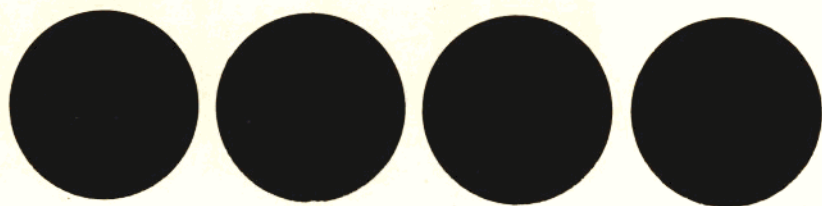


马义忠 张连珠 编



微机汉字信息处理

微机汉字信息处理

微机汉字信息处理

微机汉字信息处理

兰州大学出版社

前 言

随着科学技术的发展,计算机应用技术不断普及,尤其在信息社会时代,人们都渴望有一种先进的工具和方法来获取和处理信息。而应用计算机处理汉字信息,已成为当今计算机应用中一个十分重要的领域。不管你从事何种工作,使用它都会给你带来意想不到的收获和方便。在我们这个使用汉语的国度里,用计算机处理各种汉字信息,将是计算机得以普及的重要标志,也是计算机在我国的最大应用市场。本书就是围绕如何使用计算机处理各种汉字信息,特别是在办公室自动化和激光照排轻印刷等方面,如何对中外文资料、著作、信函、合同文件利用计算机进行处理等中心议题而编写的。对于目前一些优秀的并具有代表性的汉字信息处理系统软件和工具,书中均作了详细的介绍。全书以“学以致用”为原则,通俗易懂,简明易学,实用性很强。

本书共分七章。第一章主要介绍计算机的基本原理及系统配置;第二章介绍汉字操作系统 CC-DOS 的基本命令及使用;第三章介绍了几种常用的汉字输入方法;第四章介绍了两个优秀的汉字编辑软件;第五章,第六章详细介绍了计算机编辑排版系统的原理和“科印”微机书刊排版系统软件;第七章介绍四通 MS-2403 中外文文字处理机的使用。本书可作为文科各类本科生的教材,大约需 60 个学时左右,并附有上机实习介绍,上机实习要求 50 个学时。可根据文科各专业的不同特点,将第六、七章作为选修。本书也可作为理科和其他专业技术人员的参考书。

本书由马义忠、张连珠同志编写,在编写过程中席先觉副教授审阅了全部书稿并提出了宝贵的意见,在此表示感谢。

由于时间仓促,书中难免有错误和不妥之处,敬请广大读者给予批评指正,不胜感谢!

编者

一九九一年十二月于兰州

目 录

第一章：计算机处理文字的基本原理.....	(1)
§ 1. 1 计算机的基本原理及系统配置.....	(1)
§ 1. 2 汉字信息处理的意义和任务.....	(8)
§ 1. 3 汉字信息处理系统的构成和分类	(10)
§ 1. 4 汉字信息处理技术的发展前景	(12)
习题	(12)
第二章：汉字操作系统 CCDOS	(14)
§ 2. 1 DOS 简介	(14)
§ 2. 2 DOS 命令的类型及使用	(15)
§ 2. 3 汉字操作系统 CCDOS	(27)
§ 2. 4 磁盘文件系统	(29)
习题	(30)
第三章：计算机输入汉字的几种方法	(32)
§ 3. 1 概述	(32)
§ 3. 2 CCDOS 下的功能键介绍	(32)
§ 3. 3 区位码输入法	(34)
§ 3. 4 首尾码输入法	(35)
§ 3. 5 拼音码输入法	(37)
§ 3. 6 快速输入法	(38)
§ 3. 7 词组输入	(38)
§ 3. 8 其它键的使用	(39)
§ 3. 9 五笔字型汉字输入技术	(39)
习题	(55)
第四章：几个常用编辑软件的使用说明	(57)
§ 4. 1 汉字文字编辑软件 (Wordstar)	(57)
§ 4. 2 CCED	(67)
习题	(71)
第五章：计算机排版系统的原理	(72)
§ 5. 1 精密型照相排字机的几个发展阶段	(72)
§ 5. 2 编辑排版系统的构成	(74)
§ 5. 3 “科印”微机书刊排版系统	(76)
习题	(79)
第六章：“科印”文艺理论版版式命令.....	(80)
§ 6. 1 排版概述	(80)

§ 6. 2 操作	(83)
§ 6. 3 基本行排命令	(87)
§ 6. 4 版面处理命令	(90)
§ 6. 5 特殊排版命令	(94)
§ 6. 6 表格排版命令	(101)
习题	(108)
第七章：四通中外文文字处理机简介	(110)
§ 7. 1 键盘介绍	(110)
§ 7. 2 文章的输入方法	(112)
§ 7. 3 文章的编辑、修改技巧	(123)
§ 7. 4 制表	(148)
§ 7. 5 文章印刷	(153)
附录一：基本 ASC I 码表	(158)
附录二：在几种流行的汉字操作系统中如何使用 WS 打印输出	(160)
附录三：常见非基本字根拆分示例	(162)
附录四：五笔字型速成指南	(164)
附录五：《科印》排版软件版式命令集	(169)
附录六：《科印》版式命令错误信息表	(171)
附录七：合并文件程序使用说明	(172)

第一章 计算机处理文字的基本原理

这一章将主要介绍有关计算机的系统构成,以及计算机处理文字的基本原理,另外还将介绍与信息处理有关的一些基本概念。

§ 1.1 计算机的基本原理及系统配置

1.1.1 计算机的发展史

电子计算机是当今新技术革命的先导技术。自从 1946 年世界上第一台电子计算机问世以来,计算机已经从电子管计算机、晶体管计算机、集成电路计算机,发展到第四代大规模集成电路及超大规模集成电路计算机。下面我们看看计算机在各个发展阶段的特点。

一、电子管时代:1946 年出现了世界上第一台电子计算机“ENIAC”,该机的硬件主要由电子管组成,全机共用了 18000 个电子管,1500 个继电器(耗电 150 千瓦,每秒运算 5000 次,占地 1800 平方英尺,自从第一代计算机问世以来,计算机就以“迅猛”地速度往前发展。电子管时代持续了大约 10 年的时间(1947~1957),这一时代的计算机的特点是体积大,运算速度慢,计算机所配的软件只能是符号语言和汇编程序,软件的通用性差,仅适用于专门的科学计算。

二、晶体管时代(1958~1964):由于电子行业的发展,出现了晶体管,这时人们考虑用晶体管来替代原来的电子管,就出现了晶体管计算机,软件开始出现,多道程序设计,管理程序等软件,主要用于科学计算、数据处理、业务管理等方面。

三、集成电路(1965~1970):处于这一时代的计算机硬件主要由各种中、小规模集成电路组成,许多逻辑功能都可以集成在一个小芯片上,一个芯片可替代许多电子管或晶体管,使得计算机的体积迅速缩小,耗电量减少。运算速度提高,且计算机可以配有专门的操作系统,随着高级语言的发展,计算机软件的可移植性提高,可实现系列化、标准化,广泛应用于各个领域。

四、大规模集成电路(70 年代后):1971 年美国 Intel 公司研制出了 Intel4004 微处理器,伴随着该处理器的出现,各种大规模集成电路也应运而生,这时的集成电路的集成度很高。各种半导体存储器,各种功能的逻辑芯片的相应出现,给计算机的发展带来了革命性的变化,也就在七十年代出现了世界上第一台微型计算机,这时的计算机可以配备种应用软件,最有特色的像各种可扩充语言、数据库技术、各种电子表软件、办公室自动化软件、计算机网络、计算机辅助设计(CAD)、计算机控制、计算机绘图等,可以说计算机发展普及到今天已“无孔不入”地深入到社会生活的各个方面。

超大规模集成电路技术的出现,使计算机的速度更加提高、体积更小、成本更低,应用将更加普及。

1·1·2 计算机的基本结构原理

计算机有这样广的用途,并且已经深入到社会生活的各个方面,那么它是不是高深莫测呢?其实计算机并不神秘,最初它的运算原理还是依赖于人们用算盘解题的过程,只是由机器代替人。因此,和用算盘解题一样,它必须有如下几种设备:

①运算器:相当于算盘,在这里完成所需的各种运算。

②存储器:计算机必须要能保存和记录原始数据、运算步骤以及中间结果,也就是说需要“记忆装置”,它相当于算盘解题过程中的纸和笔。

③控制器:计算机要有一个代替人的脑和手的作用、支配机器进行自动控制的装置,这就是控制器,它是计算机的神经中枢,由它统一指挥和控制计算机各部分的联系。如什么时候取数,在什么地方取数,送到什么地方,进行什么运算,算完后的结果送到哪里去等,它相当于用算盘解题过程中人脑和手进行的工作。

④输入设备:通过该设备,人们可以将原始数据,运行程序送到计算机中去,常用的输入设备有软盘和磁带读入机、键盘等。

⑤输出设备:通过该设备来输出计算结果。常用的输出设备有打印机、阴极射线管显示器、各种磁盘等。

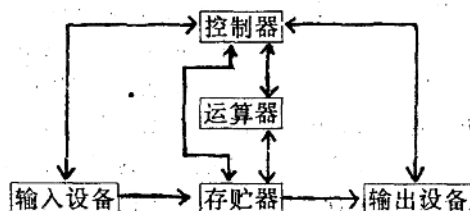


图 1-1 电子计算机组成框图

计算机的组成通常由以上五个部分完成,这五部分之间的联系及协调工作过程如图 1-1 所示。

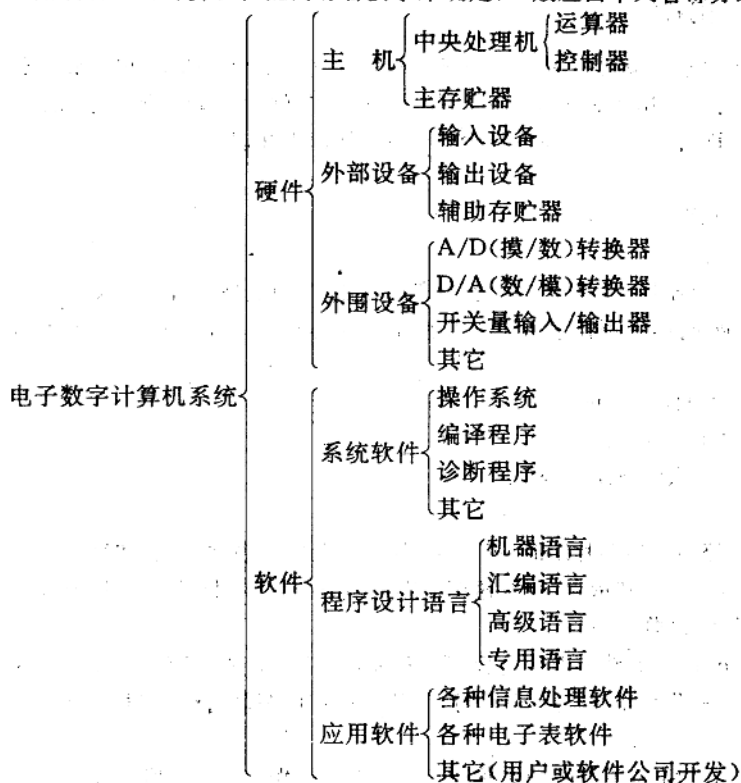
当计算机发展到集成电路时代之后,因运算器和控制器通常集成在一个芯片上,我们称该芯片为中央处理机,用 CPU(Central Processing Unit)表示。把存放在计算机内部的存储器称为内存贮器(或主存贮器)。另外由于现代电子计算机的发展以及为满足各行各业使用各种不同外部设备的需要,所以在计算机的输入和输出接口方面又做了许多扩充,称之为输入输出通道(I/O 通道)。由此看来,电子计算机是由中央处理机(CPU)、主存贮器、I/O 通道及输入输出设备等五部分组成。

1·1·3 计算机的系统结构

一个计算机实际上是由两大部分构成,即硬件和软件。所谓硬件就是指上述组成计算机的五大部件,它是计算机的实体,一个计算机系统仅有硬件还是不能工作的,我们通常称它为“裸机”,要构成一个完整的计算机系统,还必须配备有相应的系统软件、应用软件和程序设计语言。所谓系统软件是为计算机系统配置的所有软件的总和,如操作系统、编译程序、装配程序、诊断程序以及其它服务性程序等等。系统软件是为使用计算机的所有用户提供的公用“软件包”,通常由计算机生产厂家负责设计,应用软件反映的是与系统软件不同的另一类软件,它是以系统软件为工具,充分利用系

统软件为用户实现业务处理的一类软件。应用软件通常由用户或软件服务公司来编制,有些常用软件计算机厂家亦研制提供,如一些信息处理软件、通用电子表软件等。程序设计语言则是用来提供编制系统软件或应用软件的设计语言。

不同的计算机系统其硬件和软件数量是各不相同的,究竟应包括多少,应根据计算机的规模、应用场合及对计算机性能的综合要求来确定,一般应由下列各部分组成:



1.1.4 计算机的数据表示方法

我们知道,计算机中的数据和指令都是用二进制数表示的,各种数制(如十进制、十二进制等)都是按人们的习惯自然形成的,而二进制则是根据计算机内部器件的特性决定的。二进制数与其它数制之间的转换以及计算机中数值的表示方法(符号位表示法、补码表示法、定点数、浮点数等)在很多资料中均有介绍,这里只强调说明一下计算机通用的代码 ASCII 码。由于机器内部所有数据均采用二进制表示。但通过输入设备(键盘等)输入的和通过输出设备(显示器、打印机等)输出的内容却是各种各样的,包括字母、数字、功能符号、汉字等,当某个符号输入主机时必须先转换为一个二进制数,处理后输出时系统又将其还原为我们所熟悉的符号,这个符号所对应的二进制数就称为该符号的代码。ASCII 码便是目前通用的内部代码系统,它规定每个符号由 7 位二进制数表示,并可定义 128 种符号,称基本 ASCII 码。通常高位置 1,用一个字节(8 位)表示的 ASCII 码(值大于 128)称为扩展 ASCII 码,用于表示各国文字或特殊符号。用双字节表示的汉字内码便是其一。

ASCII 码表请参见附录一。

这里我们特别说明一下表示存储器容量的单位及换算公式

1 个二进制位=1 位(bit)

8 位二进制位=1 字节(byte)

1024 字节=1K 字节($=2^{10}$ 字节)

1024K 字节=1M 字节(或 1 兆字节)($=2^{20}$ 字节)

1.1.5 微机硬件系统基本配置

现在市场上各种微型计算机型号越来越多,作为用户无论选用的机型是什么档次,它们都是由一些基本配置所组成,大体可分为以下几个部分,即:主机、键盘、显示器、软盘驱动器、硬盘驱动器、打印机等等。为了保证正常使用微机,我们认为用户至少应了解掌握以下知识:

1. 主机

各种微机主机板上都由以下几部分组成:

(1)中央处理器(CPU)

它是微机的控制中心,本身由控制器、运算器、寄存器等部件组成,用以完成向计算机发送的各种指令。近几年来 CPU 型号不断更新,各项技术指标越来越高,特别体现在处理速度上,目前市场上微机的 CPU 型号有:

8088 主频 4.77MHz、8MHz、12MHz

80286 主频 8MHz、12MHz、16MHz

80386 主频 16MHz、20MHz、25MHz、33MHz

(2)内存储器(内存)

内存储器是由大规模集成电路存储器芯片组成,用于存储微机运行中的各种数据(如存放运行的程序、原始数据、运算结果等),它有着容量大、存取速度快等特点,通常内存储器分为 ROM(只读存储器)和 RAM(读写存储器)两大类,其中:

ROM——用于固化一些系统程序(始终不改变的程序),各种微机 ROM 中所固化的程序不尽相同,如 BASIC 解释程序、磁带机操作系统、磁盘引导程序、开机自检程序等等。不同微机 ROM 大小通常在 40k—128k 之间。

RAM——开机前内容为空,RAM 中没有任何数据信息,开机后由操作系统对其进行分配管理。不同机型配置的 RAM 存储器大小不等,有 256k、512k、640k 或 1M 等等,通常一部分 RAM 设计在主机板上,也可通过 I/O 扩展槽对 RAM 实现扩充。目前,高性能微机有些 RAM 已扩展到 16M 以上。

(3)I/O 扩展槽及外设接口

主机性能再好,如果不与外部设备连接通信它也只能是个摆设,毫无实用价值,它必须与外部设备相连接才能发挥作用。靠什么呢?它要靠主机板上所提供的 I/O 扩展槽和各种外设接口(8088CPU 的 I/O 槽口是 16 位的,80286 以上 CPU 的 I/O 槽口通常为 32 位),通常微机主机板上均留有 8 个扩展槽,以便用户根据需要进行扩充,但购买微机时,有些扩展槽口已被必备的外设所占用,如:为连接显示器、打印机、软盘驱动器以及硬盘所占用的槽口。主机与这些外部设备之间均需要通过 I/O 扩展槽来连接(相应的控制卡要插入某 I/O 槽口中,当然也有个别微机将某些控制卡与主机板做成一体)。尽管如此,用户微机上至少还有 3—4 个 I/O 扩展槽未使用,这些扩展槽对用户今后的扩充或使用某些系统软件和专用软件非常有用,象扩展内存、连接专用设备、网卡,尤其是当今很多优秀软件为防止解密都带一块加密卡(如各种排版软件、操作系统等),购买后必须将卡插入 I/O 扩展槽,所以 I/O 槽口越发显得重要了。

(4)主机板上还有把上述硬件组成部分连接在一起的总线结构和一些控制电路,用户知道便可,如要进一步了解,可参阅相关技术手册。

(5)关于微机电源

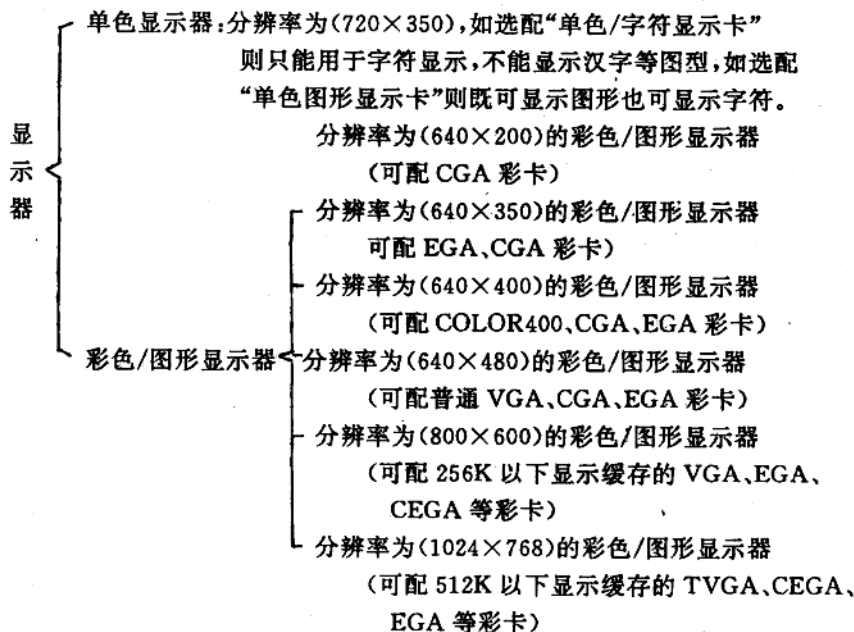
通常用户只须了解它的输出功率是多少,配接的外部设备越多,电源功率输出理应越大,但目前有些老机器的电源输出功率仅为 130w 以下,对今后扩充不便,如需扩充时电源也需更换,现在销售的微机电源功率通常在 200w 以上,比较合适。

2. 键盘

目前市场上微机所配键盘大致可分为基本键盘(83 键)、通用扩展键盘(101/102 键),专用键盘几类,各种微机支持哪种键盘也不是统一的,要视具体情况而论,用户了解即可。目前新型微机(除便携式微机外)大多采用 101/102 键盘。键盘是通过键盘连线插入主板上的键盘接口与主机相连接的。

3. 显示器

显示器是微机必不可少的外部设备之一,用于显示输出各种数据,它的内部原理与电视机基本相同,常用显示器的类型如下:



显示器与主机的连接是通过将显示器接口卡插入主机板上的某个扩展槽内,并用显示器连线将显示器与接口板连接起来便可。从上图可以看到目前市场上显示器的种类已有很多种,每种显示器可配接的显示器接口卡也不是唯一的,所以用户选择微机的显示器时,应连同显示接口卡一起考虑。如:用户在选择微机显示器时,首先要了解它的物理分辨率是多少?(即分辨率是多少×多少的),同时应明确所配显示接口卡是什么类型,例如要选择 800×600 显示器,所配显示卡可以是 EGA、VGA、CEGA 显示卡,当然可能存在大材小用之嫌,同时应注意,并不是什么机型都能任意选配显示器,如普通 PC/XT 机就不能用 VGA、COLOR400、CEGA 显示卡,而只能用 CGA、EGA、单

显等显示卡,286 以上微机基本上各种显示器都能配接。

如何识别显示器分辨率可参考显示器手册、用专用软件测试或询问经销商。

4. 软盘驱动器

和录音机上使用磁带一样,软盘只有插入软盘驱动器中才能工作,软盘驱动器工作原理这里不作介绍,我们只要知道它是微机存取软盘中数据的必需设备即可。软盘驱动器与主机的连接是通过将软盘驱动卡插入主机板中的某个扩展槽中,并用驱动卡专用连线将软盘驱动器与驱动卡连接在一起。

目前微机所配通用软盘驱动器大致有下列几种:

- (1)360KB—5.25"薄型(或全高)普通驱动器,适用于 360KB 软盘。
- (2)1.2MB—5.25"薄型高密驱动器,适用于 1.2MB 软盘。
- (3)1.44MB—3.25"薄型软盘驱动器,适用于 1.44MB 软盘。

关于软盘驱动器,有一点要注意,那就是驱动器的类型不要弄混了,3.25"与 5.25"软盘驱动器不会混,而 5.25"—1.2M 软盘驱动器与 5.25"—360K 软盘驱动器表面上看一样,在分不清的情况下存取数据可能造成数据丢失。1.2MB 格式化的软盘只能在 1.2MB 高密驱动器上进行读写,插入 360KB 普通驱动器中无效;360KB 软盘当然要在 360KB 普通驱动器上读写,也可在 1.2MB 高密软盘驱动器中进行读写操作。

不同微机所配软盘驱动器类型与个数不尽相同,通常如配置双软盘驱动器(1.2M、360K 各一)的情况下,一般将 A 驱动器设置为 1.2M,B 驱动器设为 360K,当然也可由用户自己设定,另外也可通过系统设置将 1.2M 高密软盘驱动器设置为 360K 当普通驱动器使用。

软盘驱动器也和录音机一样由使用过程中不密封,磁头易染脏物,也由于磁头长时间反复进行读写磁粉粘于磁头,可能会造成写入磁盘中的磁信号减弱,甚至会出现磁盘读写时出错等不应有的错误,为此用户应定期用高质量的清洗盘来清洗磁头,以确保正常使用。

5. 硬盘驱动器

软盘虽具有使用携带方便等特点,但其存储容量小,读写速度慢对大量数据的存储就显得力不从心,而硬盘便具有解决以上问题的全部特点。它有着软盘所不可比拟的优势,所以成为微机的主要配置之一,怎样选配硬盘、维护硬盘,怎样充分发挥其优势等一系列问题亦越发显得重要了。在这里,我们对硬盘的简单知识及使用硬盘的基本常识作一介绍。

(1)硬盘是外存储器的一种,它是由硬盘驱动器和硬盘驱动器接口卡组成,整个盘体为防尘而密封,稳定耐用,其与主机的连接是通过将硬盘驱动器接口卡插入主机扩展槽内,并用硬盘驱动器专用连线与硬盘驱动器接口卡相连接而成。

(2)硬盘通常从体积上分为 3.25"盘与 5.25"盘两种。

(3)硬盘的存储容量有 10M、20M、30M、40M、80M、……可达几百兆字节。

(4)硬盘读写速度快(要比软盘读写快得多),通常用磁头寻道时间来表示,硬盘本身相比也有快慢之分,寻道时间小于 28 毫秒的常称作高速硬盘。

(5)前期生产的硬盘不具备关机自动锁定磁头功能,必须关机前靠运行专用程序来锁定磁头(一般 DOS 系统盘上都配有此程序),此点用户必须注意,否则未锁定磁头就搬运机器可能会造成严重后果,轻则使硬盘出现坏块,重则损坏整个磁盘。而近期生产的硬盘一般都带有关机自动锁定磁头功能。

(6)目前微机上采用的各种硬盘大多为国外各个厂家所生产。各硬盘生产厂家所生产的硬盘分各种型号,每种型号都具有自己特定的物理指标(注:硬盘的几项基本物理指标是盘体内有多少磁

头、有多少柱面、每柱面上划分多少扇区等等),所以我们如果打开机箱盖去观察硬盘,通常会看到硬盘表面上标有硬盘的生产厂家、产品型号和各项物理指标,硬盘出厂后若要正常使用必须事先要对硬盘进行三项基本操作,即第一步对硬盘进行低级格式化,第二步对硬盘进行分区,第三步对硬盘进行高级格式化。这一点与软盘不同,软盘只需进行高级格式化便可使用(请注意,软盘进行高级格式化与硬盘第三步高级格式化虽然均使用FORMAT命令但内部实质有所不同)。通常在购置微机时所配硬盘已由经销商事先完成了此项工作(包括:低级格式化、硬盘分区、普通格式化FORMAT)。但因工作需要要对硬盘重新分区或因某些原因硬盘工作不正常确需重新低级格式化硬盘时,很多用户由于缺少这方面的知识便不知如何操作,有些用户甚至拉着微机四处求援,所以下面介绍一下正常处理硬盘的方法:

微机的硬盘若想正常使用必须顺利进行三项操作:

①硬盘的低级格式化

低级格式化是真正的格式化,它重新对硬盘划分磁道和扇区,并将磁盘每个扇区的内容全部清空,盘中所有数据将丢失。并不是硬盘一出问题就必须要对其进行低级格式化,要逐步检查,视具体情况而定

通常采用的工具软件有:

LOWFORM 程序:用于对 PC/XT 档次微机的硬盘进行低级格式化。

AT 机随机检测盘:用于对 286、386 档次微机的硬盘进行低级格式化。

DM 程序:用于进行综合格式化。(包括:低级格式化、硬盘分区、高级格式化 FORMAT)

专用 DM 程序:厂家为自己生产的各类硬盘所设计的专用格式化程序,如 ADM 程序、CDM 程序等,其功能基本与 DM 程序一致。

②硬盘的分区

对硬盘进行分区,实际上是为不同操作系统所占硬盘的空间进行划分。我们知道微机上可以使用不同的操作系统,如 DOS 操作系统、XENIX 操作系统等,我们既可以把整个硬盘全部划分给 DOS 操作系统使用,亦可划分一部分硬盘空间给 DOS 操作系统,而其余部分留给其它操作系统使用。

DOS 操作系统对硬盘的分区是通过 FDISK 程序完成的。

③硬盘的高级格式化

通过使用 FORMAT 命令来完成硬盘的高级格式化,实际上此步工作是检查硬盘的格式,如果发现磁盘某处有坏块,就自动将坏块标记在文件分配表之中,今后系统便不再使用这些坏块,同时将目录区和文件分配表中其它数据清空。

有关硬盘的这三步操作细节可参考有关使用手册。

6. 打印机

打印机是微机常采用的基本输出设备之一,它与主机的连接是通过并行打印接口卡插入主机板上的某扩展槽内,并用打印机专用连线将打印机与并行打印机接口卡相连接而成。打印机的种类很多,有针式打印机、激光打印机、喷墨打印机等,但由于性能价格比等原因,用户大多采用的是针式打印机,特别是 24 针打印机倍受用户喜爱。24 针打印机型号有很多种,如:M2024、M1724、TH3070、AR2463、AR3240、LQ 系列等,型号不下几十种,选择打印机型号应从以下几方面考虑:即速度、质量、噪音以及驱动软件等。有些针式打印机还配有各种点阵汉字字库,不用汉字打印驱动程序也能高速打印汉字,如 AR3240、LQ1600K 等。当然,根据工作需要还可选配更高级的打印机,如激光打印机、喷墨打印机等,它们亦可做为微型计算机的外部设备使用。

7. 异步通信接口卡

上面介绍了微机的各种常用外部设备,同时也说明了外部设备与主机之间是通过相应的控制接口卡连接而成的,异步通信接口卡也是一样,它是为微机与远程电子设备进行通信时所必需的接口卡。它插在主机板的扩展槽内,通过其后边的插座插入电缆插头,电缆的另一头可接到调制解调器(MODEM)或其它接串口的设备,该卡的作用是提供一个标准的 RS-232C 接口,通信时它将总线内部的并行数据转换成串行数据传送,再通过调制解调器将数字信号转换为音频模拟信号,从而可并入市话或长话网发送到远方,以实现微机的远程通信。

以上所述内容,只是微机的一些基本硬件常识,使用微机的同志应有所了解,以保证正常情况下使用微机。

§ 1.2 汉字信息处理的意义和任务

1.2.1 什么叫汉字信息处理:

对于“信息”一词,目前尚有多种定义,其中的一种定义是:信息(information)是各种事物所发出的信息、情报、指令、数据和信号中所包含的表征该事物的内容。随着人们对客观世界认识的日益深化,确认信息和物质、能量这三者是构成客观事物的三大要素。信息对于人类社会的重要性,表现在除了可再生资源(如动、植物)和非再生资源,(如矿物)以外,信息是维持人类生产活动、经济活动和社会活动的第二资源。信息具有多种性能,例如可传播性、可转换性、可存储性、可处理性以及可再生性等。随着科学技术的发展,信息的传输效能日益增强,其作用范围也日益宽广,例如,由于电子通信技术(特别是光纤通信和卫星通信技术)的发展,信息传输技术以电话、电报发展到传真、电视,包括声音、文字、图象和图形的传送,从而大大增强了通信效能,但是这种技术的发展只限于实现信息的传输,约在本世纪六十年代,电子计算机这项重大的科学技术成果在非数值计算领域内得到推广应用。由于电子计算机不仅能存储和控制信息,更重要的是由于它能加工和处理数据,因此相应产生了信息处理这一新概念。

随着计算机系统功能的不断提高,应用领域的迅速发展,信息处理的概念,语义、作用和涉及的范围也大大扩展了,特别是文字信息处理所包含的内容更加丰富了。例如,情报资料、图书的自动编目和检索,书刊、报纸的自动编辑和排版,事务处理、企业管理、办公室自动化,文字处理、文字翻译,医疗诊断,咨询服务,数据通信等。实际上,文字信息处理技术已渗透到人类思维、生产和生活等活动的一切方面。文字信息处理技术已逐渐同科研、生产实践、社会活动、生活环境等有了日益密切地联系,以至于人类社会的一切活动几乎都有它的用武之地。以电子计算机为基本手段的现代信息处理技术,正在促使人类的社会经济、科学技术和家庭生活发生日新月异的变革,这项技术的发展速度和应用水平已成为人类进入信息化社会、国家走向现代化的一个重要标志。

计算机最早是在西方国家发展起来的,目前这些国家已相当成熟地应用文字信息处理技术。究其原因,除了这些国家对计算机系统的设备、技术开发得较早以外,还有另一个重要原因,就是这些国家采用拼音文字,拼音文字的字母数量少,字形简单,从而容易在计算机上实现文字信息处理。我国所用的文字主要是汉字,汉字是一种表意文字,字量多,字形复杂,这两个特点使得汉字输入方法和建立汉字字模库的工作遇到不少困难。为了构成一种汉字信息处理系统,在硬件方面,除了需要

一些通用的设备(如通用电子计算机及其外围设备)外,还需专门配备汉字输入输出设备。在软件方面,要使系统软件具有对汉字处理和西文处理两者兼容的能力,而这个课题难度较大,使得汉字信息处理一度落后于西方文字处理好多年。但目前我国在汉字信息处理技术领域内已经取得了可喜的成绩。使得汉字信息处理技术在各个领域,尤其是办公自动化、情报检索、书刊报纸的排版等方面已经得到了广泛地应用。

对于计算机而言,它只认识两个字符即“0”和“1”,这是由它自身的硬件结构所决定的,因为构成计算机最基本的元件一般只有两个稳定的状态,如电压的高和低、电流的有和无、开关的开与合、二极管的截止与导通等,用它的两个稳定状态来分别表示“0”和“1”。所以一般的信息都是这些0、1组合的代码。在对西文的字母及符号的处理中,对应于26个字母(分为大写和小写)和一些常用符号,按照某种规律和约定编成一组数码,这组数码称为字符代码(如我国的国家标准GB1988,或国际标准ISO646七位代码,EBCDICII位代码等)。实际上,对文字信息的加工,就是对代码数据的加工。和西文一样,汉字亦有它自己的代码,1980年,我国颁布了《信息交换用汉字编码字符集——基本集》(即GB—2312)。GB2312是和GB1988兼容的,它用两个七位码代表一个汉字,可用作计算机系统之间汉字信息交换用的标准代码,和西文文字信息处理一样,汉字文字信息处理管理可分为三个基本阶段:

(1)信息的输入:通过输入设备把文字信息转换成代码,并送入计算机。

(2)信息的加工和处理:根据各类不同的应用,借助预先设计好的程序对输入的信息进行加工或处理,从而得到结果信息。

(3)信息的输出:通过输出设备把以数据代码表示的结果信息复原成文字。

1.2.2 汉字信息处理技术的发展过程

日本是世界上研究汉字信息处理技术较早的国家,因为日文中包括许多汉字(常用汉字就接近2000个),要解决用计算机处理日文,同样先要解决汉字处理技术。日本国内汉字系统的应用已经相当普遍,为促进日文信息处理技术的发展,日本成立了日文信息处理学会(Information Processing Society of Japan),并定期召集国内和国际会议,交流和促进这项技术的发展。

在美、加、澳等国家,亦有专门的研究汉字信息处理技术的机构,这主要是由一些外籍华人组成,特别是在美国,其研究工作具有一定的规模,如美国的中文计算机学会(Chinese Language Computer Society USA)组织,每隔一、两年定期举行学术会议,交流汉字信息处理的理论和方法。

我国台湾省和香港地区的汉字信息处理技术研究工作,始于六十年代后期,近年来汉字信息处理系统的研制和实际应用发展也很快,在汉字输入方法上比较有名的象仓额编码法,目前已发展到第四代CCG IV,它特别注意以软件的角度改进汉字的字形质量,提高输出速度。汉字系统的实际应用方面,诸如办公用计算机,文字处理机,数据库管理系统和汉字通信网络系统等,都已有很多成果。

我们大陆在六十年代末就已开始对汉字信息处理进行探索和实践。邮电部门在1968年研制成功的汉字电报译码机,是我国汉字信息输出设备的最早形式。七十年代中期,我国开始系统地研究和开发这项技术,并把“汉字信息处理系统”这个研制课题列为以电子计算机技术为中心的重点系统工程项目,在实施的最初几年内,由于受当时器材、设备条件等各方面条件的限制,进展比较缓慢。1978年以来,由于我国开始广泛应用大规模集成电路存储器和成套的微处理机芯片,因而在很大程度上促进了汉字信息处理技术的发展,而且研制了一些新型的汉字输入输出设备,在技术指标、可靠性、实用性和经济价值方面都有了很大的提高,已能用国内研制的汉字设备和计算机配制

成多种应用系统,特别是以微处理机为基础的汉字信息处理系统或用微处理机控制的汉字终端技术发展的更为迅速。

§ 1·3 汉字信息处理系统的构成和分类

1·3·1 汉字信息处理系统的构成

汉字信息处理系统和通常的计算机系统的组成情况相似,也包括硬件和软件两大部分。

1. 硬件组成

汉字信息处理系统的硬件包括主处理机,常规外部设备和汉字外部设备。主处理机是通用电子计算机。对主处理机的选择可根据所要求的处理能力(包括容量、速度和信息吞吐量)和工作方式(包括脱机成批处理和联机工作方式等),它可以是大、中、小型计算机,也可以选用微型计算机。常规外部设备主要包括外存储器,如各种类型和规格和磁盘机和磁带机,汉字外部设备包括汉字输入键盘、汉字印刷机、汉字显示终端等设备。汉字信息处理系统中,汉字显示终端和汉字字模库是重要的组成部分。

2. 汉字信息处理系统的软件:

和通常的计算机系统相似,汉字信息处理系统的软件包括系统软件和应用软件两类。

(1) 系统软件

汉字信息处理系统的软件主要包括以下几个部分:

a. 能兼容汉字和西文信息处理的操作系统。它在保留通常西文系统全部功能的条件下,还包括各种汉字设备的驱动模块,并且,它还能直接调用汉字输入输出管理程序和汉字编辑程序。此外,系统所支持的各种高级程序设计语言也能处理和识别汉字信息。

b. 汉字输入输出处理程序。它包括汉字输入输出接口程序以及汉字输入、换码、访问汉字字模库和汉字输出等程序。对于输入计算机的代码信息,当系统识别出它是汉字信息时,便根据输入码的不同编码方式,将其转换成标准汉字代码,供加工处理。当汉字输出时,先把标准码转换成汉字字模库中对应的地址码,然后读出字模信息,以供显示或打印。

c. 汉字文本编辑软件。它和西文编辑功能相似,汉字文本编辑程序除了具有对单个或多个汉字及整句或段的增、删、改的功能外,还具有换行和换段功能。对于功能强的编辑程序,还具有行首、行末、禁则处理和行对齐,以及自动成页等功能。

d. 高级程序设计语言。汉字信息处理的应用程序可以用高级程序设计语言来编写。它们在操作系统支持下,可以调用相应的编译程序。这些高级程序设计语言在保持原有功能的情况下,还能识别出应用程序中的汉字字符串和注释,直接处理汉字信息。

(2) 应用软件

汉字信息处理系统根据它的性质和用途的不同,各种应用项目都要有相应的应用程序。对于一些典型的应用程序,应提供商品化的应用程序包。由于应用项目的种类繁多,设计各种应用程序是一项工作量很大的任务。因此,要尽量利用西文系统已有的一些应用程序包,使其在经过必要的修改和改动之后,能够适合在汉字工作方式下应用,这样便可迅速扩大汉字信息处理的应用范围。

(3) 汉字文件系统

对于许多应用项目,建立文件系统是十分重要的。以汉字字符串写成的文件,种类很多,而且格式亦各异,如汉字情报检索系统或汉字数据库系统中必须建立的各种文档;汉字编辑排版系统中设计的作为排版格式依据的编辑文件;翻译系统中用作文法和句法规则的文件等。

1.3.2 汉字信息处理系统的分类

从对系统功能和输出文字质量的要求上来区分,可以把汉字信息处理系统分成两种类型:精密汉字编辑排版系统和通用型汉字系统。前一种类型用于正式出版的书刊、报纸的编辑排版;后一种类型用于汉字文件处理、统计报表、数值和数据处理等,这一类型的系统使用范围是很广的,汉字信息处理技术的推广应用很大程度上取决于这类汉字系统的发展。

(1)精密汉字编辑排版系统

这一系统最重要的技术关键是高精密汉字字模的存储和版面输出,该系统的特点对汉字字模的点阵密度要求很高,如果要求输出的分辨率达到30线/毫米以上,那么,对于一个五号字(尺寸为3.675mm),就要求其点阵密度达96×96点。对于精密汉字字模,不仅每个字的点阵信息量大,而且由于字量多,需要多种字体和字号,从而使总的字模信息处理很庞大。

精密汉字编辑排版系统所包括的主要设备有:高精密的汉字字模库,输出用的照排设备。另外还包括排版用计算机,相应的外部设备,编辑、排版用的联机汉字显示终端,校样印刷机,字模自动制作设备,图片输入设备等等。

汉字编辑排版系统要配备有大量专用的软件,它们包括:专用的操作系统、编辑排版专用语言及其编译系统,汉字文件系统,书刊、报纸等各种版式的排版应用程序,图片处理软件等。

(2)通用型汉字信息处理系统:

通用型汉字信息处理系统主要特点是用来实现数据处理或一般的汉字信息处理,其使用面广,力求系统的成本降低,不需太讲究汉字字模的质量。通用汉字系统字模点阵规格目前流行的主要有两种:15×16点、24×24点。对低于15×16点阵的字模,因质量太差,故目前已很少采用。由于目前通用型汉字系统的汉字印刷机主要是各种型号的24针打印机,所以24×24点阵的汉字字模更加流行,对于汉字打印机,今后也可以推广采用简易型激光扫描印刷机。

通用型汉字信息处理系统比精密汉字编辑排版系统的使用面要广的多,几乎只要用到汉字的地方,都有它的存在。下面举几种通用型汉字信息处理系统的例子:

a. 汉字情报检索系统(Chinese information retrieval system):用于书刊、情报资料的存储和检索的自动管理系统,进一步已经发展为汉字数据库检索系统。

b. 企业管理系统(Factory management system):它用于大型工矿企业的生产管理、计划调度、行政管理、人事工资管理、设计图纸、工艺资料管理、产品管理、合同管理以及供销计划管理等。工矿企业应用了计算机管理后,不仅可以减轻和节省人力,而且能大大提高管理效能,便于实现现代化的企业管理体制。

c. 事务处理系统(business processing system):它主要用于政府机关的计划拟定、公文档案管理、统计报表制作;银行金融机构的金融情报管理、财务管理、日常各类业务处理、报表统计;学校教育机构的数学行政管理、学生学籍档案管理;大型医院的医疗事务和病历管理;大型旅店业务管理;公安机关的户籍管理和案犯档案管理等,总之,它主要面向各种不同的事务或业务管理。

d. 办公用计算机(Office computer):它不仅用来实现办公室范围内的文件和书信处理,还可用作简易的文件档案管理。此外,还可以结合办公室业务的特点,专门设计一些应用程序进行其它业务方面的处理。对于不同性质的办公室对文字处理的不同要求,可以把一些常用的公文、书信格式、

文摘等,预先收存在系统中,以便起草文件,书信时随时调用。对于编辑好的文件、书信等可在系统中存档,以便需要时随时打印,这种系统又称为文字处理机(Word processor)、办公用计算机系统大多数采用微型机,这类设备也可以和通讯线路相连接,用作汉字电传的通信终端设备。

另外如汉字通信系统(Chinese character communication system),窗口系统(cashiersystem),文字自动翻译系统(Word automatic translation system)等,都是针对不同的应用需要而设计的专用系统。上述各类应用系统有些是为了提高准确性和可靠性;有些是为了提高工作效率,节省人力;有些是为了完成无法用人工方法实现的难度很高或极为繁琐的工作。总之,汉字信息处理系统的应用,能够大大促进在我国建立有效的管理体制和手段。

§ 1.4 汉字信息处理技术的发展前景

到目前为止,我国已经研制和设计出许多类型的汉字信息处理系统,并配有丰富的各类应用软件,而且国外的一些较好的信息处理软件都已汉化,可以用来处理汉字信息,如各种汉字编辑软件,电子表软件等,这些系统和应用软件的使用所带来的良好的效果和极高的工作效率,会使我国计算机应用展现出令人鼓舞的前景。

政府机关利用计算机进行行政管理,编选统计报表,通过通信网络汇总全国工农业生产数字和收支平衡情况,从而取代大量繁琐的人工劳动;工厂企业和物资部门利用计算机进行物资管理,建立现代化生产管理体制;科技情报管理系统配合通信网络把全国各省、市的主要图书馆科技情报研究单位联系起来,提高图书情报资料的利用率,使它们更好地为科研、生产和教学服务;利用汉字编辑排版系统,在全国有计划地集中建立计算机排版中心,各出版部门可把记录原稿信息的软盘等媒体送到排版中心制版,从而可大大提高排版效率,实现书刊出版工作的全面技术革新。各大报社、新闻通信社利用新闻通信网络和汉字编辑照排系统把新闻采集、通信、发稿、报纸编排等组成一体化作业,加快新闻宣传报导的时效;公用事业窗口系统可以改进服务质量,节省人力。现代化教育系统可以提高教育质量、改进教育方法,提高教育效率。这些系统的推广应用,将对我国的四化建设产生深远的影响。

汉字信息处理技术的发展主要依赖于计算机软硬件技术地发展,由于计算机的结构特点正向着高效率,小体积,低价格的方向发展,超大规模集成电路的出现,更促进了这一发展的速度,计算机发展到今天,已经进入了各个行业,随之汉字信息处理技术也将会有更大的发展,我国早在1980年就成立了中国中文信息研究会,并且定期组织开展这一技术领域的学术活动,而且已初见成效。目前和海外学术界的联系,学术交流已经相当广泛,这必将促进汉字信息处理技术的不断完善,也将使它在全国以至海外的应用更加普及。

第一章习题

1. 计算机是由哪几部分构成的?
2. 计算机的系统结构是什么?
3. 计算机的数据用什么方法表示?
4. 各种微机主板上都由哪些部分组成?
5. 驱动器有几类?
6. 在使用软盘时先要格式化。而想正常使用硬盘需做哪些工作?

7. 汉字信息处理系统由哪几部分组成?

8. 汉字信息处理分几类?