

全国高等职业学校·高等专科学校教材

计算机应用基础

陶增乐 杨宗源 主编

本书附盘可从本馆主页 <http://lib.szu.edu.cn/>
上由“馆藏检索”该书详细信息后下载，
也可到视听部复制

高等教育出版社

图书在版编目(CIP)数据

计算机应用基础/陶增乐主编. —北京: 高等教育出版社, 2000. 5

ISBN 7-04-008065-6

I. 计… II. 陶… III. 电子计算机—基本知识
IV. TP3

中国版本图书馆 CIP 数据核字(2000)第 62033 号

责任编辑 孔全会 封面设计 乐嘉敏
版式设计 孙鸣雷 责任印制 蔡敏燕

书 名 计算机应用基础
主 编 陶增乐 杨宗源

出版发行 高等教育出版社

社 址 北京市东城区沙滩后街 55 号

电 话 010-64054588

021-62587650

网 址 <http://www.hep.edu.cn>

邮政编码 100009

传 真 010-64014048

021-62551530

经 销 新华书店上海发行所

排 版 南京理工排版校对公司

印 刷 常熟市印刷二厂

开 本 787×1092 1/16

印 张 18

字 数 435 000

版 次 2000 年 7 月第 1 版

印 次 2000 年 7 月第 1 次

定 价 26.00 元(附光盘)

凡购买高等教育出版社图书,如有缺页、倒页、脱页等质量问题,请在所购图书销售部门联系调换。

版权所有 侵权必究

前 言

毋庸置疑,信息技术的发展正在把人类社会推进到一个崭新的时代——“信息时代”或“知识经济时代”。在这个时代,信息技术已成为经济发展的主要动力之一。

高等职业教育是培养技术型、应用型人才的,无论是高新技术发展所产生的职业岗位,还是以金融、商业、交通、房地产等为主的第三产业职业岗位,都需要信息技术的支持。本教材就是根据这个需要而编写的。

本教材力图使学生了解现代信息技术,特别是多媒体技术和网络技术的发展和应用;在内容选择上以培养信息技术应用能力为主,尽可能体现可“做”性;同时又注意对学生学习方法和能力的培养。在教材编写时,力求以“任务”为驱动,即以某个实例为先导,进而提出问题引导学生思考,通过实践,培养学生分析问题和解决问题的能力。教材中所选用的软件平台和工具,既考虑了计算机技术的最新发展,又顾及了现实的可行性。

在教学上,建议教师结合学生所学专业的内容,在熟悉计算机基本操作的基础上,启发学生动手创造自己的作品(如图文并茂的演示文稿),因材施教,鼓励创新。

这是一本基础性教材,适用范围较广,有些专业若需要进一步的计算机技术知识,可以在学完本教材的基础上,选学程序设计语言、数据结构、操作系统、数据库引论等专门课程。为了方便教学活动的开展和学生的实践操作,本教材《计算机应用基础》附有提供各种素材的光盘,并配套有《计算机应用基础上机实验与练习》一书。

本教材主编为陶增乐,副主编为杨宗源。参加编写的有:宁鲁生(第一章),汪涛波、葛友昌(第二章、第七章),殷群(第三章、第六章),杨冰(第四章),马云(第五章)。

我们在教材编写时,力求结合高等职业教育的特点编出新意,适应教育实际的需要,但由于各方面的局限性,肯定会存在不足之处,希望广大读者指正,以便再版时改进。

编 者

2000年5月

第一章 计算机基础知识

当前,人类社会已经发展进入到信息时代,计算机作为信息处理、科学计算的主要工具,开始进入千家万户,许多社会集团和部门若脱离计算机就无法正常运转。随着人们接触和使用计算机的机会日益增加,计算机知识及技术已成为现代人学习现代文明的基本内容。本章主要介绍有关计算机的基础知识,包括计算机组成及工作原理;多媒体和多媒体技术;如何使用计算机进行数据处理;程序设计的基本思想;软件开发技术以及计算机安全方面的有关知识。



1.1.1 计算机发展历史

长期以来人类一直梦想着有一台能进行自动计算的工具,帮助人们完成复杂的计算工作。在经历了许许多多科学家的不懈努力和勤奋工作,积累了丰富的经验,并为计算科学技术进行了大量理论准备和技术储备之后,电子数字通用计算机终于成为本世纪最伟大的发明之一,人类以自己不朽的科学精神实现了自己的梦想。在众多为实现自动计算而努力的科学家中最著名的有:

十七世纪

法国人帕斯卡(B. Pascal)于1642年运用齿轮技术制造了一台能完成加法运算的机器,开创了人类自动计算的先河。为纪念他的这一功绩,有一种著名的结构化程序设计语言被命名为“Pascal”语言。

德国人莱布尼兹(G. W. Leibniz)于1672年建立了基于二进制的运算法则,并提出了不用连续相加进行机械乘法运算的思想。

十九世纪

英国人巴贝奇(C. Babbage)设计了历史上第一台由齿轮式寄存器、运算装置和控制顺序三部分组成的计算器,定名为分析机(A analytical Engine),为此后人誉他为“计算机之父”。值得一提的是,在支持巴贝奇的少数人中有一位女性——英国著名诗人拜伦的女儿爱达(Ada Augusta Lovelace),她帮助巴贝奇研究分析机,建议用二进制数取代原来的十进制数,同时为某些计算开发了一些指令及编制程序。由于爱达在程序设计上的突出贡献被誉为世界上第一位程序员。为了纪念她,有一种著名的基于对象的程序设计语言被命名为“Ada”语言。

英国人布尔(G. Bool)成功地将形式逻辑归结为一种代数演算(Boolean Algebra,布尔代数),为现代逻辑电路和计算机的发展提供了重要的理论准备。

二十世纪

二、三十年代,丘奇(A. Church)、波斯特(E. L. Post)和图灵(A. M. Turing)等人先后从不同角度对计算模型进行研究。其中,图灵在他的一篇著名论文《论数字计算在决断难题

中的应用》中,以一种简单数学机器(即理论计算机)出发的研究,与现代计算机最为接近。为纪念这位先驱,国际计算机领域的最高奖被命名为“图灵奖”。

1937年斯蒂比次(G. Stibitz)组装了二进制加法器,1939年研制出能计算复数的机电式计算机。

四、五十年代计算机的设计原理逐步形成并趋于完善。

1941年德国人朱斯(K. Zuse)采用继电器制造成功一种用二进制程序控制的通用电磁计算机。该计算机有64个字节的主存储器,可以进行加、减、乘、除和开平方运算。还可以进行十进制数和二进制数的转换。

1944年美国艾肯(H. Aiken)在IBM公司的支持下,采用继电器研制成功一台通用程序控制式计算机,并在哈佛大学投入运行。这是最后一台机械计算机。

1945年美国宾夕法尼亚大学莫尔学院的莫克利(J. W. Mauchly)和艾克(W. J. Eckert)等人研制成功人类历史上第一台采用电子线路技术的电子数字计算机ENIAC(Electronic Numerical Integrator and Calculator,电子数值积分器和计算器),但其体系结构与旧式机械式计算机没有本质差别。

1945年大数学家美籍匈牙利人冯·诺伊曼(Von Neumann)在美国阿伯丁弹道实验室任顾问期间,主持了根据图灵提出的计算模型为基础而建立计算机体系结构的工作。经过多年努力,终于在1952年成功地研制成世界上第一台现代存储程序式通用电子数字计算机EDVAC(Electronic Discrete Variable Automatic Computer,电子离散变量自动计算机)。该方案确定了计算机由控制器、运算器、存储器、输入和输出这5部分组成。以后人们把具有这种体系结构的计算机系统称为冯·诺伊曼型计算机(简称为冯氏计算机或冯氏体系结构)。

自第一台电子数字式计算机(以后简称为“计算机”)问世半个多世纪以来,计算机已经历了4代的发展历程,但其每一代的更新基本上是逻辑电路和器件的更新。它们分别是电子管(electronic tube, 1946~1958年)、晶体管(transistor, 1959~1964年)、集成电路(Integrated Circuit,缩写为IC, 1965~1970年)、大规模集成电路(Large Scale Integration,缩写为LSI;超大规模集成电路,Very LSI,缩写为VLSI;极大规模集成电路,Ultra LSI,缩写为ULSI;1971年~现在)。

80年代开始,由于超大规模集成电路技术的突飞猛进,以及各种应用背景的促进,开始了非冯氏体系结构计算机和基于人工智能计算机的研制工作。

1982年4月日本宣布开始实施基于人工智能的第5代计算机系统(Fifth Generation Computer System,缩写为FGCS)的研制,并计划在90年代推出新型知识信息处理系统(Knowledge Information Processing System,缩写为KIPS)。1991年日本宣布放弃第5代计算机计划,采用基于人工神经网络(Artificial Neural Nets,缩写为ANN)的所谓第6代计算机系统。专家认为新型计算机系统应统称为“新一代计算机系统”或“未来计算机系统”为好。

1969年受到本世纪激光技术的启示,开始了光子计算机(photon computer)的研究。这是一种用光子代替电子,光互连代替导线互连,光器件代替电子器件的全新全数字化计算机。1984年6月美国IBM公司研制成功一台能在接近绝对零度条件下工作的光子计算机。1990年1月30日美国又推出能在常温条件下工作的实验性光子计算机,其速度比电子计算机快上千倍。从理论上讲,光子计算机的运行速度可达每秒一万亿次,存储容量可达一万亿二进制位。

由于本世纪中叶兴起的分子生物学有了很大发展,1970年人们发现脱氧核糖核酸(DNA)存在两种不同状态,可以用来表示逻辑值的真假和二进制数的“0”与“1”。1994年11月,美国《科学》杂志披露了南加州大学伦德教授利用DNA溶液在试管中实现了运算过程。专家认为DNA计算机运算速度快得惊人,几天的运算量可相当于人类计算机问世以来的总运算量,一立方米DNA溶液可存储一万亿亿二进制位数据,超过目前人类所有计算机的存储量,但其消耗的能量却只有一台普通计算机的十亿分之一。

我国对计算机的研究始于20世纪50年代末(1958年研制了第一台计算机(103机)),起步较晚,发展也较缓慢,60年代中期到70年代末,基本上无大的发展,直到80年代初改革开放,才迎来了计算机发展的春天。近20年来我国的计算机、计算机通信发展迅猛,实现了计算机发展模式的重大改变,形成了以应用为目标、以市场为导向、以科技为依托、以拳头产品为龙头的格局,许多产品纷纷跻身于国际市场。我国第一台PC(Personal Computer,个人计算机)与IBM公司PC在推出的时间上相差了3年,而386微机推出的时间只相差了1年;到Pentium(奔腾)PC,推出的时间仅相差3个月,这说明我国的计算机产业有了长足发展。但应当看到,我国在微电子基础工业、软件开发、计算机应用等许多领域与发达国家相比还存在着很大的差距,需要作更大的努力。

1.1.2 信息与数据

信息(information)在人类生活中扮演着极其重要的角色。它同物质资源(可构成人类原始生活资料)和能量资源(可转化为动力资源)一样,是人类3大可利用的生活资源之一。信息资源能转化为知识和智慧,促进人类对物质资源和能量资源的开发利用。人类从开始用手势和肢体语言传递信息到今天,曾经历过5次大的信息革命,它们分别是语言的产生、文字的产生、印刷术的发明、广播电视的发明和计算机及计算机网络的发明应用。每次信息革命都对人类的发展产生了空前深远的积极影响,特别是第5次的信息革命,它使人类步入了信息社会。

信息是一个抽象名词,关于这个词可有各种不同的解释和理解。在日常生活中信息常指“消息”,例如人们常说“有了某同学的消息”,也可以说“有了某同学的信息”。信息不是实体,它反映的是实体的特征、性质、变化,以及实体之间相互关系等方面的内容。

数据(data)是描述信息的物理符号,是信息的载体。例如数字、文字和各种符号都可以载荷信息。广义的数据还可以是图形、图像、声音、影像等。

本书所指的信息是可以用数据具体化的信息。因此,在一些不严格的场合,信息与数据这两个名词可以通用。例如,“信息处理”有时也说成“数据处理”。

1.1.3 信息处理

1. 什么是信息处理

信息处理(information processing)是指对信息(以各种形式存在的数据)进行收集、组织、存储、抽取、加工和传播等一系列活动的总和。

信息处理有手工、半自动(机械)和自动等不同方式。手工方式是用笔和纸人工描述信息处理的运算过程,原始信息的处理过程和处理结果都保存在纸上。半自动机械处理方式始于1880年,当时美国为了进行人口普查采用穿孔机制成的穿孔卡片文件来保存信息,用

读卡机阅读穿孔卡片文件上的信息。在发明了计算机之后,由于计算机具有运算快捷、准确和存储量大的特点,使之当然成为现代自动处理信息的首选工具。信息自动处理大致有 3 个环节,即输入信息、存储和加工信息、输出信息处理结果。

2. 计算机信息处理

计算机只能识别(接受和处理)二进制数,因此用计算机进行信息处理时,需要对现实世界(也称客观世界)的信息进行分析、归纳、提炼并转换成二进制数形式,方可被计算机进行处理。为了在这种数据转换过程中,能以统一抽象的方法研究和理解信息之间的相互关系,并组织信息结构,在现实世界的信息和机器世界的的数据之间需要引进信息世界的概念。

下面通过一个对现实世界中“学生情况”这一信息进行处理的例子,来了解信息处理的大致过程。

(1) 在手工处理的情况下,通常用二维表格的方式描述“学生情况”。如图 1-1-1 所示。

学 号	姓 名	性 别	出 生 年 月	家 庭 住 址
9701001	李国庆	男	1979.10	上海市中山北路 83 号
9701002	张小南	女	1979.12	江苏南京市新街口 106 号
9701003	黄雨生	男	1979.8	安徽芜湖市小南门 9 号
...	...	...	...	...

图 1-1-1 “学生情况”表

现实世界中称表格的第 1 行为表头,其中的每一项指出了二维表格中各列的含义,如第 1 列的内容是学生的学号,第 2 列的内容是学生的姓名等。

(2) 当使用计算机处理“学生情况”信息时,首先要区分“学生情况”中的各种数据。例如“学号”与“姓名”,前者是数值数据,后者是汉字字符数据,两者显然有不同的数据类型。为此引进属性(attribute)的概念。属性是实体(事物)某一方面的特征。例如学生可具有“学号”、“姓名”、“性别”等属性。属性可以是不同的类型,这样,计算机可以按不同数据属性类型采用不同的处理方法。

相关属性的集合表征了一个实体的类型,称为实体型(entity type)。图 1-1-2 表征了“学生情况”这个实体型,它共有 5 个属性。

学号	姓名	性别	出生年月	家庭住址
----	----	----	------	------

图 1-1-2 “学生情况”实体型

实体型中每一属性都会有若干属性值,相关属性值的集合表征了一个实体。图 1-1-3 表征了“李国庆同学的情况”这一实体。

9701001	李国庆	男	1979.10	上海市中山北路 83 号
---------	-----	---	---------	--------------

图 1-1-3 “学生情况”实体型的一个具体实体

类型相同的实体构成的集合称为实体集(entity set)。通过这样分析处理之后就完成了信息从现实世界到信息世界的转换。如图 1-1-4 所示。

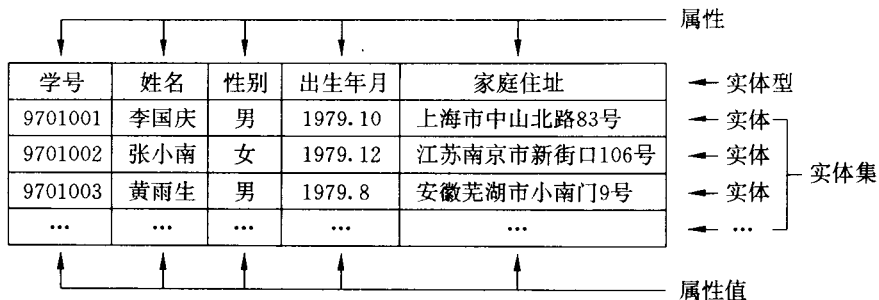


图 1-1-4 信息从现实世界到信息世界的转换

(3) 在信息世界里把现实世界中的对象抽象成信息世界的实体型、实体集、实体、属性、属性值后,可以方便地用实体联系这个概念,描述相关实体型之间的关系。例如学生实体型和课程实体型之间存在“学生选课”这样的多对多($m:n$)联系,它表明了一个学生可以学习多门课程,一门课程可以为多个学生选修这样的语义。图 1-1-5 描述了这种联系。

对于多对多联系可以用一个实体型进行描述。例如图 1-1-6 引进“学生选课”实体型构成两个一对多($1:n$)联系。

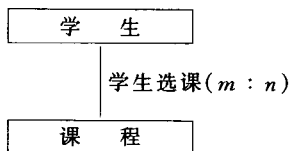


图 1-1-5 学生选课

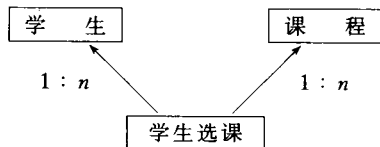


图 1-1-6 学生选课关系图

学生选课关系图表示了某个学生可以选修多门不同的课程,每门课程可由多个学生选修这样的语义。

其中,学生选课实体型用学号、课程和成绩 3 个属性表征,如图 1-1-7 所示。

学 号	课 程	成 绩
-----	-----	-----

图 1-1-7 学生选课实体型

在信息处理中常称实体型是一个关系,关系之间的联系构成了关系图(如图 1-1-6 所示)。从关系图可以看出,通过学生关系能查到某个学生所选修的所有课程及其成绩;通过课程关系能查到某课程共有多少学生选修及他们该门课的成绩。

在信息世界中分析和解决问题的好处,是可以完全不必顾忌计算机性质、性能以及数据的组织等方面的内容,而集中精力于信息逻辑关系的解析。

信息世界	机器世界
属性	数据项或字段
实体型	记录型
实体	记录
实体集	文件

图 1-1-8 从信息世界到机器世界

(4) 显然信息世界的内容(对象)仍然不能为计算机接受,还必须经过编辑、加工和编码等过程,将其转换成机器世界的对象才能被计算机处理。机器世界的对象是数据。此时对应于属性的数据是数据项或字段,对应于实体型的数据是记录类型,对应于实体的数据是记录,对应于实体集的数据是文件。如图 1-1-8 所示。

一般,信息处理过程可大致归结如图 1-1-9 所示。

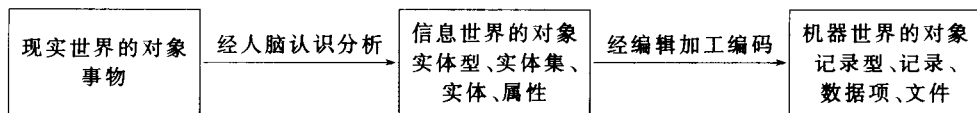


图 1-1-9 信息处理过程

值得注意的是,由于信息形式多种多样,信息处理的方法也有许多。以上方法仅是其中的一种模式,但其基本思想却带有普遍性,即信息处理时先要对现实世界的对象进行分析抽象,然后在抽象世界(例如上例中的信息世界)中讨论和分析问题,以便能理清信息之间的逻辑关系,最后再把抽象世界的对象转换成机器世界的对象,最终完成信息处理。

1.1.4 数字化和数字化技术

由于计算机只能识别二进制数,所以信息在进入机器世界时需要将其转换成二进制数据。这就涉及到数据的编码问题。同样现实世界的连续模拟信息(模拟量),在进入机器世界时也需要经过编码转换成数字化信息。这里所说的编码技术是指关于二进制数及其转换方法的技术,实际上是一种数字化技术。

1. 编码基础

(1) 数制

数制(也称计数制)是指多位数码中每一位的构成方法以及从低位到高位进位的规则。十进制(decimal, D)是人们日常最熟悉、使用最广泛的一种数制。1 234.56 用十进制的表示方法可以写成:

$$1\,234.56 = 1 \times 10^3 + 2 \times 10^2 + 3 \times 10^1 + 4 \times 10^0 + 5 \times 10^{-1} + 6 \times 10^{-2}$$

一般地,任何一个十进制数 D 都可以表示为:

$$\begin{aligned} D &= d_n d_{n-1} \cdots d_0 d_{-1} \cdots d_{-m} \\ &= d_n \times 10^n + d_{n-1} \times 10^{n-1} + \cdots + d_0 \times 10^0 + d_{-1} \times 10^{-1} + \cdots + d_{-m} \times 10^{-m} \\ &= \sum_{i=n}^{-m} d_i \times 10^i \end{aligned}$$

其中,10 为基数, 10^i 是位权, $d_i \in \{0, 1, 2, \cdots, 9\}$ 是数码, m 、 n 是正整数。

(2) 二进制

二进制(binary, B)同十进制一样也是一种计数制,只是基数为 2。一般地,任何一个二进制数 B 都可以表示成:

$$\begin{aligned} B &= b_n b_{n-1} \cdots b_0 b_{-1} \cdots b_{-m} \\ &= b_n \times 2^n + b_{n-1} \times 2^{n-1} + \cdots + b_0 \times 2^0 + b_{-1} \times 2^{-1} + \cdots + b_{-m} \times 2^{-m} \\ &= \sum_{i=n}^{-m} b_i \times 2^i \end{aligned}$$

其中,2 为基数, 2^i 是位权, $b_i \in \{0, 1\}$ 是数码, m 、 n 是正整数。

(3) 二—十进制数的转换及二进制数的加减法

① 十进制数转换成二进制数

采用求余法。例如,将十进制数 56 转换成二进制数。

2	56	0 ← (低位 b_0)
2	28	0
2	14	0
2	7	1
2	3	1
	1...	1 ← (高位 b_5)

$(56)_{10} = (111000)_2$

② 二进制数转换成十进制数

有两种方法。一种方法只需将二进制数按一般表示公式展开,再按十进制方法求和即可。例如,将二进制数 $(111000)_2$ 转换成十进制数。

$$\begin{aligned}
 (111000)_2 &= \sum_{i=5}^0 b_i \times 2^i \\
 &= 1 \times 2^5 + 1 \times 2^4 + 1 \times 2^3 + 0 \times 2^2 + 0 \times 2^1 + 0 \times 2^0 \\
 &= 1 \times 32 + 1 \times 16 + 1 \times 8 + 0 \times 4 + 0 \times 2 + 0 \times 1 \\
 &= 32 + 16 + 8 + 0 + 0 + 0 \\
 &= (56)_{10}
 \end{aligned}$$

另一种方法可以用十进制数转换成二进制数方法的逆过程,即用二进制数的高位乘以2得到的结果加上次高位,然后再用所得到的结果乘以2……。

例如,将二进制数 $(111000)_2$ 转换成十进制数。

$$\begin{aligned}
 &(((b_5 \times 2 + b_4) \times 2 + b_3) \times 2 + b_2) \times 2 + b_1) \times 2 + b_0 \\
 &= (((1 \times 2 + 1) \times 2 + 1) \times 2 + 0) \times 2 + 0 \\
 &= (56)_{10}
 \end{aligned}$$

③ 二进制数的加、减法

与十进制数的加减法一样,只不过二进制数的加、减法应采用逢二进一和逢零借一的规则。

例如:

$$(1101011)_2 + (101001)_2 = (10010100)_2$$

$$\begin{array}{r}
 1101011 \\
 +) \quad 101001 \\
 \hline
 10010100
 \end{array}$$

$$(11001011)_2 - (1101001)_2 = (1100010)_2$$

$$\begin{array}{r}
 11001011 \\
 -) \quad 1101001 \\
 \hline
 1100010
 \end{array}$$

(4) 二进制、八进制和十六进制数之间的关系与表示法

采用二进制表示数据时,书写和阅读都很不方便,为此常用八进制(octal, O)和十六进

制(hexadecimal, H)表示数据。此时1位八进制数可以表示3位二进制数(因为 $2^3 = 8$),同样1位十六进制数也可以表示4位二进制数(因为 $2^4 = 16$)。表1-1-1是这几种数制之间的对照关系。

表 1-1-1 二、八、十、十六进制数对照表

十进制(D)	二进制(B)	八进制(O)	十六进制(H)
0	0000	0	0
1	0001	1	1
2	0010	2	2
3	0011	3	3
4	0100	4	4
5	0101	5	5
6	0110	6	6
7	0111	7	7
8	1000	10	8
9	1001	11	9
10	1010	12	A
11	1011	13	B
12	1100	14	C
13	1101	15	D
14	1110	16	E
15	1111	17	F

书写时常用数字后面加上相应进制数英文单词的第一个字母大写来表示。

例如:

111000B 表示是个二进制数

507O 表示是个八进制数

567D 表示是个十进制数

3DFH 表示是个十六进制数

八进制数和十六进制数的加、减法也与十进制数和二进制数一样,只是这时应采用逢八进一或逢十六进一、逢零借一。

例如:

$$327O + 277O = 626O$$

$$\begin{array}{r} 327O \\ +) 277O \\ \hline 626O \end{array}$$

$$72E6H + 2A39H = 9D1FH$$

$$\begin{array}{r} 72E6H \\ +) 2A39H \\ \hline 9D1FH \end{array}$$

$$327O - 277O = 30O$$

$$\begin{array}{r} 327O \\ -) 277O \\ \hline 30O \end{array}$$

$$F00EH - 4C7BH = A393H$$

$$\begin{array}{r} F00EH \\ -) 4C7BH \\ \hline A393H \end{array}$$

2. 西文字符编码

(1) ASCII 码(American Standard Code for Information Interchange, 美国标准信息交换码)

在利用计算机进行信息处理时,人们通常使用字母、数字和符号(统称为字符)与计算机打交道。为了使计算机能“读懂”这些字符,需要将它们转换成二进制数据来表示;同样,对于计算机中的二进制数据,为了能让人们读懂也需要将它们还原成人们习惯使用的字符,目前国际上广泛使用的 ASCII 码就是为完成这一目标而设计的编码。

ASCII 码是一种用 7 位二进制代码编制的字符编码,共可表示 128 个字符,如表 1-1-2 所示。查表时先在表中找到要查的字符,通过这个字符就可确定它所对应的行(低 4 位: $b_3b_2b_1b_0$)和列(高 3 位: $b_6b_5b_4$),合起来即是该字符的 ASCII 码 $b_6b_5b_4b_3b_2b_1b_0$,例如字符“a”的 ASCII 码是 1100001。用一个字节(8 位二进制数表示一个字节)表示 ASCII 码时该字节的最高位为“0”,此时“a”的 ASCII 码为 01100001。

表 1-1-2 ASCII 表

	000	001	010	011	100	101	110	111
0000	NUL	DLE	SP	0	@	P		P
0001	SOH	DC1	!	1	A	Q	a	q
0010	STX	DC2	"	2	B	R	b	r
0011	ETX	DC3	#	3	C	S	c	s
0100	EOT	DC4	\$	4	D	T	d	t
0101	ENQ	NAK	%	5	E	U	e	u
0110	ACK	SYN	&	6	F	V	f	v
0111	BEL	ETB		7	G	W	g	w
1000	BS	CAN	(	8	H	X	h	x
1001	HT	EM	)	9	I	Y	i	y
1010	LF	SUB	*	:	J	Z	j	z
1011	VT	ESC	+	;	K	[	k	}
1100	FF	FS	,	<	L	\	l	
1101	CR	GS	-	=	M	]	m	}
1110	SO	RS	.	>	N	^	n	~
1111	SI	US	/	?	O	_	o	DEL

(2) 英文输入输出码

英文是一种线性拼音文字,由 26 个字母线性组合而成,用到的字符较少,只需使用键盘即可。输入码就是键盘上各键键面所印的字符。计算机用 ASCII 码作为机内码,当在键盘上按下一个键后,程序自动将其转换成相应的 ASCII 码传送给计算机。

英文字符的输出使用英文字形码,图 1-1-10 显示了用 5×7 的点阵表示的字母“E”的字形码。

●	●	●	●	●	1FH
●					10H
●					10H
●	●	●	●	●	1FH
●					10H
●					10H
●	●	●	●	●	1FH

图 1-1-10 英文字符“E”的字形码

3. 中文字符编码

(1) 中文标准信息交换用汉字编码

同英文信息交换码一样,在用计算机处理中文信息时,也需要用信息交换码。为此曾对有一百万个汉字的中文资料中每个汉字出现的频率(字频)进行统计,然后按频率由高到低的顺序排列。结果发现排列在前面的 5 177 个汉字的字频累计数为 99.99%,同时还发现排在最前面的 3 838 个汉字的字频累计数为 99.9%,其后的 1 339 个汉字的字频累计数仅为 0.09%;而排在 5 177 个汉字之后的那部分汉字(大约 5 万多个)其字频累计数仅

为 0.01%。统计分析结果显示,在总计约 6 万左右的汉字中大部分几乎不用或不常用。在此基础上,1981 年在原国家标准局的主持下,制定了 GB 2312—80《信息交换用汉字编码字符集 基本集》的国家标准。该字符集共设立了 94(01~94)区,每区设有 94(01~94)位,汉字字符就安排在这 94 区(行)×94 位(列)的阵列中。用其所对应的区位号编码就得到了该字符的区位编码,称为区位码。例如汉字“而”在 22 区 88 位的位置上,其区位码就是 22 88(如果在具有区位码输入方案的系统上键入“2288”,屏幕上会立刻显示“而”字)。

区位码可以用两个字节表示,其中第一个字节的低 7 位用来表示区号,第二个字节的低 7 位表示位号,这样就可以用两个字节表示一个字符。字符集共收录了 7 445 个汉字、字母、数字、符号和图形等字符(余下 1 391 个空位),其中 6 763 个汉字分为一级汉字 3 755 个(即常用汉字,按汉字拼音序排列)和二级汉字 3 008 个(即次常用汉字,按部首序排列)。字符集分布如下:

- | | |
|----------------|----------------------|
| 1 区 常用符号 94 个 | 8 区 汉语拼音字母、注音符 63 个 |
| 2 区 序号、数字 72 个 | 9 区 制表符号 76 个 |
| 3 区 图形字符 94 个 | 10~15 区 空白(备用) |
| 4 区 日文平假名 83 个 | 16~55 区 一级汉字 3 755 个 |
| 5 区 日文片假名 86 个 | 56~87 区 二级汉字 3 008 个 |
| 6 区 希腊字母 48 个 | 88~94 区 空白(备用) |
| 7 区 俄文字母 66 个 | |

(2) 国标码

为使二字节区位码能避开 ASCII 码的前 32 个控制符,将区位码的区号和位号分别加上 32 就得到了国标码。例如汉字“而”的区位码是 2288,分别加上 32 就得到对应的国标码 54 120(其二进制表示是 00110110 01111000)。

(3) 汉字机内码

用国标码作为汉字机内码(也称为内码)使用时会与 ASCII 码发生冲突。如汉字“而”的国标码为 00110110 01111000,但它同时也可以表示数字“6”(第 1 个字节)和英文字母“x”(第 2 个字节)的 ASCII 码,在计算机内部产生二义性。为了使计算机在处理中文与英文时能够相互兼容,一种最简单的方法就是在国标码两个字节的最高位都添加“1”(即每个字节都加上十进制的 128),以便与用单字节 7 位表示的 ASCII 码区分。这种变形的国标码称为机内码。例如汉字“而”的机内码为 10110110 11111000。机内码主要用于中文信息的存储、交换、传输等操作。

值得注意的是,在使用这种机内码时,如果遇到扩展的 ASCII 码(扩展的 ASCII 码能表示 256 个字符),两者又会发生冲突。为了能从根本上解决计算机内中西文兼容的问题,同时也为了遵循国际上为统一表示世界各国、各地区的文字而设计的“通用多 8 位编码字符集”国际标准(ISO/IEC10646),简称 UCS)编码,几年前以中国大陆为主,联合了日本、韩国和我国台湾地区、香港特别行政区等使用汉字的国家和地区,共同确定了一个由 20 902 个字符组成的字符集,经国际标准化组织正式批准为 UCS 中的汉字字符集。近年来我国开始使用的新国标 GB 13000,是我国标准化管理机构发布的与 UCS 一致的国家标准。UCS 方案用 4 个字节表示一个字符,其中 0 组 0 平面被规定为“基本多文种平面”,用来存放世界上的主要文字和符号。新国标 GB 13000 规定的字符集就在这个区域中。由于其他字符极少用到,所以“基本多文种平面”区域可以用 2 个字节表示一个字符,并以此作为机内码,实现了机内码与新国标的统一。我国标准化管理机构提出的“扩大汉字内码规范 GBK”就是以这种方式在机内表示汉字的规范。该规范能很好地解决计算机内中西文兼容问题,目前 Windows 95 和 Windows 98 中文版就使用了 GBK 方案。

(4) 汉字输入码

汉字是二维文字(即平面方块字),每个汉字由一笔或一笔以上个笔画组成。汉字的字符集十分庞大,有些字典收录的字符多达 5~6 万个。如何用英文键盘来输入为数众多的汉字,一直是人们研究的课题。汉字输入码要求易记、易学、代表每一个汉字的键数少和重码率低,这是一个难度很大的课题。所幸的是目前我国研制的汉字输入方案大约有几百种,更新更好的方法仍层出不穷。常见的汉字输入方案有标准(拼音)、双拼、全拼、智能 ABC、表形和五笔形输入码等。

(5) 汉字输出码

和英文输出码一样,汉字也可以采用点阵方式来构成字形码。如图 1-1-11 显示了用 16 × 16 的点阵表示的汉字“英”的字形码。

					●							●				04H, 10H
					●							●				04H, 10H
	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	7FH, FFH
					●							●				04H, 10H
					●							●				04H, 90H
									●							00H, 80H
				●	●	●	●	●	●	●	●	●	●	●		1FH, FCH
				●						●					●	10H, 84H
				●						●					●	10H, 84H
				●						●					●	10H, 84H
	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	7FH, FFH
							●			●						01H, 40H
						●					●					02H, 20H
					●							●				40H, 10H
				●									●			08H, 08H
	●	●	●										●	●	●	70H, 70H

图 1-1-11 汉字字形码

1.2 计算机系统

计算机系统由计算机硬件系统和计算机软件系统两部分组成。其中硬件是执行操作的器件和设备,好比人类的大脑、五官和躯体;软件(这里主要指程序)是控制操作的命令集合,好比人类的思想,是计算机的灵魂。

1.2.1 计算机硬件系统

计算机硬件系统由主机和输入/输出(Input/Output, I/O)设备(简称为 I/O 设备、外部设备或外设)组成。常用的微型计算机组成如下(图 1-2-1):

主机:它一般包括主机板/总线、接口板、电源、存储设备(其中软盘、硬盘和光盘驱动器统称为外存储器或外存),这些都安装在主机箱内。

输入设备:是外界向计算机传送信息的装置,包括键盘、鼠标、光笔、扫描仪等。

输出设备:是一种媒介装置,能接收计算机传出的信息并将其以适当的形式表示出来。输出设备一般包括显示器、绘图仪、打印机(针式、喷墨和激光式)等。

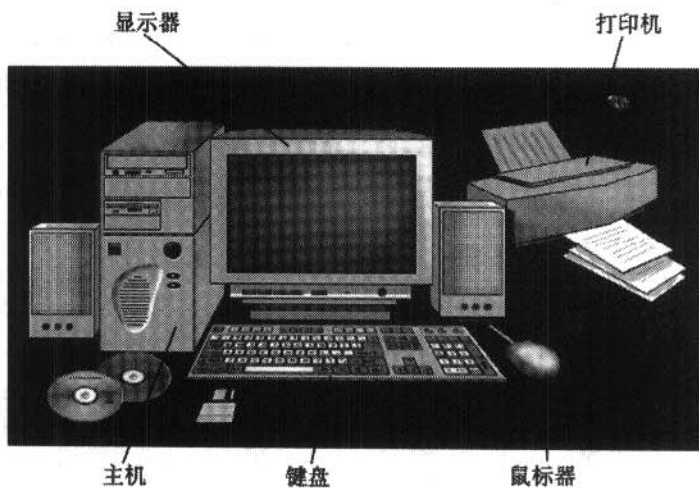


图 1-2-1 常用微型计算机的组成

1. 存储程序式计算机的体系结构

由冯·诺伊曼确定的存储程序式计算机主要由控制器(controller)、算术与逻辑运算器(Arithmetic and Logic Unit, 缩写为 ALU, 简称为运算器)、存储器(memory)、输入接口(input interface)和输出接口(output interface)组成(外部设备通过接口接入计算机)。

(1) 中央处理单元

通常将控制器与运算器合在一起称为中央处理单元(Central Processing Unit, 缩写为 CPU), 简称为 CPU。CPU 在微型计算机中称为微处理器(microprocessor)。常用微处理器的型号作为微机档次的标志, 例如奔腾 II (Pentium II)、奔腾 III (Pentium III)。

CPU 是计算机的核心部件, 负责控制和指挥其他各功能部件协调工作, 进行各种运算。

其中运算器完成对数据进行的算术与逻辑运算,它由累加器(accumulator)、加法器(adder)和寄存器阵列(registers)组成。控制器的作用是控制计算机各部件自动协调工作,主要部件有指令译码器、指令寄存器、指令指针和操作控制器等。

(2) 存储器

为了区别于磁带、磁盘和光盘等外部存储器,计算机内的存储器也称为内存。按读写方式不同可分为只读存储器(Read Only Memory,缩写为 ROM)和随机存储器(Random Access Memory,缩写为 RAM)两类。内存的作用是存放要执行的程序和数据。

① ROM

这类存储器只能读不能写,其中存放的信息是用专门编程设备根据需要预先写入的。ROM 在断电时不会丢失信息,因此常用来存放一些开机初始化信息。例如,基本输入输出驱动程序 BIOS 就存放在 ROM 中。

② RAM

这类存储器可读、可写、可随机使用。但一旦断电,RAM 中存放的信息会自动消失。

随着 CPU 速度的提高,CPU 与内存之间会出现速度不匹配的矛盾。用整体提高内存速度的办法来解决会使成本大大增加,显然不经济。于是引入了 cache(快速缓冲存储器),即把内存分为二级或三级。第一级为高速缓存,它可以直接做在 CPU 芯片内部,称为片内 cache。第二级为次高速缓存,它以芯片形式直接固定在主机板上,称为片外 cache。第三级由传统的内存构成,称为主存储器(简称主存)。主存被做成条状芯片(称为内存条),可插在主机板的内存槽口上。当 CPU 访问内存时依次访问片内 cache、片外 cache、最后才访问主存。这种多级访问的策略是基于这样的原则:当访问一个数据时,下次访问的数据最有可能是存储在它的附近的数据,于是 CPU 在访问外存时总是一次将一批数据调入内存,并分别存于片内、片外 cache 和主存中。数据交换也基于同样策略。片内 cache 的速度与 CPU 相当,片外 cache 的速度次之,主存的速度最慢,但要比外存快得多。cache 的引入提高了计算机的整体性能。

③ 微型计算机编址和存储容量的计算单位

微型计算机常用字节(byte, B)为单位编排存储单元地址,存储量也以字节为单位计量。常用单位和换算关系如下:

$$1 \text{ B} = 8 \text{ bit(位)} \quad 1 \text{ KB} = 1024 \text{ B} \quad 1 \text{ MB} = 1024 \text{ KB}$$

$$1 \text{ GB} = 1024 \text{ MB} \quad 1 \text{ TB} = 1024 \text{ GB}$$

(3) 输入输出接口

外部信息进入计算机和计算机内部信息的输出,需要由输入输出设备完成。相对于 CPU,输入输出设备由于其中有机械装置所以都是慢速设备。为了使输入输出设备的速度能与 CPU 相匹配,就需要设置中间缓冲机制,执行这种缓冲机制的部件称为接口。接口做成电路板可直接插在主机板的外部总线槽口上,CPU 通过接口与输入输出设备交换信息。例如打印机接口、显示器接口、磁盘接口等。CPU 与外设之间交换数据的策略是:在接口中设计了一个存储区(称为缓冲池),大小视输入输出设备而定。当计算机与外设交换数据时总是先送往缓冲池,然后向目的设备发出信号通知来取数据。

(4) 总线

为了能够实现 CPU、存储器、I/O 接口之间的连接,微型计算机系统一般采用总线结构。