

高等学校工科教材

80X86/Pentium 微型计算机 原理及应用

吴 宁 主编



电子工业出版社

PUBLISHING HOUSE OF ELECTRONICS INDUSTRY

URL: <http://www.phei.com.cn>

00023553

TP36
520



高等学校工科教材

80X86/Pentium 微型计算机原理及应用

吴 宁 主编



C0500588

電子工業出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

本书是依据国家教委计算机基础课程指导委员会制定的“高等学校工科非计算机专业《微机原理及应用》课程基本要求”编写的。

鉴于微型计算机的高速发展,本书以 Intel 80X86/Pentium 微处理器为背景,全面介绍了计算机基础、微处理器结构及系统、指令系统、汇编语言程序设计、存储器、输入/输出和中断技术、接口技术、微机系统与计算机网络等。

本书可作为高等学校工科非计算机类专业的教材,也可作为大专类相应专业教材或参考书,同时可作为从事计算机工作的工程技术人员自学或参考用书。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,翻版必究。

图书在版编目(CIP)数据

80X86/Pentium 微型计算机原理及应用/吴宁主编. —北京:电子工业出版社,2000.8

ISBN 7-5053-6099-X

I. 8... I. 吴... II. 微型计算机,80X86/Pentium-基本知识 N. TP36

中国版本图书馆 CIP 数据核字(2000)第 66825 号

35422/20

书 名:80X86/Pentium 微型计算机原理及应用

主 编:吴 宁

责任编辑:杨逢仪

排版制作:电子工业出版社计算机排版室

印 刷 者:北京大中印刷厂

装 订 者:三河市万和装订厂

出版发行:电子工业出版社 URL:<http://www.phei.com.cn>

北京市海淀区万寿路 173 信箱 邮编 100036

经 销:各地新华书店

开 本:787×1092 1/16 印张:26 字数:662.4 千字

版 次:2000 年 8 月第 1 版 2000 年 8 月第 1 次印刷

书 号:ISBN 7-5053-6099-X
TP·3245

印 数:6000 册 定 价:30.00 元

凡购买电子工业出版社的图书,如有缺页、倒页、脱页、所附磁盘或光盘有问题者,请向购买书店调换。若书店售缺,请与本社发行部联系调换。电话 68279077

前 言

计算机技术的飞速发展,带来了人类生活、学习和科学研究各个领域的技术革命。计算机知识和应用技能已成为人类知识经济的重要组成部分。为贯彻国家教委“面向 21 世纪教学内容改革”的精神,依据工科计算机课程教学指导委员会制定的“高等学校工科非计算机专业《微型计算机原理及应用》课程教学基本要求”,我们编写了这本《80X86/Pentium 微型计算机原理及其应用》,并选作江苏省高校教材。

从 1946 年第一台电子计算机问世到现在几十年的历史中,计算机的发展历经了电子管、晶体管、中小规模集成电路和大规模、超大规模集成电路几个发展阶段。20 世纪 70 年代初诞生的以微处理器为核心的微型计算机,也从 4 位、8 位、16 位发展到 32 位、64 位字长的、可与小型机媲美的高档微型计算机。以当今在计算机世界占有绝对优势的 80X86/Pentium 系列微型计算机为例,其 CPU 芯片经历了 8088、8086、80286、80386、80486 到 Pentium 系列(Pentium、Pentium Pro、Pentium MMX、Pentium I 及 Pentium II)的更新换代,运算速度已达 300MIPS (Million Instruction Per Second,每秒百万条指令),计算机的应用范围也从最初的科学计算发展到目前无所不及。

随着 CPU 芯片的制造工艺和性能提高,微型计算机硬件与软件产品正以几年一代的速度不断翻新。但是,计算机的基本工作原理并没有改变。对于使用者特别是应用程序的开发人员来说,所有这些 CPU 或微型计算机是一个系列,它们在升级换代过程中充分考虑了兼容性,因而在应用方面是完全兼容的。从 8088、8086 到 80386 以上直至 P II,寄存器结构仅仅是 16 位与 32 位的区分,没有本质的区别。芯片的指令系统中 80% 以上(基本指令集)是完全相同的,只是在 80286 以上的芯片增加了一些新指令(包括保护方式下的指令)以提高 CPU 的性能和功能。8086、8088 的工作方式与 80286 以上芯片的实地址工作方式几乎完全相同。所以说,8086、8088 CPU 是 Intel 80X86/Pentium 系列芯片的基础,是全面学习与掌握 80X86/Pentium 系列微机硬件技术与应用基础的很好的切入点。

根据工科非计算机专业学习与应用计算机的特点,本书从计算机硬件应用的需要出发,从 8086/8088 CPU 着手并结合 80286、80386、80486 及 Pentium 系列 CPU 的特点,系统讲述微型计算机的组成与工作原理、CPU 功能结构、寻址方式和指令系统、汇编语言程序设计、存储器及其与 CPU 连接、中断的工作原理及处理方法、微机系统典型接口芯片及 PC 机与外设的接口方法,并介绍 80X86/Pentium 系列微型计算机的系统结构、总线标准及基于总线的系统硬件扩展、计算机网络及网络互联的基础知识、Ethernet 网络的结构与编程等一系列技术,以求达到从理论和实践上对微型机的主要技术深入理解、牢固掌握、灵活应用。全书共 8 章,第 1~6 章为原理和基础部分;第 7 章为接口技术应用部分;最后一章介绍了计算机硬件系统及计算机网络方面的知识。

本书是工科非计算机专业“微机计算机原理及应用”的本科生教材,也可作为大专类相应专业的教材或参考书。书中内容是笔者在参阅了当前国内外有关微型计算机的大量资料后并总结了多年从事教学、科研工作经验综合而成。选材上注重基础、实用、先进和系统性,以反映计算机及其应用技术的最新发展;内容组织上遵循由易及难、循序渐进、宽编窄用的原则,以适

应不同专业、不同层次教学的需要;叙述上力求做到由浅入深、通俗易懂,并配有适当的习题与思考题,以利于教学和自学。全书编写力求做到原理、技术与应用并重,理论与实践结合,符合教学规律。本书教学参考学时数为60~80,使用时可根据具体情况适当增删部分内容。必须指出,“微型计算机原理及应用”是一门实践性很强的课程,应特别重视实践性环节和上机。

本书由吴宁主编并统稿,同时编写第3、4章、绪论及附录,第1、2章由易仲芳编写,第5、7章由邢岩编写,第6、8章由王敬东编写。本书在编写与出版过程中,得到了电子工业出版社和南京航空航天大学教务处的具体指导与帮助,在此表示衷心感谢。

为适应计算机技术的最新发展,本书较大幅度地更新了内容,重组了内容结构。由于笔者水平有限,书中难免有错误和不妥之处,敬请读者批评指正。

编 者

2000年2月

目 录

绪论	(1)
第 1 章 计算机基础	(4)
1.1 数据、信息、媒体和多媒体	(4)
1.2 计算机中数值数据信息的表示	(5)
1.2.1 机器数和真值	(5)
1.2.2 符号数的表示方法——原码、反码和补码	(7)
1.2.3 补码的运算	(10)
1.2.4 定点数与浮点数	(11)
1.2.5 BCD 码及其十进制调整	(15)
1.3 计算机中非数值数据信息表示	(17)
1.3.1 西文信息的表示	(17)
1.3.2 中文信息的表示	(19)
1.3.3 计算机中图、声、像的表示	(19)
1.4 微型计算机基本工作原理	(20)
1.4.1 微型计算机硬件组成	(21)
1.4.2 微型计算机软件系统	(27)
1.4.3 微型计算机中指令执行的基本过程	(28)
1.5 评估计算机主要技术指标	(30)
思考题与习题	(32)
第 2 章 80X86/Pentium 微处理器	(34)
2.1 8086/8088 CPU 内部结构	(34)
2.1.1 CPU 结构与特点	(34)
2.1.2 寄存器配置	(38)
2.2 8086/8088 引脚及功能	(41)
2.2.1 8088 CPU 引脚功能	(42)
2.2.2 8086 CPU 引脚功能	(46)
2.3 8086/8088 系统组织	(46)
2.3.1 8086/8088 支持芯片	(47)
2.3.2 单 CPU 模式(最小模式)系统	(50)
2.3.3 多 CPU 模式(最大模式)系统	(52)
2.4 8086/8088 CPU 时序	(55)
2.4.1 8086/8088 典型时序分析	(55)
2.4.2 8086 多 CPU 系统读写存储器简介	(59)
2.5 80X86/Pentium 微处理器	(60)
2.5.1 80X86/Pentium 特点与内部功能结构	(60)

2.5.2 80X86/Pentium CPU 寄存器组织	(68)
思考题与习题	(90)
第3章 80X86/Pentium 指令系统	(93)
3.1 80X86/Pentium 指令格式和寻址方式	(93)
3.1.1 指令格式	(93)
3.1.2 寻址方式	(94)
3.2 8086/8088 指令系统	(99)
3.2.1 数据传送类指令	(100)
3.2.2 算术运算类指令	(104)
3.2.3 逻辑运算与移位指令	(111)
3.2.4 串操作指令	(114)
3.2.5 控制转移类指令	(118)
3.2.6 处理器控制类指令	(124)
3.3 80X86/Pentium 指令系统	(126)
3.3.1 80286 增强与增加的指令	(126)
3.3.2 80386/80486 增强与增加的指令	(129)
3.3.3 Pentium 系列处理器增加的指令	(133)
思考题与习题	(135)
第4章 汇编语言程序设计	(140)
4.1 程序设计语言概述	(140)
4.2 汇编语言的程序结构与语句格式	(141)
4.2.1 汇编语言源程序的框架结构	(141)
4.2.2 汇编语言的语句	(143)
4.3 汇编语言的伪指令语句	(147)
4.3.1 基本伪指令语句	(148)
4.3.2 80X86/Pentium 扩展伪指令	(160)
4.4 汇编语言程序设计基础	(165)
4.4.1 程序设计的基本过程	(165)
4.4.2 程序的基本结构	(166)
4.4.3 程序设计基本方法	(167)
4.4.4 子程序设计与调用技术	(175)
4.5 模块化程序设计技术	(182)
4.5.1 概述	(182)
4.5.2 程序中模块间的关系	(183)
4.5.3 模块化程序设计举例	(185)
4.6 实用程序设计举例	(188)
思考题与习题	(195)
第5章 半导体存储器	(199)
5.1 概述	(199)
5.1.1 半导体存储器的性能指标	(199)

5.1.2 存储器的分类	(199)
5.2 随机存取存储器(RAM)	(200)
5.2.1 静态 RAM(SRAM)	(200)
5.2.2 动态 RAM(DRAM)	(204)
5.2.3 高速 RAM	(206)
5.2.4 高集成度 RAM(IRAM)	(208)
5.3 只读存储器(ROM)	(212)
5.3.1 掩膜 ROM 和 PROM	(212)
5.3.2 EPROM(可擦除的 PROM)	(213)
5.4 存储器连接与扩充	(216)
5.4.1 存储器芯片选择	(216)
5.4.2 存储器容量扩充	(218)
5.5 8086/8088 与存储器连接	(219)
思考题与习题	(221)
第 6 章 输入/输出和中断技术	(223)
6.1 输入输出及接口	(223)
6.1.1 I/O 信息的组成	(223)
6.1.2 I/O 接口的作用	(223)
6.1.3 I/O 端口的编址方法	(224)
6.1.4 简单的 I/O 接口	(225)
6.2 输入和输出的传送方式	(228)
6.2.1 程序控制的输入和输出	(228)
6.2.2 中断控制的输入和输出	(231)
6.2.3 直接数据通道传送	(232)
6.3 中断技术	(234)
6.3.1 中断及中断处理过程	(234)
6.3.2 中断优先权	(237)
6.4 80X86/Pentium 中断系统	(240)
6.4.1 中断结构	(240)
6.4.2 中断向量表	(242)
6.4.3 中断响应过程	(243)
6.4.4 IBM PC/AT 中断分配	(246)
6.4.5 高档微处理器的中断	(248)
6.5 8259A 可编程中断控制器	(249)
6.5.1 8259A 的内部结构与引脚	(249)
6.5.2 8259A 的工作过程	(250)
6.5.3 8259A 的工作方式	(251)
6.5.4 8259A 命令字	(253)
6.5.5 8259A 应用举例	(260)
6.6 中断程序设计	(263)
6.6.1 设计方法	(263)
6.6.2 中断程序设计举例	(266)
思考题与习题	(269)

第7章 微型机接口技术	(271)
7.1 概述	(271)
7.2 可编程定时/计数器	(272)
7.2.1 概述	(272)
7.2.2 可编程定时器/计数器 8253	(273)
7.2.3 可编程定时器/计数器 8254	(280)
7.3 可编程并行接口	(280)
7.3.1 可编程并行接口芯片 8255A	(281)
7.3.2 并行打印机接口	(289)
7.3.3 键盘和显示器接口	(292)
7.4 串行通信和串行接口	(296)
7.4.1 串行通信的基本概念	(296)
7.4.2 可编程异步通信接口 INS8250	(301)
7.4.3 可编程串行通信接口 8251A	(311)
7.4.4 串行打印机接口	(317)
7.5 模拟通道接口	(318)
7.5.1 概述	(318)
7.5.2 并行和串行 D/A 转换器	(319)
7.5.3 并行和串行 A/D 转换器	(325)
思考题与习题	(333)
第8章 微型计算机系统与计算机网络	(336)
8.1 80X86/Pentium 系列微机硬件系统	(336)
8.1.1 IBM PC/XT 微机硬件系统	(337)
8.1.2 IBM PC/AT 微机硬件系统	(341)
8.1.3 386、486 微机的硬件特点	(344)
8.1.4 Pentium 以上微机的硬件特点	(344)
8.2 微机系统的内存结构	(345)
8.2.1 内存分层	(345)
8.2.2 高速缓存 Cache	(345)
8.2.3 虚拟存储器	(347)
8.3 微机系统的总线标准	(348)
8.3.1 PC 总线	(348)
8.3.2 ISA 总线	(350)
8.3.3 EISA 总线	(352)
8.3.4 PCI 局部总线	(353)
8.3.5 USB 通用串行总线	(356)
8.4 总线基础上的系统硬件扩展	(356)
8.4.1 I/O 接口扩展	(356)
8.4.2 存储器扩展	(358)
8.5 计算机网络	(359)
8.5.1 网络基本知识	(359)

8.5.2 局域网通信协议	(362)
8.5.3 Ethernet 网卡结构及编程	(367)
8.5.4 网络互联与 Internet	(370)
思考题与习题	(372)
附录 1. 8086/8088 指令系统	(374)
附录 2. 系统中断	(390)

绪 论

电子数字计算机的出现,是近代重大科学成就之一。自从1946年第一台电子计算机问世至今几十年的历史中,它的发展经历了电子管、晶体管、中小规模集成电路和大规模、超大规模集成电路几个历史阶段。如今,在个人计算机(Personal Computer)上操作的人们,很难想象到第一台电子计算机ENIAC(Electronic Numerical Integrator And Computer)是怎样的一个庞然大物:它的体积有85立方米,占地面积150平方米,重30吨,有1800多只电子管,耗电140千瓦,价值40多万美元,而其性能上与现代一些通用微型机相比相差甚远,它只能存储750条指令,每秒钟只能进行360次乘法运算。

20世纪70年代初期,由于微电子技术和超大规模集成技术的发展,导致了以微处理器为核心的微型计算机的诞生。微型计算机(Microcomputer)与其他计算机的区别在于它的中央处理器CPU(Central Processing Unit)是采用超大规模技术集成在一块硅片上,又称其为微处理器(Microprocessor)。微处理器包含了冯·诺依曼计算机体系结构中的运算器(算术逻辑部件ALU+寄存器组RS)和控制器(Control Unit),是计算机的核心部件。随着超大规模集成技术的发展,微处理器中所集成的内容越来越多,除运算器和控制器外,还有协处理器、高速缓冲存储器、各种接口和控制部件等。今天,在PC机中广泛使用的CPU是Intel公司的Pentium系列微处理器。

微型计算机由微处理器、存储器、输入输出设备与接口和其他支持逻辑部件组成,完全包含了冯·诺依曼计算机体系结构中的五个部件,它们彼此通过系统总线(地址总线AB、数据总线DB和控制总线CB)连接起来。将微型计算机配置相应的系统软件、应用软件及外部设备等,则构成一个完整的微型计算机系统(Microcomputer System)。

微型计算机的发展是以微处理器的发展为表征的,以字长和功能来计,微处理器的发展经历了如下几个阶段。

1971~1973年为4位或8位低档微处理器和微型计算机时代。典型CPU产品是Intel 4004和Intel 8008。基本特点是采用PMOS工艺,集成度低(1200~2000晶体管/片),系统结构与指令都比较简单,且速度慢(基本指令执行时间为10~20 μ s)。主要应用于家用电器和简单控制场合。

1974~1977年为8位中档微处理器和微型计算机时代。典型CPU产品有Intel 8080、Zilog-Z80和Motorola的MC6800,它们的显著特点是采用了NMOS工艺,集成度提高约4倍(5000~9000晶体管/片),速度提高了10~15倍(基本指令执行时间为1~2 μ s)。系统结构与指令都比较完善,广泛应用于电子仪器、现金出纳机和打印机等。

1978~1984年为16位微处理器和微型计算机时代。典型CPU产品有8086、8088、Z8000和MC68000。这一代CPU的主要特点是采用HMOS工艺,其集成度(达20 000~70 000晶体管)和速度(基本指令执行时间为0.5 μ s)都比8位微处理器提高了一个数量级。体系结构与指令更为完善与丰富,采用了多级中断、多种寻址方式、段式寄存器等结构。

Intel 80X86/Pentium系列CPU以Intel公司1978年首先推出16位的8086,次年推出外部数据总线为8位的8088(这主要是便于和大部分8位外设相连接)为标志。8086/8088CPU

都有 20 条地址线,内存直接寻址范围为 $2^{20}=1\text{MB}$,主频为 4.77MHz 以上。它们首先在芯片内部引入了流水线技术(将 CPU 分成执行部件 EU 和总线接口部件 BIU 两部分),在存储器管理上引入内存分段管理的概念。此外,它们还可以和同期推出的供浮点运算的 8087 数学协处理器配套使用。8086/8088 功能强大的指令集在条数上已超过小型计算机。IBM 公司最早推出的 IBM PC 和 IBM PC/XT 机都采用 8088 主板。

1982 年 Intel 公司还推出了性能更高的 16 位 CPU80286(以 80287 作为它的协处理器),它有 24 条地址线,内存寻址范围为 16MB,主频为 6MHz 以上。它将 CPU 中的 BIU 分成地址单元 AU、指令单元 IU 和总线单元 BU 三部分,并利用 IU 进行预译码来进一步提高速度。在存储器管理方面引入保护虚地址方式,并可提供 $2^{30}=1\text{GB}$ 的虚拟内存空间,将部分外存信息有条件地与内存信息交换。从使用角度看,大大扩大了有限的内存容量。同时利用有效的特权保护可使由 286 CPU 构成的 IBM PC/AT(286 机)支持多用户。286 机具有实地址和保护虚地址两种工作方式。

1985 年~1992 年为 32 位微处理器和微型计算机时代。与 16 位微处理器相比,32 位微处理器从体系结构设计上有了概念性的改革与革新。典型 CPU 产品是 Intel 80386、80486 和 Motorola 公司的 MC68030、68040 等。这一代微处理器大多采用了 HMOS 或 CMOS 工艺,其集成度每片芯片高达 100 万只晶体管,基本指令执行时间一般在 25MIPS,为微型计算机带来了小型机的性能。它们具有 32 条地址线,内存寻址范围为 4GB。Intel 80386 工作主频在 16MHz 以上,以 80387 为协处理器。为了与 16 位外设兼容,1988 年 Intel 还推出了数据总线内 32 位外 16 位的 80386SX,仍用 80287 作协处理器,其他的结构则与 386 相同。Intel 80386 有实地址、保护虚地址和虚拟 8086(即可在机器上同时运行实地址、保护虚地址等不同方式的程序)三种工作方式。此外,为加快存储器操作还引入了高速缓冲存储器 Cache,这样可将具体数据运算从慢速的动态 RAM(DRAM)调整到 SRAM 中进行。

1989、1990 和 1992 年 Intel 公司相继推出 80486DX、80486SX 和 80486DX2 CPU,其工作主频提高到了 50MHz 以上。在 486DX 内部集成了 80386、80387、8KB 的指令/数据 Cache 和高速缓存控制逻辑。为提高处理速度,它采用了精简指令集计算(RISC)技术以减少指令执行时间,将芯片内的浮点运算完全和常规算术逻辑运算并列进行,综合性能要比 386 高 2~4 倍。486SX 和 486DX 的不同只是它内部不包含 387 数学协处理器。而 DX2 则是在芯片内外采用两种主频工作,内部主频为外部的两倍。

从 1993 年 Intel 公司推出外部数据总线 64 位的 Pentium(又称 P5,俗称 586)CPU 始,进入了 64 位高档微处理器和微型计算机时代。Pentium 采用了 0.6 微米的静态 CMOS 工艺,芯片内集成了 310 万只晶体管。其地址线为 32 位,寻址范围是 4GB,工作主频 60MHz 以上。Pentium CPU 芯片在 486 基础上采用了全新的体系,它重新设计了增强型的浮点运算器,速度比 486 提高 3~5 倍,将程序和数据 Cache 分开(各为 8KB),以减少等待及移动数据的次数和时间。它还增加了分支目标缓冲器,以预测分支指令结果提前安排指令执行顺序。最重要的是采用了超标量流水线结构,允许多条指令同时执行来提高效率。具体设置有两条指令流水线和独立的超标量执行单元,在同一时钟内可同时发两条整数指令或一条浮点(某些情况还能再送一条整数)指令,并将常用指令固化以硬件速度执行。

1995 年 Intel 推出的 Pentium Pro(又称 P6,俗称高能奔腾)也是一种 64 位 CPU,其中集成了 550 万只晶体管。地址线为 36 条,寻址范围为 64GB,其主频已提高到 133MHz 以上,具有两倍 P5 的性能。它的主要改进表现在两个方面:一是采用了动态执行技术。除了 P5 具有的转

移指令预测功能外,还通过提前对指令间数据流的相互关系进行分析,对指令流进行优化重排,保证了超标量执行单元能满负荷工作。二是将二级 Cache(以加快存储器的操作,PC Pentium 机中除了主芯片内含有 Cache 外,在主板上又安装了 256~512KB 的二级 Cache)也集成在同一块芯片上,从而在芯片内形成双重独立总线,有效地提高了性能。

随着多媒体技术的融入,在 1996 年~1997 年间 Intel 公司相继推出了基于 P5 和 P6 芯片并附加多媒体声像处理指令(共 57 条)的 CPU,称为“具有 MMX 技术的 Pentium 和 Pentium Pro”,其型号分别为 P55C 和 Pentium II(简称 P II),它们均采用 0.35 微米工艺。P55C 在结构上比 P5 又有所改进,例如,它将内部 Cache 从 8KB+8KB 增加到 32KB,指令预测功能也有所提高。而 P II 除了对 P6 性能作了改善外,还在外观上采用了新的封装技术,先将芯片固定到基板上再将它密封到金属盒中,它可直接插到主板插槽中去。由 P55C 和 P II 构成的 PC 机分别称为多能奔腾机和 P II(即奔腾二代)机,它们比较适用于多媒体应用领域。

1999 年 2 月,Intel 公司再次推出 64 位的 CPU Pentium III(简称 P III),主频 450MHz 以上,具有 32KB 一级 Cache,512KB 二级 Cache。它针对网络功能进行了优化,并且新增 70 条 SSE(Streaming SIMD Extensions,单指令多数据流扩展)指令,以提高 CPU 处理连续数据流的效率、浮点运算速度并加强多媒体功能。

微处理器的出现,经过多年来市场的激烈竞争,目前市场占有率最高的当属 Intel 80X86/Pentium 系列。相信随着微电子技术的发展,功能更强的 CPU 还会不断问世,并被不断用于提高 PC 系列微机的性能。

计算机技术的发展对社会进步产生了巨大的影响。今天,计算机及其应用技术的发展速度、深度及其广度,都远远超过了历史上任何一种技术手段和装备,在国防、科学研究、政治经济、教育文化等方面无所不及。计算机应用技术不仅引起社会各领域的巨大变革,反过来又推动计算机技术本身不断向前发展。

第 1 章 计算机基础

电子数字计算机是一个自动化信息处理机,其中心任务是处理“信息”。为此,本节先介绍信息处理领域中有关数据、信息和媒体的概念。

1.1 数据、信息、媒体和多媒体

1. 数据

日常生活中人们所说的数据,多半是指可比较其大小的一些数值,但信息处理领域中,数据的概念要比它广泛得多。

国际标准化组织(ISO)对数据的定义是:“数据是对事实、概念或指令的一种特殊表达形式,这种特殊的表达形式可以用人工的方法或者用自动化的装置进行通信、翻译转换或者进行加工处理。”在这个定义中,首先强调的是数据表达了一定的内容,即“事实、概念或指令”,同时指出,数据是一种“特殊的表达形式”,它不仅由人工加工处理,而且这种“特殊表达形式”更适合于计算机系统高效率地加工处理、通信传递以及翻译转换。

根据这个定义,通常意义下的数字、文字、图画、声音、活动图像等都可以认为是数据,因为人们可以对它们进行各种人工方式的处理,如通信、加工、转换等。例如在办公室中涉及的大量公文、报表,都可以认为是办公过程中办公人员需加工处理的数据。

对于计算机而言,数字、文字、图画、声音、活动图像等都不能直接由计算机进行处理,它们必须采用“特殊的表达形式”才能由计算机进行通信、转移或加工处理。这种特殊的表达形式就是二进制编码形式,即采用二进制编码表示的数字、文字、图画、声音和活动图像才能由计算机进行处理(包括通信、转移及存储)。所以,计算机系统所指的数据,一般均是以二进制编码形式出现的。

通常,在计算机内部又把数据区分为数值型数据和非数值型数据。数值型数据是指日常生活中经常接触到的数字类数据,它主要用来表示数量的多少,可以比较大小;而把定义中其他的数据统称为非数值型数据。在非数值型数据中又有一类最常用的数据,称为字符型数据,它可以方便地表示文字信息,供人们直接阅读和理解。其他的非数值型数据则主要用来表示图画、声音和活动图像,它们在许多应用场合使用。

用计算机进行数据处理指的是对数据进行加工、转换、储存、合并、分类、排序和计算的过程。数据处理的目的就是从原始数据或基础数据生成或转移得到对使用者有用的数据。

2. 信息

尽管上面多次提到信息与数据,但是,要很严格地区分它们,并不十分容易。

根据 ISO 的定义,可以通俗地认为:信息是对人有用的数据,这些数据将可能影响到人们的行为与决策。

上面说过,数据是客观存在的事实、概念或指令的一种可供加工处理的特殊表达形式,而信息却强调是对人有用的数据。例如,某公司的一张报表,对于该公司及有关部门或者其同行

可能是极为宝贵的信息;而对于大多数读者可能是毫无用处的,根本不认为这些报表是有用的信息。

计算机信息处理,实质上就是由计算机进行数据处理的过程。也就是说,通过数据的采集和输入,有效地把数据组织到计算机中,由计算机系统对数据进行相应的存储、建库、处理、加工、转换、合并、分类、计算、统计、汇总、传送等操作过程。经过计算机对数据的加工处理后,向人们提供有用的信息,这个全过程就是信息处理。更通俗地说,信息处理的本质就是数据处理,数据处理的主要目的是获取有用的信息。

在许多场合,如果不引起混淆,信息和数据往往不严格加以区分。但是,就计算机应用系统的分类而言,人们俗称的信息系统(或称信息管理系统、数据处理系统),常常特指一类数据密集型的应用系统。

3. 媒体

媒体又称媒介、媒质,来源于英文 medium 的释音,指承载信息的载体。根据国际电信联盟(ITU)下属的国际电报电话咨询委员会(CCITT)的定义,与计算机信息处理有关的媒体有 5 种:

①感觉媒体 即使人们的听觉、视觉、嗅觉、味觉和触觉器官直接产生感觉的一类媒体,如声音、文字、图画、气味等,它们是人类使用信息的有效形式。

②表示媒体 为了使计算机能有效地加工、处理、传输感觉媒体而在计算机内部采用的特殊表示形式,即声、文、图、活动图像的二进制编码表示。

③存储媒体 用于临时或永久性存放,以便计算机加工处理或互相交换信息的物理实体,如磁盘、光盘、半导体存储器等。

④表现媒体 用于把感觉媒体转移成表示媒体、把表示媒体转移为感觉媒体的物理设备,前者是计算机的输入设备,如键盘、扫描仪、话筒等,后者是计算机的输出设备,如显示器、打印机、音箱等。

⑤传输媒体 用来将表示媒体从一台计算机传送到另一台计算机的通信载体,如同轴电缆、光纤、电话线等。

4. 多媒体

多媒体技术中的多媒体,指的是多种感觉媒体。所谓多媒体技术,是指能够交互式地综合处理多种不同感觉媒体(语言、音乐、文字、数值、图画、活动图像,其中至少包含声音或活动图像)的信息处理技术;具有这种功能的计算机就是多媒体计算机;具有这种能力的通信系统就是多媒体通信;能够有效地存储、管理、检索多种感觉媒体的数据库系统就是多媒体数据库系统。多媒体技术的发展,使计算机更有效地进入了人类生活的各个领域,促进了全新的信息产业制造业与信息服务业的繁荣兴旺,促使人与计算机之间建立起更为默契与更加融洽的新型关系。

1.2 计算机中数值数据信息的表示

1.2.1 机器数和真值

在计算机中,只能表示 0 和 1 两种数码。为了表示正数和负数,专门选择一位符号位来表示数的符号,通常选择最高位作为符号位,当该位为“0”时,表示正号,为“1”时,表示负号,即最

高位为符号位:符号位为 0,表示正数的符号“+”;符号位为 1,表示负数的符号“-”。

这就是说,数的符号在计算机中也数码化了。我们把一个数在机器(计算机)中的表示形式称为机器数,而把原来实际数本身值叫真值(机器数真值)。真值可以用二进制数表示,也可用十进制数表示,但根据习惯,常用十进制数表示,如图 1-1 所示。

机器数								真值
D₇ D₀ N₁ 0 1 1 0 1 0 1 1 ↑ 符号位 数值位								N ₁ = +1101011 = +107
N₂ 1 1 1 0 1 0 1 1 ↑ 符号位 数值位								N ₂ = -1101011 = -107

图 1-1 机器数真值

机器数有如下特点:

1. 前已述及,机器数的正负号已经数值化了,这是它的第一个特点。

2. 机器数所能表示数的范围受到机器(计算机)字长的限制。那么什么是计算机字长呢?我们先来解释几个计算机中的常用术语。

(1)位(bit):位是计算机所能表示的最小数据单位,它只能有二种状态“0”和“1”。要想表示更大的数,就得把更多的位组合起来作为一个整体,每增加一位,所能表示的数就增大一倍。

(2)字节(Byte):一个 8 位二进制数称为一个字节,是计算机处理数据的基本单位,即以字节为单位存储和解释信息。微型计算机存储器是由一个个存储单元构成的,每个存储单元的大小就是一个字节。所以存储器均以字节为单位编址,存储容量大小也以字节数来度量。常使用四种度量单位,即 KB、MB、GB 和 TB,其大小分别为:

$$1\text{KB} = 2^{10} = 1024\text{B}$$

$$1\text{MB} = 2^{10} \cdot 2^{10} = 1024\text{KB}$$

$$1\text{GB} = 2^{10} \cdot 2^{10} \cdot 2^{10} = 1024\text{MB}$$

$$1\text{TB} = 2^{10} \cdot 2^{10} \cdot 2^{10} \cdot 2^{10} = 1024\text{GB}$$

(3)字(Word):计算机处理数据时,CPU 通过数据总线一次存取、加工或传送的数据称为字。一个字通常由一个字节或若干字节组成。

(4)字长(Word Length):指字的二进制的位数。8 位微处理器的字长为 8 位,每个字由一个字节构成;而在 16 位微处理器中,每个字由 2 个字节构成,如图 1-2 所示。由于字长是计算机一次所能处理的实际位数的长度,所以字长是衡量计算机性能的一个重要标志。字长越长,性能越强,精度越高。常用字长有 8 位、16 位、32 位和 64 位。

3. 小数点不能直接标出,需要按一定方式约定小数点的位置。

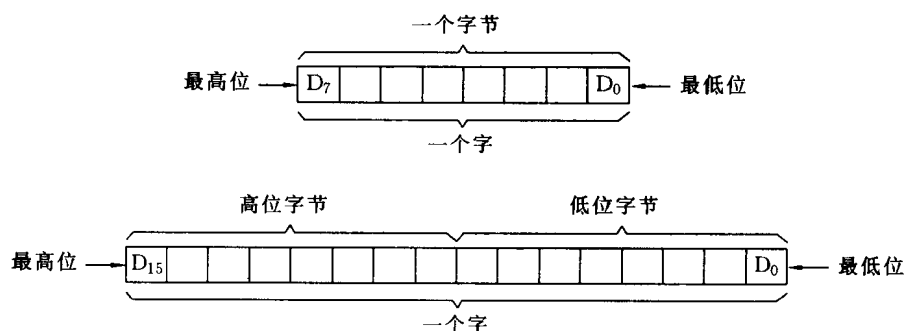


图 1-2 字和字节

1.2.2 符号数的表示方法——原码、反码和补码

既然一个数的数值和符号全都是数码,那么当对这种机器数进行运算时,符号位怎样处理?能不能也同数值位一道参加运算呢?为妥善解决这些问题,引出了机器数 x 的三种不同的编码形式,即原码、反码和补码。分别记作: $[x]_{\text{原}}$ 、 $[x]_{\text{反}}$ 和 $[x]_{\text{补}}$ 。

一、原码

设 $x = x_1 x_2 \cdots x_{n-1}$, 其中 x_i 为一位二进制数, $i = 1, 2, \cdots, (n-1)$
则

$$[x]_{\text{原}} = \begin{cases} 0 & x_1 x_2 \cdots x_{n-1} \quad \text{当 } x \geq 0 \\ 1 & x_1 x_2 \cdots x_{n-1} \quad \text{当 } x \leq 0 \end{cases}$$

即一个数的原码,就是数值部分不变,仅仅用 0 和 1 分别来表示数的符号“+”和“-”的机器数。

例: $x_1 = 67 = +1000011\text{B}$ $[x_1]_{\text{原}} = 01000011$

$x_2 = -67 = -1000011\text{B}$ $[x_2]_{\text{原}} = 11000011$ 在原码表示法中,根据定义,数 0 的原码有两种不同形式(设字长为 8 位):

$$[+0]_{\text{原}} = 00000000$$

$$[-0]_{\text{原}} = 10000000$$

原码表示简单易懂,而且与真值的转换方便。但原码表示的数不便于计算机运算,因为在两原码数运算时,首先要判断它们的符号,然后再决定用加法还是用减法。致使机器的结构相应地复杂化或增加机器的运算时间。为解决上述弊病,引入反码和补码表示法。

二、反码

设 $x = x_1 x_2 \cdots x_{n-1}$, 其中 x_i 为一位二进制数, $i = 1, \cdots, (n-1)$ 。
则

$$[x]_{\text{反}} = \begin{cases} 0 & x_1 x_2 \cdots x_{n-1} \quad \text{当 } x \geq 0 \\ 1 & \bar{x}_1 \bar{x}_2 \cdots \bar{x}_{n-1} \quad \text{当 } x \leq 0 \end{cases}$$

式中

$$\bar{x}_i = \begin{cases} 0 & \text{当 } x_i = 1 \\ 1 & \text{当 } x_i = 0 \end{cases}$$