

全国计算机等级考试最新规划教材
National Computer Rank Examination

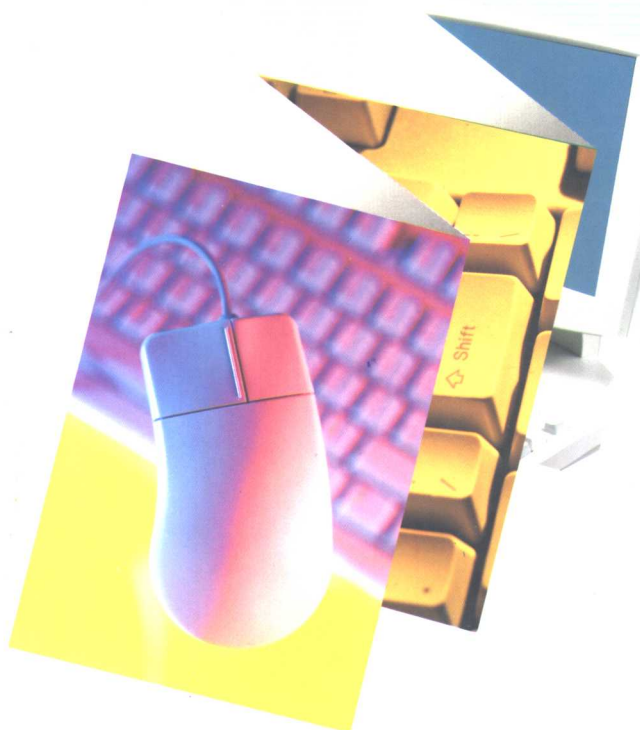
全国计算机等级考试



二级 数据库技术

重点难点 例题精析 精编四合一
上机指导 全真试卷

教育考试研究中心 组编



中国和平出版社

全国计算机等级考试最新规划教材

三级数据库技术 精编四合一教程

重点难点 · 例题精析 · 上机指导 · 全真试卷

教育考试研究中心 组编

中国和平出版社

图书在版编目(CIP)数据

全国计算机等级考试最新规划教材/教育考试研究中心组编—北京：
中国和平出版社,2003.1

ISBN 7-80154-658-X

I.全... II.教... III.电子计算机-水平考试-教材 IV.TP3

中国版本图书馆 CIP 数据核字(2002)第 089597 号

版权所有·翻印必究

全国计算机等级考试最新规划教材
三级数据库技术·精编四合一教程
教育考试研究中心 组编

出版发行:中国和平出版社

经 销:全国新华书店

印 刷:巩义市前程印刷有限公司

开 本:850×1168 毫米 1/16

印 张:290

字 数:7460 千字

版 次:2003 年 1 月第 1 版 2003 年 1 月第 1 次印刷

书 号:ISBN 7-80154-658-X/G·644

社 址:北京市东城区和平里东街民旺甲 19 号 邮政编码:100013

(本书如有缺页或倒装,请与本书销售部门联系退换)

定 价:408.00 元

前言

作为长期从事大学计算机教学和等级考试辅导、评卷与组织管理的教师,我们一直在苦苦思索和探讨如何在有限时间内使考生切实掌握应考知识与操作技能,熟悉考试方法与技巧,进一步提高考试通过率。

令人欣慰的是,功夫不负有心人,通过多年来孜孜不倦的研究探索,我们终于摸索出一套行之有效的辅导培训方法,积累了数以百万字的辅导讲义、典型习题和考题、考试分析及相关著述。这些有形和无形的成果使得我们所辅导的考生大多在各级考试中取得优异成绩,通过率连年大幅提升。

参加全国计算机等级的许多人普遍感到,这种考试与传统考试不同,分为笔试和上机考试,目前除指定教材外,缺少考试要点、例题解析、上机指导和模拟试题四合一的精编辅导书。为此,我们根据 2002 年国家教育部最新颁布的考试大纲和指定教材,将多年来积累的经验重新梳理和汇集,编写了这套丛书。

在本书编写过程中,我们全面研究了近年来计算机等级考试的一系列要点与试题,深入分析了考生容易失误的方面和原因,使得本书突出了以下几个特色:

● **极具参考价值** 本书系计算机等级考试各考点高校中长期从事此项工作的教师和专家学者丰富的考试与辅导经验的精萃集要。他们对大纲领会透彻,对考试把握准确,信息灵、经验多,因此使得本书极具参考价值。

● **知识系统全面** 本书内容涵盖了大纲中所有的笔试和机试知识点与考核点,突出重点与难点。在编排体例上考试要点、例题解析、上机指导与模拟试题恰当组合,针对学习规律有的放矢,全方位训练,从而有效提高实战能力。

● **针对性强** 针对历年来考生容易失误的地方,本书精选典型例题与习题,并作出精辟讲解分析,给读者以前车之鉴。而书的模拟试题则是在把握历年来考试脉络与趋势的基础上,针对下次考试而作出的具有前瞻性与预测性的模拟考试,建议读者按照实际考试要求而训练。

● **精心编撰、打造精品** 作者力争在有限的篇幅内把自己的知识、经验与爱心完美地呈现给读者,奉献于社会。因此,在编写之初就牢固确立起精品意识——精选内容、精心编排、精心排错、打造精品。本书所用程序均经过上机测试,排除差错,给读者以洁净的阅读空间。

本书着力把笔试要点、机试要点与模拟试题浓缩于一册,既可以节约考生的购书开支,又可以使考生开卷即获考试精要,从而节约考生的时间与精力,提高学习效率。我们的目标是使本书读者“一书在手,精华全有;一册在握,顺利通过!”

在将本书奉献给广大读者之际,我们也热切盼望广大同行和读者将自己的经验和体会以及对本书的意见和建议及时反馈给我们,以利于促进和改善我们的工作。

编者

目录

第一部分 笔试指导/1

第一章 基础知识/1

考试要点/1

例题精析/24

标准习题自测/35

答案要点精解/40

第二章 数据结构与算法/42

考试要点/42

例题精析/67

标准习题自测/80

答案要点精解/88

第三章 操作系统/89

考试要点/89

例题精析/130

标准习题自测/137

答案要点精解/145

第四章 数据库技术基础/147

考试要点/147

例题精析/153

标准习题自测/157

答案要点精解/164

第五章 关系数据库系统/166

考试要点/166

例题精析/172

标准习题自测/174

答案要点精解/182

第六章 关系数据库标准语言—SQL/184

考试要点/184

	例题精析/191
	标准习题自测/195
	答案要点精解/204
第七章	关系数据库的规范化理论与数据库设计/207
	考试要点/207
	例题精析/215
	标准习题自测/218
	答案要点精解/228
第八章	数据库管理系统/231
	考试要点/231
	例题精析/236
	标准习题自测/238
	答案要点精解/240
第九章	事务管理与数据库安全性/241
	考试要点/241
	例题精析/246
	标准习题自测/247
	答案要点精解/255
第十章	新一代数据库应用开发工具/257
	考试要点/257
	例题精析/262
	标准习题自测/263
	答案要点精解/268
第十一章	数据库技术的发展/270
	考试要点/270
	例题精析/275
	标准习题自测/277
	答案要点精解/283

第二部分 上机指导 / 286

考试要求 / 286

考试环境 / 286

Turbo C 系统的上机操作 / 287

标准习题自测 / 326

答案要点精解 / 334

第三部分 模拟试题 / 338

笔试模拟试题(一) / 338

笔试模拟试题(一)参考答案 / 346

笔试模拟试题(二) / 347

笔试模拟试题(二)参考答案 / 354

笔试模拟试题(三) / 355

笔试模拟试题(三)参考答案 / 362

机试模拟试题(一) / 363

机试模拟试题(一)参考答案 / 365

机试模拟试题(二) / 366

机试模拟试题(二)参考答案 / 368

机试模拟试题(三) / 369

机试模拟试题(三)参考答案 / 370

附 录 / 371

附录 A 三级数据库设计等级考试大纲 / 371

附录 B 2002 年 9 月全国计算机等考试二级笔试试卷
及参考答案 / 373

第一部分 笔试指导

第一章 基础知识

考试要点

一、计算机的发展

自从 1946 年 2 月现代电子计算机的鼻祖 ENIAC(electronic numerical integrator and computer)在美国宾夕法尼亚大学问世以后,短短 50 年里,计算机技术经历了巨大的变革。

学术界经常使用器件(硬件)划分计算机的发展史,如第一代电子管计算机(1947~1957),第二代晶体管计算机(1958~1964),第三代集成电路计算机(1964~1972),第四代大规模集成电路计算机(1972~),目前提出了所谓的第五代(或新一代)计算机。

从 1946 年到 50 年代后期(1946~1957)为电子管计算机时期。计算机的元器件主要由电子管(vacuum tube)组成。其特点是体积庞大、功耗高、运算速度较低。如 ENIAC 占地 170m^2 ,重达 30 吨,功耗为 140kW ,有 18000 多个电子管,每秒钟能进行 5000 次加法计算。这一阶段,计算机主要用于军事、国防等尖端技术领域。除了 ENIAC 以外,1945 年左右,冯·诺依曼等人在研制 EDVAC(electronic discrete variable computer)时,提出了存储程序(stored-program)概念,奠定了以后计算机发展的基石。IBM 公司 1954 年 12 月推出的 IBM650 是第一代计算机的代表。

从 20 世纪 50 年代后期到 60 年代中期(1958~1964)为晶体管计算机时期。自从 1947 年晶体管(transistor)在贝尔实验室诞生后,引发了一场影响深远的电子革命。体积小、功耗低、价格便宜的晶体管取代了电子管,不仅提高了计算机的性能,也使计算机在科研、商业等领域内得到广泛地应用。第二代计算机不仅采用了晶体管器件,而且存储器改用速度更快的磁芯存储器;与此同时高级编程语言和系统软件的出现,也大大提高了计算机的性能和拓宽了其应用领域。这一时期计算机的代表主要有 DEC 公司 1957 年推出的 PDP-1、IBM 公司于 1962 年推出的 7094 以及 CDC 公司 1964 年研制成功的 CDC6600。1969 年 CDC 公司研制的 DCD7600 平均速度达到每秒千万次浮点运算。

从 20 世纪 60 年代中期到 70 年代初期(1965~1972)为集成电路计算机时代。第一代和第二代计算机均采用分离器件(discrete component)组成。集成电路(integrated circuit)的出现,宣告了第三代计算机的来临。由于采用了集成电路,使得计算机的制造成本迅速下降;同时因为逻辑和存储器件集成化的封装,大大提高了运行速度,功耗也随之下降;集成电路的使用,使得计算机内各部分的互联更加简单和可靠,计算机的体积也进一步缩小。这一时期的代表为 IBM 的 system/360 和 DEC 的 PDP-8。

113531/14

从 20 世纪 70 年代初期到 70 年代后期(1972~1978)为大规模集成电路(LSI)计算机时代。20 世纪 70 年代初半导体存储器的出现,迅速取代了磁芯存储器,计算机的存储器向大容量、高速度的方向飞速发展。存储器芯片从 1kbit,4kbit,16kbit,64kbit,256kbit,1Mbit,4Mbit 发展到 16Mbit(1992 年)。

接着就进入了超大规模集成电路(VLSI)计算机时代。随着技术的日新月异,软件和通信的重要性也逐步上升,成为和硬件一样举足轻重的因素。同时系统结构的特点对计算机的性能也有巨大的影响(中断系统、Cache 存储器、流水线技术等等)。实际上在第三代计算机以后,就很难找到一个统一的标准进行划分。

也可以从应用的观点来划分计算机的发展史。最早的应用是军事上的需要,如炮弹弹道计算,核武器的设计等;其次是广泛地用于科学计算,工程设计计算;第三阶段是大量用于管理,现在计算机的 80% 以上用于管理;再接着是计算机辅助设计(CAD)和辅助制造(CAM);进入 90 年代,计算机的应用已趋向于综合化和智能化,例如在一个企业里,计算机不仅用于科学计算、辅助设计和辅助制造,还用于辅助管理和辅助决策(MIS 与 DSS),以及办公自动化(OA)等等,使设计、生产自动化和管理自动化融为一体,形成所谓计算机集成制造系统(CIMS-Computer Integrated Manufacturing System),再发展下去就是工厂自动化(Factory Automation)或称无人工厂。DSS(Decision Support System)/ES(Expert System)利用人工智能(AI——Artificial Intelligence)技术,让计算机代替人判断、推理,寻找最优方案,以辅助决策者决策。

目前更流行的是认为计算机的发展经过了三次浪潮(wave)。

计算机的发展第一个浪潮是单个主机(Mainframe)的时期,以 IBM360、370 为代表的大型机的出现,其特点是以批处理为主,主要用于大规模科学计算。

第二次浪潮为客户机/服务器(Client/Server)的时期,这时期出现了小型机、微型机和局域网。其特点是多用户分时处理。

第三个浪潮是 70~80 年代的微型计算机 PC(Personal Computer)的出现。现在正处于第三次浪潮,网络计算机的时期,即以网络为中心或以网络为基础的计算机时期。

目前计算机向综合的方向发展,将各种计算机的特点和优点综合起来,并结合了多媒体技术、通信技术等,把人类带入了网络社会。

二、计算机的分类及其应用

计算机分类的方法大致可分如下几种:

1. 按信息的形式和处理方式分类

计算机按信息的形式和处理方式可分为数字计算机、模拟计算机以及数字混合计算机。

2. 按计算机的用途分类

计算机按用途可分为通用计算机和专用计算机。

3. 按计算机规模分类

计算机按规模可划分为巨型机、大型机、中型机、小型机、微型机等。

计算机的应用如下:

1. 在科学计算中的应用
2. 在实时控制中的应用
3. 在数据处理中的应用

4. 计算机在辅助设计和辅助制造(CAD/CAM)中的应用

5. 办公自动化系统中的应用

三、计算机硬件结构

实际应用的计算机系统是由计算机硬件系统、软件系统以及通信网络系统组成的一个整体系统。计算机硬件系统是指构成计算机的所有实体部件的集合,通常这些部件由电路(电子元件)、机械等物理部件组成,它们都是看得见摸得着的,故通常称为“硬件”。计算机硬件结构也可以称为冯·诺伊曼结构,它由五大部件组成:主机部分由运算器、控制器、存储器组成,外设部分由输入设备和输出设备组成,其中核心部分部件是运算器。

计算机硬件之间的连接线路分为网状结构与总线结构,这里主要介绍总线(BUS)结构。总线结构有如下几种形式:

1. 以 CPU 为中心的双总线结构

所谓总线实际上是一组并行的导线,导线的数目和计算机字长相同,数据和指令通过总线传送。

2. 以存储器为中心的双总线结构

3. 单总线结构

主要部件功能:

1. 运算器

运算器是完成二进制编码的算术或逻辑运算的部件。运算器由累加器(用符号 L_A 表示)、通用寄存器(用符号 L_B 表示)和算术逻辑单元(用符号 ALU 表示)组成,核心是算术逻辑单元。

2. 存储器

在计算机中的存储器包括内存储器(又叫主存储器或随机存储器,简称内存或主存)、外存储器、只读存储器 and 高速缓冲存储器以及寄存器等。随机存储器是按地址存取数据的,若地址总线共有 20 条($A_0 \sim A_{19}$),即有 20 个二进制位,可形成 $2^{20} = 1048576$ 个地址(1 兆地址)。

3. 控制器

控制器由三大部件组成,它们是指令部件、时序部件和操作控制部件。

(1) 指令部件

指令部件包括程序计数器 PC,指令寄存器 IR 和指令译码器 ID。

(2) 时序部件

时序部件产生定时节拍,一般由时钟信号源、节拍发生器及微操作电路组成。

4. 输出寄存器

输出寄存器用于存放输出结果,以便由它通过必要的接口(输出通道),在输出设备上输出运算结果。

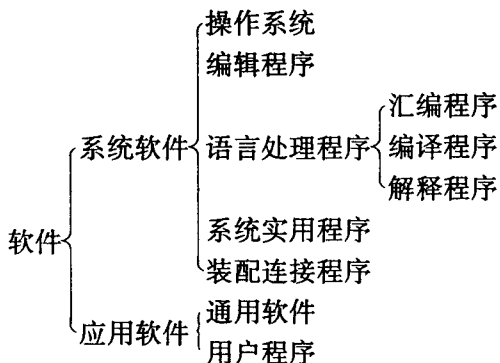
5. 输入设备

目前主要通过 CRT 终端和键盘实现人机对话。磁性设备阅读机、光学阅读机等也可作为输入设备。

四、计算机软件的功能及分类

所谓软件是指为运行、维护、管理、应用计算机所编制的所有程序的总和。软件分为系统软件和应用软件。

系统软件包括计算机操作系统(Operation System)、计算机的各种管理程序、监控程序、调试程序、编辑程序以及各种语言的编译或解释程序等。应用软件是为解决各种实际问题而设计的程序。



1. 操作系统

操作系统具有三大功能:管理计算机硬、软件资源,使之有效使用;组织协调计算机的运行,以增强系统的处理能力;提供人机接口,为用户提供方便。

操作系统具有的管理:

- (1)作业管理。
- (2)资源管理。
- (3)中断处理。
- (4)I/O 处理。
- (5)调度。
- (6)错误处理。
- (7)保护和保密处理。
- (8)记帐。

操作系统的基本类型:

- (1)批处理操作系统。
- (2)分时系统。
- (3)实时系统。

操作系统的管理功能主要内容:

- (1)处理机管理。
- (2)存储管理。
- (3)文件管理。
- (4)设备管理。

2. 数据库管理系统

数据库管理系统既可以认为是一个系统软件也可以认为是一个通用的应用软件。

目前有三种类型的数据库管理系统,故可存放三种模型的数据,这三种数据库管理系统分别为层次数据库、网状数据库和关系数据库。

3. 计算机网络软件

计算机网络系统是通过通信线路连接的硬件、软件与数据集合的一个计算机系统。从硬件来说,除计算机作为网络的结点以外,还有如服务器(也可是一台计算机),网络适配器,终端控制器以及网络连接器等硬件设备;从软件来说,有网络操作系统,网络通信及协议软件,网络数据库管理系统等。

4. 高级语言及语言处理器

用户用高级语言编写的程序称源程序,源程序不能由计算机直接执行,必须翻译成机器能执行的语言——机器语言,这种翻译是由机器自动翻译的,“译员”称编译程序或编译器,当源程序输入计算机后,调用编译程序编译成机器语言(称目标程序),然后执行。还有一种语言处理程序叫解释程序,输入一条语句,翻译一条。现在已出现了第4代语言(4GL)和计算机辅助软件工具CASE。

5. 常用的通用软件

在数据处理、事务处理、报表处理中有许多通用软件,如字处理软件WPS、WORD,报表处理软件LOTUS 1-2-3等。

五、计算机数据表示

1. 二进制计数制

引入二进制数字系统的计算机结构和性能具有如下的优点:

(1)技术实现容易。

(2)二进制运算规则简单。

(3)计算机中二进制的0、1数码与逻辑代数变量值0与1吻合,所以二进制同时可以使计算机方便地进行逻辑运算。

(4)二进制数和十进制数之间的关系亦不复杂。

2. 任意进制计数制和十进制计数制的相互转换

十进制数转换成二进制数:

十进制数据转换为二进制数时,因整数部分与小数部分转换算法不同,需要分别进行。

(1)整数转换方法——除基取余法

十进制整数除以2取余数作最低位系数 k_0 ,再取商的继续除以2取余数作高一位的系数,如此继续直到商为0时停止,最后一次的余数就是整数部分最高有效位的二进制系数,依次所得到的余数序列就是转换成的二进制数。因为除数2是二进制的基数,所以这种算法称作“除基取余”法。

(2)小数转换方法——乘基取整法

把十进制小数乘以2,取其积的整数部分作为对应二进制小数的最高位系数 k_{-1} ,再取积的纯小数部分乘以2,新得积的整数部分又作下一位的系数 k_{-2} ,再取其积的纯小数部分继续乘2, ..., 直到乘积小数部分为0时停止,这时乘积的整数部分是二进制数最低位系数,每次乘积得到的整数序列就是所求的二进制小数,这种方法每次乘以基数取其整数作系数。所以叫乘基取整法。需要指出的是并不是所有十进制小数都能转换成有限位的二进制小数并出现乘积的小数部分0的情况,有时整个换算过程会无限进行下去,此时可以根据要求并考虑计算机字长,取一定长度的位数后四舍五入,这时得到的二进制数是原十进制数的近似值。

一个既有整数部分又有小数的数送入计算机后,由机器把整数部分按“除基取余”法,小数部分按“乘基取整”法分别进行转换,然后合并。

任意进制数转换成十进制数:

任意一种进位计数制的数转换成十进制数的方法都是一样的。把任意进制数按权展开成多项式和的形式,把各位的权与该位上的数码相乘,乘积逐项相加,其和便是相应的十进制数。

十进制数转换成任意进制数:

十进制数转换成任意进制数与十进制数转换成二进制数的方法完全相同,即整数部分用除基取余的算法,小数部分用乘基取整的方法,然后将整数与小数拼接成一个数作为转换的最后结果。

3. 数的机器码表示

符号数的机器码表示:

(1) 机器数和真值

数在计算机中的表示形式统称为机器数。机器数有两个基本特点,其一,数的符号数值化。实用的数据有正数和负数,因为计算机只能表示 0、1 两种状态,数据的正号“+”或负号“-”,在机器里就用一位二进制的 0 或 1 来区别。通常这个符号放在二进制数的最高位,称符号位,以 0 代表符号“+”,以 1 代表符号“-”,这样正负符号就被数值化了。因为有符号占据一位,数的形式值就不等于真正的数值,带符号位的机器数对应的数值称为机器数的真值。

机器数的另一个特点是二进制的位数受机器设备的限制。机器内部设备一次能表示的二进制位数叫机器的字长,一台机器的字长是固定的。字长 8 位叫一个字节(Byte),现在机器字长一般都是字节的整数倍,如字长 8 位、16 位、32 位、64 位。

符号位数值化之后,为能方便的对机器数进行算术运算、提高运算速度,计算机设计了多种符号位与数值一起编码的方法,最常用的机器数表示方法有三种:原码、反码和补码。

(2) 原码表示法和反码表示法

一个机器数 X 由符号位和有效数值两部分组成。设符号位为 X_0 , X 真值的绝对值 $|X| = X_1X_2\cdots X_n$, X 的机器数原码表示为:

$$[X]_{\text{原}} = X_0X_1X_2\cdots X_n$$

$$\text{当 } X \geq 0 \text{ 时, } X_0 = 0$$

$$\text{当 } X < 0 \text{ 时, } X_0 = 1$$

原码表示很直观,但原码加减运算时符号位不能视同数值一样参加运算,运算规则复杂,运算时间长,而计算机大量的数据处理工作是加减运算,原码表示就很不方便了。

一个负数的原码符号位不动,其余各位取相反码就是机器数的另一种表示形式——反码表示法。正数的反码与原码相同。

$$\text{设 } [X]_{\text{原}} = X_0X_1X_2\cdots X_n$$

$$\text{当 } X_0 = 0 \text{ 时, } [X]_{\text{反}} = X_0X_1X_2\cdots X_n$$

$$\text{当 } X_0 = 1 \text{ 时, } [X]_{\text{反}} = X_0\bar{X}_1\bar{X}_2\cdots\bar{X}_n$$

(3) 补码表示法(complement)

设计补码表示法的目的是:①使符号位能和有效数值部分一起参加数值运算从而简化运算规则,节省运算时间。②使减法运算转化成加法运算,从而进一步简化计算机中运算器的线路

设计。

计算机是一种有限字长的数字系统,因此都是有模运算,超过模的运算结果都将溢出。 n 位二进制整数的模是 2^n 。

一个数的补码记作 $[X]_{\text{补}}$, 设模是 M , X 是真值, 补码定义如下:

$$[X]_{\text{补}} = \begin{cases} [X]_{\text{原}} & X \geq 0 \\ M + X & X < 0 \end{cases}$$

从这个定义出发就能求得一个数的补码。

对于二进制数还有一种更加简单的方法由原码求得补码。

①正数的补码表示与原码一样, $[X]_{\text{补}} = [X]_{\text{原}}$

②负数的补码是将原码符号位保持“1”之后其余各位取相反的码, 末位加 1 便得到补码, 即取其原码的反码再加 1: $[X]_{\text{补}} = [X]_{\text{反}} + 1$ 。

真值 $+0$ 和 -0 的补码表示是一致的, 但在原码和反码表示中具有不同的形式。8 位补码机器数可以表示 -128 , 但不存在 $+128$ 的补码, 由此可知 8 位二进制补码能表示数的范围是 $-128 \sim +127$ 。应该注意, 不存在 -128 的 8 位原码和反码形式。

根据互补的概念, 一个补码机器数再求一次补就得到机器数的原码了。

定点数与浮点数:

(1) 定点数(fixed-point number)

计算机处理的数据不仅有符号, 而且大量的数带有小数, 小数点不占有二进制位而是隐含在机器数里某固定位置上。通常采用两种简单的约定: 一种是约定所有机器数的小数点位置隐含在机器数的最低位之后, 叫定点纯整数机器数, 简称定点整数。

另一种约定是所有机器数的小数点位置隐含在符号位之后、有效数值部分最高位之前, 叫定点纯小数机器数, 简称定点小数。

计算机采用定点数表示时, 对于既有整数又有小数的原始数据, 需要设定一个比例因子, 数据按比例因子缩小成定点小数或扩大成定点整数再参加运算, 结果输出时再按比例折算成实际值。 n 位原码定点整数的表示范围是 $-(2^{n-1}-1) \leq X \leq 2^{n-1}-1$, n 位原码定点小数的表示范围是 $-(1-2^{-(n-1)}) \leq X \leq 1-2^{-(n-1)}$ 。当机器数小于定点数的最小值时, 被当作 0 处理, 超出定点数的最大值时, 机器无法表达, 称作“溢出”, 此时机器将停止运算, 屏幕显示溢出警告。

定点数表示方法简单直观, 不过定点数表示数的范围小, 不易选择合适的比例因子, 运算过程容易产生溢出。

(2) 浮点数(floating-point number)

计算机采用浮点数来表示数值, 它与科学计算法相似, 把任意一个二进制数通过移动小数点位置表示成阶码和尾数两部分:

$$N = 2^E \times S$$

其中: E —— N 的阶码(exponent), 是有符号的整数;

S —— N 的尾数(mantissa), 是数值的有效数字部分, 一般规定取二进制定点纯小数形式。

浮点数运算必须化成规格化形式。所谓规格化, 对于原码尾数应使最高数字位 $S_1 = 1$, 如果不是 1, 且尾数不是全为 0 时就要移动尾数直到 $S_1 = 1$, 阶码相应变化, 保证 N 值不变。如果

尾数是补码,当 N 是正数时, S_1 必须是 1,而 N 是负数时, S_1 必须是 0,才称为规格化的形式。

4. 数字编码

十进制数在机内转换成二进制数时,有时也以一种中间数字编码形式存在,它把每一位十进制数用四位二进制编码表达,每一组只表达 0~9 的数值运算时,有专门的电路在每四位二进制间按“十”进位处理,故称为二进制编码的十进制数——BCD 码(Binary Coded Decimal)或称二一十进制数。其编码种类很多,如格雷码、余 3 码等,最常用的叫 8421 BCD 码,4 个二进制位自左向右每位的权分别是 8、4、2、1。0~9 的 8421 码与通常的二进制一样进位,十分简单,当计数超过 9 时,需要采取办法自动向十进制高位进一,即要进行“十进制调整”才能得到正确结果。

5. 校验码

由于器件质量不可靠、线路工艺不过关、远距离传送带来的干扰或受来自电源、空间磁场影响等因素,使得信息在存取、传送和计算过程中难免会发生诸如“1”误变为“0”的错误,计算机一旦出错,要能及时检测并纠正错误,其中一种方法是对数据信息扩充,加入新的代码,它与原数据信息一起按某种规律编码后具有发现错误的能力,有的甚至能指出错误所在的准确位置使机器自动纠正,能起这种作用的编码叫“校验码”(check code)。

奇偶校验码:

将每个数据代码扩展一个二进位作校验位(parity bit),这个校验取 0 还是取 1 的原则是:若是奇校验(odd parity),编码是含“1”的个数连同校验位的取值在内共有奇数个“1”;若是偶校验(even parity),连同校验位在内编码里含“1”的个数是偶数个。

交叉校验:

计算机进行大量字节传送时一次传送几百甚至更多字节组成的数据块,如果不仅每一个字节有一个奇偶校验位——称横向校验,而且全部字节的同一位也设置了一个奇偶校验位——称纵向校验,对数据块代码的横向纵向同时校验,这种情况叫交叉校验。

循环冗余校验码——CRC 码(Cyclic Redundancy Check):

计算机信息传向远方终端或传到另一个计算中心时,信息沿一条通信线路一位位传送,这种通信方式叫串行通信。循环冗余码(简称 CRC 码)就是一种检验能力很强,在串行通信中广泛采用的校验编码。

(1)CRC 码

串行传送的信息 $M(X)$ 是一串 k 位二进制序列,在它被发送的同时,被一个事先选择的“生成多项式”相除,“生成多项式”长 $r+1$ 位,相除后得到 r 位余数就是校验位,它拼接到原 k 位有效信息后面即形成 CRC 码。CRC 码到达接收方时,接收方的设备一方面接收 CRC 码,一方面用同样的生成多项式相除,如果正好除尽,表示无信息差错,接收方去掉 CRC 码后面 r 位校验位,收下 k 位有效信息;当不能除尽时,说明有信息的状态位发生了转变,即出错了。一般要求重新传送一次或立即纠错。

(2)CRC 码计算

传送信息时生成 CRC 码以及接收时对 CRC 码校验都要与“生成多项式”相除,这里除法是“模 2 运算”,即二进位运算时不考虑进位和借位。作模 2 除法时,取商的原则是当部分余数首位为 1 时商取 1,反之商取 0,然后按模 2 减,求部分余数。这个余数不计高位。当被除数逐位除完时,最后余数的位数比除数少一位。该余数就是校验位。它拼接在有效信息后面组成

CRC 码。因为校验位扩充了传送部分的代码,所以这是一种基于“冗余校验”的思想的校验办法。

(3)生成多项式

CRC 码是 $M(X)$ 除以某一个预先选定的多项式后产生的,所以这个多项式叫生成多项式。并不是任何一个 $r+1$ 位的编码都可以作生成多项式用,它应能满足当任何一位发生传送错误时都能使余数不为 0,并且不同位发生错误时应当使余数也不同,这样不但能检错而且能推断是哪一位出错,从而有利于准确的纠错。有两个生成多项式,其检错率很高。

$$X^{16} + X^{15} + X^2 + 1$$

$$X^{16} + X^{12} + X^6 + 1$$

6. 非数值数据的表示方法

计算机中数据的概念是广义的,机内除有数值数据之外,还有文字、符号、图象、语言和逻辑信息等等,因为它们也都是 0、1 形式存在,所以称为非数值数据。

(1)字符数据

字符数据主要指数字、字母、通用符号、控制符号等,在机内它们都被变换成计算机能够识别的二进制编码形式。国际上普遍采用的一种编码是美国国家信息交换标准代码(American Standard Code for Information Interchange),简称 ASCII 码。

ASCII 码选择了四类共 128 种常用的字符:

①数字 0~9。

②字母。

③通用符号。

④动作控制符。

(2)逻辑数据

逻辑数据是指计算机不带符号位的一位二进制数。

逻辑数据在计算机中虽然也是“0”或“1”的形式,但是与数值有很大区别:

①逻辑数据的取值只有“0”和“1”两个值,不可能再有其他值,而数值数据 0 与 1 的不同组合可以反映很多不同数值。

②逻辑数据的“0”和“1”代表两种成对出现的逻辑概念,与一般数学中代表“0”和“1”的数值概念截然不同。

③逻辑数据和逻辑数据运算可以表达事物内部的逻辑关系,而数值数据表达的是事物的数量关系。

汉字:

(1)汉字字音编码

(2)汉字字形编码

(3)汉字音形编码

(4)电报码

(5)整字编码

为了能在不同的汉字系统之间交换信息、高效率高质量共享汉字信息,近年来国家推出了一系列有关中文信息处理的标准。比如 1981 年我国制定推行的 GB2312-80 国家标准信息交

换用汉字编码字符集(基本集)——简称国标码,以及若干辅助集。国标码收集、制定的基本图形字符有 7 千余个,其中常用汉字 3755 个,次常用汉字 3008 个,共 6763 个汉字,还有俄文字母、日语假名、拉丁字母、希腊字母、汉语拼音,每字节内占用 7 bit 信息,最高位补 0,例如汉字“啊”的国际码,前一字节是 01100000,后一字节是 00100001,编码为 3021H。

汉字内部码是汉字在计算机内部存储、运算的信息代码,内部码的设计要求与西文信息处理有较好的兼容性,当一个汉字以某种汉字输入方案送入计算机后,管理模块立刻将它转换成两字节长的 GB2312-80 国标码,如果给国标码的每字节最高位加“1”,作为汉字标识符,就成为一种机器内部表示汉字的代码——汉字内部码。

汉字内部码的特点十分明显:

①汉字内部码结构简短。一个汉字内部码只占两个字节,两字节足以表达数千个汉字和各种符号图形,且又节省计算机存储空间。

②便于和西文字符兼容。西文字符的 ASCII 码占一个字节,两字节的汉字内码可以看成是它扩展的字符代码,在同一个计算机系统中,只要从最高位标识符就能区分这两种代码。标识符是“0”,即是 ASCII 码;标识符是“1”,则是汉字内部码。

7. 语音识别及语言表示原理

语音产生机理的研究表明,每一种语言的语音都有自己特定的音素特征,语音是不同频率振动的结果。分析语音的音素特点,找出音素的基频和高次频率成分,就能在计算机中建立发音系统的模型,在实施中对语音采样,通过滤波器分解提取频率信息,由模/数转换设备转换成数字输入计算机,与机内的语言模型比较,由此达到识别语音的目的。与此相反,如果选择已知音素的参数,应用语音系统模型,就能得到指定的音素,进一步按照一定的规则合成语言。

六、运算器

1. 运算器的组成

多功能算术/逻辑运算单元(ALU):

(1) 基本思想

关于一位全加器(FA)的逻辑表达式为:

$$F_i = A_i \oplus B_i \oplus C_i$$

$$C_{i+1} = A_i B_i + B_i C_i + C_i A_i$$

式中 F_i 是第 i 位的和数, A_i 、 B_i 是第 i 位的被加数和加数, C_i 是第 i 位的进位输入, C_{i+1} 为第 i 位的进位输出。

一位算术/逻辑运算单元的逻辑表达式为:

$$F_i = X_i \oplus Y_i \oplus C_{n+i}$$

$$C_{n+i+1} = X_i Y_i + Y_i C_{n+i} + C_{n+i} X_i$$

上式中,进位下标用 $n+i$ 代替原来一位全加器中的 i , i 代表集成在一片电路上的 ALU 的二进制位数,对于四位一片的 ALU, $i=0,1,2,3$ 。 n 代表若干片 ALU 组成更长的运算器时每片电路的进位输入。

(2) 逻辑表达式

ALU 的某一位逻辑表达式如下: