

高等学校教材

# 大学计算机基础

主 编 郝兴伟  
副主编 巩裕伟

高等教育出版社

## 内 容 提 要

本书是根据教育部非计算机专业计算机基础课程教学指导分委员会提出的《关于进一步加强高校计算机基础教学的意见》中有关“大学计算机基础”课程教学要求编写的。本书内容是在计算机文化基础上发展而来的,主要包括计算机基础知识、操作系统、办公自动化软件、计算机网络基础、Internet 与 WWW、网页设计和 HTML 语言、软件开发、常用工具软件和国民经济与社会信息化等。本书加强了计算机的基本概念、基本工作原理的讲解,力求为计算机知识的进一步学习和应用打好基础。同时,又兼顾了内容的实用性,以提高学生的计算机应用水平。本书在内容安排上,密切结合计算机技术的最新发展,具有较宽的知识面、适中的讲解深度和较好的可操作性。

本书可作为高等学校大学计算机基础课程教材,同时,对于那些需要学习计算机知识及应用、计算机网络和 Internet 的读者,也是一本很好的参考书。

# 前 言

随着计算机应用的快速普及和社会信息化进程的快速发展,计算机基础教育正在发生两个明显的变化,一是计算机文化基础课程的教学内容已经下移,即在中学普遍开设了“信息技术”课程;二是社会对各方面人才计算机应用水平的要求不断提高。在这样的社会背景下,教育部高等学校非计算机专业计算机基础课程教学指导分委员会和教育部高等学校文科计算机基础教学指导委员会分别组织编写的《关于进一步加强高校计算机基础教学的意见》和《高等学校文科类专业大学计算机教学基本要求(2003年版)》,对高等学校的计算机基础教育提出了指导性的意见。

在深入理解和领会教育部对高等学校计算机教育新的要求下,结合多年来计算机文化基础的教学经验,在《计算机文化基础》第二版的基础上,对教学内容进行了进一步的规划和组织,形成本书。

新的大学计算机基础继续贯彻 what + how + why 教学理念,将从更深、更专业的角度全面讲解计算机应用的基本知识,为学生下一步的计算机应用打好理论和技术基础。全书分成9个大的模块,分别是:

第1章,计算机基础知识。取消了一些常识性的内容,增加和补充了关于数据编码、计算机硬件结构和数据存储方面的基本知识。

第2章,操作系统。介绍操作系统的基本原理、分类、功能,同时简单地介绍了 Windows 2000 操作系统的使用。

第3章,办公自动化软件。主要是考虑到学生的计算机操作需要,继续介绍了常用的办公软件,如 Word、Excel、PowerPoint 和 Access,但是在讲解方式上做了重要的调整,将 Office 作为一个统一的整体考虑,大大简化了具体的操作步骤介绍,突出软件基本功能和数据交换,适当增加常用的操作技巧。可以看成是计算机文化基础或中学信息技术相关内容的提高。

第4章,计算机网络基础。全面介绍网络的基本概念、功能、网络模型、TCP/IP 协议、网络设备及其功能,以及网络的架设。

第5章,Internet 与 WWW。介绍了 Internet 的发展历史,Internet 的常用服务,WWW 的基本原理,中国互联网络信息中心(CNNIC),国内主要的骨干网,IE 浏览器以及电子邮件服务的基本原理。

第6章,网页设计和 HTML 语言。对网页制作和管理工具 FrontPage 进行了介绍。

第7章,软件开发。介绍软件开发的基本概念、一般过程、算法与数据结构以及软件工程,同时介绍了 C 语言程序设计,并安排了大量的例题和习题,目的是使学生基本掌握 C 语言程序设计。

第8章,常用工具软件。在计算机应用的过程中,有许多非常优秀的软件,在这一章中简单地介绍了一些常用软件的功能和用法,包括压缩软件、多媒体制作软件、常用网络服务

软件( MSN、Netmeeting 等 )。

第9章 国民经济与社会信息化。主要介绍了信息化的基本概念,国内外信息化的发展现状、信息技术与信息产业、电子政务、电子商务、企业信息化以及数字城市和数字地球的概念。

在长期的计算机基础教学中,我们深切地体会到,随着学生计算机知识的提高,计算机基础教学不再是手把手的教学,更重要的是教会学生计算机应用的基本原理和技术。同时,基础教学又不能按照计算机专业来要求,要做到两者合理的融合是很困难的。无论是教育部计算机基础教育的3个层次要求,还是新的计算机基础教育指导性意见,都反映了上述精神。几年来,我们对不同层次的学生需求分别编写了《计算机文化基础》、《计算机技术基础》和《计算机技术及应用》3本教材。并且,在教学过程中根据学生提出的问题和发展的实际,做到每两年左右进行改版,保证内容的先进性和实用性,力争精益求精。

在教学过程中,我们收集了大量的教学素材,制作了完整的课件,课件采用浏览器方式制作,内容丰富,便于网络教学和自学,需要教学课件光盘的教师可与作者联系。同时,我们建立了计算机基础教学网络平台,教师和学生可以针对学习中遇到的相关问题进行讨论,网站域名为 <http://jcjy.sdu.edu.cn/>。

最后,我们对工作在计算机基础教学一线的广大老师表示我们深深的敬意,你们任劳任怨,为我国计算机应用的普及做出了重大的贡献。同时,对这4年多来为教材的出版付出辛勤劳动的高等教育出版社的领导和编辑表示衷心的感谢,没有他们的支持我们将无法完成本书,也不能高质量地完成《计算机文化基础》(第一版、第二版)、《计算机技术基础》以及《计算机技术及应用》的出版。

由于作者水平有限,本书可能存在许多不足之处,希望广大老师和同学们提出宝贵的意见。作者 E-mail 地址为 [hwx@sdu.edu.cn](mailto:hwx@sdu.edu.cn)。

编 者

2004年3月6日

# 目 录

第 1 章 计算机基础知识 .....	2.1.3 操作系统的分类 .....
1.1 二进制与数据的编码 .....	2.2 Windows 2000 概述 .....
1.1.1 数的进制与转换 .....	2.2.1 Windows 2000 特性 .....
1.1.2 原码、反码与补码 .....	2.2.2 Windows 2000 的安装和配置 .....
1.1.3 数的定点表示和浮点表示 .....	2.2.3 登录到计算机 .....
1.1.4 数据的存储单位 .....	2.2.4 关机 .....
1.1.5 字符编码 .....	2.3 基础知识 .....
1.2 微型计算机的组成和基本原理 .....	2.3.1 基本概念 .....
1.2.1 微型计算机系统的 3 个层次 .....	2.3.2 鼠标与键盘操作 .....
1.2.2 微型计算机系统的硬件结构 .....	2.3.3 应用程序 .....
1.2.3 微型计算机硬件的各个组成 部分 .....	2.3.4 Windows 窗口 .....
1.2.4 微型计算机的指令 .....	2.3.5 对话框和控件 .....
1.2.5 微型计算机的工作过程 .....	2.3.6 剪贴板 .....
1.2.6 微型计算机的主要性能指标 .....	2.4 Windows 桌面 .....
1.3 计算机软件系统 .....	2.4.1 任务栏 .....
1.3.1 操作系统 .....	2.4.2 我的文档 .....
1.3.2 程序设计语言 .....	2.4.3 我的电脑 .....
1.3.3 数据库管理系统 .....	2.4.4 网上邻居 .....
1.3.4 应用软件 .....	2.4.5 回收站 .....
1.4 数据存储 .....	2.4.6 Internet Explorer 浏览器 .....
1.4.1 常用的外存储器 .....	2.5 控制面板 .....
1.4.2 磁盘的结构和有关的规定 .....	2.5.1 显示 .....
1.4.3 文件和文件目录的概念 .....	2.5.2 系统 .....
1.5 计算机与信息的安全 .....	2.5.3 添加/删除程序 .....
1.5.1 计算机病毒及其防治 .....	2.5.4 文件夹选项 .....
1.5.2 网络及信息安全 .....	2.5.5 用户和密码 .....
1.6 相关法律法规与伦理道德问题 .....	2.5.6 打印机 .....
1.6.1 我国涉及计算机与信息安全的 法律法规 .....	2.6 文件和文件夹操作 .....
1.6.2 网络使用中的伦理道德问题 .....	2.6.1 资源管理器 .....
习题一 .....	2.6.2 新建文件或文件夹 .....
第 2 章 操作系统 .....	2.6.3 打开文件或文件夹 .....
2.1 操作系统的基本概念 .....	2.6.4 选择文件或文件夹 .....
2.1.1 什么是操作系统 .....	2.6.5 复制或移动文件或文件夹 .....
2.1.2 操作系统的基本功能 .....	2.6.6 删除文件或文件夹 .....
	2.6.7 重命名文件或文件夹 .....
	2.6.8 发送文件或文件夹 .....

2.6.9 查找文件或文件夹 .....	放映 .....
2.6.10 文件或文件夹属性 .....	3.4 Access 数据库应用 .....
2.6.11 快捷方式 .....	3.4.1 数据库的基本概念 .....
2.7 区域选项、字体及中文输入 .....	3.4.2 数据库和表的建立与修改 .....
2.7.1 Unicode 字符集 .....	3.4.3 查询的建立与修改 .....
2.7.2 区域选项 .....	3.4.4 窗体的设计 .....
2.7.3 字体 .....	3.4.5 报表及其设计 .....
2.7.4 中文输入法 .....	习题三 .....
2.8 附件 .....	<b>第4章 计算机网络基础</b> .....
2.8.1 记事本 .....	4.1 计算机网络概述 .....
2.8.2 画图 .....	4.1.1 网络的功能 .....
2.8.3 媒体播放机 .....	4.1.2 网络标准 .....
2.9 用户账户和组 .....	4.1.3 网络的分类 .....
2.9.1 用户账户 .....	4.1.4 网络拓扑结构 .....
2.9.2 用户组 .....	4.1.5 网络协议 .....
2.9.3 创建和管理用户账号 .....	4.2 网络传输介质 .....
2.9.4 创建和管理用户组 .....	4.2.1 双绞线 .....
2.10 帮助 .....	4.2.2 同轴电缆 .....
习题二 .....	4.2.3 光纤 .....
<b>第3章 办公自动化软件</b> .....	4.3 网络硬件和网络设备 .....
3.1 文字处理 .....	4.3.1 网卡 .....
3.1.1 Word 2000 的界面和视图 .....	4.3.2 中继器 .....
3.1.2 文档的建立、编辑与保存 .....	4.3.3 集线器 .....
3.1.3 文档的格式化 .....	4.3.4 桥连接器 .....
3.1.4 Word 2000 的表格处理 .....	4.3.5 交换机 .....
3.1.5 对象的插入与格式化 .....	4.3.6 路由器 .....
3.1.6 文档的版式设计及打印 .....	4.3.7 机柜 .....
3.2 电子表格 .....	4.4 网络的架设 .....
3.2.1 Excel 2000 概述 .....	4.4.1 以太网(Ethernet) .....
3.2.2 工作簿的建立与管理 .....	4.4.2 10BaseT 网络的架设 .....
3.2.3 数据输入与填充 .....	4.5 网络操作系统及其配置 .....
3.2.4 工作表的格式化 .....	4.5.1 TCP/IP 协议 .....
3.2.5 公式与函数 .....	4.5.2 Windows 2000 中的网络配置 ...
3.2.6 Excel 图表 .....	4.6 使用网上邻居 .....
3.2.7 数据清单的管理 .....	习题四 .....
3.2.8 打印工作表 .....	<b>第5章 Internet 与 WWW</b> .....
3.3 演示文稿的制作 .....	5.1 Internet 的发展历程 .....
3.3.1 PowerPoint 2000 概述 .....	5.2 WWW 概述 .....
3.3.2 演示文稿的建立与编辑 .....	5.2.1 Web 是什么 .....
3.3.3 文本编辑和对象插入 .....	5.2.2 超文本、HTML 与 Web 页 .....
3.3.4 文本与对象的格式化 .....	5.2.3 Web 的工作机理 .....
3.3.5 定义动画与动作设置 .....	5.2.4 浏览器 .....
3.3.6 演示文稿的放映设置与	5.2.5 几个主要概念 .....

5.3 国内 Internet 骨干网及 ISP 介绍 .....	6.1.3 管理 FrontPage Web 站点 .....
5.3.1 中国互联网络信息中心 .....	6.2 网页的编辑 .....
5.3.2 中国科技网 .....	6.2.1 创建新网页 .....
5.3.3 中国公众计算机互联网 .....	6.2.2 保存网页 .....
5.3.4 中国教育和科研网 .....	6.2.3 文本编辑和格式化 .....
5.3.5 中国金桥网 .....	6.2.4 插入图片和图片格式的设置 .....
5.3.6 Internet 服务供应商 .....	6.2.5 插入表格 .....
5.4 连接到 Internet .....	6.2.6 超链接 .....
5.4.1 Internet 接入技术 .....	6.3 网页属性和框架网页 .....
5.4.2 通过局域网连接 .....	6.3.1 网页属性 .....
5.4.3 通过电话线和调制解调器连接 .....	6.3.2 框架网页 .....
5.5 使用 IE 浏览器 .....	6.4 HTML 语言概述 .....
5.5.1 Internet Explorer 5.0 概述 .....	6.4.1 HTML 语言的产生与发展 .....
5.5.2 访问 Web 站点 .....	6.4.2 什么是 HTML .....
5.5.3 在 Internet 上搜索信息 .....	6.4.3 HTML 标记和文档结构 .....
5.5.4 使用收藏夹 .....	6.5 常用的 HTML 标记及其属性 .....
5.5.5 保存和打印信息 .....	6.5.1 标记的语法 .....
5.5.6 使用历史记录 .....	6.5.2 常用标记及属性 .....
5.5.7 脱机浏览 Web .....	6.6 HTML 中的列表、表格与框架 .....
5.5.8 使用 FTP 搜索引擎和访问 BBS .....	6.6.1 列表 .....
5.6 IE 浏览器的高级设置 .....	6.6.2 表格设计 .....
5.6.1 Internet 选项常规设置和 IE 主页 .....	6.6.3 框架的应用 .....
5.6.2 Internet 选项连接设置与代理服务 .....	习题六 .....
5.6.3 Internet 选项高级设置与网络实名 .....	第 7 章 软件开发 .....
5.6.4 清除私人信息 .....	7.1 程序设计的相关概念 .....
5.7 收发电子邮件与阅读新闻 .....	7.1.1 数据结构和算法 .....
5.7.1 主要概念 .....	7.1.2 程序和程序设计语言 .....
5.7.2 邮件传输的基本原理 .....	7.1.3 软件、软件危机和软件工程 .....
5.7.3 Outlook Express .....	7.2 程序设计语言——C .....
5.7.4 使用 Outlook Express 阅读新闻 .....	7.2.1 C 语言开发环境 .....
5.7.5 使用 IE 浏览器收发邮件 .....	7.2.2 C 语言程序结构 .....
5.8 国内外常见的门户网站 .....	7.2.3 C 语言基本符号 .....
习题五 .....	7.3 数据、数据类型、类型定义和标准函数 .....
第 6 章 网页设计和 HTML 语言 .....	7.4 运算符、表达式、赋值运算和赋值表达式 .....
6.1 FrontPage 2000 概述 .....	7.5 程序语句 .....
6.1.1 FrontPage 2000 的主窗口 .....	7.5.1 赋值语句和输入/输出 .....
6.1.2 显示模式 .....	7.5.2 分支语句 .....
	7.5.3 循环控制语句 .....
	7.6 函数 .....
	7.6.1 函数定义的一般形式 .....

7.6.2 函数调用 .....	第9章 国民经济与社会信息化 .....
7.6.3 局部变量和全局变量 .....	9.1 社会信息化概述 .....
7.6.4 变量的存储类别 .....	9.1.1 信息化的概念 .....
7.7 预处理命令(宏定义、文件包含和 条件编译).....	9.1.2 信息化的意义 .....
7.8 文件操作 .....	9.1.3 信息化指标体系 .....
习题七.....	9.2 国内外信息化的发展现状 .....
第8章 常用工具软件.....	9.2.1 美国的信息化发展现状 .....
8.1 网络上的软件资源 .....	9.2.2 世界其他主要国家信息建设 的现状 .....
8.1.1 自由软件 .....	9.2.3 我国的信息化建设 .....
8.1.2 免费软件 .....	9.3 信息技术与信息产业 .....
8.1.3 共享软件 .....	9.3.1 信息技术 .....
8.1.4 商业软件 .....	9.3.2 信息产业 .....
8.1.5 开放源码软件 .....	9.4 信息化领域及内容 .....
8.2 压缩软件 .....	9.4.1 电子政务 .....
8.2.1 WinRAR .....	9.4.2 电子商务 .....
8.2.2 WinZip .....	9.4.3 企业信息化 .....
8.3 通信软件 .....	9.4.4 数字城市 .....
8.3.1 MSN Messenger .....	9.4.5 数字地球 .....
8.3.2 NetMeeting .....	习题九.....
8.4 网关代理软件 WinGate .....	参考文献.....
习题八.....	

# 第 1 章 计算机基础知识

计算机无疑是人类社会 20 世纪最伟大的发明之一,它的出现使人类迅速进入了信息社会,彻底改变了人们社会文化生活,并且对人类的整个历史发展都有着不可估量的影响。在当今的信息社会中,计算机已经成为人们在社会生活中不可缺少的工具。

## 1.1 二进制与数据的编码

最初计算机的设计目的仅用于数值计算,后来发展成可以处理文字数据、声音数据、图像数据等。那么数据是怎样在计算机中保存的呢?

### 1.1.1 数的进制与转换

在日常生活中,人们习惯了十进制的记数方法,而在计算机中采用十进制的数据表示方式显然是不方便的。十进制的 10 个数码要用物质或物理上的十种状态表示出来。大家知道,数“量”的大小与表示它的进制无关。对于计算机中的数据也是这样,同一个数据,用一种与十进制相比更简单的进制表示,自然对数字的实际大小不会产生影响。

#### 1. 数的进制

数制( Numbering System )即表示数值的方法,有非进位数制和进位数制两种。表示数值的数码与它在数中的位置无关的数制称为非进位数制,如罗马数字就是典型的非进位数制。按进位的原则进行计数的数制称为进位数制,简称“进制”。对于任何进位数制,它有以下的基本特点:

① 数制的基数确定了所采用的进位计数制。表示一个数字时所用的数字符号的个数称为基数( Radix ),如十进制数制的基数为 10 ;二进制的基数为 2。对于  $N$  进位数制,有  $N$  个数字符号,如十进制中有 10 个数字符号 0 ~ 9 ;二进制有 2 个符号 0 和 1 ;八进制有 8 个符号 0 ~ 7 ;十六进制共有 16 个符号 0 ~ 9、A ~ F。

② 逢  $N$  进一。如十进制中逢 10 进 1 ;八进制中逢 8 进 1 ;二进制中逢 2 进 1 ;十六进制中逢 16 进 1。

四种常用进制表示如表 1.1 所示。

③ 采用位权表示法。任何一个  $r$  进制具有有限位小数的正数,都可以表示为:

$$( a_n a_{n-1} \dots a_1 a_0 . b_1 b_2 \dots b_{m-1} b_m )_r \quad (1-1)$$

其中  $a_i, b_j \in \{k | k=0, 1, \dots, r-1\}$  ;  $i=0, 1, 2, \dots, n$  ;  $j=1, 2, \dots, m$ 。

对于数字的  $n+1$  位整数部分,可以用以下的数学公式描述:

$$( a_n a_{n-1} \dots a_1 a_0 )_r = a_0 \times r^0 + a_1 \times r^1 + \dots + a_{n-1} \times r^{n-1} + a_n \times r^n = \sum_{i=0}^n a_i r^i \quad (1-2)$$

表 1.1 0~16 之间整数的四种常用进制表示

十进制	二进制	八进制	十六进制	十进制	二进制	八进制	十六进制
0	0	0	0	8	1000	10	8
1	1	1	1	9	1001	11	9
2	10	2	2	10	1010	12	A
3	11	3	3	11	1011	13	B
4	100	4	4	12	1100	14	C
5	101	5	5	13	1101	15	D
6	110	6	6	14	1110	16	E
7	111	7	7	15	1111	17	F

同理,对于数字的  $m$  位的小数部分,可以用以下的数学公式描述:

$$(0.b_1b_2\dots b_m)_r = b_1 \times r^{-1} + b_2 \times r^{-2} + \dots + b_m \times r^{-m} = \sum_{i=1}^m b_i r^{-i} \quad (1-3)$$

由以上公式可知,处在不同位置上的数码  $a_i$  和  $b_j$  所代表的值不同,一个数字在某个位置上所表示的实际数值等于该数值与这个位置的因子  $r^i$ 、 $r^{-j}$  的乘积,  $r^i$ 、 $r^{-j}$  由所在位置相对于小数点的距离  $i$ 、 $j$  来确定,简称为位权(Weight)。因此,任何进制的数字都可以写出按位权展开的多项式之和。

在数的各种进制中,二进制是最简单的一种计数进制,一是它的数码只有两个 0 和 1,在自然界中,具有两种状态的物质俯拾皆是,如电灯的“亮”与“灭”,电磁场的 N 极和 S 极等,如果用物质的这两种状态分别表示“0”和“1”,按照数位进制的规则,采用一组同类物质可以很容易地表示出一个数据;二是二进制的运算规则很简单:

$$0+0=0 \quad 0+1=1 \quad 1+1=10$$

这样的运算很容易实现,在电子电路中,只要用一些简单的逻辑运算元件就可以完成。因此,在计算机中数的表示全部用二进制,并采用二进制的运算规则完成数据间的计算。

尽管在计算机中数据一律用二进制表示,但是在数据的输入/输出、数据处理程序的编写中仍然大量地采用其他进制,如在屏幕上看到的数据及计算结果都是十进制数据。这是因为数据进制的转换工作已经由计算机代劳了。在应用计算机的过程中,不用考虑数据在机器内部的表示及底层的处理方式、处理过程。

在输入/输出数据时,可以用数据后加一个特定的字母来表示它所采用的进制:字母 D 表示数据为十进制(也可以省略);字母 B 表示数据为二进制;字母 O 表示数据为八进制;字母 H 表示数据为十六进制。例如,567.17D(十进制的 567.17)、110.11(十进制的 110.11,省略了字母 D)、110.11B(二进制的 110.11)、245O(八进制的 245)、234.5BH(十六进制的 234.5B)、234.5B(错误的表示方法)。

## 2. 不同进制数之间的转换

### (1) 二进制数转换为十进制数

根据公式(1-1)、(1-2)、(1-3),对于一个二进制数,如果希望求出它对应的十进制数,可以写出该数的位权展开式,从而很容易地算出它所对应的十进制数。例如:

$$11010101B = 1 \times 2^0 + 0 \times 2^1 + 1 \times 2^2 + 0 \times 2^3 + 1 \times 2^4 + 0 \times 2^5 + 1 \times 2^6 + 1 \times 2^7 = 213D$$

$$0.1101B = 1 \times 2^{-1} + 1 \times 2^{-2} + 0 \times 2^{-3} + 1 \times 2^{-4} = 0.8125D$$

## (2) 十进制数转换为二进制数

一个十进制数转换为二进制数,需要整数部分和小数部分分别转换:

① 对公式(1-2)稍做分析知,整数部分的转换可采用“除基数取余法”,即用基数2多次去除被转换的十进制数,记下余数的值,直到商为0。将每次所得到的余数按逆序排列,就是转换后的二进制数。

例 1.1  $158D = ?B$

2	158	0
2	79	1
2	39	1
2	19	1
2	9	1
2	4	0
2	2	0
2	1	1
	0	

得  $158D = 10011110B$

② 分析公式(1-3),小数部分的转换可采用“乘基数取整法”,即用基数2多次乘十进制的小数部分,每次相乘后取整数部分按正序排列,就是所对应的二进制数。

例 1.2  $0.8125D = ?B$

0.8125	0.6250	0.2500	0.5000
2	2	2	2
1.6250	1.2500	0.5000	1.0000
1	1	0	1

得  $0.8125D = 0.1101B$

对于十进制转换为八进制、十六进制的方法,与上述转换为二进制的方法相同。

## (3) 二进制转换为八进制、十六进制

大家知道  $8 = 2^3$ 、 $16 = 2^4$ ,也就是说,1个八进制位等于3个二进制位,1个十六进制位等于4个二进制位。因此,可以很容易实现二进制数与八进制数,二进制数与十六进制数之间的转换。

例 1.3  $1101001111001.11011B = ?O = ?H$

001	101	001	111	001	. 110	110
1	5	1	7	1	. 6	6

得  $1101001111001.11011B = 15171.66O$

0001	1010	0111	1001	. 1101	1000
1	A	7	9	. D	8

得  $1101001111001.11011B = 1A79.D8H$

从以上例子可以看到,从二进制转换成八进制(十六进制)的方法是,从小数点开始,整数部分向左每3位(4位)一组划分,当不足3位(4位)时在前面补0;小数部分向右每3位(4位)一组划分,不足3位(4位)时在后面补0;然后每一组再转换成一个8位(16位)数符

即可完成。读者自然地就会想出八进制(十六进制)数据转换为二进制数的方法。

### 1.1.2 原码、反码与补码

在计算机中数据的表示和运算都是以二进制的形式进行的。通常规定,一个数的最高位为符号位,用 0 表示正数,用 1 表示负数,称作数符。一个数在计算机内部表示称为机器数。机器数所真正表示的数值称为真值。机器数常采用原码、反码和补码表示法。

#### 1. 原码

原码是机器数的一种简单的表示法,其符号用 0 表示正号,用 1 表示负号,如以下的二进制数:

$$X_1 = +1010101, \quad X_2 = -1010101$$

其原码记作:

$$[X_1]_{\text{原}} = [+1010101]_{\text{原}} = 01010101$$

$$[X_2]_{\text{原}} = [-1010101]_{\text{原}} = 11010101$$

原码机器数的表示范围因字长而定,采用 8 位二进制原码表示时,其真值的表示范围为  $[-127, 127]$ ,即二进制的取值范围为  $[11111111, 01111111]$ 。应该注意的是:对数字 0 的表示有二种原码形式 00000000 和 10000000。

#### 2. 反码

机器数的反码可以由原码得到。如果机器数为正数,则该机器数的反码和原码相同;如果机器数为负数,则其反码是对原码除符号位以外的所有数位取反,如以下的二进制数:

$$X_1 = +1010101, \quad X_2 = -1010101$$

其反码记作:

$$[X_1]_{\text{反}} = [[+1010101]_{\text{原}}]_{\text{反}} = [01010101]_{\text{反}} = 01010101$$

$$[X_2]_{\text{反}} = [[-1010101]_{\text{原}}]_{\text{反}} = [11010101]_{\text{反}} = 10101010$$

#### 3. 补码

机器数的补码可以由原码得到,如果机器数是正数,则该机器数的补码与原码相同,如果机器数是负数,则该机器数的补码是对它的原码除符号位外的各位取反,并且在末位上加 1 得到的,如以下的二进制数:

$$X_1 = +1010101, \quad X_2 = -1010101$$

其补码记作:

$$[X_1]_{\text{补}} = [[+1010101]_{\text{原}}]_{\text{补}} = [01010101]_{\text{补}} = 01010101$$

$$[X_2]_{\text{补}} = [[-1010101]_{\text{原}}]_{\text{补}} = [11010101]_{\text{反}} + 1 = 10101010 + 1 = 10101011$$

机器数的补码表示范围因字长而定,采用 8 位二进制补码表示时,其真值的表示范围为  $[-128, 127]$ ,即二进制整数取补的取值范围为  $[10000000, 01111111]$ 。而对于数字 0 的补码表示只有一种形式 00000000。

**例 1.4** 已知  $X_1 = +1010101$  和  $X_2 = +0011101$ ,通过其补码表示法计算  $X_1 - X_2$  的值。

**解**  $X_1 - X_2 = X_1 + (-X_2)$

$$[X_1]_{\text{补}} = [[+1010101]_{\text{原}}]_{\text{补}} = [01010101]_{\text{补}} = 01010101$$

$$[-X_2]_{\text{补}} = [ [ -0011101 ]_{\text{原}} ]_{\text{补}} = [ 10011101 ]_{\text{反}} + 1 = 11100010 + 1 = 11100011$$

$$[X_1 - X_2]_{\text{补}} = [X_1]_{\text{补}} + [-X_2]_{\text{补}} = 01010101 + 11100011 = 00111000 \text{ (超出字长的进位丢弃)}$$

由  
知

$$[X_1 - X_2]_{\text{补}}]_{\text{原}} = [00111000]_{\text{原}} = 00111000$$

$$X_1 - X_2 = 00111000$$

从该例子可以看到,通过机器数的补码表示,可以将减法运算转化为加法运算来完成。因此在计算机运算器的设计中只考虑加法运算即可。

### 1.1.3 数的定点表示和浮点表示

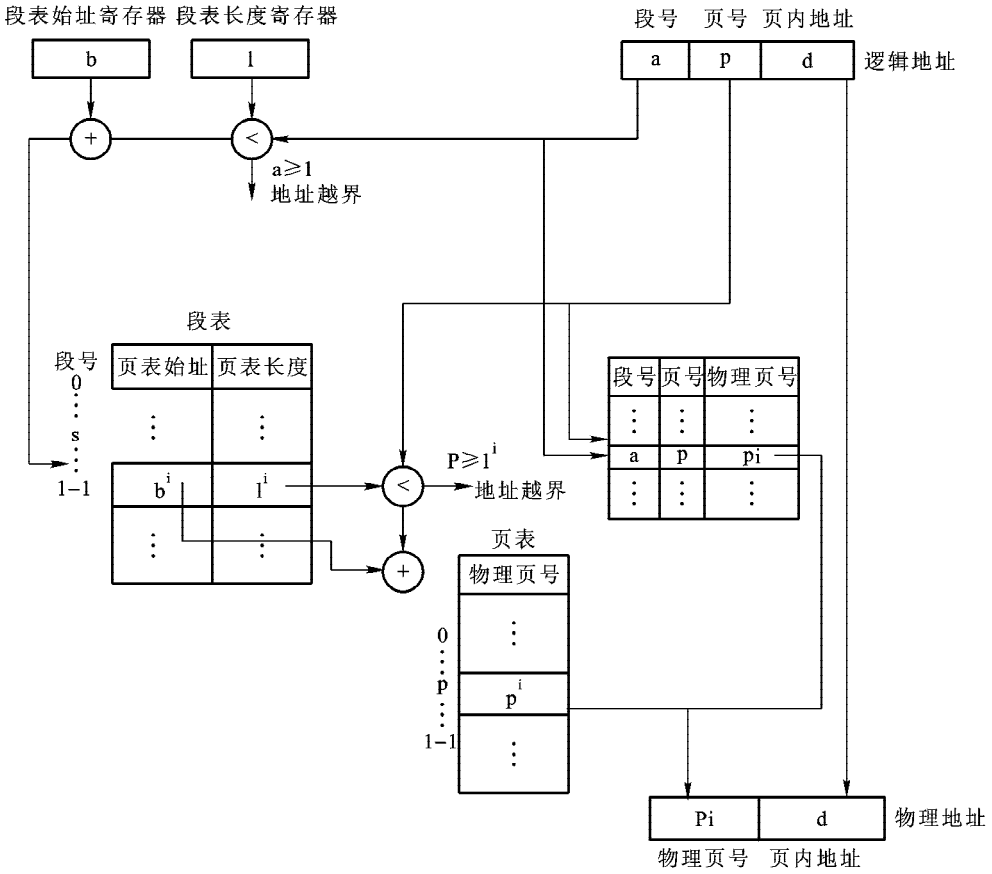
在计算机中,一般用若干个二进制位表示一个数或一条指令,把它们作为一个整体来处理、保存或传送。这样一个作为整体来处理的二进制字串称为计算机字。表示数据的字称为数据字,表示一条指令的字称为指令字。这个二进制字所占的位数称为字长。

字长的大小是由数据的处理设备和数据的类型所决定的,同类型的数据字长相同。一般来说,字长为8的倍数,如8位字长、16位字长、32位字长等。不同字长的数据字,取值范围差别很大,字长越大,它可表示的数的范围也越大。当表示的数超出其所能表示的范围时,将会发生“溢出”的错误,从而使数据的处理失败。

对于数值数据,有定点表示和浮点表示两种表示方法,采用定点表示的数称为定点数,采用浮点表示的数称为浮点数。

#### 1. 定点数

数的定点表示是指数据字中小数点的位置固定不变。一般用来表示一个纯小数(不含整数位的数或者整数)。当表示一个纯小数时,小数点固定在符号位之后。当表示一个整数时,小数点固定在数据字最后一位之后,如字长为16时,数据“ $-2^{-15}$ ”和“ $+32767$ ”表示如图1.1所示。



4-15

图 1.1 16 位字长数据字的定点表示

数的定点表示其数值的取值范围有限,为了扩大数的表示范围,也可以通过编程技术,采用多个字节表示一个定点数,如4字节或8字节等。

## 2. 浮点数

数的浮点表示法是指表示一个数时,其小数点的位置是浮动的,它实际上是数的指数记数法在计算机中的具体实现,如“123.456”可以写成“ $0.123456 \times 10^3$ ”、“ $-0.0000123456$ ”可以写成“ $-0.12345 \times 10^{-4}$ ”一样。从而解决定点表示中数的取值范围过窄的问题。

在数的浮点表示中,一个数由两部分组成:其一是阶码部分(表示数的指数记数法中的指数,记为 $E$ );其二是尾数部分(相当于指数记数法中的尾数,记为 $M$ ),因此对于一个数 $N$ ,通过浮点表示法可以表示(注意 $E$ 和 $M$ 中都包含有各自的符号位)为:

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

尾数 $M$ 的小数点位置位于尾数部分的数符位之后。 $M$ 为一纯小数,并且最高位从数据中第一个非零数位开始,阶码 $E$ 为一整数,如数据“0.00000111011”的 $M$ 值为“0.111011”;阶码 $E$ 为“-101”。

在计算机中保存一个浮点数时,阶码 $E$ 的长度和尾数 $M$ 的长度都是固定的。当数的指数位数小于阶码长度减1时在前面补零;当数的尾数位长度小于 $M$ 的长度减1时在后面补零。反之,如果数的尾数位长度大于 $M$ 的长度减1,则多出的位自动丢弃;如果数的指数位数大于阶码长度减1,则数的大小超出了浮点表示的范围,发生“溢出”错误。

假设机器字长为32位,则其阶码占8位,尾数占24位,各部分的规定如图1.2所示。

### 1.1.4 数据的存储单位

在计算机中,数据存储的最小单位为比特(bit),1比特为1个二进制位。

由于1比特太小,无法用来表示出数据的信息含义,所以又引入了“字节”(Byte, B;注意这里B作为数据量大小的单位,不要和数的表示中表示为二进制数的B混淆)作为数据存储的基本单位。在计算机中规定,1个字节为8个二进制位。除字节外,还有千字节(KB)、兆字节(MB)、吉字节(GB)、太字节(TB),它们的换算关系是:

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

$$1 \text{ MB} = 1024 \text{ KB} = 1048576 \text{ B} = 2^{20} \text{ B}$$

$$1 \text{ GB} = 1024 \text{ MB} = 1048576 \text{ KB} = 1073741824 \text{ B} = 2^{30} \text{ B}$$

$$1 \text{ TB} = 1024 \text{ GB} = 2^{40} \text{ B}$$

### 1.1.5 字符编码

计算机除进行数值计算以外,大多还是进行各种数据的处理,其中字符数据的处理占有相当大的比重。从键盘或其他途径输入的字符数据在计算机中怎样保存就成了问题。人们所看到的字符,实际上是一个个的图形符号,直接保存这些图形符号不但要占用大量的存储空间,而且也给数据的处理带来很大的麻烦。人们自然想到了为字符编码的方法,这样既可以节省存储空间,数据处理的过程也很容易完成。

在日常处理的字符数据中,有西文字符和中文字符两种,因两种字符本身的差别所致,编码的方法也不相同。

## 1. ASCII 码

在计算机中,最常用的是英文字符,它的编码为 ASCII 码(American National Standard Code for Information Interchange,美国国家信息交换标准码),如表 1.2 所示,它原为美国的国家标准,1967 年确定为国际标准。在 ASCII 码中,用 7 个二进制位表示 1 个字符,共可以表示 128 个字符,其中 95 个可打印或显示的字符,其他的则为不可打印或显示的字符。在 ASCII 码的应用中,也经常用十进制表示,如空格 32,数字 0~9 48~57,大写字母 A~Z 65~90,小写字母 a~z 97~122。这样,英文中的每一个字符都有一个固定的编码,保存字符时只需保存它的 ASCII 码即可。

从表 1.2 中可以看到,一个 ASCII 码的长度不超过 8 个二进制位。因此,保存一个 ASCII 码只需一个字节。由于一个字节的内内容可以用一个 2 位的十六进制数来表示,所以在书写字符的 ASCII 码时,也常使用十六进制,如 20H 为空格,41H 为字母 A 的 ASCII 码。读者会注意到,ASCII 码只占了一个字节中低端的 7 位,最高位(第 8 位)为 0。

表 1.2 ASCII 码表

高4位 低4位	0000	0001	0010	0011	0100	0101	0110	0111
0000	NULL	DLE	空格	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	BELL	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. 汉字的编码

汉字与西文字符相比,其特点是量多而且字形复杂,这两个问题的解决,也是依靠对汉字的编码来实现的。下面看汉字的编码问题。

### (1) 区位码

为了解决汉字的编码问题,1980 年我国公布了 GB2312-80 国家标准,在此标准中共含有 6 763 个简化汉字和 682 个汉字符号。在该标准的汉字编码表中,汉字和符号按区位排列,共分成了 94 个区,每个区有 94 个位,例如,在 16 区中所有汉字的排列如表 1.3 所示。

一个汉字的编码由它所在的区号和位号组成,称为区位码,如“啊”字区位码为“1601”,

“白”的区位码是“1655”。

区位码中规定,1~15区(其中有些区没有被使用)为汉字符号区;16~94区为汉字区。在汉字区中,根据汉字的使用频度分成了两级:使用频度最高的汉字共3755个,为一级汉字,按汉语拼音排序,占用了16~55区;其余的3008个汉字为二级汉字,按部首排序,占用了56~87区。这样在区位码表中,包括了使用频度占99.99%的汉字。在区位码表的汉字符号区,包括西文字母、日文假名和片假名、俄文字母、数字、制表符及一些特殊的图形符号。

表 1.3 第 16 区汉字区位码表

	0	1	2	3	4	5	6	7	8	9
0		啊	阿	埃	挨	哎	唉	哀	皑	癌
1	蔼	矮	艾	碍	爱	隘	鞍	氨	安	俺
2	按	暗	岸	胺	案	肮	昂	盎	凹	敖
3	熬	翱	袄	傲	奥	懊	澳	芭	捌	扒
4	叭	吧	笆	八	疤	巴	拨	跋	靶	把
5	耙	坝	霸	罢	爸	白	柏	百	摆	佰
6	败	拜	裨	斑	班	搬	扳	般	颁	板
7	版	扮	拌	伴	瓣	半	办	绊	邦	帮
8	梆	榜	膀	绑	棒	磅	蚌	镑	傍	谤
9	苞	胞	包	褒	剥					

## (2) 汉字的机内码

保存一个汉字的区位码要占用两个字节,区号、位号各占一个字节。区号、位号都不超过94,所以这两个字节的最高位仍然是“0”。为了避免汉字区位与ASCII码无法区分,汉字在计算机内的保存采用了机内码,也称汉字的内码。目前占主导地位的汉字机内码是将区码和位码分别加上数A0H作为机内码。如“啊”字的区位码的十六进制表示为1001H,而“啊”字的机内码则为B0A1H。这样汉字机内码的两个字节的最高位均为“1”,很容易与西文的ASCII码区分。以GB2312-80国家标准制定的汉字机内码也称为GB2312码,它和国标区位码的换算关系是:

$$\text{机内码} = \text{区位码} + \text{A0A0H}$$

像英文字符一样,汉字在排序时所依据的大小关系也是根据它的编码大小来确定的,即在不同区里的汉字由机内码的第一字节决定大小,在同一区中的汉字则由第二字节的大小来决定。由于汉字的内码都大于128,所以汉字无论是高位内码还是低位内码都是大于ASCII码的(仅对GB2312码而言)。

需要说明的是,在我国的台湾省,目前广泛使用的是“大五码(BIG-5)”,对于这种内码一个汉字也是用两个字节表示,共可以表示13053个汉字。

## (3) 汉字输入码

由于汉字具有字量大、同音字多的特点,怎样实现汉字的快速输入也是应解决的重要问题之一。为此,不少个人或团体发明了多种多样的汉字输入方法,如全拼输入法、双拼输入法、智能ABC输入法、表形码输入法、五笔字型输入法等。对于任何一种汉字输入法,都有一套对汉字的编码,称为汉字输入码。汉字输入码实际上是输入汉字时所使用的代码,它按