

高等院校计算机应用系列教材

C 语言程序设计

(微课版)

焉德军 辛慧杰 编著

清华大学出版社

北 京

内 容 简 介

本书以全国计算机等级考试二级考试大纲为指导,介绍了二级 C 语言等级考试所要求掌握的全部内容。全书共分为 11 章,包括计算机基础知识、C 语言概述、基本数据类型与常用库函数、运算符和表达式、C 语言的控制结构、数组、函数、编译预处理、指针、结构体与共用体、文件等内容。

本书例题丰富,习题解答详细,注重利用 C 语言解决实际问题能力的培养,既可作为高等院校 C 语言程序设计课程的教学用书,也可作为全国计算机等级考试(NCRE)备考人员的参考资料。

本书配套的电子课件、习题答案和实例源文件可以到 <http://www.tupwk.com.cn/downpage> 网站下载,也可以扫描前言中的二维码获取。扫描前言中的“看视频”二维码可以直接观看教学视频。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。举报:010-62782989, beiqinquan@tup.tsinghua.edu.cn

图书在版编目(CIP)数据

C 语言程序设计:微课版 / 焉德军, 辛慧杰编著.

北京:清华大学出版社, 2024. 7. -- (高等院校计算机应用系列教材). -- ISBN 978-7-302-66483-3

I. TP312.8

中国国家版本馆 CIP 数据核字第 2024XM6504 号

责任编辑:胡辰浩

封面设计:高娟妮

版式设计:茆博文化

责任校对:马遥遥

责任印制:丛怀宇

出版发行:清华大学出版社

网 址: <https://www.tup.com.cn>, <https://www.wqxuetang.com>

地 址:北京清华大学学研大厦 A 座 邮 编:100084

社总机:010-83470000 邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

印 装 者:北京联兴盛业印刷股份有限公司

经 销:全国新华书店

开 本:185mm×260mm 印 张:17.75 字 数:455 千字

版 次:2024 年 8 月第 1 版 印 次:2024 年 8 月第 1 次印刷

定 价:69.00 元

产品编号:106157-01

前言

C 语言是一种非常优秀的程序设计语言，既具备高级语言的特点，又具有直接操纵计算机硬件的能力，并因丰富灵活的控制性和数据结构、简洁而高效的语句表达、清晰的程序结构和良好的可移植性而拥有大量的使用者。目前，国内几乎所有的高等院校都开设了“C 语言程序设计”课程。人们对计算机知识的需求使得 C 语言不仅成为计算机专业学生的必修课，也成为广大非计算机专业学生和计算机爱好者首选的程序设计语言，而且全国计算机等级考试也将 C 语言列为重要的考试内容之一。

本书面向程序设计新手，在编写过程中，作者力求使本书体现如下特点：

(1) 对 C 语言中的重点、难点进行分解，并将其分散编排，使学生的学习过程中能够循序渐进掌握；

(2) 对 C 语言中生僻、不常用的内容不做过多描述；对在实践中使用较多、需要牢固掌握的内容进行详细叙述，并给出大量示例；

(3) 本书在介绍 C 语言基本知识的同时，还有意识地训练读者逐步养成良好的程序编写习惯和程序设计风格。

全书共分为 11 章。第 1 章介绍计算机系统与工作原理、数制与编码；第 2 章介绍 C 程序的基本构成和简单的 C 程序示例；第 3 章介绍基本数据类型、常用的输入/输出函数；第 4 章介绍各种运算符和表达式；第 5 章介绍 C 语言的控制结构；第 6 章介绍数组的使用及常用字符串处理函数；第 7 章介绍函数的使用及变量的存储类别；第 8 章介绍编译预处理命令；第 9 章介绍指针的使用方法；第 10 章介绍结构体及链表；第 11 章介绍文件的类型和操作。

为了方便读者学习，作者录制了相应的教学视频，并对书中习题做了详细解答。

本书作者长期从事高等学校“C 语言程序设计”课程的教学工作，在总结多年教学经验的基础上，编写了本教材。由于作者水平有限，书中难免存在不足，恳请读者批评指正。我们的邮箱是 992116@qq.com，电话是 010-62796045。

本书配套的电子课件、实例源文件和习题答案可以通过 <http://www.tupwk.com.cn/downpage> 网站下载，也可以扫描下方二维码获取。扫描下方二维码可以直接观看教学视频。

扫描下载



配套资源

扫一扫



看视频

编者
2024 年 4 月

目 录

第 1 章 计算机基础知识	1
1.1 计算机系统与工作原理	1
1.1.1 计算机系统的组成	1
1.1.2 计算机硬件系统	2
1.1.3 计算机软件系统	4
1.1.4 计算机的工作原理	6
1.2 数制与编码	7
1.2.1 数制的基本概念	7
1.2.2 常用的数制	7
1.2.3 数制间的转换	11
1.2.4 数据在计算机中的表示方式	13
1.2.5 字符编码	16
1.2.6 汉字编码	18
1.2.7 存储单位	21
1.3 个人计算机的基本配置及性能指标	22
1.3.1 硬件	22
1.3.2 操作系统	24
1.4 习题	25
第 2 章 C 语言概述	28
2.1 C 语言的发展历史	28
2.2 C 语言的特点	29
2.3 C 语言程序设计初步	30
2.3.1 简单的 C 程序示例	30
2.3.2 C 程序结构	32
2.4 C 程序在计算机上的执行步骤	33
2.5 习题	34
第 3 章 基本数据类型与常用库函数	36
3.1 字符集与标识符	36

3.2 数据类型与基本数据类型	37
3.2.1 数据类型	37
3.2.2 基本数据类型	38
3.3 常量与变量	39
3.3.1 常量	39
3.3.2 符号常量	41
3.3.3 变量	41
3.4 输入/输出函数	42
3.4.1 字符输入/输出函数	42
3.4.2 格式输出函数 printf	43
3.4.3 格式输入函数 scanf	46
3.5 常用库函数	48
3.6 习题	49
第 4 章 运算符和表达式	52
4.1 算术运算符和算术表达式	52
4.2 自增和自减运算	55
4.3 位运算	55
4.4 赋值运算	59
4.5 关系运算与逻辑运算	60
4.5.1 关系运算	60
4.5.2 逻辑运算	62
4.6 条件表达式与逗号表达式	63
4.6.1 条件表达式	63
4.6.2 逗号表达式	64
4.7 习题	65
第 5 章 C 语言的控制结构	67
5.1 算法与程序	67
5.1.1 算法	67
5.1.2 程序	68
5.2 顺序结构	69

5.3 选择结构	69	7.1.4 函数声明	118
5.3.1 if语句	70	7.1.5 函数间的参数传递	119
5.3.2 switch语句	74	7.2 函数的嵌套调用和递归调用	121
5.4 循环结构	78	7.2.1 函数的嵌套调用	121
5.4.1 while语句	78	7.2.2 函数的递归调用	122
5.4.2 do-while语句	80	7.3 局部变量和全局变量	126
5.4.3 for语句	81	7.3.1 局部变量	126
5.4.4 break语句	82	7.3.2 全局变量	126
5.4.5 continue语句	83	7.4 变量的存储类别	129
5.4.6 单重循环结构	84	7.4.1 自动变量	129
5.4.7 多重循环结构	86	7.4.2 静态变量	130
5.5 习题	88	7.4.3 寄存器变量	131
第6章 数组	92	7.5 内部函数和外部函数	131
6.1 一维数组	92	7.6 习题	132
6.1.1 一维数组的定义	92	第8章 编译预处理	137
6.1.2 一维数组的存储形式	93	8.1 宏定义	137
6.1.3 一维数组的初始化	93	8.1.1 不带参数的宏定义	137
6.1.4 一维数组的引用	94	8.1.2 带参数的宏定义	138
6.1.5 一维数组应用举例	94	8.1.3 终止宏定义的作用域	140
6.2 二维数组	98	8.2 文件包含	140
6.2.1 二维数组的定义	98	8.3 条件编译	141
6.2.2 二维数组的存储形式	99	8.4 习题	143
6.2.3 二维数组的初始化	99	第9章 指针	145
6.2.4 二维数组的引用	99	9.1 指针概述	145
6.2.5 二维数组应用举例	100	9.1.1 指针的概念	145
6.3 字符数组	102	9.1.2 指针变量	146
6.3.1 字符数组的定义	102	9.1.3 指针运算	148
6.3.2 字符数组的初始化	102	9.1.4 将指针作为函数的参数	149
6.3.3 字符数组的输入/输出	103	9.2 指针与一维数组	151
6.3.4 常用字符串处理函数	104	9.2.1 一维数组元素的表示	151
6.3.5 二维字符数组	107	9.2.2 将数组名作为函数的参数	153
6.3.6 字符串应用举例	110	9.3 指针与二维数组	156
6.4 习题	111	9.3.1 二维数组中的指针	156
第7章 函数	114	9.3.2 指向二维数组元素的指针变量	157
7.1 函数定义和函数调用	114	9.3.3 将二维数组的行指针作为函数的参数	159
7.1.1 函数定义	114	9.4 指针与字符串	161
7.1.2 return语句	116	9.4.1 字符串的表示形式	161
7.1.3 函数调用	117		

9.4.2 将字符串指针作为函数的参数	163	10.10 习题	199
9.5 指针与函数	164	第 11 章 文件	204
9.5.1 指向函数的指针	164	11.1 文件的基本概念	204
9.5.2 返回指针的函数	165	11.2 文件的打开与关闭	205
9.6 指针数组与多级指针	166	11.2.1 文件类型的指针变量	205
9.6.1 指针数组的概念	166	11.2.2 文件的打开	205
9.6.2 多级指针的概念	167	11.2.3 文件的关闭	206
9.7 习题	168	11.3 文件的读写	207
第 10 章 结构体与共用体	173	11.3.1 字符读写函数fgetc和fputc	207
10.1 结构体类型的定义	173	11.3.2 字符串读写函数fgets和fputs	209
10.2 结构体变量	174	11.3.3 数据块读写函数fread和fwrite	210
10.2.1 结构体变量的定义	174	11.3.4 格式化读写函数fscanf和fprintf	212
10.2.2 结构体变量的引用	176	11.4 文件的随机读写	213
10.3 结构体数组	177	11.4.1 文件的定位	213
10.3.1 结构体数组的定义与初始化	177	11.4.2 进行文件的随机读写	214
10.3.2 结构体数组的引用	178	11.5 习题	215
10.4 结构体和指针	179	参考文献	217
10.5 结构体和函数	181	附录 A C 语言中的关键字	218
10.5.1 结构体作为函数的参数	181	附录 B 字符与 ASCII 码对照表	219
10.5.2 返回结构体的函数	183	附录 C 运算符的优先级与结合性	221
10.6 链表	185	附录 D 常用库函数	223
10.6.1 简单链表	185	附录 E 习题参考解答	229
10.6.2 处理动态链表所需的库函数	186	附录 F 微课视频目录	273
10.6.3 单向链表的基本操作	188		
10.7 共用体	195		
10.8 枚举	197		
10.9 使用typedef定义类型别名	198		

计算机基础知识

20 世纪 40 年代诞生的电子数字计算机是 20 世纪最伟大的发明之一，是人类科学技术发展史上的一个里程碑。半个多世纪以来，计算机科学技术飞速发展，计算机的性能越来越卓越、价格越来越便宜、应用越来越广泛。时至今日，计算机已经被广泛应用于国民经济以及社会生活的各个领域，特别是互联网技术的发展和智能手机的普及，极大地改变了人们的生活方式，并对人们的生产、生活产生越来越大的影响。计算机科学技术的发展水平、计算机的应用程度已经成为衡量一个国家现代化水平的重要标志。

本章介绍计算机的基础知识，包括计算机系统的组成与工作原理、数据在计算机中的表示，以及个人计算机的基本配置及性能指标。

1.1 计算机系统与工作原理

计算机在诞生初期主要用于进行科学计算，因此被称为“计算机”。然而，现在计算机的处理对象已经远远超出“计算”的范围，可以对数字、文字、声音及图像等各种形式的数据进行处理。实际上，计算机是一种能够按照事先存储的程序，自动、高速地对数据进行输入、处理、输出和存储的系统。本节介绍计算机系统的组成和工作原理。

1.1.1 计算机系统的组成

完整的计算机系统包括硬件系统和软件系统两部分，如图 1-1 所示。硬件是指肉眼看得见的机器部件，就像计算机的“躯体”。人们通常看到的计算机都配有机箱，里边是各式各样的电子元器件，机箱外部还有键盘、鼠标、显示器和打印机等设备，它们是计算机工作的物质基础。不同种类的计算机其硬件组成各不相同，但无论什么类型的计算机，都可以将其硬件划分为功能相近的几大部分。软件则是计算机的“灵魂”，是程序、数据及相关文档的总称。程序是由一系列指令组成的，每条指令都能指挥机器完成相应的操作。当程序执行时，其中的各条指令就依次发挥作用，指挥机器按指定顺序完成特定任务，并把执行结果按照各种格式输出。

计算机系统是一个整体，既包括硬件也包括软件，两者缺一不可。计算机若没有软件的支持，也就是在没有装入任何程序之前，被称为“裸机”，裸机是无法完成任何处理任务的。反之，若没有硬件设备的支持，单靠软件本身，软件也就失去了其发挥作用的物质基础。计算机系统的软硬件系统相辅相成，共同完成任务。

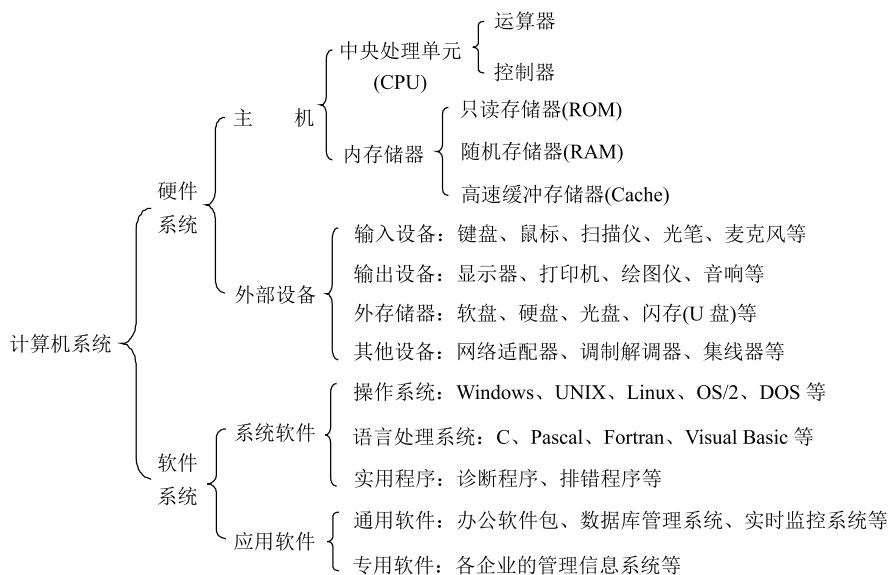


图 1-1 计算机系统的组成结构

1.1.2 计算机硬件系统

计算机硬件(computer hardware)又称硬件平台,是指计算机系统所包含的各种机械的、电子的、磁性的装置和设备。每个功能部件各尽其职、协同工作,缺少其中任何一个就不能构成完整的计算机系统。

计算机能够处理存储的数据。可以说,存储和处理是一个整体:存储是为了处理,处理需要存储。“存储和处理的整体性”的最初表达是美国普林斯顿大学的冯·诺依曼于 1946 年提出的计算机体系结构思想,一般称为“存储程序思想”。计算机从 1946 年问世至今都以这种思想为基本依据,主要特点可归结为以下 3 点。

- (1) 计算机由 5 个基本部分组成:运算器、控制器、存储器、输入设备和输出设备。
- (2) 程序和数据存放在存储器中并按地址访问。
- (3) 程序和数据用二进制表示。与十进制相比,实现二进制运算的结构更简单,更容易控制。

如今,半个多世纪过去了,计算机的系统结构发生了很大改变,就结构原理来说,仍然是冯·诺依曼型计算机,结构如图 1-2 所示,图中带箭头的实线表示数据流,带箭头的虚线表示控制流。

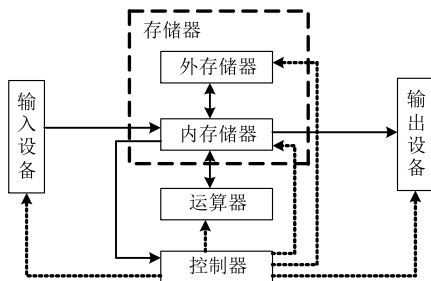


图 1-2 冯·诺依曼型计算机的结构

硬件是计算机系统的物质基础。计算机的性能,如运算速度、存储容量、计算精度、可靠性等在很大程度上取决于硬件的配置。下面简单介绍计算机的 5 个基本组成部分。

1. 运算器

计算机最主要的工作是运算,大量的数据运算任务是在运算器中进行的。

运算器由算术逻辑单元(Arithmetic Logic Unit, ALU)和寄存器等组成。算术逻辑单元的主要功能是进行算术运算和逻辑运算。算术运算是指加、减、乘、除等基本运算;逻辑运算是指逻辑判断、关系比较,以及与、或、非等其他的基本逻辑运算。但不管是算术运算还是逻辑运算,都只是基本运算。也就是说,运算器只能做这些最简单的运算,复杂的运算都要通过基本运算一步步实现。然而,运算器的运算速度却快得惊人,因而计算机才具有高速的信息处理能力。运算器中的寄存器用来暂存参与运算的操作数和中间结果,常用的寄存器有累加寄存器、暂存寄存器、标志寄存器和通用寄存器等。

运算器的主要技术指标是运算速度,单位是 MIPS(百万指令每秒)。由于执行不同的指令所花费的时间不同,因此一台计算机的运算速度通常是按照一定的频率执行各类指令的统计值。微型计算机一般采用主频来描述运算速度,主频越高,运算速度就越快。目前,个人计算机的运算速度已达数十亿次每秒,而超级计算机的运算速度通常以万亿次每秒计算,例如,我国自行研制的超级计算机“天河一号”,其系统峰值性能为 1206 万亿次双精度浮点运算每秒。

2. 控制器

控制器是计算机的神经中枢和指挥中心,只有在它的控制之下整个计算机才能有条不紊地工作,自动执行程序。控制器的主要特点是采用内存程序控制方式。也就是说,在使用计算机时必须预先编写(或由编译程序自动生成)由计算机指令组成的程序并存入内存储器,由控制器依次读取并执行。

控制器由程序计数器(PC)、指令寄存器(IR)、指令译码器(ID)、时序控制电路及微操作控制电路等组成。其中:程序计数器用来对程序中的指令进行计数,使得控制器能够依次读取指令;指令寄存器在指令执行期间暂存正在执行的指令;指令译码器用来识别指令的功能,分析指令的操作要求;时序控制电路用来生成时序信号,以协调指令执行周期内各部件的工作;微操作控制电路用来产生各种操作控制命令。

运算器和控制器合称为中央处理单元(Central Processing Unit, CPU)。

3. 存储器

存储器的主要功能是存放程序和数据。使用时,可以从存储器中取出信息,不破坏原有的内容,这种操作称为存储器的读操作;也可以把信息写入存储器,原来的内容被抹掉,这种操作称为存储器的写操作。

存储器分为程序存储区、数据存储区和栈。程序存储区存放程序中的指令,数据存储区存放数据。CPU 通过地址总线发出相应的地址,找到存储器中与该地址对应的存储单元,然后通过数据总线操作存储单元中的数据。

存储器通常分为内存储器和外存储器。

1) 内存存储器

内存存储器简称内存(又称主存),是计算机中信息交流的中心。用户通过输入设备输入的程序和数据最初被送入内存,控制器执行的指令和运算器处理的数据取自内存,运算的中间结果和最终结果保存在内存中,输出设备输出的信息来自内存,内存中的信息如要长期保存,则应送到外存储器中。总之,内存要与计算机的各个部件打交道,进行数据交换。因此,内存的存取速度直接影响计算机的运算速度。

2) 外存储器

外存储器设置在主机的外部,简称外存(又称辅存),主要用来长期存放暂时不用的程序和数据。通常,外存不和计算机的其他部件直接交换数据,只和内存交换数据,而且不是按单个数据进行存取,而是成批地进行数据交换。常用的外存有磁盘、磁带和光盘等。

4. 输入设备

输入设备用于接收用户输入的原始数据和程序,并将它们转换为计算机可以识别的形式(二进制代码)后存放到内存中。常用的输入设备有键盘、鼠标、扫描仪、光笔、数字化仪和麦克风等。

5. 输出设备

输出设备用于将存放在内存中的由计算机处理的结果转换为人们所能接受的形式。常用的输出设备有显示器、打印机、绘图仪和音响等。

1.1.3 计算机软件系统

计算机软件(computer software)是相对于计算机硬件而言的,包括计算机运行所需的各种程序、数据,以及相关的技术文档资料。在裸机上只能运行机器语言程序,使用很不方便,效率也低。硬件是软件赖以运行的物质基础,软件是计算机的灵魂,是发挥计算机功能的关键。

计算机软件通常可分为系统软件和应用软件两大类。用户与计算机系统各层次之间的关系如图 1-3 所示。

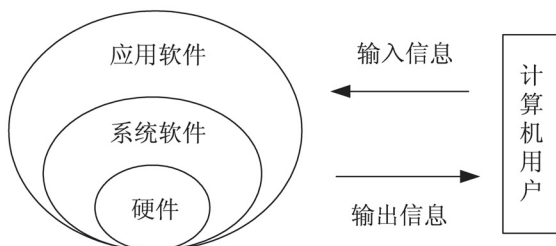


图 1-3 用户与计算机系统各层次之间的关系

1. 系统软件

系统软件是管理、监控和维护计算机资源的软件,用来扩展计算机的功能,提高计算机的工作效率,方便用户使用计算机的软件,包括操作系统、语言处理系统、数据库管理系统、系统辅助处理程序等。

1) 操作系统

在计算机软件中,最重要且最基本的就是操作系统(Operating System, OS)。操作系统是最底层的软件,控制在计算机上运行的所有程序并管理整个计算机的资源,在裸机与应用程序及用户之间架起了一座沟通的桥梁。没有操作系统,用户就无法自如地应用各种软件或程序。

目前,常见的计算机操作系统有 Windows、UNIX、Linux 和 macOS 等。

2) 语言处理系统

计算机语言大致分为机器语言、汇编语言和高级语言。

(1) 机器语言:机器语言是以二进制代码表示的指令集合,是计算机能直接识别和执行的计算机语言。其优点是执行效率高、速度快;但直观性差,可读性不强,因而给计算机的推广和使用带来了极大的困难。

(2) 汇编语言:汇编语言是符号化的机器语言,这种语言使用助记符来表示指令中的操作码和操作数。汇编语言相比机器语言前进了一步,助记符比较容易记忆,可读性也较强,但编写程序的效率不高、难度较大、维护较困难,属于低级语言。

(3) 高级语言:高级语言就是接近人类自然语言和数学语言的计算机语言,是第三代计算机语言。高级语言的特点是与计算机的指令系统无关,因而从根本上摆脱了对机器的依赖,使之独立于机器。该语言面向过程,进而面向用户。由于易学易记、便于书写和维护,因此提高了程序设计的效率和可靠性。目前广泛使用的高级语言有 Java、C、C++、PHP 和 Python 等。

将那些计算机无法直接执行的使用非机器语言编写的程序,翻译成计算机能直接执行的机器语言的翻译程序,我们称之为语言处理程序。

使用汇编语言和高级语言编写的程序称为源程序,计算机不能直接识别和执行源程序。把计算机本身不能直接读懂的源程序翻译成机器能够识别的机器指令代码后,计算机才能执行它们,这种翻译后的程序称为目标程序。

计算机将源程序翻译成机器指令的方法有两种:编译方式和解释方式。编译方式与解释方式的工作过程如图 1-4 所示。

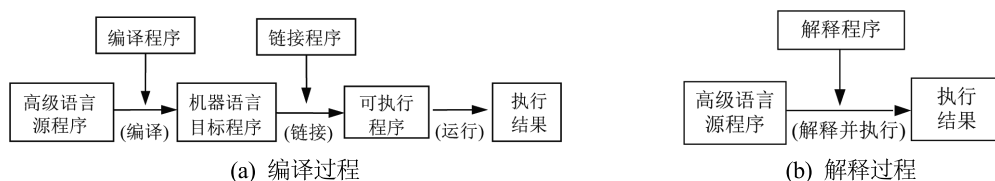


图 1-4 将源程序翻译成机器指令的过程

由图 1-4 可以看出,编译方式会使用相应的编译程序把源程序翻译成机器语言的目标程序,然后链接成可执行程序,运行可执行程序后得到结果。目标程序和可执行程序都以文件方式存放在磁盘上,再次运行程序时,只需要直接运行可执行程序,而不必重新编译和链接。

解释方式则把源程序输入计算机,然后使用相应的解释程序对代码进行逐条解释、逐条执行,执行后只能得到结果,而不能保存解释后的机器代码,下次运行程序时需要重新解释并执行。

3) 数据库管理系统

数据库管理系统是能够对数据进行加工、管理的系统软件。常见的数据库管理系统有 Visual FoxPro、Oracle、Access 和 SQL Server 等。

4) 系统辅助处理程序

系统辅助处理程序又称“软件研制开发工具”“支持软件”或“工具软件”，主要包括编辑程序、调试程序、装配和连接程序以及测试程序等。

2. 应用软件

应用软件是用户利用计算机及系统软件，为解决实际问题而开发的软件的总称。应用软件一般分为两大类：通用软件和专用软件。

- 通用软件支持最基本的应用，如文字处理软件(Word)、表格处理软件(Excel)等。
- 专用软件是专门为某一专业领域开发的软件，如财务管理系统、计算机辅助设计(CAD)软件和仅限于某个部门使用的应用数据库管理系统等。

1.1.4 计算机的工作原理

如前所述，计算机能自动、连续地工作，主要是因为内存中存储了程序。计算机在执行程序时，在控制器的控制下，逐条从内存中取出指令、分析指令、执行指令，从而完成相应的操作。

1. 指令和程序

1) 指令

指令(instruction)是控制计算机操作的代码，又称指令码(instruction code)。一条指令就是给计算机下达的一道命令，用于告诉计算机执行什么样的操作、参与此项操作的数据来自何处、操作结果又将送往哪里。所以，一条指令包括操作码(operation code)和操作数两部分，操作码控制执行何种操作，操作数指出参与操作的数据或数据存放的位置。

通常，一台计算机能够完成多种操作，也就是执行多条指令，计算机所能执行的所有指令的集合称为计算机的指令系统(instruction system)。由于不同计算机的硬件结构不同，因而它们的指令系统也不同。计算机指令系统在很大程度上决定了计算机的处理能力，是衡量计算机性能的重要指标之一。

2) 程序

程序就是为完成某项任务而由若干指令组成的有序集合。编写程序称为程序设计。人们通过编写程序，发挥计算机的作用，从而解决工作和生活中的各种问题。那么计算机是怎样执行程序的呢？

2. 指令和程序在计算机中的执行过程

为了执行程序，首先需要将程序和程序所要操作的数据放入内存。执行程序就是依次执行组成程序的一条条指令。计算机在执行程序中的每条指令时，先将需要执行的指令从内存中取出，之后放到 CPU 内并通过控制器对这些指令进行译码分析，判断这些指令所要完成的操作，随后向相关部件发出控制信号以完成这些操作。由此可见，计算机执行程序的过程，就是在控

制器的控制下，逐条从内存中读取程序中的指令并进行分析，而后向相关部件发出控制信息以执行指令，从而实现程序的运行。

1.2 数制与编码

在计算机系统中，数字和符号都是使用电子元件的不同状态来表示的，也就是以电信号进行表示。根据计算机的这一特点，我们会提出下面这个问题：数值在计算机中是如何表示和运算的？这就是本节将要讨论的“数制”问题。

1.2.1 数制的基本概念

通过一组固定的数字(数码符号)和一套统一的规则来表示数值的方法称为数制(number system)，也称为计数制。数制的种类很多，除了十进制，还有二十四进制(24 小时为一天)、六十进制(60 秒为 1 分钟、60 分钟为 1 小时)、二进制(手套、筷子等两只/支为一双)等。

计算机系统采用二进制的主要原因是电路设计简单、运算方便、工作可靠且逻辑性强。

不论是哪一种数制，它们的计数和运算都有着一些共同的规律和特点。

1. 逢 R 进一

R 是数制中所需数字字符的总数，称为基数(radix)。例如，十进制使用 0、1、2、3、4、5、6、7、8、9 这 10 个不同的数制符号来表示数值。在十进制中，基数是 10，表示逢十进一。

2. 位权表示法

位权(简称权)是指一个数字在某个位置上所代表的值，处在不同位置的数字所代表的值不同，每个数字的位置决定了这个数字的值或位权。例如，在十进制数 586 中，5 的位权是 100(即 10^2)。

位权与基数的关系是：位权的值是基数的若干次幂。因此，使用任何一种数制表示的数都可以写成按位权展开的多项式之和。例如，十进制数 256.07 可以用如下形式表示：

$$(256.07)_{10}=2\times 10^2+5\times 10^1+6\times 10^0+0\times 10^{-1}+7\times 10^{-2}$$

位权表示法的原则是数字的总数等于基数，每个数字都要乘以基数的幂次，幂次是由每个数字所在的位置决定的。排列方式是以小数点为界，整数自右向左依次为 0 次方、1 次方、2 次方，以此类推；小数自左向右依次为负 1 次方、负 2 次方，以此类推。

1.2.2 常用的数制

在使用计算机解决实际问题的过程中，我们往往使用的是十进制数，而计算机内部使用的是二进制数。另外，在计算机应用中又经常需要使用十六进制数或八进制数，二进制数因为与十六进制数和八进制数正好有倍数关系，如 2^3 等于 8、 2^4 等于 16，所以十分便于在计算机应用中表示其他进制数。

1. 十进制数(decimal)

按“逢十进一”的原则进行计数，每一位计满 10 时向高位进 1。对于任意十进制数，都可以使小数点分成整数部分和小数部分。

十进制数的特点是：数字的个数等于基数 10，逢十进一，借一当十；最大数字是 9，最小数字是 0，有 10 个数制符号——0、1、2、3、4、5、6、7、8、9；在具体表示时，每个数字都要乘以基数 10 的幂次。

例如，在十进制数 436.82 中，小数点左边第一位的 6 代表个位，值为 6×10^0 ；小数点左边第二位的 3 代表十位，值为 3×10^1 ；小数点左边第三位的 4 代表百位，值为 4×10^2 ；小数点右边第一位的值为 8×10^{-1} ；小数点右边第二位的值为 2×10^{-2} 。由此可见，处于不同的位置，代表的数值也是不同的。

十进制数的性质是：小数点向右移动一位，数值扩大 10 倍；反之，小数点向左移动一位，数值缩小为原来的十分之一。

2. 二进制数(binary)

按“逢二进一”的原则进行计数，每一位计满 2 时向高位进 1。

1) 二进制数的特点

二进制数的特点是：数字的个数等于基数 2；最大数字是 1，最小数字是 0，只有两个数制符号——0 和 1；在数值的表示中，每个数字都要乘以 2 的幂次，也就是每一位的位权。第一位的位权是 2^0 ，第二位的位权是 2^1 ，第三位的位权是 2^2 ，后面以此类推。表 1-1 给出了二进制的位权和十进制数值之间的对应关系。

表 1-1 二进制的位权与十进制数值之间的对应关系

二进制位数	7	6	5	4	3	2	1	-1	-2	-3	-4
位权	2^6	2^5	2^4	2^3	2^2	2^1	2^0	2^{-1}	2^{-2}	2^{-3}	2^{-4}
十进制数值	64	32	16	8	4	2	1	0.5	0.25	0.125	0.0625

以二进制数 1101.11 为例，可以表示为：

$$(1101.11)_2 = 1 \times 2^3 + 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 + 1 \times 2^{-1} + 1 \times 2^{-2} = (13.75)_{10}$$

二进制数的性质是：小数点向右移动一位，数值就扩大为原来的两倍；反之，小数点向左移动一位，数值就缩小一半。例如，把二进制数 110.101 的小数点向右移动一位，变为 1101.01，数值就变为原来的两倍；把 110.101 的小数点向左移动一位，变为 11.0101，数值相比原来就缩小了一半。

2) 二进制运算

二进制运算与十进制运算类似，同样可以进行算术运算，不仅操作简单、直观，而且更容易实现。

二进制求和法则如下:

$$0+0=0$$

$$0+1=1$$

$$1+0=1$$

$$1+1=10(\text{逢二进一})$$

二进制求差法则如下:

$$0-0=0$$

$$1-0=1$$

$$10-1=1(\text{借一当二})$$

$$1-1=0$$

二进制求积法则如下:

$$0\times 0=0$$

$$0\times 1=0$$

$$1\times 0=0$$

$$1\times 1=1$$

二进制求商法则如下:

$$0\div 1=0$$

$$1\div 1=1$$

例如,在对两个二进制数进行相加时,首先写出被加数和加数,然后按照由低位到高位
的顺序,根据二进制求和法则把两个数字逐位相加即可,就像将两个十进制数相加一样。

例 1.1 求 $1101.01+1001.11$ 。

$$\begin{array}{r} \text{解:} \quad 1101.01 \\ + 1001.11 \\ \hline 10111.00 \end{array}$$

计算结果: $1101.01+1001.11=10111.00$

例 1.2 求 $1101.01-1001.11$ 。

$$\begin{array}{r} \text{解:} \quad 1101.01 \\ - 1001.11 \\ \hline 0011.10 \end{array}$$

计算结果: $1101.01-1001.11=11.10$

例 1.3 求 1101×110 。

$$\begin{array}{r} \text{解:} \quad 1101 \\ \times 110 \\ \hline 0000 \\ 1101 \\ 1101 \\ \hline 1001110 \end{array}$$

计算结果: $1101\times 110=1001110$

例 1.4 求 $100111 \div 1101$ 。

解：

$$\begin{array}{r}
 11 \\
 1101 \overline{) 100111} \\
 \underline{1101} \\
 1101 \\
 \underline{1101} \\
 0
 \end{array}$$

计算结果： $100111 \div 1101 = 11$

3. 八进制数(octal)

八进制数的进位规则是“逢八进一”，基数 $R=8$ ，采用的数制符号是 0、1、2、3、4、5、6、7，每一位的位权是 8 的幂次。例如，八进制数 376.4 可表示为：

$$\begin{aligned}
 (376.4)_8 &= 3 \times 8^2 + 7 \times 8^1 + 6 \times 8^0 + 4 \times 8^{-1} \\
 &= 3 \times 64 + 7 \times 8 + 6 + 0.5 \\
 &= (254.5)_{10}
 \end{aligned}$$

4. 十六进制数(hexadecimal)

十六进制数的特点如下。

(1) 采用的 16 个数制符号是 0、1、2、3、4、5、6、7、8、9、A、B、C、D、E、F。符号 A~F 分别代表十进制中的 10~15。

(2) 进位规则是“逢十六进一”，基数 $R=16$ ，每一位的位权是 16 的幂次。例如，十六进制数 3AB.11 可表示为：

$$\begin{aligned}
 (3AB.11)_{16} &= 3 \times 16^2 + 10 \times 16^1 + 11 \times 16^0 + 1 \times 16^{-1} + 1 \times 16^{-2} \\
 &\approx (939.0664)_{10}
 \end{aligned}$$

5. 常用数制的对应关系

1) 常用数制的基数和数制符号

常用数制的基数和数制符号如表 1-2 所示。

表 1-2 常用数制的基数和数制符号

数制	十进制	二进制	八进制	十六进制
基数	10	2	8	16
数制符号	0~9	0, 1	0~7	0~9, A, B, C, D, E, F

2) 常用数制的对应关系

常用数制的对应关系如表 1-3 所示。

表 1-3 常用数制的对应关系

十进制	二进制	八进制	十六进制
0	0	0	0
1	1	1	1
2	10	2	2
3	11	3	3
4	100	4	4
5	101	5	5
6	110	6	6
7	111	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
16	10000	20	10

1.2.3 数制间的转换

将数由一种数制转换成另一种数制称为数制间的转换。由于计算机采用二进制，而日常生活中人们习惯使用十进制，因此计算机在进行数据处理时必须把输入的十进制数转换成计算机所能接收的二进制数，计算机运行结束后，再把二进制数转换成人们习惯的十进制数并输出。这两个转换过程完全由计算机系统自动完成。

1. 二进制数与十进制数之间的转换

1) 将二进制数转换成十进制数

将二进制数转换成十进制数的方法前面已经讲过，只需要将二进制数按位权展开，然后将各项数值按十进制数相加，便可得到等值的十进制数。例如：

$$(10110.11)_2 = 1 \times 2^4 + 1 \times 2^2 + 1 \times 2^1 + 1 \times 2^{-1} + 1 \times 2^{-2} = (22.75)_{10}$$

同理，为了将任意进制数转换为十进制数，只需要将数 $(N)_R$ 写成按位权展开的多项式，然后按十进制规则进行运算，便可求得相应的十进制数 $(N)_{10}$ 。

2) 将十进制数转换成二进制数

在将十进制数转换成二进制数时，需要对整数部分和小数部分分别进行转换。

(1) 整数转换。

整数转换采用除 2 取余法。

例 1.5 将 $(57)_{10}$ 转换为二进制数。

解：设 $(57)_{10}=(a_n a_{n-1} \cdots a_2 a_1 a_0)_2$

采用除 2 取余法，得：

$$\begin{array}{r}
 2 \overline{) 57} \quad \text{余数} \\
 2 \overline{) 28} \quad \cdots \cdots 1 = a_0 \\
 2 \overline{) 14} \quad \cdots \cdots 0 = a_1 \\
 2 \overline{) 7} \quad \cdots \cdots 0 = a_2 \\
 2 \overline{) 3} \quad \cdots \cdots 1 = a_3 \\
 2 \overline{) 1} \quad \cdots \cdots 1 = a_4 \\
 0 \quad \cdots \cdots 1 = a_5
 \end{array}$$

结果： $(57)_{10}=(111001)_2$

(2) 小数转换。

小数转换采用乘 2 取整法。

例 1.6 将 $(0.834)_{10}$ 转换成二进制数。

解：设 $(0.834)_{10}=(0.a_{-1}a_{-2}a_{-3}a_{-4})_2$

采用乘 2 取整法，得：

$$\begin{array}{r}
 0.834 \quad \text{整数} \\
 \times 2 \\
 \hline
 1.668 \quad \cdots \cdots 1 = a_{-1} \\
 0.668 \\
 \times 2 \\
 \hline
 1.336 \quad \cdots \cdots 1 = a_{-2} \\
 0.336 \\
 \times 2 \\
 \hline
 0.672 \quad \cdots \cdots 0 = a_{-3} \\
 \times 2 \\
 \hline
 1.344 \quad \cdots \cdots 1 = a_{-4}
 \end{array}$$

结果： $(0.834)_{10} \approx (0.1101)_2$

由例 1.6 可见，在对小数部分乘 2 取整的过程中，不一定能使最后的乘积为 0，因此存在转换误差。通常，当二进制数的精度已经达到预定的要求时，运算便可结束。

在将带有整数和小数的十进制数转换成二进制数时，必须对整数部分和小数部分分别按除 2 取余法和乘 2 取整法进行转换，之后再将两者的转换结果合并起来即可。

同理，要将十进制数转换成任意的 R 进制数 $(N)_R$ ，整数转换可采用除 R 取余法，小数转换可采用乘 R 取整法。

2. 二进制数与八进制数、十六进制数之间的转换

八进制数和十六进制数的基数分别为 $8=2^3$ 、 $16=2^4$ ，所以三位的二进制数恰好相当于一位的八进制数，四位的二进制数则相当于一位的十六进制数，它们之间的相互转换是很方便的。

将二进制数转换成八进制数的方法是：从小数点开始，分别向左、向右将二进制数按每三位一组的形式进行分组(不足三位的补 0)，然后写出与每一组二进制数等值的八进制数。

例 1.7 将二进制数 100110110111.00101 转换成八进制数。

解：

100	110	110	111	.	001	010
↓	↓	↓	↓	↓	↓	↓
4	6	6	7	.	1	2

结果： $(100110110111.00101)_2 = (4667.12)_8$

将八进制数转换成二进制数的方法恰好与将二进制数转换成八进制数相反：从小数点开始，分别向左、向右将八进制数的每一位数字转换成三位的二进制数。例如，可对例 1.7 按相反的过程进行转换。

$$(4667.12)_8 = (100110110111.00101)_2$$

将二进制数转换成十六进制数的方法和二进制数与八进制数之间的转换方法相似，从小数点开始，分别向左、向右将二进制数按每四位一组的形式进行分组(不足四位的补 0)，然后写出与每一组二进制数等值的十六进制数。

例 1.8 将二进制数 1111000001011101.0111101 转换成十六进制数。

解：

1111	0000	0101	1101	.	0111	1010
↓	↓	↓	↓	↓	↓	↓
F	0	5	D	.	7	A

结果： $(1111000001011101.0111101)_2 = (F05D.7A)_{16}$

类似地，在将十六进制数转换成二进制数时，可按例 1.8 展示的相反过程进行操作。

1.2.4 数据在计算机中的表示方式

计算机中处理的数据可分为数值型数据和非数值型数据两类。数值型数据是指数学中的代数值，具有量的含义，如 235、-328.45 或 $3/8$ 等；非数值型数据是指输入计算机中的所有文本信息，没有量的含义，如数字 0~9、大写字母 A~Z 或小写字母 a~z、汉字、图形、声音及一切可印刷的符号+、-、!、#、%等。

由于计算机内部使用的是二进制数，因此这些数据在计算机内部必须以二进制编码的形式表示。也就是说，一切输入计算机中的数据都是由 0 和 1 两个数字进行组合的。数值型数据有正负之分，数学中分别使用符号+和-来表示正数和负数，但在计算机中，数的正负符号也要使用 0 和 1 来表示。

1. 有符号数的表示方法

在计算机中，有符号的数通常使用原码、反码和补码 3 种形式表示，这么做的主要目的是

解决减法运算的问题。任何正数的原码、反码和补码形式都完全相同，而负数的原码、反码和补码形式则不同。

1) 数的原码表示

正数的符号位用 0 表示，负数的符号位用 1 表示，有效值部分用二进制绝对值表示，这种表示法称为原码。原码对 0 的表示方法不唯一，分为正的 0(000...00)和负的 0(100...00)。

例如，如果

$X=+76$; $Y=-76$

那么

$(X)_{\text{原}}=$	0	1001100
$(Y)_{\text{原}}=$	<u>1</u>	<u>1001100</u>
	↑	↑
	符号位	数值

2) 数的反码表示

正数的反码和原码相同，负数的反码则是对原码的除符号位外的其他各位取反：将 0 变成 1，而将 1 变成 0。

例如：

$(+76)_{\text{原}}=(+76)_{\text{反}}=01001100$

$(-76)_{\text{原}}=11001100$

$(-76)_{\text{反}}=10110011$

可以验证，任意数的反码的反码即为原码本身。

3) 数的补码表示

正数的补码和原码相同，负数的补码则需要对反码加 1。

例如：

$(+76)_{\text{原}}=(+76)_{\text{反}}=(+76)_{\text{补}}=01001100$

$(-76)_{\text{原}}=11001100$

$(-76)_{\text{反}}=10110011$

$(-76)_{\text{补}}=10110100$

可以验证，任意数的补码的补码即为原码本身。引入补码的概念之后，减法运算就可以使用加法来实现。

例如：

$$\begin{aligned}
 & (1)_{10} - (2)_{10} \\
 &= (1)_{10} + (-2)_{10} \\
 &= (00000001)_{\text{补}} + (11111110)_{\text{补}} \\
 &= (11111111)_{\text{补}} \\
 &= (11111110)_{\text{反}} \\
 &= (10000001)_{\text{原}} \\
 &= (-1)_{10}
 \end{aligned}$$

结果正确。设计补码的目的如下。

- 使符号位能与有效值部分一起参与运算，从而简化了运算规则。

- 使减法运算转换为加法运算，从而简化计算机中运算器的线路设计。

注意，这些转换都是在计算机的最底层进行的，而我们在高级语言中使用的都是原码。

2. 定点数与浮点数

数值除分正负之外，还有带小数点的数值。当所要处理的数值含有小数部分时，计算机就必须解决数值中的小数点的表示问题。在计算机中，通常采用隐含规定小数点的位置这种方式来表示带小数点的数值。

根据小数点的位置是否固定，数值的表示方法可以分为定点整数、定点小数和浮点数 3 种类型。定点整数和定点小数统称为定点数。

1) 定点整数

定点整数是指小数点隐含固定在整个数值的最后，符号位右边的所有位数表示的是整数。如果用 4 位表示定点整数，那么 0110 表示二进制数+110，也就是十进制数+6。

2) 定点小数

定点小数是指小数点隐含固定在某个位置的小数。人们通常将小数点固定在最高数据位的左边。如果用 4 位表示定点小数，那么 0110 表示二进制数+0.110，也就是十进制数+0.75。

由此可见，定点数可以表示纯小数和整数。定点整数和定点小数在计算机中的表示并没有什么区别，小数点完全靠事先约定而隐含在不同位置，如图 1-5 所示。

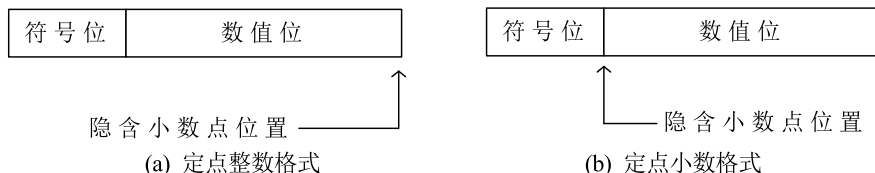


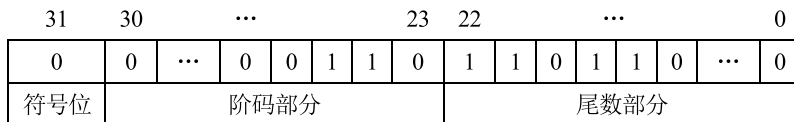
图 1-5 定点数格式

3) 浮点数

浮点数是指小数点位置不固定的数，浮点数既有整数部分又有小数部分。在计算机中，通常把浮点数分阶码(也称为指数)和尾数两部分进行表示。其中：阶码用二进制定点整数表示，尾数用二进制定点小数表示，阶码的长度决定数值的范围，尾数的长度决定数值的精度。为保证不损失有效数字，通常还会对尾数进行规格化处理，从而保证尾数的最高位为 1。实际数值可通过阶码进行调整。

浮点数的格式多种多样。例如，某计算机用 32 位表示浮点数，阶码部分为 8 位补码的定点整数，尾数部分为 24 位补码的定点小数。浮点数的最大特点在于比定点数表示的数值范围大。

例如，二进制的+110110 等于 $2^6 \times 0.11011$ ，阶码为 6，也就是二进制的+110，尾数为+0.11011。单精度浮点数的表示形式如图 1-6 所示。



1.2.5 字符编码

计算机是以二进制编码方式组织并存放信息的,信息编码就是对输入计算机中的各种数值数据和非数值数据使用二进制方式进行编码。对于不同机器、不同类型的数据来说,编码方式也是多种多样的。为了使信息的表示、交换、存储或加工处理更方便,计算机系统通常采用统一的编码方式,因此人们制定了编码的国家标准或国际标准,比如位数不等的二进制码、BCD码、ASCII码等。计算机使用这些编码在计算机内部和键盘等终端设备之间以及计算机之间进行信息交换。

在输入过程中,系统自动将用户输入的各种数据按编码的类型转换成相应的二进制形式并存入计算机的存储器。在输出过程中,再由系统自动将二进制编码的数据转换成用户可以识别的数据形式并输出给用户。

字符是计算机中使用最多的非数值型数据,是人与计算机进行通信和交互的重要媒介,国际上广泛使用的字符编码有美国信息交换标准码(American Standard Code for Information Interchange, ASCII)、BCD码和Unicode码。

1. ASCII码

ASCII码有7位码和8位码两种形式。7位的ASCII码是用7位的二进制数进行编码的,所以可以表示128个字符。这是因为1位的二进制数可以表示 $2^1=2$ 种状态——0或1;2位的二进制数可以表示 $2^2=4$ 种状态——00、01、10、11;以此类推,7位的二进制数可以表示 $2^7=128$ 种状态。每种状态都唯一对应一个7位的二进制码,这些二进制码可以排列成十进制序号0~127,参见附录B。

ASCII码表中的128个符号是这样分配的:第0~32号和第127号(共34个)为控制字符,主要包括换行、回车等功能字符;第33~126号(共94个)为字符,其中第48~57号为0~9的10个数字符号,第65~90号为26个大写英文字母,第97~122号为26个小写英文字母,其余的为一些标点符号、运算符号等。例如,大写字母A的ASCII码值为1000001,对应十进制数65;小写字母a的ASCII码值为1100001,对应十进制数97。这些字符已基本能够满足各种程序设计语言、西文文字及常见控制命令的需要。

为了方便使用,在计算机的存储单元中,字符的ASCII码占1字节(8位),并且最高位只用作奇偶校验位,如图1-7所示。

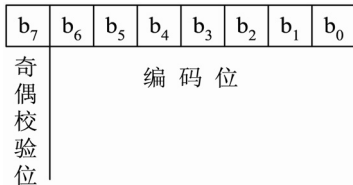


图1-7 字符的ASCII码占1字节

在代码传送过程中,奇偶校验是用来检验是否出现错误的一种方法。一般分为奇校验和偶校验两种。奇校验规定,如果代码正确,那么单字节中1的个数必须是奇数,若是非奇数,则在最高位b₇处补1来满足;偶校验规定,如果代码正确,那么单字节中1的个数必须是偶数,若是非偶数,则在最高位b₇处补1来满足。

例 1.9 将“COME”中的 4 个字符使用带奇校验位的 ASCII 码进行存储。

解：首先由附录 B 查出十进制的 ASCII 码，然后转换成二进制的 ASCII 码，最后根据奇校验的规定在左边补上奇校验位，如表 1-4 所示。

表 1-4 ASCII 码的应用示例

字 符	十进制的 ASCII 码	二进制的 ASCII 码	带奇校验位的 ASCII 码
C	67	1000011	01000011
O	79	1001111	01001111
M	77	1001101	11001101
E	69	1000101	01000101

例 1.10 当 ASCII 码值为 101010 时，表示的是什么字符？当采用偶校验时，奇偶校验位 b_7 等于什么？

解： $(101010)_2 = (42)_{10}$ ，由附录 B 得知表示的字符为*；根据偶校验规则，应使单字节中 1 的个数为偶数，所以在奇偶校验位补 1，使 $b_7=1$ 。

ASCII 码虽然是最常用的编码，但由于最高位 b_7 用来作为校验位，因此只能表示 128 个不同的字符。如果最高位 b_7 也用来编码，则称为扩展的 ASCII 码，扩展的 ASCII 码可用来表示 256 个不同的字符。

目前，还有一种编码在许多环境中也得到了应用，这就是 Unicode 码。Unicode 码使用 16 位的二进制进行编码，最多可以表示 65 536 个不同的字符。通过把高 8 位置为 0，并保持原来的编码不变，Unicode 码把 ASCII 码和扩展的 ASCII 码也吸收了进来。例如，字母 S 的 3 种编码如表 1-5 所示。

表 1-5 字母 S 的 3 种编码

ASCII 码	扩展的 ASCII 码	Unicode 码
01010011	01010011	00000000 01010011

2. BCD 码

BCD(Binary Coded Decimal)码是二进制编码的十进制数，有 4 位的 8421 BCD 码、6 位的 BCD 码和扩展的 BCD 码 3 种。

1) 8421 BCD 码

8421 BCD 码使用 4 位的二进制数来表示十进制数字，4 位的二进制数从左到右依次为 8、4、2、1，这只能表示十进制中的字符 0~9。为了对位数更多的十进制数进行编码，就需要提供与十进制数的位数一样多的 4 位组。

2) 扩展的 BCD 码

由于 8421 BCD 码只能表示 10 个十进制数，因此人们在原来的 4 位 BCD 码的基础上又产生了 6 位的 BCD 码。6 位的 BCD 码能表示 64 个字符，其中包括 10 个十进制数、26 个英文字母和 28 个特殊字符。但在某些场合，英文字母仍需要区分大小写，于是提出了扩展的 BCD 码。扩展的 BCD 码由 8 位组成，可表示 256 个符号，其全称为 Extended Binary Coded Decimal

Interchange Code, 简称为 EBCDIC。EBCDIC 码是常用编码之一, IBM 及 UNIVAC 计算机均采用这种编码。

1.2.6 汉字编码

为了在计算机中表示汉字, 用计算机处理汉字, 同样也需要对汉字进行编码。计算机对汉字信息的处理过程实际上是各种汉字编码间的转换过程。这些编码主要包括汉字信息交换码、汉字输入码、汉字内码、汉字字形码及汉字地址码等。下面分别对各种汉字编码进行介绍。

1. 汉字信息交换码

汉字信息交换码是用于在汉字信息处理系统之间或汉字信息处理系统与通信系统之间进行信息交换的汉字代码, 简称交换码。汉字信息交换码是为了能够在系统、设备之间交换信息时采用统一的形式而制定的。

我国于 1981 年颁布了国家标准——《信息交换用汉字编码字符集 基本集》(GB/T 2312—1980), 也就是国际码。了解国际码的下列概念, 对使用和研究汉字信息处理系统十分有益。

1) 常用汉字及其分级

国际码规定了进行一般汉字信息处理时使用的 7445 个字符编码, 其中包括 682 个非汉字图形符号(如序号、数字、罗马数字、英文字母、日文假名、俄文字母等)和 6763 个汉字代码。汉字代码中又有一级常用字 3755 个, 二级常用字 3008 个。一级常用字按汉语拼音字母顺序排列, 二级常用字按偏旁部首排列, 部首依笔画多少排序。

2) 国际码需要使用 2 字节来表示

由于 1 字节只能表示 2^8 共 256 种编码, 这显然无法表示汉字的国际码, 因此国际码需要使用 2 字节来表示。

3) 国际码的编码范围

为了中英文兼容, GB/T 2312—1980 规定, 在国际码中, 所有字符(包括符号和汉字)的每一字节的编码范围与 ASCII 码表中的 94 个字符编码一致, 所以国际码的编码范围是 2121H~7E7EH (可表示 94×94 个字符)。

4) 国际码是区位码

类似于 ASCII 码表, 国际码也有一张区位码表。简单而言, 就是将 7445 个国际码放置在 $94 \text{ 行} \times 94 \text{ 列}$ 的阵列中。阵列的每一行称为汉字的“区”, 用区号表示; 每一列称为汉字的“位”, 用位号表示。显然, 区号的范围是 1~94, 位号的范围也是 1~94。这样, 一个汉字在区位码表中的位置就可以使用这个汉字所在的区号与位号来确定。区号与位号的组合就是汉字的“区位码”。区位码的形式是高两位为区号, 低两位为位号。如“中”字的区位码是 5448, 也就是 54 区 48 位。区位码与汉字之间具有一一对应的关系。国际码在区位码表中的安排是: 1~15 区是非汉字图形符区; 16~55 区是一级常用字区; 56~87 区是二级常用字区; 88~94 区是保留区, 可用来存储自造字。实际上, 区位码也是一种输入法, 最大优点就是能够实现一字一码的无重码输入, 最大缺点是难以记忆。

2. 汉字输入码

为了将汉字输入计算机而编制的代码称为汉字输入码，也叫外码。

目前，汉字主要是经标准键盘输入计算机的，所以汉字输入码由键盘上的字符或数字组合而成。例如，为了使用全拼输入法输入“中”字，就需要输入字符串 zhong(然后选字)。汉字输入码是根据汉字的发音或字形结构等多种属性及有关规则编制的，目前流行的汉字输入码的编码方案有许多种，如全拼输入法、双拼输入法、自然码输入法、五笔输入法等，可分为音码、形码、音形结合码3类。全拼输入法和双拼输入法是根据汉字的发音进行编码的，称为音码；五笔输入法是根据汉字的字形结构进行编码的，称为形码；自然码输入法是以拼音为主、辅以字形字义进行编码的，称为音形结合码。

可以想象，对于同一个汉字，不同的输入法有不同的输入码。例如，“中”字的全拼输入码是 zhong，双拼输入码是 vs，而五笔输入码是 kh。不管采用何种输入方法，输入的汉字都会被转换成对应的机内码并存储在介质中。

3. 汉字内码

汉字内码是为了在计算机内部对汉字进行存储、处理和传输而设置的汉字编码。一个汉字在输入计算机后，必须首先转换为内码，然后才能在计算机内部进行传输和处理。汉字内码的形式也是多种多样的，目前对应于国际码，汉字内码也使用2字节来表示，并把每一字节的最高二进制位置1，以此作为汉字内码的标识，以免与单字节的ASCII码混淆并产生歧义。换言之，对于国际码来说，只要将每一字节的最高位置1，即可转换为内码。

4. 汉字字形码

目前，汉字信息处理系统在产生汉字时，大多以点阵的方式形成汉字。汉字字形码是用来确定汉字字形点阵的编码，也叫字模或汉字输出码。

汉字是方块字，将方块等分为 n 行 n 列的格子，简称点阵。凡笔画所到的格子点为黑点，用二进制数1表示；否则为白点，用二进制数0表示。这样，汉字的字形就可以使用一串二进制数表示了。例如， 16×16 的汉字点阵有256个点，需要使用256个二进制位才能表示一个汉字的字形码。图1-8展示了“中”字的 16×16 点阵的字形示意图。

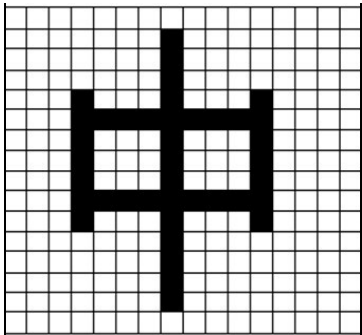


图1-8 “中”字的 16×16 点阵的字形示意图

在计算机中，8个二进制位可组成1字节，字节是对存储空间进行编址的基本单位。因此， 16×16 点阵的字形码需要 $16\times 16/8=32$ 字节的存储空间；同理， 24×24 点阵的字形码需要

$24 \times 24 / 8 = 72$ 字节的存储空间。例如,使用 16×16 点阵的字形码存储“中国”两个汉字,需要占用 $2 \times 16 \times 16 / 8 = 64$ 字节的存储空间。

显然,点阵中的行列划分越多,字形的质量越好,锯齿现象也就越轻微,但汉字字形码占用的存储空间却更大了。汉字字形通常分为通用型和精密型两类。通用型汉字字形点阵分为如下 3 种。

- 简易型: 16×16 点阵。
- 普通型: 24×24 点阵。
- 提高型: 32×32 点阵。

精密型汉字字形用于常规的印刷排版,由于信息量较大(字形点阵一般在 96×96 点阵以上),人们通常采用信息压缩存储技术。

汉字的点阵字形在输出汉字时需要经常使用,为此,可以把各个汉字的字形码固定地存储起来。存放各个汉字字形码的实体称为汉字库。为满足不同需要,业内还出现了各种各样的字库,如宋体字库、仿宋体字库、楷体字库、黑体字库和繁体字库等。

汉字的点阵字形的缺点是放大后会出现锯齿现象,很不美观。中文的 Windows 操作系统广泛采用了 TrueType 类型的字形码,由于采用数学方法来描述汉字的字形码,因此汉字可以无限放大而不会产生锯齿现象。

5. 汉字地址码

汉字地址码是指汉字库(这里主要是指字形的点阵式字模库)中用来存储汉字字形信息的逻辑地址码。在汉字库中,字形信息都按一定顺序(大多数按国际码中汉字的排列顺序)连续存放在存储介质中,所以汉字地址码大多也是连续有序的,而且与汉字内码有着简单的对应关系,以简化从汉字内码到汉字地址码的转换。

6. 各种汉字编码之间的关系

汉字的输入、处理和输出过程,实际上也就是汉字的各种编码之间的转换过程,或者说汉字编码在系统相关部件之间的传输过程。图 1-9 展示了汉字编码在汉字信息处理系统中的位置以及它们之间的关系。

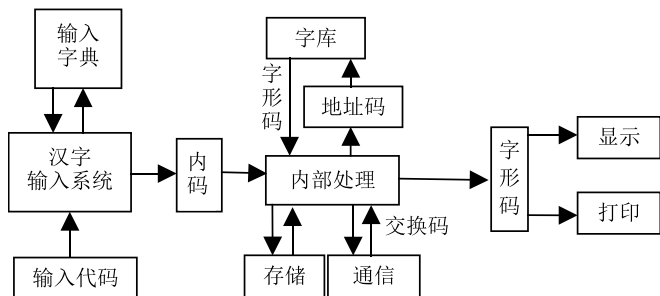


图 1-9 汉字编码在汉字信息处理系统中的位置以及它们之间的关系

汉字输入码向内码的转换是通过使用输入字典(又称索引表,是汉字外码与内码的对照表)实现的,计算机系统往往具有多种输入方法,每种输入方法都有各自的索引表。

在计算机的内部处理过程中,汉字信息的存储和各种必要的加工以及向磁盘存储汉字信息

等，都是以汉字内码的形式进行的。

在通信过程中，汉字信息处理系统将汉字内码转换为适合于通信的交换码(国际码)以实现通信处理。

在汉字的显示和打印输出过程中，汉字信息处理系统根据汉字内码计算出汉字地址码，再根据汉字地址码从字库中取出汉字字形码，从而实现汉字的显示或打印输出。

1.2.7 存储单位

各种各样的数据在计算机内部都以二进制形式存储，在计算存储空间的大小时，需要用到不同的存储单位。

1. 位

位(bit)是计算机的最小存储单位，符号为 b，表示二进制中的一位。位又称为“比特”，比特是 bit 的音译。二进制中的位只能表示两种状态，也就是只能存放二进制数 0 或 1。

2. 字节

字节(Byte)是计算机存储数据的基本单位，符号为 B，表示二进制中的 8 位， $1\text{ B} = 8\text{ b}$ 。

常用的存储单位有 KB、MB、GB、TB。KB 读作“千字节”，表示 2 的 10 次方字节，等于 1024 字节；MB 读作“兆字节”，表示 2 的 20 次方字节， $1\text{ MB} = 1024\text{ KB}$ ；GB 读作“吉字节”或“千兆字节”，表示 2 的 30 次方字节， $1\text{ GB} = 1024\text{ MB}$ ；TB 读作“太字节”，表示 2 的 40 次方字节， $1\text{ TB} = 1024\text{ GB}$ 。

3. 常用计量单位

在描述计算机的存储容量时，对于常用的计量单位 KB、MB、GB、TB 来说，前缀的含义如表 1-6 表示。

表 1-6 计算机存储容量的常用计量单位

符 号	汉文名	英文名	注 释
KB	千字节	Kilobyte	$2^{10}\text{ B} = 1024\text{ B}$
MB	兆字节	Megabyte	$2^{20}\text{ B} = 1\,048\,576\text{ B}$
GB	吉字节	Gigabyte	$2^{30}\text{ B} = 1\,073\,741\,824\text{ B}$
TB	太字节	Terabyte	$2^{40}\text{ B} = 1\,099\,511\,627\,776\text{ B}$

下面举例说明常用存储容量的计算方法。

例 1.11 为了满足一台 1024×768 像素的单色显示器的存储要求，需要使用多少字节？

解：这台显示器具有 1024 列，每列包含 768 像素。由于是单色显示器，每像素需要 1 位的存储空间，也就是说，1 字节可以存放 8 像素，因此总共需要

$$\frac{1024 \times 768}{8} \text{ 字节} = 98\,304 \text{ 字节}$$

例 1.12 假设显示器的每一像素都要用 2 字节来存储, 一台 1024×768 像素的显示器需要多少存储空间?

解: 需要的存储空间为

$$\frac{1024 \times 768 \times 2}{1024} \text{KB} = 1536 \text{KB}$$

例 1.13 计算机动画对计算机的存储需求一直在增长。假设有一种格式的计算机动画需要把一系列图片存储到计算机中, 图片则显示在屏幕上的一个 1024×768 像素的长方形矩阵中, 而且每一像素需要 1 字节的存储空间。为了存储一段包含 32 幅图片的动画, 需要多少存储空间?

解: 一幅图片包含的像素总数为

$$1024 \times 768 \text{ 像素} = 2^{10} \times (3 \times 2^8) \text{ 像素} = 3 \times 2^{18} \text{ 像素}$$

由于每一像素需要 1 字节的存储空间, 因此 32 幅图片所需要的总字节数为

$$32 \times 3 \times 2^{18} \text{ 字节} = 3 \times 2^{23} \text{ 字节}$$

也就是

$$3 \times 2^{23} \text{ B} \times \frac{1 \text{ MB}}{2^{20} \text{ B}} = 3 \times 2^3 \text{ MB} = 24 \text{ MB}$$

注意: 本书中使用 MB 表示 Megabyte(百万字节), 而使用 Mb 表示 Megabit(百万位)。

1.3 个人计算机的基本配置及性能指标

个人计算机是人们日常工作和生活中的重要设备, 怎样衡量个人计算机的性能是人们所关心的一个重要问题。

1.3.1 硬件

硬件的性能是基础性指标, 包括 CPU、内存、硬盘、显示器等核心部件的性能。

1. CPU

CPU 是计算机的心脏, 其性能好坏对整个计算机的性能起决定作用。CPU 的基本指标包括:

- 时钟频率。计算机的所有操作都是按照严格的时序进行的, 时钟相当于操作的节拍, 节拍越快, 操作速度越快。目前常见的时钟频率单位是 GHz($1 \text{ GHz} = 1\,000\,000\,000 \text{ Hz}$), 频率越高, 意味着运算速度越快。
- 运算器的字长。运算器一次运算可以处理的二进制位数叫字长, 字长越大, 处理能力越强。目前, 个人计算机中 CPU 的典型字长是 32 位和 64 位。
- 地址总线宽度和数据总线宽度。地址总线宽度是指内存空间的二进制位数, 典型的宽度是 32 位和 64 位, 位数越多, 内存空间越大。数据总线宽度是指 CPU 和外部交换数据时一次可以传输的二进制数据的位数, 位数越大, 数据传输效率越高。一般所说的

32 位 CPU 是指字长、地址总线宽度、数据总线宽度都是 32 位的 CPU；64 位 CPU 则是指字长、地址总线宽度、数据总线宽度都是 64 位的 CPU。

- 内核数量。内核数量是指 CPU 芯片内封装的 CPU 数量。多核处理器一般比单核处理器具有更强的处理能力。

2. 内存

内存是计算机中暂存数据的部件，是计算机各部件沟通的桥梁。内存的基本指标包括：

- 容量。内存是以字节为单位的，1 字节包含 8 个二进制位，可以整体读写。内存容量是指内存的字节总数。目前常见的内存容量单位是吉字节(1 吉字节 $\approx 1\,000\,000\,000$ 字节，实际上等于 $1024 \times 1024 \times 1024$ 字节)。内存容量越大越好。
- 读写速度。内存的读写速度快，CPU 的处理能力才能充分发挥出来。目前，主流内存为 DDR SDRAM(Double Data Rate SDRAM)，比较常见的是 DDR3、DDR4 等，衡量内存速度的指标是内存主频，单位为 MHz(1 MHz = 1 000 000 Hz)。内存主频越高越好。

3. 硬盘

硬盘是典型的外存设备。与内存不同的是，硬盘数据不能直接被 CPU 读写，从硬盘读出的数据必须先保存到内存中，然后才能被应用程序使用。同样，写入硬盘的数据也必须先在内存中准备好，然后才能写入硬盘。

硬盘的主要指标包括硬盘容量、转速和数据传输速率等。

- 硬盘容量是指硬盘可以存储的总字节数，个人计算机的硬盘容量一般为 1~2 TB。
- 转速是指磁盘盘片每分钟旋转的转数。一般而言，转速越高，磁盘的读写速度越快。典型个人计算机的硬盘转速一般为 5400r/min 或 7200r/min。
- 数据传输速率是指硬盘与内存之间的数据交换速率，因为取决于硬盘接口性能，所以又称为接口传输率。目前主流的硬盘接口类型为 SATA(Serial Advanced Technology Attachment)，SATA 2.0 可以达到 3 Gb/s，SATA 3.0 则可以达到 6 Gb/s。

4. 显示器

显示器是个人计算机的典型输出设备。目前流行的显示器是液晶显示器，主要性能指标包括屏幕尺寸、屏幕比例(宽高比)、分辨率、功耗等。

- 屏幕尺寸是指屏幕显示区域的对角线长度，以英寸为单位。目前，对于个人计算机而言，显示器的屏幕尺寸大多为 20 英寸左右。
- 屏幕比例是指屏幕显示区域的宽度和高度的比值。目前，宽屏显示器的屏幕比例一般为 16:9 或 16:10，普通显示器的屏幕比例一般为 4:3 或 5:4。
- 分辨率是指屏幕的水平及垂直像素点数，如 1920×1200、1920×1080、1680×1050、1440×900 等。一般尺寸较大的屏幕分辨率也较高。
- 功耗是指显示器消耗的电能，一般为数十瓦左右。

显示器的显示效果与显卡的性能密切相关。显卡是个人计算机中处理信息显示的部件，对于图形、高速视频图像处理应用而言，应精心选择显卡及显示器。普通应用对显卡及显示器的

要求不高。

5. 其他设备

个人计算机的硬件还包括键盘、鼠标及光盘驱动器等。键盘和鼠标是个人计算机必不可少的设备,因其性能对计算机的整体性能影响不大且价格不高,因此可以忽略。光盘驱动器是可选设备,在移动存储设备日益普及的当下,光盘驱动器的作用已越来越弱。

1.3.2 操作系统

操作系统(Operating System, OS)是一种用来综合管理和控制计算机运行的计算机程序,只有安装了操作系统,普通用户才能操作和使用计算机。目前,个人计算机大多采用 Windows 操作系统。这里仅介绍常用的 Windows 操作系统。

购买和安装 Windows 操作系统需要注意 32 位版本和 64 位版本的区别。如果个人计算机上的 CPU 为 64 位的,应选择安装 64 位版本的 Windows,这样才能充分发挥 CPU 的性能。32 位版本的 Windows 可以访问的实际内存大小为 3.5 GB 左右,即使安装更多的内存,多余的内存空间也会被白白浪费,根本得不到利用。64 位版本的 Windows 可以访问的实际内存大小在 100 GB 以上,足以满足个人计算机用户的需求。

1. Windows 10

2015 年 1 月 22 日,微软公司正式发布了 Windows 10,主推跨平台融合。Windows 10 在推广期是免费安装的,任何正版 Windows 7、Windows 8 用户都可免费升级到 Windows 10。2015 年 5 月 14 日,微软公司正式公布 Windows 10 家族阵容,一共包含 7 大版本。

- 家庭版(Windows 10 Home),主要面向消费者和个人计算机用户,适合个人或家庭用户。
- 专业版(Windows 10 Pro),主要面向个人计算机用户,并且还面向大屏平板电脑、笔记本电脑、PC 平板二合一变形本等桌面设备。
- 移动版(Windows 10 Mobile),主要面向小尺寸的触摸设备,比如智能手机、小屏平板电脑等移动设备。
- 企业版(Windows 10 Enterprise),在专业版的基础上,增加了专门为满足大中型企业需求而开发的高级功能,适合企业用户。
- 教育版(Windows 10 Education),主要基于企业版进行开发,能够满足学校教职工、管理人员、教师和学生需求。
- 移动企业版(Windows 10 Mobile Enterprise),主要面向使用智能手机和小尺寸平板电脑的企业用户,能为他们提供最佳的操作体验。
- Windows 10 IoT Core,主要面向低成本的物联网设备。

Windows 10 需要的硬件环境如下。

- CPU 时钟频率在 1 GHz 以上。
- 内存至少 1 GB(32 位)或 2 GB(64 位)。
- 硬盘空间至少 16 GB(32 位)或 20 GB(64 位)。
- 显卡需要支持微软的 DirectX 9 图形设备。

2. Windows 8

Windows 8 于 2012 年 10 月 26 日正式推出, 可选版本包括:

- 标准版, Windows 8 中文版便是标准版。标准版适用于台式机、笔记本电脑和平板电脑, 是普通家庭用户的最佳选择。
- 专业版, 简称 Windows 8 Pro, 是面向专业技术人员及开发人员的高级版本。
- 企业版, Windows 8 Enterprise 的批量授权版本, 其功能与专业版相同。
- Windows RT, 一种具有专门用途的 Windows 8 版本, 不适合普通用户使用。

Windows 8 需要的硬件环境如下。

- CPU 时钟频率在 1 GHz 以上。
- 内存至少 1 GB(32 位)或 2GB(64 位)。
- 硬盘空间至少 16 GB(32 位)或 20 GB(64 位)。
- 显卡需要支持微软的 DirectX 9 图形设备。
- 网络及显示器分辨率: 为了访问 Windows 应用商店, 需要有效的 Internet 连接以及至少 1024×768 像素的屏幕分辨率。为了拖曳程序, 屏幕分辨率至少要达到 1366×768 像素。
- 其他: 为了使用触控功能, 需要提供支持多点触控的平板电脑或显示器。

3. Windows 7

Windows 7 的可选版本如下。

- 家庭普通版, 对应的英文版本是 Starter。
- 家庭高级版, 对应的英文版本是 Home Premium。
- 专业版, 对应的英文版本是 Professional。
- 旗舰版, 对应的英文版本是 Ultimate。

4. Windows Vista

Windows Vista 是 Windows 7 的上一代产品, 目前已停止销售, 微软公司只对现有产品提供技术支持。

Windows Vista 的家族成员有家庭普通版、家庭高级版、商用版和旗舰版等, 功能依次递增。

5. Windows XP

Windows XP 是比 Windows Vista 更早的产品, 目前已很少见到。与 Windows Vista 相比, Windows XP 取得了巨大成功, 其普及程度和流行时间远远高于 Windows Vista。

1.4 习题

一、选择题

1. 完整的计算机系统包括_____两大部分。
 - A) 控制器和运算器
 - B) CPU 和 I/O 设备
 - C) 硬件和软件
 - D) 操作系统和计算机设备

2. 计算机硬件系统包括_____。
- A) 内存储器 and 外部设备 B) 显示器、机箱、键盘
C) 主机 and 外部设备 D) 主机 and 打印机
3. 计算机软件系统包括_____。
- A) 操作系统 and 语言处理系统 B) 数据库软件 and 管理软件
C) 程序 and 数据 D) 系统软件 and 应用软件
4. 银行的储蓄程序属于_____。
- A) 表格处理软件 B) 系统软件
C) 应用软件 D) 文字处理软件
5. 系统软件中最重要的是_____。
- A) 解释程序 B) 操作系统
C) 数据库管理系统 D) 工具软件
6. 计算机能直接执行_____。
- A) 使用高级语言编写的源程序 B) 机器语言程序
C) 英语程序 D) 十进制程序
7. 将高级语言翻译成机器语言的方式有_____两种。
- A) 解释方式和编译方式 B) 文字处理和图形处理
C) 图像处理和翻译 D) 语音处理和文字编辑
8. “程序存储思想”是由_____提出来的。
- A) 丹尼尔·里奇 B) 冯·诺依曼
C) 贝尔 D) 马丁·理查德
9. $(10110110)_2 + (111101)_2 = (\quad)_2$ 。
- A) 110101 B) 11110011 C) 11001100 D) 11010111
10. $(10010100)_2 - (100101)_2 = (\quad)_2$ 。
- A) 11110101 B) 10010011 C) 1101111 D) 1100111
11. $(1101)_2 \times (101)_2 = (\quad)_2$ 。
- A) 1000001 B) 1010011 C) 1011100 D) 1101111
12. $(10010)_2 \div (11)_2 = (\quad)_2$ 。
- A) 1010 B) 111 C) 1100 D) 110
13. 将如下补码转换为十进制数: $(11110110)_{\text{补}} = (\quad)_{10}$ 。
- A) 8 B) -9 C) -10 D) 11
14. 已知字符 8 的 ASCII 码值是 56, 那么字符 5 的 ASCII 码值是_____。
- A) 52 B) 53 C) 54 D) 55
15. 1 KB 表示_____。
- A) 1024 位 B) 1000 位 C) 1000 字节 D) 1024 字节
16. 指令存储在存储器的_____存储区。
- A) 程序 B) 数据 C) 栈 D) 堆

二、填空题

1. 计算机由如下 5 个基本部分组成：运算器、控制器、_____、_____和输出设备。
2. 运算器的主要功能是算术运算和_____。
3. 存储器通常分为内存储器和_____。
4. 计算机能够直接识别和执行的计算机语言是_____。
5. 中央处理器是决定计算机性能的核心部件，由_____组成。
6. $(254)_{10} = (\rule{1cm}{0.4pt})_2 = (\rule{1cm}{0.4pt})_8 = (\rule{1cm}{0.4pt})_{16}$ 。
7. $(3.40625)_{10} = (\rule{1cm}{0.4pt})_2 = (\rule{1cm}{0.4pt})_8 = (\rule{1cm}{0.4pt})_{16}$ 。
8. $(125)_{10} = (\rule{1cm}{0.4pt})_{\text{原}} = (\rule{1cm}{0.4pt})_{\text{反}} = (\rule{1cm}{0.4pt})_{\text{补}}$ 。
9. $(-25)_{10} = (\rule{1cm}{0.4pt})_{\text{原}} = (\rule{1cm}{0.4pt})_{\text{反}} = (\rule{1cm}{0.4pt})_{\text{补}}$ 。
10. 十进制数 3527 的 8421 BCD 码的表示形式为_____。
11. 已知字符 a 的 ASCII 码值是 97，那么字符 f 的 ASCII 码值是_____。