



普通高等教育“十一五”国家级规划教材
中国高等教育学会医学教育专业委员会规划教材

全国高等医学院校教材
供基础、临床、预防、口腔医学类专业用

计算机应用基础

(第6版)

主编 郭永青 李祥生 黎小沛

*Computer
Application Foundation*



北京大学医学出版社

普通高等教育“十一五”国家级规划教材
中国高等教育学会医学教育专业委员会规划教材
全国高等医学院校教材

供基础、临床、预防、口腔医学类专业用

计算机应用基础

Computer Application Foundation

(第6版)

主 编 郭永青 李祥生 黎小沛

编 者 (按姓名汉语拼音排序)

郭建光 (北京大学医学部)

刘 群 (华中科技大学)

郭永青 (北京大学医学部)

齐惠颖 (北京大学医学部)

胡 彬 (华中科技大学)

王 静 (北京大学医学部)

胡加立 (北京大学医学部)

王路漫 (北京大学医学部)

黄轶旻 (桂林医学院)

杨俊丽 (山东医科大学)

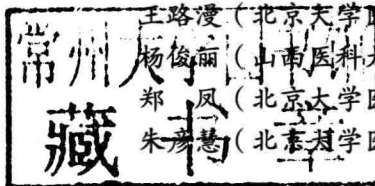
兰顺碧 (华中科技大学)

郑 凤 (北京大学医学部)

黎小沛 (天津医科大学)

朱彦慧 (北京中医药大学)

李祥生 (山西医科大学)



北京大学医学出版社

JISUANJI YINGYONG JICHU

图书在版编目 (CIP) 数据

计算机应用基础/郭永青, 李祥生, 黎小沛主编. —6 版. —北京:
北京大学医学出版社, 2013. 12
ISBN 978-7-5659-0736-4

I. ①计… II. ①郭…②李…③黎… III. ①电子计算机—高等学校—教材
IV. ①TP3

中国版本图书馆 CIP 数据核字 (2013) 第 308808 号

计算机应用基础 (第 6 版)

主 编: 郭永青 李祥生 黎小沛

出版发行: 北京大学医学出版社 (电话: 010-82802230)

地 址: (100191) 北京市海淀区学院路 38 号 北京大学医学部院内

网 址: <http://www.pumpress.com.cn>

E - mail: booksale@bjmu.edu.cn

印 刷: 北京朝阳新艺印刷有限公司

经 销: 新华书店

责任编辑: 张其鹏 责任校对: 金彤文 责任印制: 苗 旺

开 本: 850mm×1168mm 1/16 印张: 20 字数: 568 千字

版 次: 2013 年 12 月第 6 版 2013 年 12 月第 1 次印刷

书 号: ISBN 978-7-5659-0736-4

定 价: 37.00 元

版权所有, 违者必究

(凡属质量问题请与本社发行部联系退换)

全国高等医学院校临床专业本科教材评审委员会

主任委员 王德炳 柯 杨

副主任委员 吕兆丰 程伯基

秘 书 长 陆银道 王凤廷

委 员 (按姓名汉语拼音排序)

白咸勇 曹德品 陈育民 崔慧先 董 志

郭志坤 韩 松 黄爱民 井西学 黎孟枫

刘传勇 刘志跃 宋焱峰 宋印利 宋远航

孙 莉 唐世英 王 宪 王维民 温小军

文民刚 线福华 袁聚祥 曾晓荣 张 宁

张建中 张金钟 张培功 张向阳 张晓杰

周增桓

序

北京大学医学出版社组织编写的全国高等医学院校临床医学专业本科教材（第2套）于2008年出版，共32种，获得了广大医学院校师生的欢迎，并被评为教育部“十二五”普通高等教育本科国家级规划教材。这是在教育部教育改革、提倡教材多元化的精神指导下，我国高等医学教材建设的一个重要成果。为配合《国家中长期教育改革和发展纲要（2010—2020年）》，培养符合时代要求的医学专业人才，并配合教育部“十二五”普通高等教育本科国家级规划教材建设，北京大学医学出版社于2013年正式启动全国高等医学院校临床医学专业（本科）第3套教材的修订及编写工作。本套教材近六十种，其中新启动教材二十余种。

本套教材的编写以“符合人才培养需求，体现教育改革成果，确保教材质量，形式新颖创新”为指导思想，配合教育部、国家卫生和计划生育委员会在医药卫生体制改革意见中指出的，要逐步建立“5+3”（五年医学院校本科教育加三年住院医师规范化培训）为主体的临床医学人才培养体系。我们广泛收集了对上版教材的反馈意见。同时，在教材编写过程中，我们将与更多的院校合作，尤其是新启动的二十余种教材，吸收了更多富有一线教学经验的老师参加编写，为本套教材注入了新鲜的活力。

新版教材在继承和发扬原教材结构优点的基础上，修改不足之处，从而更加层次分明、逻辑性强、结构严谨、文字简洁流畅。除了内容新颖、严谨以外，在版式、印刷和装帧方面，我们做了一些新的尝试，力求做到既有启发性又能引起学生的兴趣，使本套教材的内容和形式再次跃上一个新的台阶。为此，我们还建立了数字化平台，在这个平台上，为适应我国数字化教学、为教材立体化建设作出尝试。

在编写第3套教材时，一些曾担任第2套教材的主编由于年事已高，此次不再担任主编，但他们对改版工作提出了很多宝贵的意见。前两套教材的作者为本套教材的日臻完善打下了坚实的基础。对他们所作出的贡献，我们表示衷心的感谢。

尽管本套教材的编者都是多年工作在教学第一线的教师，但基于现有的水平，书中难免存在不当之处，欢迎广大师生和读者批评指正。

王德炳 柯杨

2013年11月

前 言

《计算机应用基础》是面向医药类院校非计算机专业学生的教材，从1994年第一版到这次第六版的出版发行，每次再版都凝聚了编者们的辛勤汗水和智慧的结晶。

当我们回顾20年来我国社会信息化的发展，会感叹我们正处在一个高速发展的信息时代，对于从事大学计算机基础教育的教师，面临着信息技术和计算机知识的快速更新，因此，认真地思考在计算机基础教育中如何针对医学生的信息素养进行能力培养，清晰地认识计算机应用能力的培养需求是十分重要的。

本次教材编写是在2010年教育部高等学校教学指导委员会发布的《高等学校计算机基础教学发展战略研究报告暨计算机基础课程教学基本要求》的整体原则框架下进行的，在再版编写中增加了新的信息技术内容和信息技术在医学中应用的案例导引，力图引导学生以计算思维的思维方式去认识、分析问题，培养学生使用计算机，利用计算机技术来延伸自己处理信息的能力；编写同时也考虑了医药类学生的特点与实际情况，调整了部分章节，压缩了文本编辑等内容，增加了“医学信息学概论”一章，探索了计算机基础教学与医学信息学教学进一步的融合。

本书共分为七章，第一章计算机与网络（郭永青、胡彬、李祥生编写），第二章计算机硬件系统（郭建光编写），第三章软件系统（王静编写），第四章常用应用软件（刘群、李祥生、杨俊丽、黄轶旻、王路漫编写），第五章多媒体技术应用（兰顺碧、王路漫、刘群编写），第六章数据管理（齐惠颖编写），第七章医学信息学概述（李祥生、朱彦慧编写）。在每一章节后面，我们都精心编写了习题和上机练习题，便于读者自测练习和上机操作练习。全书由郭永青、胡加立统稿，郑凤负责排版。

最后，特别感谢北京大学医学部及北京大学医学出版社为计算机基础教学的教材建设所提供的大力支持。

由于水平有限，书中如有错误之处，敬请读者批评指正。

编 者

2014年1月3日

目 录

| | | | |
|---------------------------|-----|--------------------------------------|-----|
| 第一章 计算机与网络 | 1 | 第二节 声音处理软件 | 182 |
| 第一节 电子计算机概述 | 1 | 第三节 Flash 多媒体动画素材的编辑与制作 | 187 |
| 第二节 计算机的组成 | 9 | 第四节 Photoshop 图像处理软件 .. | 197 |
| 第三节 信息数字化与信息技术 | 11 | 第五节 视频编辑 | 209 |
| 第四节 网络应用基础 | 14 | 习题五 | 222 |
| 第五节 计算机与网络安全 | 29 | 第六章 数据管理 | 225 |
| 习题一 | 41 | 第一节 数据库技术概述 | 225 |
| 第二章 计算机硬件系统 | 44 | 第二节 构建数据库 | 233 |
| 第一节 微型计算机硬件系统结构 | 44 | 第三节 查询数据 | 241 |
| 第二节 微型计算机的核心组件 | 45 | 第四节 窗体 | 250 |
| 第三节 输入输出设备 | 56 | 第五节 报表概述 | 256 |
| 习题二 | 58 | 第六节 宏 | 259 |
| 第三章 软件系统 | 60 | 第七节 VBA 编程 | 262 |
| 第一节 计算机软件概述 | 60 | 第八节 应用举例——医院门诊收费系统 | 271 |
| 第二节 操作系统概述 | 61 | 习题六 | 278 |
| 第三节 Windows 7 操作系统 | 67 | 第七章 医学信息学概述 | 282 |
| 第四节 软件开发基础知识 | 84 | 第一节 医学信息的分类与编码 | 282 |
| 第五节 医学相关软件简介 | 92 | 第二节 医学信息系统 | 283 |
| 习题三 | 94 | 第三节 医疗服务类信息系统 | 284 |
| 第四章 常用应用软件 | 97 | 第四节 公共卫生信息系统 | 301 |
| 第一节 文档编辑 | 97 | 习题七 | 306 |
| 第二节 电子表格处理 | 115 | 主要参考书目 | 307 |
| 第三节 幻灯片制作软件 PowerPoint .. | 152 | 名词索引 | 308 |
| 习题四 | 167 | | |
| 第五章 多媒体技术应用 | 177 | | |
| 第一节 多媒体概述 | 177 | | |

第一章 计算机与网络

电子计算机 (electronic computer) 是一种用电来进行各种信息加工的机器, 它可以按照预先编好的程序自动执行各种操作, 以完成信息的输入、存取、加工处理及输出。在当今信息化时代, 计算机是信息自动化处理的最基本、最有效的工具。特别是 20 世纪 40 年代以来, 以电子、通信、计算机和网络技术为标志的第三次技术革命, 计算机技术与通信技术的结合促使计算机网络产生, 计算机网络迅速发展, 给世界带来了很大的变化, 已成为人们生活和工作不可或缺的部分。在以网络为基础的信息社会里, 人们的行为方式、思想方式甚至社会形态都发生了显著的变化, 计算机处理能力正在从多方面为科研工作打开新局面。

第一节 电子计算机概述

回顾计算机发展历史, 自 1946 年世界第一台电子计算机问世, 经历了电子管、半导体、集成电路、大规模集成电路、超大规模集成电路等几代的发展, 其性能提高程度以指数形式增长。超级计算机、光计算机、生物计算机等的研制开发取得了显著成果。计算机技术的发展, 使人类利用“0”和“1”编码技术, 来实现对一切声音、文字、图像和数据的编码和解码, 使各类信息的采集、处理、储存和传输实现了标准化和高速处理。

一、电子计算机发展史

人类的计算技术有着悠久的历史, 我国的祖先发明的算盘至今在某些领域还在使用。在 19 世纪, 由于西方国家生产力的发展, 使普通的计算工具难以完成计算的需要, 因此, 人类一直在寻求新的计算技术。19 世纪 50 年代, 英国数学家乔治·布尔 (George Boole) 创立了逻辑代数, 用二进制进行运算, 是当前电子计算机的数学基础。英国科学家图灵 (Alan Mathison Turing) (图 1-1) 于 1936 年发表的传世论文《论可计算数及其在判定问题中的应用》, 首次提出逻辑机的通用模型, 即“图灵机”, 建立了算法理论, 为计算机的出现提供了重要的理论依据, 被称为计算机之父。1966 年美国计算机协会 (Association for Computing Machinery, ACM) 设立了“A. M 图灵奖”, 专门奖励在计算机科学研究中做出创造性贡献、推动了计算机技术发展的杰出科学家。图灵奖是计算机界最负盛名的奖项, 有“计算机界诺贝尔奖”之称。这个奖项的设立目的之一就是为纪念这位现代计算机奠基者。



图 1-1 Alan Mathison Turing

1946 年 2 月世界上第一台电子计算机在美国宾夕法尼亚大学诞生, 被命名为“电子数字积分计算器” (electronic numerical integrator and calculator, ENIAC)。图 1-2 所示的是 ENIAC 的部分场景。它共使用了 18000 只电子管, 耗电 150 千瓦, 重约 30 吨, 占地约 170 平方米, 每秒能进行 5000 次加法运算。ENIAC 的问世表明了电子计算机时代的到来, 它的出现具有划时代的意义。鉴于 ENIAC 的缺点, 美籍匈牙利数学家冯·诺依曼 (John Von Neumann) (图 1-3) 于 1946 年 6 月发表了《电子计算机装置逻辑结构初探》的论文, 提出三个要

点：其一是计算机所有数据和程序都采用二进制；其二是将程序和指令顺序存放在内存存储器，且能自动依次执行指令；其三是计算机由输入设备、输出设备、内存存储器、运算器和控制器五大部分组成。采用以上结构的计算机，被称为冯·诺依曼结构计算机。

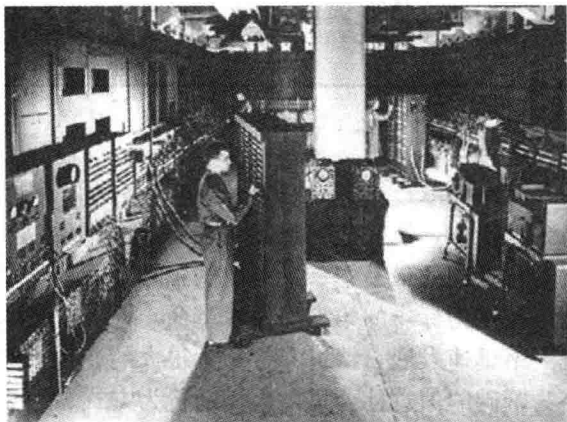


图 1-2 第一台电子计算机 ENIAC



图 1-3 John Von Neuman

英国剑桥大学威尔克斯教授 (M. V. Wilkes) 在 1946 年接受了冯·诺依曼的存储程序计算机结构原理后，在剑桥大学设计了 EDSAC (electronic delay storage automatic computer) 计算机，于 1949 年 5 月研制成功并投入运行。它是世界上首台“存储程序”的电子计算机。

1951 年 6 月 14 日，第一台计算机作为商品交付使用，从此计算机从实验室走向社会，标志着人类进入计算机时代。

至今，所有的计算机仍然没有摆脱冯·诺依曼的理论，目前使用的仍是冯·诺依曼结构计算机。

从计算机的诞生到现在，尤其是晶体管与集成电路的发明，使信息技术发生了根本性的变化，21 世纪计算机与通信结合的网络革命，极大地推动了人类社会网络化、信息化的进程。随着移动互联网、物联网等互联网的应用不断深入，人与计算机的关系已经从“人围绕着计算机转”发展到“计算机围绕着人转”，人们完全生活在网络之中。

我国的计算机发展起步较晚，但发展极为迅速，1956 年国家制定 12 年科学规划时，把发展计算机、半导体等技术定为重点学科。1958 年我国组装调试成第一台电子管计算机 (103 机)，1959 年研制成大型通用电子管计算机 (104 机)，其运算速度为 10 000 次/秒。1964 年我国推出了第一批晶体管计算机，其运算速度为 10 万~20 万次/秒。1971 年我国研制成功第三代集成电路计算机。1982 年采用大、中规模集成电路研制成功 16 位计算机 DJS-150。

1983 年长沙国防科技大学推出向量运算速度达 1 亿次的银河 I 巨型计算机。目前世界上只有很少几个国家能生产巨型机，我国是其中之一。我国已形成了相当规模的计算机产业。

2009 年 10 月 29 日，随着我国第一台千万亿次超级计算机——“天河一号”的亮相，中国拥有了历史上计算速度最快的工具，使中国成为继美国之后世界上第二个能够自主研发千万亿次超级计算机的国家。安装在国家超级计算天津中心的天河一号用户单位达到 600 多个，涵盖了石油勘探、地震数据处理、高端装备制造、土木工程设计、航空航天、生物医药、天气预报与气候研究、海洋环境研究、新能源、新材料、基础科学研究、动漫与影视渲染等应用领域。

2013 年 6 月 17 日我国研制成功世界上首台 5 亿亿次 (50PFLOPS) 超级计算机——“天河二号” (图 1-4)。这是国家 863 计划在“十二五”高效能计算机重大项目的阶段性成果。天

河二号的双精度浮点运算峰值速度已达到了每秒 5.49 亿亿次。天河二号与天河一号相比，不仅应用领域更加广阔，而且在计算规模、计算精度和计算效率等方面也将大大超过天河一号。以 500 人规模的全基因组信息关联性分析为例，华大基因利用自建系统需 1 年时间，而利用天河二号只需 3 个小时，在有效的运行时间内，天河二号可以满足百万人量级的全基因组分析。另外，天河二号还能在智慧城市、电子商务、云计算与大数据等领域发挥重要作用。

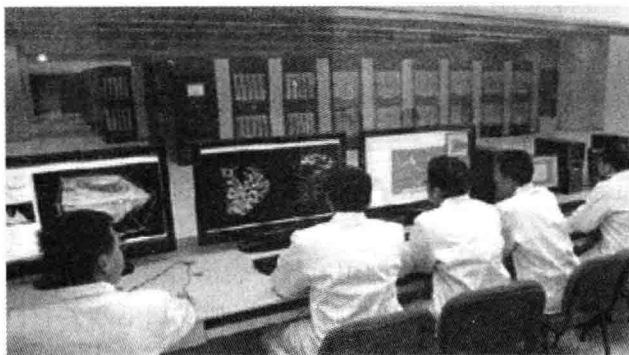


图 1-4 天河二号

二、电子计算机的分类

按用途分类，计算机分为专用计算机和通用计算机。专用计算机一般用于对其他设备的控制，比如工业自动生产流水线、医疗设备的控制分析等。通用计算机就是我们日常所用的可以对各种数据进行加工处理的计算机。

按其规模分类，计算机分为巨型机、大型机、小巨型机、小型机、工作站以及微型计算机（微机）。巨型机和大型机一般用于尖端科学领域，它的功能是最强的，速度和精度也最高。小型机一般用于大中型企业以及比较大的科研单位，功能仅次于巨型机和大型机。工作站一般用于计算机辅助设计。在这些机种中，微机的功能是最弱的，但应用领域最为广泛，大到企、事业单位，小到普通家庭，渗透于各领域。微机又分为台式机、便携机（笔记本电脑），近年来又出现了手持计算机（掌上电脑）、平板电脑、与通讯技术结合的产品智能手机等，总之，这一类计算机越来越小，便于携带甚至穿戴，而功能也越来越多，图 1-5 所示的为各种可穿戴计算机。

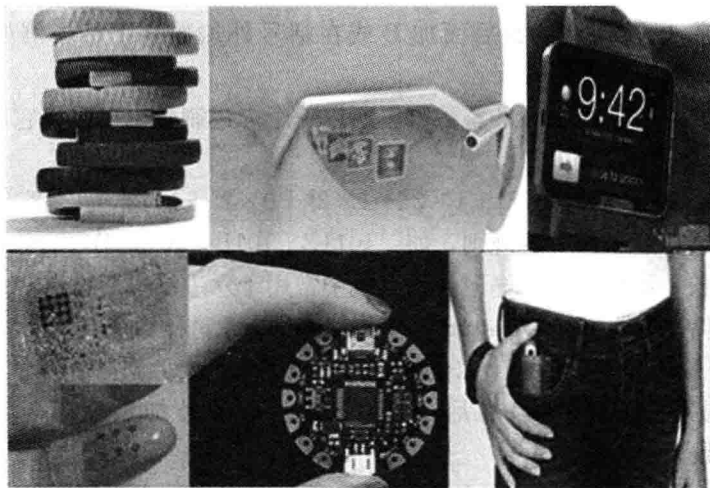


图 1-5 各种可穿戴式计算机

由于微型计算机的发展,使过去由一台大型计算机带若干终端的集中化使用模式向人手一机、独立使用的分散化模式转变。近年计算机网络快速发展,就是由于微机的出现才使得家庭上网得以实现,计算机的计算模式从主机时代到个人计算机时代再到网络时代。这也促进了多媒体技术的发展。

三、计算机中采用的计数制

在日常生活中我们习惯使用十进制,对于十进制可以用 0, 1, 2, …, 9 这十个数码表示,把这些数码的个数称为基数,即十进制的基数为十。采用逢基数进一的规则,则称为进位计数制。

除了十进制,人们也使用其他进制,有时还采用十六进制、六十进制,比如在很早以前使用过十六两为一斤的秤。

由于计算机内部使用的是数字电路,即用电脉冲表示信号,而脉冲信号只有两种状态,电压的有无(即高电平和低电平)、灯光的亮与暗(灯泡加没加电),两种状态都可以用数码 0、1 来表示。而两种状态的电路最容易实现,而且稳定、可靠,用开关的开启和关闭即可实现,只不过开关是由电子开关完成。如果出现三种以上的稳定状态,电路上实现起来就复杂了。所以计算机中处理的各种信息都是用二进制代码来表示的,现在对这些计数制作一简单介绍。

1. 十进制数制(decimal number system) 十进制使用 0~9 这十个数码表示,它的基数是十,它的计数规律为逢十进一。

十进制数的书写规则是将该数后面加 D 或在括号外加数字下标,例如 237.68D 或 $(237.68)_{10}$ 都是表示十进制的 237.68。通常都省略。

在十进制中,数值的大小不仅和其所用的代码有关,还与其所在的位置有关,比如 262.84 这个数,六个代码中出现了两个 2,但它的大小是不一样的,小数点左边的 2 代表 2,最左位上的 2 代表 200,同样是数码 2,但在不同的位置它具有不同的值,我们称之为位权,也称权重(weight)。为便于观察,我们可以把该数展开,即:

$$262.84 = 2 \times 10^2 + 6 \times 10^1 + 2 \times 10^0 + 8 \times 10^{-1} + 4 \times 10^{-2}$$

10^n 、 10^{n-1} 、…、 10^0 、…、 10^{-m+1} 、 10^{-m} 就是我们所说的位权。可见,位权是数码在该位置所具有的值。

2. 二进制数制(binary number system) 二进制使用 0、1 两个数码,基数为 2,计数规则为逢二进一。

二进制数的书写规则是将该数后面加 B 或在括号外加数字下标,例如 $(1011.101)_2$ 或 1011.101B 都表示该数为二进制数。

与十进制相仿,它的位权为 2 的整数幂。所以一个二进制数也可以将它展开,展开后各项值的和是十进制表示的值。例如:

$$101101.11B = 1 \times 2^5 + 0 \times 2^4 + 1 \times 2^3 + 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 + 1 \times 2^{-1} + 1 \times 2^{-2}$$

二进制相加,遵照逢二进一的规则,如: $1011B + 101B = 10000B$

二进制数书写长,不好读,不好记。计算机中常用十六进制来对二进制进行“缩写”。

3. 十六进制数制(hexadecimal number system) 十六进制数基数为十六,使用 0, 1, 2, 3, …, 9, A, B, C, D, E, F 这十六个数码,其规则为逢十六进一。

十六进制的书写方法为将该数后面加 H 或在括号外加数字下标,例如 13D2H 或 $(13D2)_{16}$ 都表示该数为十六进制。

与十进制相仿,它的位权为 16 的整数幂。一个十六进制数也可以将它展开,展开后各项值之和是十进制表示的值。例如: $(13D8)_{16} = 1 \times 16^3 + 3 \times 16^2 + 13 \times 16^1 + 8 \times 16^0$

4. 各种进制数之间的转换

(1) 将二进制数转换为十进制数：将二进制数展开，然后用十进制运算规则计算每一项的值再相加，即可转换为十进制数，例如：

$$(1011.101)_2 = 1 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 1 \times 2^0 + 1 \times 2^{-1} + 0 \times 2^{-2} + 1 \times 2^{-3} \\ = 8 + 0 + 2 + 1 + 0.5 + 0 + 0.125 = (11.625)_{10}$$

(2) 将十进制数转换为二进制数：可以将整数部分和小数部分分开，整数部分采用除 2 取余逆排法；小数部分采用乘 2 取整顺排法。例如：将 $(13.375)_{10}$ 转换为二进制，方法为：

整数部分：

小数部分：

| | | | | | | |
|---|----|-------|----------------|------------------|---------|-----------------|
| 2 | 13 | 余...1 | 2 ⁰ | 0.375 | | |
| | 6 | 余...0 | 2 ¹ | $\times \quad 2$ | | |
| | 3 | 余...1 | 2 ² | 0.750 | 取整数...0 | 2 ⁻¹ |
| | 1 | 余...1 | 2 ³ | $\times \quad 2$ | | |
| | 0 | | | 1.500 | 取整数...1 | 2 ⁻² |
| | | | | $\times \quad 2$ | | |
| | | | | 1.000 | 取整数...1 | 2 ⁻³ |

即： $(13.375)_{10} = (1101.011)_2$

(3) 将二进制数转换为十六进制数：通过表 1-1 可以找出规律：一位十六进制数可以用四位二进制数表示，因此二进制的整数部分转换为十六进制时，只需从二进制数的小数点往左，每四位为一组，与一位十六进制数相对应，最后若不够四位，可以在其左端用 0 补齐。二进制小数部分转换为十六进制数时，则以小数点开始往右，每四位一组，最后若不够四位，在其右端用 0 补足。

例如：将二进制数 $(10110101101.10101)_2$ 转换为十六进制数，方法为：

$$\frac{0101 \ 1010 \ 1101}{5 \quad A \quad D} . \frac{1010 \ 1000}{A \quad 8}$$

因此， $(10110101101.10101)_2 = (5AD.A8)_{16}$

表 1-1 十进制、二进制和十六进制关系对照表

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

(4) 将十六进制数转换为二进制：将十六进制数的每位用四位二进制表示，转换方法是：对整数部分，小数点以左，每一位十六进制数用相应的四位二进制数表示，不足四位时，在其左端添“0”补足。小数部分则是从小数点开始往右，用四位二进制数表示一个十六进制数。

例如：将 $(5A3B.AF)_{16}$ 转换为二进制数，其方法如下：

$$\begin{array}{cccccc} 5 & A & 3 & B & . & A & F \\ \hline 0101 & 1010 & 0011 & 1011 & . & 1010 & 1111 \end{array}$$

因此： $(5A3B.AF)_{16} = (101101000111011.10101111)_2$

(5) 十六进制和十进制数之间的转换：整数部分可以采用除十六取余法，小数部分采用乘 16 取整法，但较为复杂，建议通过二进制作为一个桥梁进行转换。

四、计算机中的数据信息的表示

1. 符号数据的表示 众所周知计算机不仅可以处理数值数据，还可以处理非数值数据，如字符、汉字、图形图像、音频视频等。在计算机中，所有的数据都要有自己的编码，只不过是二进制代码表示，以美国信息交换标准码 ASCII 码 (American Standard Code for Information Interchange) 为例，比如英文字母“A”的代码为 01000001，“B”的代码为 01000010，就像学生有学号一样。ASCII 码是一种用于信息交换的美国标准代码，使用 7 位二进制数来表示所有的大写和小写字母、数字 0~9、标点符号，以及在美式英语中使用的特殊控制字符，一共有 128 种 (0~127) 组合，见表 1-2。

计算机存储通常用 8 位二进制代码作为一个存储单位，即 8 个二进制位称为一个字节，用 Byte 表示，它就是后面章节要讲到的计算机存储器容量的单位“字节”。8 位二进制代码其中的某一位，不管它是 0 或是 1，我们把它记作 1 个信息单位，称为 1 个比特，用英文 bit 表示。在计算机的存储单元中，一个 ASCII 码值占一个字节，其最高位 (b_7) 为空，用作校验位，后 7 位用于字符编码，即 $b_7 b_6 b_5 b_4 b_3 b_2 b_1 b_0$ ，其中 $b_7=0$ 。

表 1-2 英文字符 ASCII 表

| b_7 b_6 b_5 b_4 | b_3 b_2 b_1 b_0 | 000 | 001 | 010 | 011 | 100 | 101 | 110 | 111 |
|-------------------------|-------------------------|-----|-----|-----|-----|-----|-----|-----|-----|
| 0 0 0 0 | | NUL | DLE | SP | 0 | @ | P | ` | p |
| 0 0 0 1 | | SOH | DC1 | ! | 1 | A | Q | a | q |
| 0 0 1 0 | | STX | DC2 | " | 2 | B | R | b | r |
| 0 0 1 1 | | ETX | DC3 | # | 3 | C | S | c | s |
| 0 1 0 0 | | EOT | DC4 | \$ | 4 | D | T | d | t |
| 0 1 0 1 | | ENQ | NAK | % | 5 | E | U | e | u |
| 0 1 1 0 | | ACK | SYN | & | 6 | F | V | f | v |
| 0 1 1 1 | | BEL | ETB | ' | 7 | G | W | g | w |
| 1 0 0 0 | | BS | CAN | (| 8 | H | X | h | x |
| 1 0 0 1 | | HT | EM |) | 9 | I | Y | i | y |
| 1 0 1 0 | | LF | SUB | * | : | J | Z | j | z |
| 1 0 1 1 | | VT | ESC | + | ; | K | [| k | { |
| 1 1 0 0 | | FF | FS | , | < | L | \ | l | |
| 1 1 0 1 | | CR | GS | - | = | M |] | m | } |
| 1 1 1 0 | | SO | RS | . | > | N | ↑ | n | ~ |
| 1 1 1 1 | | SI | US | / | ? | O | _ | o | DEL |

在表 1-2 中上横栏为 ASCII 码的高四位，由于 b_7 为校验位，所以未标出，左面竖栏为低四位，ASCII 码可以用十进制或十六进制表示。比如：数字 0，从表 1-2 中可以得到其 $b_6 b_5 b_4$ 为 011，而 $b_3 b_2 b_1 b_0$ 为 0000，所以它的 ASCII 码为 0110000，它的机内码是在 ASCII 码最高位前加校验位 0 构成一个字节，即为 00110000，其十进制表示为 48。

2. 汉字编码 英文字符数量相对较少而且它们本身有序，所以对它们进行数字化编码是较为容易的。而对汉字字符进行数字化编码难度要大得多，因为汉字既多又复杂，几千个汉字

要对应几千个编码,除了这些,汉字的字形比起其他国家的文字要复杂得多,不同的汉字有不同的形状,即便是同一汉字,又有宋体、楷体等多种字体,每个汉字之间又缺乏联在性,所以对汉字进行编码要考虑很多因素,比如汉字的排列顺序、汉字如何输入以及汉字字形如何在计算机中表示等。

(1) 常用的汉字信息编码标准:在汉字信息编码标准中,常用的是简体中文 GB2312、GB18030、繁体中文 Big5 码等。

GB2312 码是中华人民共和国国家汉字信息交换用编码,全称《信息交换用汉字编码字符集基本集》,由国家标准总局发布,1981年5月1日实施,简称国标码。通行于中国内地,新加坡等地也使用此编码。GB2312 收录简化汉字及符号、字母、日文假名等共 7445 个图形字符,其中汉字占 6763 个。GB2312 规定“对任意一个图形字符都采用两个字节表示,每个字节均采用七位编码表示”,习惯上称第一个字节为“高字节”,第二个字节为“低字节”。该字符集是几乎所有的中文系统和国际化的软件都支持的中文字符集,这也是最基本的中文字符集。

由于 GB2312-80 仅收汉字 6763 个,这大大少于现有汉字。随着计算机的广泛应用,国标 GB2312-80 已不能适应发展需要,为了解决这些问题,以及配合电脑业界组织的 UNICODE 的实施,全国信息化技术委员会于 1995 年 12 月 1 日发布了《汉字内码扩展规范》,之后原信息产业部和国家质量技术监督局于 2000 年 3 月 17 日发布了两项新的国家标准:GB18030-2000 和 GB18031-2000。GB18030-2000《信息技术信息交换用汉字编码字符集基本集的扩充》(简称 GBK),共收录了 27 484 个汉字,具体规定了图形字符的单字节编码和双字节编码,并对四字节编码体系结构做出了规定。该标准是一个强制性标准。与现有的绝大多数操作系统、中文平台在计算机内码一级兼容,能够支持现有的应用系统。

Big5 是中国台湾地区财团法人信息工业策进会于 1984 年制定的,CNS 11643-1992 (Chinese National Standard) 是其扩展版本。

Hong Kong GCCS (Government Chinese Character Set) 是中国香港地区政府在 Big5 基础上增加了 3049 个字符;之后又制定了 HKSCS 标准,它包括了 Big5 和 ISO10646 的编码,也可以说 HKSCS 是 GCCS 的增强版。

一般来讲,计算机内汉字编码中包括机内码、输入码和汉字输出码。

(2) 机内码:以国标码 GB2312 码为例,在计算机内,如果直接采用国标码,势必会造成与 ASCII 码混淆,例如:汉字“大”的国标码为 00110100 01110011,而数字“4”和“s”的 ASCII 码分别为 00110100、01110011,如果不加以指定,计算机会把 0011010001110011 当成两个英文字符 4s 来处理。鉴于以上情况,必须把国标码变成机内码才可以让计算机处理,方法是将每个字节的最高位置 1,只要每个字节的最高位为 1 即为汉字,这样就构成了汉字机内码。无论是国标码还是机内码,书写时都可用十六进制。还是以汉字的“大”为例,它的国标码、机内码和 ASCII 码有如下对应关系:

| 名称 | 编码(十六进制) | 编码(二进制) | |
|---------|----------|-------------------|----------|
| 国标码 | 3473 | 00110100 01110011 | |
| 机内码 | B4F3 | 10110100 11110011 | |
| ASCII 码 | 3473 | 00110100 01110011 | 代表英文“4s” |

由此可见,如果用十六进制表示,将国标码转为机内码的方法只需将国标码加 8080 即可。即:

$$\begin{array}{r} 3473 \quad \text{国标码} \\ +8080 \\ \hline B4F3 \quad \text{机内码} \end{array}$$

(3) 输入码:由于汉字的独立性,使汉字的输入变得较为复杂,常用的汉字输入法基本分

为两大类:

1) 编码汉字输入: 编码汉字输入现基本分三类, 以音为主的拼音输入、以形为主的笔形输入以及音形结合的输入方法。它们各有特色, 拼音易学好记, 但相同读音的汉字太多, 即重码率高, 检字困难, 典型的拼音输入法有全拼输入法和微软拼音输入法。由于相同形状的汉字很少, 所以笔形输入重码率低, 但掌握困难, 典型的笔形输入是五笔输入编码。无论采用哪一种, 它们都称为输入码, 也称外码。当向计算机输入外码时, 一般都要转换成机内码后才能进行存储和处理, 当然这是各种汉字操作系统所要解决的问题, 使用者只需输入汉字的外码, 剩下的由计算机自己处理。

2) 非编码汉字输入: 近年人们发明了不少用于汉字输入的设备, 如手写板输入和语音输入。手写板输入需要一特定的硬件(手写板), 用户可以将汉字手工写在手写板上, 计算机自动将手写体识别成可编辑的文本, 这种方法只要会写汉字就可以向计算机输入汉字, 如果所写汉字不是很潦草, 识别率还是很高的。语音输入是指用麦克风按正常的说话速度朗读, 计算机通过声卡和识别软件, 将语音自动识别成可编辑的文本, 但这种方法存在个体差异, 即由于每个人的声调、发音均不同, 会造成识别错误, 因此要对计算机进行训练, 让计算机逐渐能够适应用户的发音, 才能有较高的识别率。除上述两种非编码输入之外, 用扫描仪将书报上的文稿以图像的形式扫到计算机中, 再通过识别软件(optical character recognition, OCR)进行识别, 还原成可编辑的文本, 如果原稿比较清楚, 其识别率可达90%以上。

(4) 汉字输出码: 汉字输出码是地址码、字形存储码和字形码的统称。

地址码: 是指汉字字形信息在汉字字模库中存放的首地址。每个汉字在字库中占有一个固定大小的连续区域, 其中首地址即是该汉字的地址码。

字形存储码: 不同的字体有不同的字库, 如黑体、仿宋体等, 点阵的点数越多字的质量越高, 越美观。

由于汉字都是方块字, 每个汉字看做是一个有 M 行 N 列点组成的矩阵, 称为汉字的点阵字模, 简称点阵。如果用二进制数 1 代表点阵中的黑点, 用 0 表示无黑点。一个汉字若用 16×16 点阵表示, 则共有 256 个点, 如图 1-6 所示。

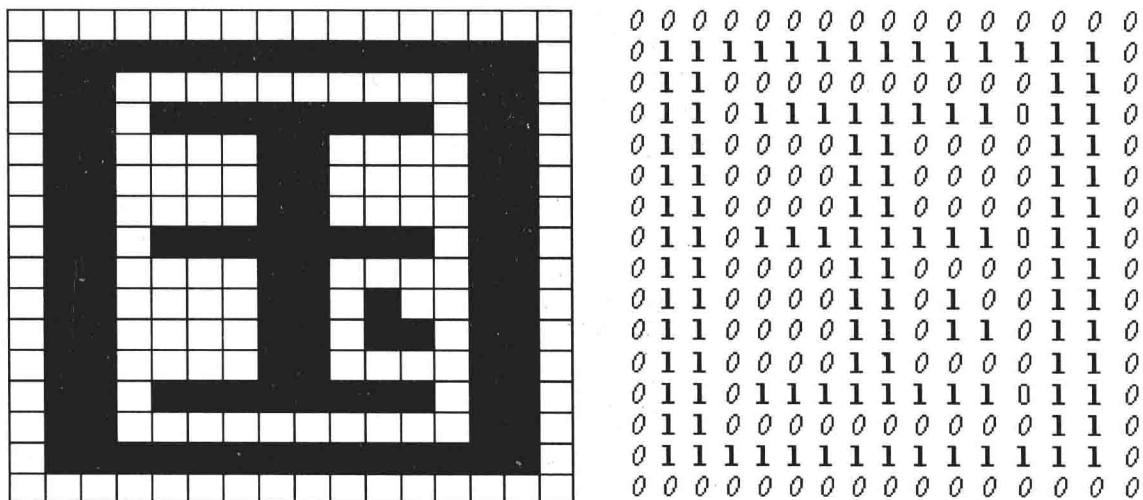


图 1-6 汉字字形码示意图

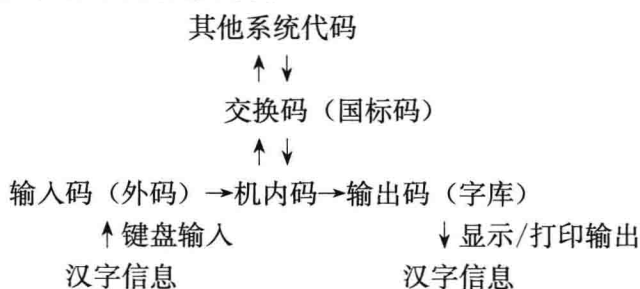
由图 1-6 可以看出, 一个汉字可以用 16 行二进制代码表示, 一行为 16 位, 正好为 2 个字节, 所以一个 16 点阵字库要占 $16 \times 2 = 32$ 个字节。对于 16 点阵字库要用于打印, 其质量显而易见, 所以 16 点阵字库主要用于显示, 真正用于打印, 应采用 24 点阵以上的字库, 而 24 点

阵字库每个汉字要占 $24 \times 3 = 72$ 个字节。

汉字字形库直接存储点阵码时占用的存储空间大，为了减少字库所占的容量，采用了数据压缩技术。使用较多的字库压缩方法有哈夫曼树法、矢量法和字根压缩法。近年来开发的新的汉字操作系统中常使用矢量汉字。所谓矢量汉字就是经过矢量法把基本点阵字模进行压缩后得到的汉字。这些汉字信息存在矢量字库中，显示和打印时要经过相应的转换程序进行还原和变换，得到不同的字体。

字形码：指在输出设备上输出汉字时所送出的汉字字形点阵码。点阵数据的组织是按照输出设备的特性及输出字体的一些特点（如倾斜角度、放大倍数）进行的，是对基本字库中数据进行变换得到的。

以上所介绍的各种汉字编码之间的关系为：



第二节 计算机的组成

一个完整的计算机系统是由硬件系统和软件系统两大部分组成。所谓硬件就是构成计算机实体的所有器件，而软件是指那些看不到、摸不着却又实实在在存在的那些计算机中存储的数据、程序等。软件系统是计算机的灵魂，以硬件系统为依托，对硬件设备进行控制和管理。只有硬件、软件系统相互结合，才能发挥计算机系统的强大功能。

一、计算机硬件

从原理上讲，计算机是由输入设备、输出设备、存储器、控制器和运算器五大部分组成，如图 1-7 所示。下面简单地讲述各部分的功能。

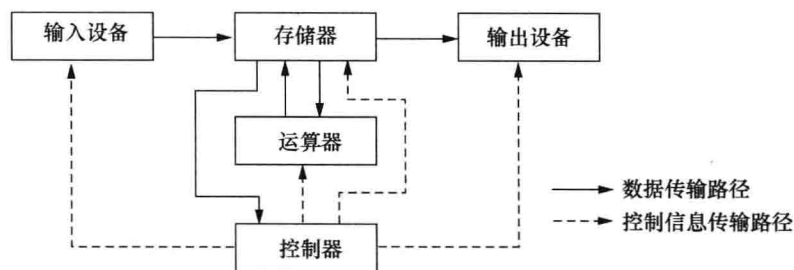


图 1-7 计算机硬件的基本组成

1. 运算器 (calculator) 运算器也称为算术逻辑运算单元 (arithmetic and logic unit, ALU)，计算机中所有的算术运算、逻辑运算和信息传送都在这里进行，它由加法器、移位电路、逻辑部件、信息传送部件以及寄存器等电路组成。由于任何数学运算最终可以用加法和移位这两种基本操作来完成，因而加法器是 ALU 的核心部件。寄存器则用来暂时存放参与运算的操作数和运算结果。

2. 控制器 (controller) 控制器也称控制电路 (control circuit), 它就像一个部队的指挥部, 是整个计算机的控制中心, 其任务是按预定的顺序不断取出指令进行分析, 然后根据指令的要求向运算器、存储器等各部分发出控制信号让其完成指令所规定的操作。指令是一条命令, 就是让计算机做什么, 是由一串二进制代码组成, 不同的代码表示不同的命令, 是给计算机约定好的。控制电路由指令计数器、指令寄存器、指令译码器和操作控制部件等组成。指令计数器用于提供指令的存放地址。指令寄存器可以把从存储器取出的指令暂存起来。指令译码器是把取到的指令译成操作控制部分所能识别的信号, 使其完成指令所规定的操作。

3. 存储器 (memory) 存储器是计算机存放数据的地方, 它由一片片连续的存储单元组成, 每个单元都赋有编号, 称为地址, 就像楼房的房间号一样, 每个单元都可以存放一组二进制代码, 就像房间里住的是谁。信息存入内存的过程称为写入, 取出的过程称为读出。存储器的基本指标是容量和读写速度。存储器分内部存储器和外部存储器, 内部存储器可以由 CPU 直接访问, 内存的读写速度快但其存储空间是有限的, 外部存储器作为内部存储器的扩展存储, 存储容量大但读写速度相对较慢。存储器的计量单位是字节 (Byte), 由于计算机存储器容量很大, 所以用字节很不方便, 一般用千字节 (KB)、兆字节 (MB) 等表示, 其换算如下:

$$1\text{KB}=1024\text{Byte} (2^{10})$$

$$1\text{MB}=1024\text{KB} (2^{20})$$

$$1\text{GB}=1024\text{MB} (2^{30})$$

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

4. 输入设备 (input device) 用于将数据和信息输入到计算机的设备称为输入设备, 键盘和鼠标是最基本的输入设备, 此外还有扫描仪、数码相机、触摸屏等。

5. 输出设备 (output device) 用于将计算机内的数据输出的设备称为输出设备, 显示器、打印机是计算机最基本的输出设备, 其他输出设备还有绘图仪、投影机等。

二、计算机软件

计算机软件是指在硬件设备上运行的各种程序和有关资料。程序是计算机完成指定任务指令的集合。用户使用程序时不仅需要程序, 还需要关于它的说明和其他资料, 这些资料通常称为文档, 因此, 软件包括程序和文档。

软件分系统软件和应用软件两大类。

1. 系统软件 用于管理、监控和维护计算机硬件资源和软件资源的软件称为系统软件。系统软件包括操作系统、语言处理系统和数据库管理系统, 如常用的 Windows 操作系统、Linux 操作系统等。

2. 应用软件 应用软件是针对某一个专门目的而开发的软件, 如文字处理软件、表格处理软件、图形处理软件、财务管理系统、辅助教学软件、用于各种科学计算的软件包等。

目前广泛使用的应用软件有: 办公软件 WPS, 文字处理软件 Word、电子表格软件 Excel、图形处理软件 Photoshop、计算机辅助设计软件 AutoCAD、动画处理软件 3ds Max 和 Flash5、多媒体制作软件 Authorware、卫生统计分析软件包 SAS 和 SPSS 等。

此外, 还有一些常用的工具软件, 使用起来很方便, 如图 1-8 所示的是某网站提供的可以下载的常见的工具软件。