第1章

1.1 信息与计算机

虽然人们很早就开始自觉或不自觉地使用和处理信息，也产生了一些原始简单的信息处理技术，但真正意义上的信息技术（即现代信息技术），却是在最近几十年内迅速发展起来的。众所周知，现代信息技术的发展彻底改变了人们工作、学习和生活的方式。在这一改变中，计算机起了举足轻重的作用，无论是从信息的获得和存储，还是从信息的加工、传输和发布来看，它都是最主要的处理工具。本节主要讨论信息的基本概念以及计算机的发展和主要应用领域。

1.1.1 信息与信息技术

一、什么是信息

人们由于研究目的和角度不同，对信息的理解和解释也不尽相同。控制论的创始人维纳认为，信息是人们在适应外部世界并且将这种适应反作用于世界的过程中，同外部世界进行交换的内容的名称；接收信息和使用信息的过程，就是我们适应外部偶然性的过程。信息论的创始人香农说，信息是用以消除不确定性的东西。决策学的代表人物西蒙则提出，信息是影响人们改变对决策方案的期待或评价的外界刺激。

在信息技术应用领域，一般认为信息是经过加工、具有一定含义的且对决策有价值的数据。由此也可看出，信息的表达是以数据为基础的。例如，“42%”是一项数据，但这一数据除了数字上的意义外，并不表示任何内容，而“张三得到 42%的选票”对接收者是有意义的，接收者知道“42%”是表示客观实体张三的得票率这一属性值。因此，“张三得到 42%的选票”不仅仅有数据，更重要的是对数据有一定的解释，从而使接收者得到了客观实体张三的得票率信息。若再加一条信息“得票率大于 40%，即可当选委员”。综合以上两条信息之后可以得出一条抽象程度更高的信息，即“张三可当选委员”。由此可见，数据和信息是密不可分的，而信息之间的联系又可以得到抽象层次更高的信息。

二、信息的基本特征

1．普遍性和无限性

由于信息反映的内容主要是事物的状态和特征，而这两者都是普遍存在的，这就决定了信息的普遍性。人类认识、接收和利用信息是无止境的，这又决定了信息的无限性。

2．寄载性

寄载性有两个层次的含义，一是信息必须借助于某种特定的“符号”来表示，如文字、图形和图像等；二是信息符号必须寄载于一定的物理介质上，如纸张、磁盘等。信息的这一特征

需要全本请访问 www.ertongbook.com

要求我们设计和选择恰当的信息载体，并对其进行科学的编码，才能使信息的传递、保存、加工与利用更加方便有效。

3．时效性

客观事物（或系统）都是在不断发展变化的，信息只有及时、新颖才有价值，才能发挥巨大的作用。换句话说，一条信息在某一时刻可能有很高的价值，但过了这一时刻它的价值可能大大降低，甚至没有。

4．可处理性

可处理性是指信息的内容是可以被识别的，信息的形式是可以转换或变换的。信息可以被各种方法多环节地加工和处理，而经过某些处理（如分析、综合和提炼）后的信息可以比原始信息更具有价值。

5．共享性

信息可无限扩散，信息本身不会因为知道的人数增加而减少。但却可能会因信息被分享，而使信息的所有者蒙受损失，如企业的技术专利、军事动态等信息方面的共享。为了避免信息共享给信息的所有者造成损失，信息共享往往是有范围（区域上、时间上）和有条件的。

6．层次性

信息的层次性表现在 3 个方面：一是一般信息与重要信息；二是上层信息与下层信息；三是表层信息与深层信息。不同层次的信息，其作用是不同的。收集不同层次信息的目的是能够准确、深入地把握信息所反映事物的真实内涵。只有合理地确定层次，才能正确地确定信息需求的范围和信息的价值，并有效地进行信息处理。

7．传递性

信息在时间和空间上都具有传递性。从时间的延续性讲，信息可以借助于各种载体被代代相传，信息在时间上的传递叫做信息的存储；从空间转移的角度讲，信息也可以从一个位置传递到另一个位置，信息在空间上的传递叫做通信。

除了以上特征外，信息还有许多其他特征，如主观性、抽象性、整体性、滞后性和不完全性等。

三、信息技术

信息技术，最简单的理解就是人们处理信息的相关技术。它是随着人类的出现而出现，随着人类文明的进步而不断发展起来的。尤其是在最近二三十年，科学技术得到了有史以来最迅猛的发展，各种高新技术如雨后春笋般纷纷出现，借助于这些高新技术，信息技术也得到了前所未有的发展，而且已经成为当代新技术革命最活跃的领域。

我们现在所讲的信息技术一般是指最近几十年刚刚发展起来的现代信息技术，它是指利用电子计算机和现代通信手段实现信息的获取、传递、存储、处理、显示和分配等相关技术。

具体来讲，信息技术主要包括以下几方面：

1．感测与识别技术

感测技术包括传感技术和测量技术，它的作用是扩展人类获取信息的感觉器官的功能，尤其是由传感技术、测量技术和通信技术相结合而产生的遥感技术，更使人感知信息的能力得到进一步的加强。信息识别包括文字识别、语音识别和图形识别等，识别技术的实现通常要借助

于一种叫作“模式识别”的方法。

2．信息传递技术

它的主要功能是实现信息快速、可靠、安全的转移。各种通信技术都属于这个范畴。广播技术也是一种传递信息的技术。由于存储、记录可以看成是从“现在”向“未来”或从“过去”向“现在”传递信息的一种活动，因而也可将它看作是信息传递技术的一种。

3．信息处理与再生技术

信息处理包括对信息的编码、压缩、加密等。在对信息进行处理的基础上，还可形成一些新的更深层次的决策信息，即信息的“再生”。信息的处理与再生都依赖于现代电子计算机的超凡功能。

4．信息施用技术

它是信息过程的最后环节，包括控制技术、显示技术等。

从以上分析可以看出，现代信息技术是一种综合性技术，它的支柱技术主要包括计算机（Computer）技术、通信（Communication）技术和控制（Control）技术，即“3C”技术。

当今时代，各个领域处理的信息量越来越大，处理的信息形式越来越复杂，这就要求各类专业技术人员能利用信息技术开发出各种实用的信息产品应用于这些领域，加快这些领域信息处理的效率，而信息技术的发展又为它的全面应用铺平了道路。目前信息技术的应用领域主要包括：办公、工业、科研教育、医疗保健、军事和家庭等。

1.1.2 信息化与信息社会

一、信息化

信息化是近年来世界各国都非常关注的并具有深远影响的战略课题。与此相应，有关未来信息社会的种种构想与预测也在不同的媒体中出现，以不同的方式被公众所了解。

信息化也称国民经济和社会信息化，是指在国民经济和社会各个领域，不断推广和应用计算机、通信、网络等信息技术和其他相关智能技术，以达到全面提高经济运行效率、劳动生产率、企业核心竞争力和人民生活质量的目的。信息化是工业社会向信息社会的动态发展过程。在这一过程中，信息产业在国民经济中所占比重上升，工业化与信息化的结合日益密切，信息资源成为重要的生产要素。

信息化的内容包括信息的生产和应用两大方面。信息生产要求发展一系列高新信息技术及其产业，既涉及微电子产品、通信器材和设施、计算机软硬件、网络设备的制造等领域，又涉及信息和数据的采集、处理、存储等领域。信息技术的应用主要表现在用信息技术改造和提升农业、工业、服务业等传统产业上。

信息化已经成为当今世界经济和社会发展的趋势，这种趋势主要表现在：

（1）信息技术突飞猛进，成为新技术革命的领头羊。

（2）信息产业高速发展，成为经济发展的强大推动力。

（3）信息网络迅速崛起，成为社会和经济活动的重要依托。目前，世界各国正在积极应对信息化的机遇和挑战。

目前，世界各国正在积极应对信息化的机遇和挑战。

二、信息社会

信息社会也称为信息化社会，一般是指这样一种社会：信息产业高度发达且在产业结构中占据优势，信息技术高度发展且在社会经济发展中广泛应用，信息资源充分开发利用且成为经济增长的基本资源。

信息社会是信息化的必然结果，但信息社会与工业社会并没有一个严格的界限，应该是一个渐进的、逐步的演化过程。尽管如此，信息社会还是会表现出一些与工业社会不同的地方，这些特征主要包括：

（1）由于人类脑力劳动的相当部分可以由计算机信息系统来代替，从而极大地增强了人类处理信息的能力。

（2）社会信息交换将很大程度上围绕信息网络及其服务中心展开。

（3）信息产业将成为社会的主要支柱产业之一。

（4）专业技术阶层成为社会的主导阶层和中坚力量。

（5）在社会生产中，体力劳动和资源的投入相对减少，脑力劳动和科学技术的投入相对增加。

1.1.3 信息处理的历史与计算机

一、信息处理的历史

信息作为一种社会资源自古就有，一直以来人类也在不断地利用各种信息资源。要利用信息资源，就需要对它进行加工和处理。我们可以把人类处理信息的历史大致分为 4 个阶段。

1．信息处理的原始阶段

在这一阶段，人类主要是通过语言来传播和继承知识及信息。除此以外，人类也已通过结绳记事、筹码计算等手段，开始超出大自然所赋予的器官与功能，借助各种“身外之物”来帮助自己处理信息。因此，可以说信息技术的萌芽在这里已经显现。原始人类在利用工具“延长”自己的手臂，利用钻木取火获取能量的同时，也已经开始利用各种原始的技术协助自己处理信息。

2．信息处理的手工阶段

信息处理手段的第一次质的飞跃是文字的产生与使用。文字的出现使人们在信息的存储方面有了重大的突破。作为一个整体、一个民族、一个部落有了独立于个别人的头脑之外的、可靠稳定的、不受时间与空间限制的、共同的信息存储形式。这与只靠语言来传播和继承知识与信息的时代相比，无疑是一个极大的进步。与此同时，信息的存储、加工、传递和显示等方面，在这一阶段也有了相应的进步与发展。如纸张的产生与印刷术的进步使信息的存储和传递有了很好的载体；同样，从古代的筹算到流传至今的算盘，也是早期信息处理技术的典型例子；遍布全国的烽火台系统和驿道系统同样表现出我们的祖先为加快信息传递速度而做出的巨大努力；我国古代发明的指南针则是原始的感测技术和显示技术。

3．信息处理的机电阶段

以蒸汽机的出现为标志，工业革命在物质和能量的使用方面开创了一个全新的时代。在信息处理方面，工业革命的思想与技术同样产生了一系列成果。例如，帕斯卡发明的机械计算机，这种设备可以在一定程度上帮助人们从事大量数据的累加、乘除等运算。以其为原形发展起来的手摇计算机直到 20 世纪 60 年代初还在许多地方使用。在信息的加工与传递上，由于电的使

用，人类又发明了一系列新的技术，如电报（包括有线的与无线的）和电话。这些技术与设备使人类信息处理能力有了进一步的提高。这个时期可以称为机械与电气为主要手段的机电时代。

4. 信息处理的现代阶段

20 世纪中叶，由于生产社会化程度的空前提高，人类在信息处理方面也进入了一个全新的阶段，我们可以称为信息处理的现代阶段，或信息处理的电子时代。所谓现代信息技术，就是指在这几十年内迅速发展起来并迅速普及的一系列技术（其核心是计算机技术、现代通信技术和控制技术）。正是这些技术构成了现代信息处理的基础。

二、计算机的发展和应用

计算机技术作为现代信息技术的核心之一，它的发展和应用从根本上改变了人类收集、加工、处理和利用信息的方式方法。要理解信息技术和信息处理的实质，必须首先对计算机有一个完整而全面的了解，下面首先介绍一下计算机的发展历史和主要应用领域。

1. 计算机的发展

1946 年 2 月 15 日，世界上第一台通用电子数字计算机 ENIAC（Electronic Numerical Integrator And Calculator）宣告研制成功。ENIAC 的研制成功，是人类信息技术发展史上的一座里程碑，是在发展计算技术的历程中达到的一个新高度，同时也是一个新起点。



图 1.1 第一台电子计算机

ENIAC 计算机的最初设计方案是由 36 岁的美国工程师莫奇利于 1943 年提出的。总工程师由年仅 24 岁的埃克特担任。ENIAC 共使用了 18000 个电子管，1500 个继电器以及其他器件，安装在面积为 $9\times15m^2$ 的室内，如图 1.1 所示。ENIAC 的存储量很小，只能存放 20 个 10 位的十进制数，运算速度为 5000 次/s 加法运算。虽然 ENIAC 的能力根本无法与现在的计算机相比，但它开启了人类用计算机处理信息的崭新一页。

计算机从最初用电子管作为元器件，发展到今天用超大规模集成电路作为元器件，已走过了近 60 年的历程。习惯上，人们根据计算机所用的逻辑元器件的种类不同对计算机进行了分类，大致上分成 4 个发展阶段。

（1）第一代计算机。以使用电子管为特点的第一代电子计算机在 20 世纪 40 年代末和 50 年代初获得重大发展。除 ENIAC 外，大多数计算机都是依照“存储程序和程序控制”原理设计制造的，其主要代表机型有 IBM-701 和 UNIVAC-1 等。其实，后来（直到现在）的计算机也一直是依照‘存储程序和程序控制”原理设计制造的，由于这原理是由美籍匈牙利数学家冯·诺依曼提出来的，所以习惯上把这一类计算机叫作“冯·诺依曼机”。电子管体积大、功耗高、反应速度慢且寿命短，导致第一代计算机体积大、耗电量大、可靠性差、维护困难且计算速度慢。这个时期的计算机采用延迟线或磁鼓作为内存储器，外存储器开始使用磁带机，存储容量有限。在软件方面，这个时期的计算机只使用机器语言和符号语言，没有高级语言，更没有操作系统等系统软件。

（2）第二代计算机。第二代电子计算机于 20 世纪 50 年代中期问世，以晶体管代替电子管，并增加了浮点运算。第二代计算机的运算速度和工作可靠性都较第一代有明显改善。内存储器采用磁芯，磁盘开始作为外存储器使用，容量较第一代计算机有较大增加。在软件方面，产生

需要全本请访问 www.ertongbook.com

了如 FORTRAN、COBOL和 ALGOL60 等高级程序设计语言，也有了操作系统的雏形——系统管理软件。

（3）第三代计算机。1958 年，世界上第一个集成电路诞生了，它包括一个晶体管、两个电阻和一个电阻与电容的组合。后来集成电路工艺日趋完善，1964 年 4 月 IBM360 系统问世，它成为使用集成电路的第三代电子计算机的著名代表。发展到 20 世纪 70 年代初期，大部分电路元件都已经以集成电路的形式出现，甚至在约 $1cm^2$ 的芯片上，就可以集成上百万个电子元件。因为它看起来只是一块小小的硅片，因此人们常把它称为芯片。与第二代相比，第三代计算机的速度和稳定性有了更大程度的提高。内存储器普遍采用半导体器件，存储容量进一步提高，可靠性和存取速度也有了明显的改善。高级程序语言进一步发展，产生了标准化的高级程序设计语言和人机会话式的 BASIC语言，系统管理程序上升为操作系统。

（4）第四代计算机。在 1967 年和 1977 年，分别出现了大规模集成电路和超大规模集成电路，并于 20 世纪 70 年代中期在电子计算机上得到了应用。由大规模和超大规模集成电路组装成的计算机，就被称为第四代电子计算机。第四代计算机体积更小，可靠性更强，寿命更长。计算机软件的配置空前丰富，软件系统开始工程化、理论化、程序设计部分自动化。

早在 20 世纪 70 年代，人们就开始研制第五代计算机，设想中的第五代计算机应该是具有高智能的，它不仅具有存储和记忆功能，而且应该有学习和掌握知识的机制，并能模拟人的感觉、行为和思维等。尽管至今没有出现真正意义上的第五代计算机，但计算机技术并没有停止不前。这一时期，计算机的硬件性能不断得到提高，软件也得到了空前的发展。

2. 计算机的分类

计算机按用途可分为专用计算机和通用计算机。

专用与通用计算机在其效率、速度、配置、结构复杂程度、造价和适应性等方面是有区别的。

专用计算机针对某类问题能显示出最有效、最快速和最经济的特性，但它的适应性较差，不适于其他方面的应用。我们在导弹和火箭上使用的计算机大部分就是专用计算机。但这些计算机再先进，也不能用它来玩游戏。

通用计算机适应性很强、应用面很广，但其运行效率、速度和经济性依据不同的应用对象会受到不同程度的影响。

通用计算机按其规模、速度和功能等又可分为巨型机、大型机、中型机、小型机、微型机及单片机。这些类型之间的基本区别通常在于其体积大小、结构复杂程度、功率消耗、性能指标、数据存储容量、指令系统以及设备和软件配置等。

一般来说，巨型计算机的运算速度很高，可达每秒执行几亿条指令甚至更高，数据存储容量很大、规模大结构复杂、价格昂贵，主要用于大型科学计算。它也是衡量一国科技实力的重要标志之一。单片计算机则只由一片集成电路制成，其体积小、重量轻、结构十分简单。性能介于巨型机和单片机之间的就是大型机、中型机、小型机和微型机，它们的性能指标和结构规模则相应地依次递减。

3. 计算机的特点

计算机之所以在信息处理中起了至关重要的作用，是与其处理问题的特点分不开的，计算机的主要特点包括：

（1）处理速度快。

（2）存储容量大，存储时间长久。

（3）计算精确度高。

（4）具有逻辑判断能力。

（5）应用领域广泛。

4. 计算机的应用领域

随着计算机硬件性能的不断提升和软件产品的不断丰富，计算机的应用变得越来越普及，它已经从最初单纯的科学计算，逐渐渗透到人们工作、学习和生活的各个领域，其中最主要的应用有以下几个方面。

（1）科学计算。计算机最早应用于科学计算方面，世界上第一台计算机就是用于研制原子弹而制造的。除了用于解决科学实验和工程技术中所提出的数学问题，计算机在物理、化学、生物和材料等领域的数据测算方面所起的作用也非常明显，航天技术中卫星轨道的计算更是离不开计算机。我们每天收看到的天气预报，也要通过计算机对大量的数据作快速的计算处理，用巨型计算机计算，能快速、及时、准确地获得计算结果。

（2）信息处理。信息处理主要是指非数值形式的数据处理。计算机信息处理在社会和经济发展中的作用越来越为人们所重视。信息处理包括对数据资料的收集、存储、加工、分类、排序、检索和发布等一系列工作。计算机信息处理包括办公自动化（OA）、企业管理、情报检索、报刊编排处理等。计算机数据处理的特点是信息处理及时、数据量大、处理速度快，并能给出各种形式的输出格式。目前计算机应用已深入到经济、金融、保险、商业、教育、档案、公安、法律、行政管理、医疗、社会普查等各个方面。目前计算机在信息处理方面的应用已经远远超过它在其他方面的应用，已占所有应用的 80%左右。

（3）过程控制。在科学技术、军事、工业、农业以至于我们日常生活的各个领域都应用到过程控制。用于过程控制的计算机，先将模拟信息如压力、速度、电压、温度等模拟量转换成数字量，然后再由计算机进行处理。计算机处理后输出数字量结果，再将其转换成模拟量去控制对象。过程控制一般都是实时控制，有时对计算机运算速度的要求不高，但要求可靠性高、响应及时，这样才能保证被控制对象的准确动作。

（4）计算机辅助系统。计算机辅助系统有计算机辅助教学(CAI)、计算机辅助设计(CAD)、计算机辅助制造（CAM）、计算机辅助测试（CAT）和计算机集成制造（CIMS）等系统。

1.1.4 信息系统及其应用

信息系统是一种专门的系统，是一种通过对各种数据进行采集、处理、传播，产生能解决某方面问题的数据和信息，并按照一定的要求产生决策信息，以实现预期目标的有组织的应用工程。信息系统一般分为 3 种：事务处理系统（Transaction Processing System）、管理信息系统（Management Information System ）和决策支持系统（Decision Support System）。

事务处理系统是用来记录完成某商业交易中的人员、过程、数据以及设备的人机系统。如证券实时交易系统就是一个事务处理系统。过去，证券交易只能通过手工填单或电话委托方式将委托数据报给交易所，这种人工报单方式速度慢、容量小、功能少和实时性差。现在，用户可以直接在系统的客户端进行交易委托，完成交易。另外，系统还可以提供资金存取、成交交

割和查询功能，从而大大提高了证券交易的效率。其他常见的事务处理系统还有航空订票系统、高校招生录取系统和图书借阅系统等。

管理信息系统主要是一个以人为主导，利用计算机、网络通信设备和其他办公设备，对数据进行收集、传输、加工、存储、更新和维护，支持企事业中高层决策、中层控制、基层运作的集成化的人机系统。以办公室信息管理系统为例，过去由人工管理职工的档案、工资明细等，工作量大，查找麻烦，易出错。而通过办公室管理系统管理职工档案、工资信息，则大大方便了查询、修改和更新信息，简化了办公手续，提高了办公效率。现在已经有许多企事业单位开发出针对本单位的管理信息系统，如企业单位的职工管理系统、事业单位的效能监督系统和各类学校的学生管理系统等。

决策支持系统是一种以计算机为基础和工具，应用决策科学和其他有关的理论和方法，辅助决策者进行决策的信息系统。如证券分析系统就是一个决策支持系统。过去，投资人通过手工记录每日股票行情或凭经验买卖股票，这样盲目操作，投资风险大、收益低。现在，各种证券分析系统将每日行情写入决策数据库，根据历史数据以及相关决策理论和方法来分析和预测未来股票走势。投资人可以根据决策系统提供的各种买卖指示信息，及时买卖股票，减少了投资的盲目性，降低了投资风险，提高了投资回报。其他常见的决策支持系统还有企业客户资源分析系统、政府部门各类预警决策系统等。

除了以上 3 种，还有一种非常重要的信息系统叫作专家系统，它是以知识库、推理引擎及其接口为基础组成的信息系统，其目的在于对某一特定领域的问题作出判断、解释及认知。医院专家系统是应用最为广泛的一种专家系统，目前已经开发出的一些针对疑难疾病的专家系统，记录了以往各个专家对某一种疾病的临床经验，它可以为那些需要向专家咨询的医生提供帮助，以确定病人患何种疾病以及如何治疗该疾病。

1.1.5 信息技术的发展趋势

回顾 20 世纪，信息技术的迅猛发展，推进了世界信息化的进程，也推动了信息产业的持续高速发展，创造出了众多的新产品、就业机会和新的财富，改变了人们的工作、生活、消费以及思维方式，提供了经济增长的新模式。展望 21 世纪，一个全新的信息社会的形态给信息技术的应用提供了无限的创造和发展空间。特别是由于信息技术和信息产业主要依靠知识、高技术的投入，具有技术含量高、增长率高、附加值高的特点，随着世界经济由投资拉动型向知识、技术推动型转换，信息技术的发展将成为推动世界经济高速增长的强大动力，信息产业也将成为全球最大的产业。以下就几个主要的方面，简单说明一下信息技术发展的趋势。

（1） 电话网、有线电视网、计算机网（主要指 Internet）“三网合一”将成为信息技术发展的必然。目前已经发展的 IP 电话就是基于 Internet 的数字通信进行的，人们购买了 IP 电话卡就可以利用普通的电话机以更便宜的价格进行通话，这时通信的信道主要是 Internet，传输的是数字信号。将来的发展将是集电视、录放、计算机和 Internet 数据接收等综合功能于一体的独特电视——IP 电视。IP 电视将打破电视机、录放机与计算机之间的界限，实现在宽带网上同时进行数据、图像、语言多媒体信息的高速互动传播和处理，并使各种业务活动都可以在网上方便快捷地进行。因此有人预言：在 21 世纪，IP 电视将会成为各种政治、社会、经济和文化信息交流平台。

（2）数字化将是信息技术发展的又一个必然趋势。一方面数字化设备便于大规模生产，生产一台模拟设备需要花很多时间，模拟电路每一个单独部分都需要进行单独设计和调试，而数字设备是单元式的，设计非常简单，便于大规模生产，可以大大降低成本。另一方面数字化有利于综合，每一个模拟电路其电路物理特性区别都非常大，而数字电路由二进制电路组成，非常便于综合。

（3）集成电路技术、软件技术、计算机技术、通信技术和广播电视技术等多专业技术彼此联系、相互结合、互为支撑的趋势日益明显。

（4）信息技术发展将导致更加高速、更大容量的信息化产品不断出现。无论是计算机的核心 CPU，还是各种存储设备和网络通信设备的速度都将越来越快，容量也将越来越大。

（5）计算机软件将更加开放。各种软件将遵循统一的标准，不同的软件之间可以相互访问，异构数据库也可以互联互访。大量的软件被开发成软件构件形式，从而可以像工业产品一样随意组装。

（6）信息技术的发展将进入后 PC 时代，即将微电子技术、微处理器技术等应用到所有的家用电器上，使家用电器进一步智能化，同时还可在家庭内部建立局域网，使所有的家电都上网。如使家里的空调、冰箱、洗衣机、电视机、移动电话、电话、电饭煲、微波炉等都连接到网络上。这样不仅方便调控，而且还可以在异地通过网络来远程控制自己家里的电器，使人们可以更高效地利用时间。

1.2 信息数字化的方法与技术

由于二进制编码具有运算简单、电路实现方便、成本低廉等优点，所以无论是什么类型的信息，在计算机内部都是以二进制编码形式表示的。这些信息包括数值、文本、图形图像、声音和视频等。本节主要介绍这些信息如何在计算机中被表示及一些相关技术。

1.2.1 数值信息的表示

一、进位计数制

所谓进位计数制，就是按进位的方法进行计数。下面只介绍我们习惯使用的十进制以及与计算机密切相关的二进制和十六进制。

1．十进制

十进制中使用 0、1、2、3、4、5、6、7、8、9 这 10个符号来计数，称为 10 个数码，采用的计数规则是：逢十进一。由此可见，各相邻位之间的“权”之比都是 10。所谓“权”是指在进位数制中，为了确定一个数位的实际值而需乘上的一个因子。如 3468.795 这个数中的 4 就表示 $4\times10^2=400$，这里把 10^2 称作位权（即我们所说“百”位），简称为“权”，十进制数权的一般形式为 10^n（n＝……，2，1，0，-1，-2，……）。每个十进制数都可以表示成按“权”展开的多项式。例如：

$$3468.795=3\times10^3+4\times10^2+6\times10^1+8\times10^0+7\times10^{-1}+9\times10^{-2}+5\times10^{-3}$$

2．二进制

计算机中的数据是以二进制形式存放的，二进制数的数码是用 0 和 1 这两个符号来表示的。

权为 2^n（n=……，2，1，0，-1，-2，……），二进制数的计数规则是：逢二进一。

对于一个二进制数，也可以表示成按权展开的多项式。例如：

$10110.101=1\times2^4+0\times2^3+1\times2^2+1\times2^1+0\times2^0+1\times2^{-1}+0\times2^{-2}+1\times2^{-3}$

3．十六进制

十六进制数的数码是用 0、1、……、9、A、B、C、D、E、F 这 16 个符号来表示的。权为 16^n（n = ……, 2, 1, 0, -1 ,-1, ……），十六进制数的计数规则是：逢十六进一。十六进制同样可以表示成按权展开的多项式。例如：

$1CB.D8=1\times16^2+12\times16^1+11\times16^0+13\times16^{-1}+8\times16^{-2}$

为了区分一个数是几进制表示的数，可以使用两种方法。一种方法是在数的后面加一个大写字母，其中，B 表示二进制数，D 表示十进制数，H 表示十六进制数，O 表示八进制数。如 101B 表示二进制数 101，而 101H表示十六进制数 101。另一种方法是将要表示的数用小括号括起来，然后用一个下标表示该数是几进制数，如 $(10110.101)_2$ 表示一个二进制数，而$(10110.101)_{10}$ 表示一个十进制数。

二、数制之间的转换

1．二进制、十六进制转换为十进制

要将一个二进制数或十六进制数转换为十进制数，可以统一将这个数按权位展开成多项式，然后计算该多项式的值即为十进制数。

例 1.1 将二进制数$(1011.101)_2$转换成等值的十进制数。

$(1011.101)_2=1\times2^3+0\times2^2+1\times2^1+1\times2^0+1\times2^{-1}+0\times2^{-2}+1\times2^{-3}=(11.625)_{10}$

例 1.2 将十六进制数$(3D.B)_{16}$转换成等值的十进制数。

$(3D.B)_{16}=3\times16^1+13\times16^0+11\times16^{-1}=(61.6875)_{10}$

2．十进制数转换为二进制数

要将一个十进制数转换为二进制数，首先要把这个十进制数分为整数部分和小数部分，然后对整数部分采用“除 2 取余”法得到相应二进制数的整数部分，小数部分则采用“乘 2 取整”法得到相应的二进制数的小数部分，最后将两部分合起来就得到了与原十进制数等值的二进制数。具体的转换如例 1.3。

例 1.3 将十进制数 26.8125 转换为等值的二进制数。

如图 1.2 所示，整数部分 26 采用“倒除法”转换为二进制数是 11010，这里“取余值”要从下往上取，小数部分 0.8125 转换为二进制数是 0.1101，这里“取整值”则要从上往下取，最后将两部分合起来得到的结果是$(26.8125)_{10}=(11010.1101)_2$。同理，十进制数转换为十六进制数可以用整数部分“除十六取余”法和小数部分“乘十六取整”法求得。

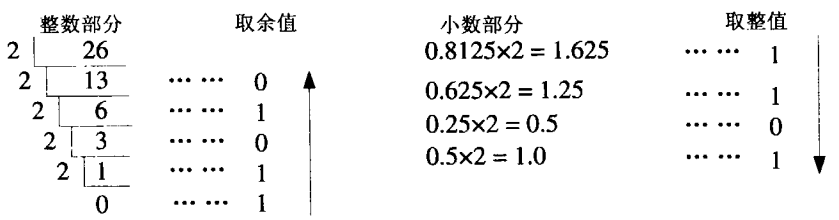


图 1.2 十进制数转换为二进制数

3. 二进制数与十六进制数之间的相互转换

将二进制数转换为十六进制数的方法是：将要转换的二进制数从小数点开始分别向左（整数部分）和向右（小数部分）每 4 位二进制数码分成一组，在最左或最右不足 4 位的用 0 补足，补 0 的方法是：整数部分补在左边，小数部分补在右边。分别将 4 位二进制数码转换成一位十六进制数码后，将这些十六进制数码连接起来就是相应的十六进制数了。

例 1.4 将二进制数 11010011010.11101010011 转换为等值的十六进制数。

如图 1.3 所示，此二进制数对应的十六进制数为 69A.EA6。

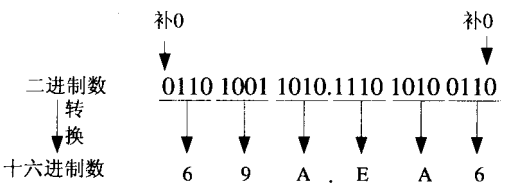


图 1.3 二进制数转换为十六进制数

将一个十六进制数转换为二进制数的方法是：每位十六进制数码用对应的 4 位二进制数替换，并去掉整数部分最左边的 0 和小数部分最右边的 0。

三、二进制数的运算

在计算机中，对二进制数可作两种基本运算：算术运算和逻辑运算。其中二进制数的算术运算包括加、减、乘、除，具体规则如下：

加法：0 + 0 = 0　1 + 0 = 1　0 + 1 = 1　1 + 1 = 0（有进位）

减法：0 − 0 = 0　0 − 1 = 1（有借位）　1 − 0 = 1　1 − 1 = 0

乘法：0 × 0 = 0　0 × 1 = 0　1 × 0 = 0　1 × 1 = 1

除法：0 ÷ 1 = 0　1 ÷ 1 = 1

二进制数的逻辑运算包括与、或、非 3 种基本运算，它们的具体运算规则如下：

与运算：$0 \wedge 0 = 0$　$0 \wedge 1 = 0$　$1 \wedge 0 = 0$　$1 \wedge 1 = 1$

或运算：$0 \vee 0 = 0$　$0 \vee 1 = 1$　$1 \vee 0 = 1$　$1 \vee 1 = 1$

非运算：$\overline{0} = 1$　$\overline{1} = 0$

四、数据单位

位（bit）也称为比特，是计算机存储数据的最小单位，是二进制数据中的一个位，一位表示二进制信息 0 或 1。例如，10010011 一共有 8 位。

字节（Byte）简记为 B，一个字节由 8 个位组成，即 1B = 8bits。字节是计算机数据处理的基本单位，例如“10001100”是一个字节。通常一个字节可以存放一个 ASCII 码，两个字节可以存放一个汉字机内码。

在计算机中，其他经常使用的度量单位主要还有：KB（千字节）、MB（兆字节）、GB（吉字节）和 TB（太字节）等，下面是它们之间的换算关系：

$$1KB = 2^{10}\ Bytes = 1024Bytes$$

需要全本请访问 www.ertongbook.com

$1MB=2^{10}KB=1024KB$

$1GB=2^{10}MB=1024MB$

$1TB=210GB=1024GB$

五、数值在计算机中的表示

我们已经知道，数值信息在计算机中是以二进制形式表示的，但它们究竟是如何被表示的呢？是不是直接存放它们的二进制值呢？事实上，我们除了要表示一个数的值以外，还要考虑它的正负号如何表示，小数点如何表示，甚至也要考虑如何表示更有利于计算机实现，如何设计表示的范围更大，如何表示精度更高等。

1. 整数的表示

表示一个整数除了要表示其绝对值外，还要表示其正负号。由于计算机内部采用二进制，所以计算机最终只能用 0 或 1 表示正负号。历史上曾经出现过的表示整数的编码方案主要有：原码、反码和补码，现在的计算机基本上采用补码。这几种方案都是将最高位作为负号位，其中 0表示正数，1 表示负数。下面以 8 位表示一个整数为例，介绍计算机中整数的表示。

（1）原码。最高位为符号位，其余为数值部分，数值部分用二进制数的绝对值表示整数的方法称为原码表示法，如+57 和-57 的原码表示为：

$[+57]_{原}=00111001$　　$[-57]_{原}=10111001$

用原码表示时，虽然数的真值和它的原码表示之间的对应关系简单，但原码本身的运算逻辑复杂，所以不常使用。

（2）反码。正数的反码与原码相同，负数的反码则是保持符号位为 1，然后将其原码的数值部分按位取反得到的，如+57 和-57 的反码表示为：

$[+57]_{反}=00111001$　　$[-57]_{反}=11000110$

虽然计算机中得到反码非常方便，但是反码用于计算还是比较麻烦，因此实际中很少使用，通常只是将反码作为求补码过程的中间形式。

（3）补码。补码的表示法是：正数的补码与原码相同，负数的补码是将其原码除符号位外按位取反后加 1 得到的。如+57 和-57 的补码表示为：

$[+57]_{补}=00111001$　　$[-57]_{补}=11000111$

由于补码可以用相同的运算规则来进行各种运算，使运算简化，并且运算逻辑的实现也比较容易，所以大多数计算机系统都采用补码来表示整数。

2. 定点数

定点数采用整数的表示逻辑来存放带小数点的数。定点数是指小数点的位置固定，为了处理方便，一般分为定点纯整数和定点纯小数。其中，定点纯整数隐含小数点位置固定在数值部分的最后，定点纯小数隐含小数点位置固定在数值部分的最高位与符号位之间。

3. 浮点数

与小数点位置固定的定点数相对应，浮点数是指小数点的位置是可以浮动的。在计算机中，这种数的表示法往往用来表示一般的实数。我们知道一个实数总可以表示成一个乘幂和一个纯小数之积，例如：

$123.45 = 10^3 \times (\ 0.12345)$

$-0.0034574 = 10^{-2} \times (-0.34574)$

乘幂中的指数部分我们称其为“阶码”（这是一个整数），而括号中的部分我们称其为“尾数”（这是一个纯小数），尾数和阶码都有正负之分。同理，任何一个二进制数也可以表示成相同的形式，例如：

$1001.011 = 2^{100} \times (\ 0.1001011\ \ )$

$-0.0010101 = 2^{-10} \times (\ -0.10101\ )$

所不同的是这里的阶码和尾数都是用二进制数表示的。这样只要我们能表示出一个二进制数的阶码和尾数，那么也就惟一地表示出了这个数。在计算机内部，浮点数通常表示成如图 1.4 所示格式。



图 1.4 浮点数格式

其中，符号位占一位，用来表示尾数的正负号，阶码用整数表示，尾数用定数小数表示。阶码和尾数的位数可以根据对表示精度和范围的要求不同而不同，也会因整个数的长度不同而不同。以一个 32 位浮点数为例，若规定阶码长 8 位，则尾数长为 23 位。但由于阶码可以选用不同的编码（原码、补码等），尾数的格式和小数点位置也可以有不同的规定，因此，不同的计算机中浮点数的表示方法可以互不相同。

2.2 文本信息的表示

文本是文字信息在计算机中表示的统称，它是基于特定字符集的具有上下文相关性的一个字符流，每个字符均使用二进制编码表示。文本在计算机中的处理包括输入、编辑、存储、传输和显示等过程，每个环节对文本的编码都可能不同，下面介绍几种主要编码方案。

一、西文字符的编码

常用字符的集合叫作“字符集”。西文字符集由英文字母、数字、标点符号和一些特殊符号组成。字符集中的每一个符号都有一个代码，即字符的二进制编码。目前计算机中使用得最广泛的西文字符集是 ASCII 字符集，其编码称为 ASCII 码，即美国标准信息交换码（American Standard Code for Information Interchange），同时对应的国际标准是 ISO646。基本 ASCII 码采用一个字节表示一个字符，但只使用字节的低 7 位，所以可以表示为 $0\sim2^7-1$ 共 128 个字符。其中 96 个是可打印字符，32 个是控制字符，表 1.1 为 ASCII 字符表，$b_3\ b_2\ b_1\ b_0$ 为 ASCII 的低 4 位，$b_6\ b_5\ b_4$ 为 ASCII 的高 3 位。

由于标准 ASCII 字符集只有 128 个字符，在很多应用中无法满足要求，为此国际标准化组织又制定了 ISO2002 标准，称为《七位字符集的代码扩充技术》，它规定了在保持与 ISO646 兼容的前提下，将 ASCII 字符扩充为 8 位代码的统一方法，如表 1.1 所示。

表 1.1 ASCII 码字符编码表

$b_3 b_2 b_1 b_0$ \ $b_6 b_5 b_4$	000	001	010	011	100	101	110	111
0000	NUL	DLE	SP	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	BEL	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

二、汉字的编码

中文文本的基本组成单位是汉字。目前我国汉字的总数超过 6 万字。数量大、字形复杂、同音字多，异体字多，这就给汉字在计算机内部的表示与处理、传输与交换、输入与输出带来了一系列问题。

1. GB2312-80 汉字编码

汉字编码方案有多种，GB2312-80 是应用最为广泛的一种。GB2312-80 是我国于 1981 年颁布的一个国家标准——信息交换用汉字编码字符集·基本集（GB2312-80），简称国标字符集，其编码称为国标码。国标码用两个字节编码一个汉字，并且规定每个字节只用低 7 位，即高位都为 0。它对 6763 个常用汉字和 682 个其他基本图形字符进行了编码。

GB2312 国标字符集由 3 部分组成。第一部分为字母、数字和各种符号；第二部分为一级常用汉字，共 3755 个，按汉语拼音排列；第三部分为二级常用字，共 3008 个，因为不太常用，所以按偏旁部首排列。

GB2312 国标字符集构成一个二维平面，由 94 行和 94 列组成，其中的行号称为区号，列号称为位号。每一个汉字或符号都有一个惟一的区号和位号，将区号和位号放在一起，就构成了汉字或符号的区位码。例如，“文”字的区号是 46，位号是 36，所以它的区位码是 4636。如图 1.5 所示是 GB2312-80 字符集及区位分布情况。

区位码中的区号和位号都是十进制数，区位码与国标码之间的关系如下：

国标码首字节 = 区号（二进制表示）+ 00100000B

国标码尾字节 = 位号（二进制表示）+ 00100000B

根据“文”字的区位码“4636”，可以计算出它的国标码是 01001110 01000100，用十六进制数表示为 4E44H 具体计算过程如下：

“文”国标码首字节 = 00101110B（区号 20 的二进制表示）+ 00100000B = 01001110B

“文”国标码尾字节 = 00100100B（位号 20 的二进制表示）+ 00100000B = 01000100B

由于文本中的汉字与西文字符经常是混合在一起使用的，在计算机内部，汉字信息如不予特别的标识，它与单字节的 ASCII 码就会混淆。为解决这一问题，采用的方法是将国标码的两个字节的最高位都置为 1 来作为汉字的机内表示，这样就可以将一个双字节汉字与两个单字节 ASCII 字符区分开来。这种汉字编码称为汉字的机内码，例如由“文”字的国标码，可以马上得到汉字的机内码是 11001110 11000100，用十六进制数表示为 CEC4H。

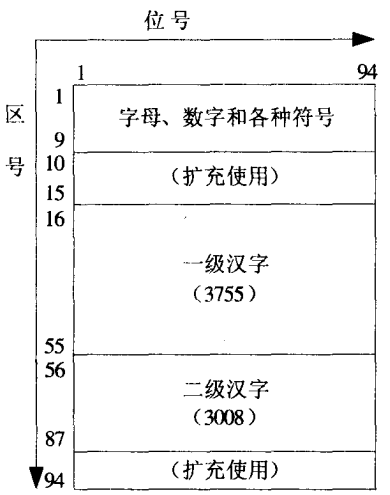


图 1.5 GB2312-80 字符集及区位分布

2. GBK 和 GB18030

GB2312 表示的汉字比较有限，因此一些偏僻汉字在 GB2312 中无法表示。随着计算机应用的普及，这个问题日渐突出，我国的信息标准化委员会就对标准进行了扩充，得到了扩充后的汉字编码方案 GBK。它一方面向上兼容 GB2312，另一方面将常用的繁体字填充到原编码标准中留下的空白码段中，使汉字数增加到 20902 个。值得注意的是 GBK 不是一个国家标准，而只是一个规范，随着 GB18030 国家标准的发布，它将完成它的历史使命。GB18030 采用变长编码，其中两字节部分与 GBK 完全兼容，共收录了 27484个汉字，总的编码数超过 150 万个码位。

3. Unicode

随着互联网的迅速发展，进行数据交换的需求越来越大，不同的编码体系越来越成为信息交换的障碍，而且多种语言共存的文档不断增多，单靠 ANSI 代码页已很难解决这些问题，于是 Unicode 应运而生。

Unicode 采用两个字节编码体系，因此它允许表示 65536 个字符，这已能满足目前大多数场合的需要。前 128 个 Unicode 字符是标准 ASCII 字符，接下来的 128 个扩展的 ASCII 字符，其余的字符供不同语言的文字和符号使用。其版本 V3.0 于 2000 年公布，内容包括字母和符号 10236 个、汉字 27786 个、韩文拼音 11172 个、造字区 6400 个、保留 20249 个，控制符 65 个。在 Unicode 中，ASCII 字符也采用两个字节表示，这样，ASCII 字符与其他字符的处理就统一起来了，这就大大简化了处理的过程。

三、汉字字库

汉字信息存储在计算机内采用机内码，但输出时必须转换成字形码，以人们熟悉的汉字形式输出才有意义。因此对每一个汉字，都要有对应的字的模型（简称字模）储存在计算机内，字模的集合就构成了字模库，简称字库。汉字输出时，需要先根据内码找到字库中对应的字模，再根据字模输出汉字。

构造汉字字形有两种方法：向量（矢量）法和点阵法。

向量法是将汉字分解成笔画，每种笔画使用一段的直线（向量）近似地表示，这样每个字形都可以变成一连串的向量。

点阵方式又称“字模点阵码”。每一个汉字以点阵形式存储在记录介质上，有点的地方为“1”，空白的地方为“0”。例如，可以将“字”字画在图 1.6 所示的 64×64 的方格上，每一行为 64 位（8 个字节），共 64 行组成一个汉字的字形码，即共需要 8（字节）×64=512 字节。

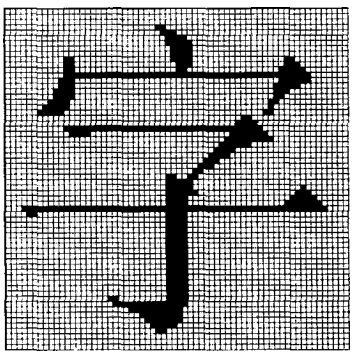


图 1.6 用点阵组成的汉字字形

如果对显示或打印的质量要求不是很高，也可以采用 16×16 点阵或 24×24 点阵。表示一个 16×16 点阵的汉字的字形只要 2（字节）×16=32 字节。点阵规模越大，每个汉字存储的字节数就越多，字库也就越庞大。但字形分辨率越好，字形也越美观。

大多数汉字信息处理系统把汉字字库存放在磁盘上，这样的字库称为“软字库”，使用时全部或部分调入内存储器，并通过专门的软件实现从汉字机内码转变为对应的汉字字模点阵码的地址，根据地址找到相应的字形码。

相对于“软字库”，一般将固化在 EPROM 或 MASK-ROM 的芯片中的汉字字库称为“硬字库”。一般打印机等设备中都安装带有固化汉字库的集成电路芯片，以提高输出汉字的速度。

1.2.3 图像与图形信息处理技术

计算机中的数字图像按其生成方法可以分为两大类，一类是从现实世界中通过数字化设备获取的图像，它们称为取样图像（Sampled Image）、点阵图像（Dot Matrix Image）和位图图像（Bitmap Image），以下简称图像（Image）；另一类是计算机合成的图像（Synthetic Image），它们称为矢量图像（Vector Image），以下简称图形（Graphics）。

一、图像的数字化

从现实世界中获取数字图像的过程称为图像的获取（Capturing），所使用的设备称为图像获取设备，如扫描仪、数码相机和数字摄像机等。图像获取的过程实质上是模拟信号的数字化过程，如图 1.7 所示它的处理步骤大体分为 3 步：取样、分色和量化。

1．取样

将画面划分为 M×N 个网格，每个网格称为一个取样点。这样，一幅模拟图像就转换为 M×N 个取样点组成的一个阵列。

2．分色

将彩色图像取样点的颜色分解成 3 个基色（如 R、G、B 三基色），即每个取样点用 3 个亮

度值来表示，称为 3 个颜色分量；如果不是彩色图像（即灰度图像或黑白图像），则每一个取样点只有一个亮度值。

3．量化

对取样点的每个分量进行 A/D 转换，把模拟量的亮度值使用数字量（一般是 8～12 位正整数）来表示。

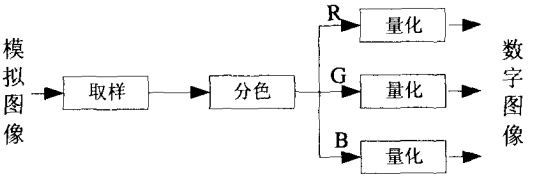


图 1.7　图像的数字化过程

一幅不经压缩的图像数据量可按下面的公式计算（以字节为单位）：

图像数据量=图像水平分辨率×图像垂直分辨率×像素深度/8

所谓像素的深度是指表示每个取样点的颜色值所采用的数据位数。一幅 640×480 分辨率(即 640×480 采样点）的图像若分成红、绿、蓝 3 色，并且每一颜色分量的亮度用 8 位二进制数来表示，则可以计算出图像的数据量。由于每一个采样点有 3 个颜色分量，且每个颜色分量用 8 位来量化，所以每一个采样点的数据位数为 8+8+8=24bits，这就是像素深度。根据以上公式，此图像数据量为 640×480×24/8=900KB 。表 1.2 列出了若干不同参数图像的数据量。

表 1.2　几种常见图像的数据量

像素深度 / 图像分辨率	8 位（256 色）	16 位（65535 色）	24 位（真彩色）
640×480	300KB	600KB	900KB
1024×768	768KB	1.5MB	2.25MB
1280×1024	1.25MB	2.5MB	3.75MB

二、图像的压缩编码

由表 1.2 可知一幅未经压缩的图像的数据量是比较大的。这既不方便存储（浪费空间），也不利于图像在 Internet 上传输（传输速度慢、通信费用大），因此进行图像压缩就显得特别重要。

由于数字图像中的数据相关性很强，或者说数据的冗余度很大，因此对数字图像进行大幅度压缩是完全可能的。而且，人眼的视觉有一定的局限性，即使压缩前后的图像有一定的失真，只要限制在人眼无法察觉的误差范围之内，也是允许的。

数据压缩可分成两种类型，一种是无损压缩，另一种是有损压缩。无损压缩是指把压缩以后的数据进行图像还原（即解压缩）时，重建的图像与原始图像完全相同，例如行程长度编码（RLC）、哈夫曼（Huffman）编码等。有损压缩是指对压缩后的数据进行图像重建时，重建的图像与原始图像有一定的误差，但这种误差应不影响人们对图像含义的正确理解。常见的有损压缩编码有变换编码和矢量编码等。

为了便于在不同的系统间交换图像数据，人们为计算机中使用的图像压缩编码方法制订了一些国际标准和工业标准。其中最著名的是 JPEG，它是一个静止图像数据压缩编码国际标准。

需要全本请访问 www.ertongbook.com