

计 算 机

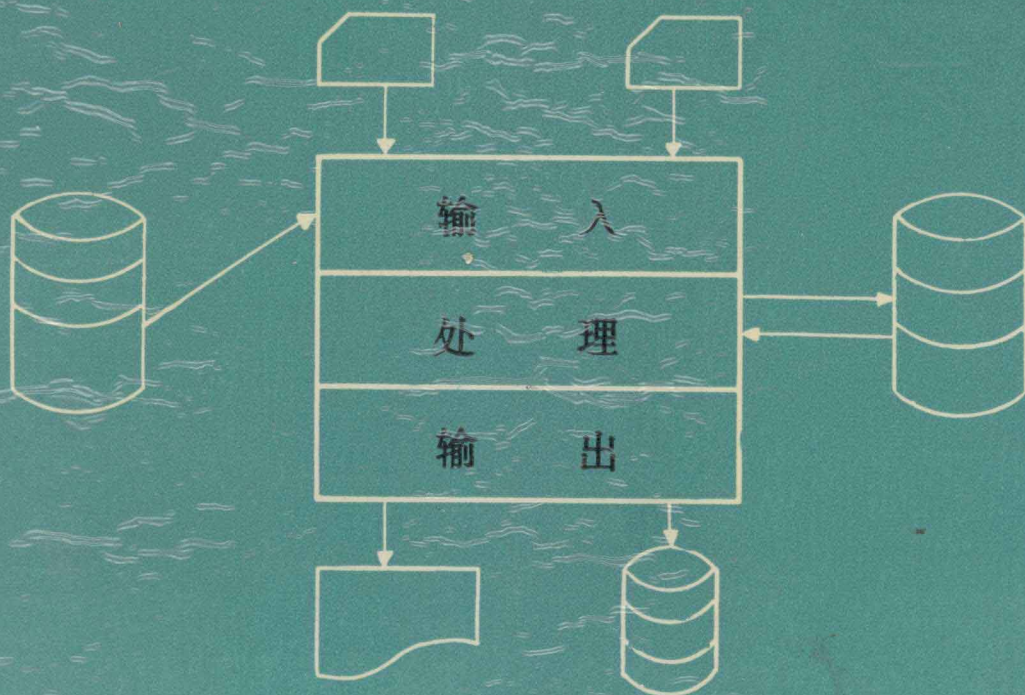
数据处理技术与工程预算系统

王振武 编著

韩双林

主审

许开甲



JISUANJI SHUJUCHULI JISHU YU GONGCHENGYUSUAN XITONG

东北财经大学出版社

计 算 机

数据处理技术与工程预算系统

王振武 编著

韩双林 主审
许开甲

东北财经大学出版社

计算机数据处理技术与工程预算系统

王振武 编著

韩双林 主审

许开甲

东北财经大学出版社出版发行 (大连黑石礁)

朝阳新华印刷厂分厂印刷

开本: 787×1092 1/16 印张: 15 字数: 346 000

1991年2月第1版 1991年2月第1次印刷

责任编辑: 高晓明

责任校对: 张剑宇

印数: 1—5 000

ISBN 7-81005-397-3/F·289 定价: 6.50元

出版社登记证: (辽)第10号

内 容 简 介

本书内容分为两部分，第一部分共七章，着重介绍计算机数据处理技术。包括数据的编码、输入、存储、排序、检索与输出技术；数据文件的组织与处理技术；以及应用系统的开发方法与步骤。第二部分共三章，完整地介绍了一个典型的建筑工程预算系统的分析、设计与使用实例。

本书内容新颖具体，深入浅出，可作为高等财经院校非计算机专业特别是投资经济专业的教材，也可供从事计算机应用工作的技术人员及从事建筑企业管理的有关人员参考。

前 言

计算机科学技术的发展和应用程度,标志着一个国家的现代化水平。深入研究和大力推广计算机应用技术,是发展计算机科学,促进四化建设极有意义的工作。近几年随着我国改革开放事业的深入发展,各行业各领域电子计算机尤其微机应用工作正在迅速推广和普及,取得了许多重大而可喜的成果,但是就计算机的装机水平与取得的成就相比,尚不尽如人意。形成这种局面固然有许多因素,但其中一个很重要的因素是计算机应用工作的从业人员中的许多人对计算机数据处理技术知之甚少,导致计算机从业人员整体素质较低,这恐怕与我国计算机应用科学方面的书籍资料比较贫乏不无关系。鉴于上述种种,我们将多年的实践经验和教学成果编著成书,奉献给广大读者。

本书初稿经东北财经大学投资经济系试用五届,在此过程中其结构和内容经过了多次调整与完善。用过此书的毕业生、研究生和在校本科生对本书反映一致很好。书中数据处理技术部分的主要内容也曾在为社会上从事计算机应用工作的人员举办的培训班上讲授过,他们一致认为书中介绍的许多方法和技术有较高的实用价值。作者已将书中几个典型的算法写成论文发表在《微计算机应用》杂志上。

本书原则上不想拘于某种具体的语言来阐述有关的技术问题。但为使其内容不致太笼统太抽象,在每种数据处理的技术中还是尽可能地以最广为人知的BASIC语言给出具体的算法与程序,使每种技术更接近于直接使用的程度。因此,凡通晓一门计算机高级语言特别是BASIC语言的读者都能从书中学到一些有用的东西。

书中前七章内容适用于一切有志从事计算机应用软件工作的人员,对于他们,后三章可以不读;而对于那些从事建筑工程预算及其工程预算电算化的工作人员,后三章无疑是较有价值的参考资料。

由于作者水平所限,书中肯定会存在一些缺点和不足,恳切希望广大读者提出宝贵意见和建议,这对推动我国计算机应用工作的水平会有一定的作用。

本书经韩双林、许开甲两位教授审阅,提出了许多宝贵的修改意见,书中全部程序流程图均由张军达副教授审绘,在此对他们以及其他对本书给予大力支持和帮助的同志们一并表示衷心的感谢。

编 著

1989年6月于大连

目 录

一章 概 论	1
§ 1.1 电子计算机的基本知识	1
§ 1.2 信息与数据	6
§ 1.3 数据的编码、组织与存储	9
§ 1.4 数据处理	16
第二章 数据结构	19
§ 2.1 概 述	19
§ 2.2 数组和顺序线性表	20
§ 2.3 栈和队	24
§ 2.4 链表	31
§ 2.5 树	39
§ 2.6 字符串	50
§ 2.7 记录结构	53
第三章 程序设计技术	55
§ 3.1 分支	55
§ 3.2 循环	57
§ 3.3 通用表格输出程序设计	60
§ 3.4 内存覆盖技术	67
第四章 排序	71
§ 4.1 内部排序	71
4.1.1 插入排序法	71
4.1.2 冒泡排序法	73
4.1.3 延迟交换排序法	75
4.1.4 希尔排序法	76
4.1.5 堆排序法	79
4.1.6 快速排序法	82
4.1.7 直接二路归并排序法	83
4.1.8 树形排序法	86
§ 4.2 外部排序	87
4.2.1 多路归并外部排序算法	88

4.2.2 堆溢流外排序法·····	91
第五章 查找与检索·····	96
§ 5.1 概 述·····	96
§ 5.2 查 找·····	97
5.2.1 简单顺序查找·····	97
5.2.2 快速顺序查找·····	98
5.2.3 有序文件的顺序查找·····	99
5.2.4 二分法查找·····	100
5.2.5 插入查找·····	102
§ 5.3 检 索·····	104
5.3.1 单辅助键检索·····	104
5.3.2 多辅助键逻辑复合检索·····	105
第六章 文件组织与处理·····	109
§ 6.1 基本概念·····	109
§ 6.2 顺序文件·····	110
§ 6.3 索引文件·····	115
§ 6.4 直接存取文件·····	125
§ 6.5 多重链表文件·····	133
§ 6.6 倒排文件·····	136
§ 6.7 文件系统与数据库·····	142
第七章 数据处理系统开发·····	146
§ 7.1 概 述·····	146
§ 7.2 系统分析·····	150
§ 7.3 系统设计·····	156
§ 7.4 系统实施·····	166
第八章 工程预算系统分析·····	175
§ 8.1 预算业务使用计算机的必要性·····	175
§ 8.2 建筑安装工程费用项目的组成与工程项目的划分·····	177
§ 8.3 工程预算系统的功能及其设计原则·····	181
第九章 工程预算系统设计·····	183
§ 9.1 概 述·····	183
§ 9.2 总体设计·····	184
§ 9.3 定额库结构设计·····	186
§ 9.4 工程量计算公式的选择·····	194
§ 9.5 输入表格设计·····	196
§ 9.6 周转文件设计·····	199
§ 9.7 输出报告设计·····	201
§ 9.8 系统程序设计·····	202

第十章 工程预算系统实施.....	210
§ 10.1 依施工图填输入表	210
§ 10.2 系统运行——试算	214
§ 10.3 预算报告书	216
附 录：三个程序.....	222
1. 多路归并外排序程序.....	222
2. 堆溢流法外排序程序.....	224
3. 多辅助键逻辑复合检索程序.....	225

第一章 概 论

§ 1.1 电子计算机的基本知识

近几年来,在我国电子计算机介入管理工作以后,已经取得了十分可喜的成就,人们发现了它在管理领域中的巨大潜力。在新技术革命的影响下,很多企事业单位积极购置计算机,以便更深入更广泛地开展计算机的应用,开发以计算机为中心的管理信息系统。然而随着计算机的普及,在一些人的心目中还存在着各种不正确的看法。譬如,有些人对计算机有种神秘感,认为计算机的学问太深奥,因而自己不敢接触,从而产生计算机与自己无关,应用计算机靠专业人员的观点。另一些人则把计算机想得太简单,认为只要有了计算机,管理工作就可以信息化,就可以及时向管理者提供所需之信息。

事实上,计算机既不是深奥得凡人不能介入的神物,也不是一按电钮要啥有啥的万能机器。计算机只是一种工具,它并不能代替人去研制出一个符合需要的管理信息系统,也不能人想要什么,它就会自动地为您干什么。在目前的技术条件下,还需要人去把要解决的问题的各个方面搞清楚并编出相应的程序,计算机才能按人的意图去完成处理工作。因此只有人明白到某种程度,计算机才能干到那种程度,简单说还是人脑指挥电脑。

存在上述错误认识的原因可能有多种,但其中很重要的一条就是人们对计算机的基本知识还缺乏一定的了解。因此本书的开始,我们对计算机的基本知识先作些简单介绍。

一、计算机系统的基本组成

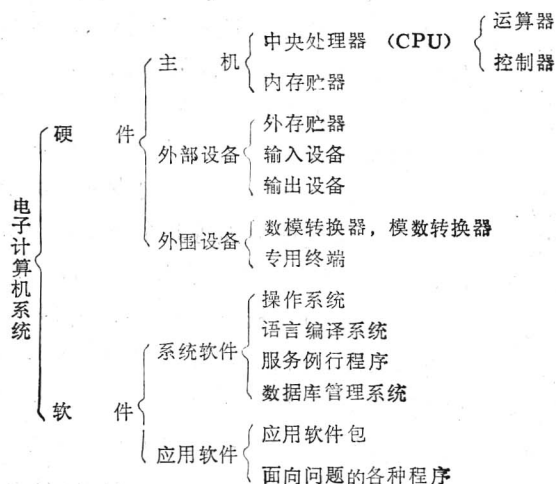


图1.1 电子计算机系统的组成

概括地说，电子计算机系统是由两大部分组成的，即硬件部分和软件部分。它们的大致内容如图1.1所示。

计算机的硬件，系指构成计算机的全部物理实体，主要由电子线路，元器件和机械部件等构成，一般包括运算器、控制器、存储器（内存和外存），输入设备和输出设备五个通用部分。图1.2是通用计算机结构示意图。

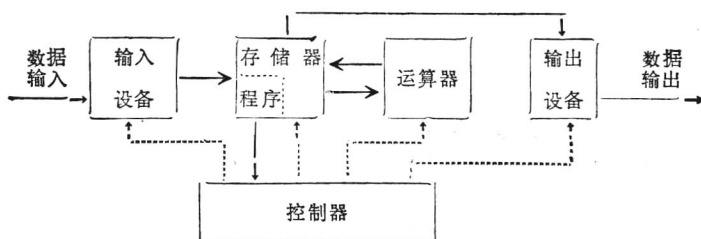


图1.2 通用计算机结构

计算机软件，是指使用计算机必备的各种程序的全体。软件的任务多种多样，其中系统软件是随同机器由厂家提供或从专门服务部门购买的。它并不具体地解决实际问题，而是为方便程序设计，保证程序正确运行，控制和管理机器各部件协调工作提供服务。应用软件可以由用户自己编写，也可以委托专门人员代为编写，或从软件商店直接购买。它们是针对某个或某类具体问题的，由于解决的问题不同，应用软件也就千差万别。

上述可见，硬件是物质产品，是构成计算机系统的“物质基础”，软件是精神产品，是建立在物质基础之上的“上层建筑”。二者同等重要，缺一不可。

二、计算机发展的过程与趋势

世界上第一台电子计算机是美国宾夕法尼亚大学于1946年研制成功的ENIAC机。自ENIAC机问世至今，虽然只有42年的历史，计算机却获得了相当迅速的发展，最明显的变化有以下几点：逻辑元件由真空管到晶体管到集成电路到大规模集成电路；存储器由磁鼓到磁芯到半导体存储器；软件由机器语言到汇编语言到高级语言到数据库管理系统。与上述变化相适应的是计算机的功能迅速增强，体积迅速缩小，成本迅速降低。

现在人们通常把计算机的发展分成四代，那主要是以其逻辑元件为依据。也就是电子管，晶体管，集成电路和大规模集成电路。第五代计算机正在研制中，它的逻辑元件将采用超大规模集成电路，其软件主要面向人工智能。随着电子电路的集成度不断提高，计算机的主机与外部设备的体积之比将越来越小，目前广泛使用的微机就足以说明这一点。

当前，计算机正向着“巨、微、网、智”四个方向发展。其中：“巨”是指巨型化，其特征主要体现在速度高，容量大，解题能力强等几个方面，并非指体积之大。例如美国的ILLIAC—IV机，平均运算速度每秒一亿五千万次，除了内存容量为 8.4×10^6 位之外，还有存储量几千兆位的外存储器。巨型机主要用于军事部门和尖端科学研究，如高能物理，导弹火箭设计，宇宙飞船导航，遗传工程等。“微”是指微型化，微型机具有体积小，功耗小，价格低和可靠性高等优点，所以得到广泛的普及和应用。“网”是指网络化，即多台分立的计算机由通信线路联结成网，共享软硬件和信息资源。计算机网络化已经形成较强的发展趋势。例如，日本KDD公司所提供的国际计算机查询服务网络系统，

使得不出日本国土，就可以在自己的终端上通过卫星直接获得有关美国计算机的信息。

“智”是指智能模拟，就是用计算机模仿人类的智能，使其象人一样具有思维、联想、启发、推理等能力，具有对文字、图象和语音的识别能力，具有人一样的“热情”。智能模拟是建立在控制论、计算机、仿生学、心理学等学科基础上的边缘科学。

展望未来，随着新技术的不断涌现，新型计算机如激光计算机、超导计算机也将问世。激光计算机是使用叫波带片的光学零件来进行信息处理和数据运算的，速度可高达每秒10000亿次，存储容量比现在的大型计算机还要大10000亿倍。到那时一台计算机就能顶现在的几千台计算机的工作，体积还要比现在的一台计算机小得多。目前，激光计算机正处于探索阶段，主要用于测试雷达信号处理，石油勘探，地震预报和电子对抗等几方面，已经显示出它的巨大优越性。超导计算机是利用器件在低温下的超导特性制成，超导开关器件的作用与半导体开关器件的作用相仿，但超导器件没有电阻，所以它的动作更快，功耗更小。目前，各国正在争相研究超导体，超导计算机也在初期试制中。

三、计算机的种类

关于计算机的种类，采用不同的划分方法就有不同的类型。如果按工作原理分，可分为：

1. 电子数字计算机

它采用离散量进行计算，其特征是速度快、精确度高、稳定性好。

2. 电子模拟计算机

它采用连续量，例如：电压、长度、角度等进行模拟处理，其特征是速度快，而精确度差。

3. 混合式计算机

它是数字和模拟技术的结合，采用两种技术之长，摒弃两种技术之短研制而成的。

如果按用途分，可分为：

1. 通用电子计算机

这种计算机硬件齐全，软件完善，具有较强的系统性，可用于解决各类应用问题。

2. 专用电子计算机

这种计算机不论硬件还是软件都是针对某一类具体问题研制和配备，为特定部门服务的，因此其结构比较简单，但效率较高。

如果从规模和功能划分，可分为：

巨型、大型、中型、小型和微型。

其实根据计算机的功能和规模对计算机进行分类，还没有一个统一的标准，特别是随着计算机技术的发展，许多大型机引入了微型机的技术，许多微型机也具有了一些大型机的功能，这就使得它们之间本来就不太明显的界限愈加模糊起来。这里我们仅就人们的经验与习惯作法作一些简单介绍。

1. 巨型机

它比大型机功能更强，国外有人为巨型机下的定义是“三个一千万以上的计算机”，也就是：计算速度在每秒1000万次以上，内存容量在1000万位以上，价格在1000万美元以上。例如美国的ILLIAC—IV机和我国的“银河”机都应属于巨型机的范畴。

2. 大型机

大型机以高速度、大容量和众多的外部设备为其主要特征。其速度一般在每秒百万次以上,字长在32位以上,内存自几百KB到一兆KB,而且一般都采用多道程序运算,即一台主机同时可为多个用户服务。大型机主要用于大型科学计算和数据处理。如火箭发射,人口普查。大型机的典型产品有IBM—370, CDC—6000等。

3. 中型机

顾名思义,中型机就是在速度、容量、功能和价格等方面都较大型机略小的计算机。但是由于计算机硬件水平的迅速提高,许多新的中型机已经完全达到了大型机的水平。例如,美国的IBM—360/50, PDP10, 日本的M150和罗马尼亚的FELIX机都属中型机范畴。

4. 小型机

小型机与微型机相比,主要特点在于它具有较高的性能/价格比。它的结构简单,体积较小,生产周期短,造价低,可靠性好,所以在微型机出现之前它是最受推崇使用也是大量生产的一种机型。美国的PDP—11和后来换代产品VAX—11以及我国的DJS—100系列都是小型机的代表。

5. 微型机

提到微型机一般有三种理解:一是指CPU本身,也叫微处理机;另一种指CPU与较小的内存,以及很少外围接口结合在一起而形成的主机(它包括单板机和微型机系统的主机);再一种含义是指主机再加上屏幕、打印机、驱动器、键盘等外围设备而形成的微型机系统。目前占统治地位的微处理机有INTEL公司的8086和8088, ZILOG公司的Z80和Z8000。单板机主要用于过程控制。对于企业管理而言,更关心的还是微型机系统。前几年我国常见的微型机系统有Apple—II型, Cromemco公司的Cromemco System—III, 目前最为普遍的微机系统是83年以后迅速崛起的IBM—PC机,及其兼容机,如鹰牌机、TCS机和国产的长城0520机等。

四、衡量计算机性能的几个技术指标

衡量一台计算机性能的优劣,是根据一组综合指标,即由它的系统结构、指令系统、硬件组成、外设多少及软件丰富程度等多方面因素决定的,不能只根据一两项技术指标就得出结论。综合技术指标主要有:

1. 字长

字长指内存中一个存储单元有多少个二进制位,字长标志着计算机的精确程度。

计算机内存中有若干个小的存储单元,每个单元可存放一个具有一定意义的二进制代码,即做为整体来处理的一组二进制数字,这样的一组二进制数字称为字,字长就是一个单元所能存储的二进制数的位数。不同的机器字长是不同的,字长越大,精度越高。一般巨型机字长达128位,大型机字长一般为32~64位,中型机32位,小型机16~32位,微型机4~16位(目前多为8位机)。

2. 内存容量

内存具有的存储单元的总数称为内存容量。内存容量越大,机器的功能越强。

以存储单元数量的多少来衡量内存容量,有时还不够准确,这是因为不同的机器之间存储单元的大小即字长是不相等的。因此,人们把存储单元的二进制位以每8位分成一节,

称为字节 (Byte)，一般以字节为单位来标识内存容量，形式为： $n \times \text{字长} / 8 (\text{Byte})$ ，其中： $k = 1024 \text{Byte}$ 。

目前，投入运行的大型机和巨型机内存已达2000KB~3000KB，微机大多在64KB~512KB之间 (1KB=1024字节)。

其实，对一台机器而言，内存容量也不是固定不变的，因为用户可在一定范围内对内存进行扩充。例如，IBM-PC/XT机就可以由128KB扩充到1兆B。

由于许多重要软件在机器运行过程中必须始终放在内存里，而且内存的存取速度远远大于外存，所以，内存容量是衡量计算机性能的重要指标。

3. 存取周期

把信息代码放入内存单元称为写入，把信息代码从内存单元取出称为读出，对内存单元进行一次完整的读写操作所需要的全部时间称为存取周期。存取周期的长短与制作内存的材料有关，一般半导体存储器一个周期只需一百到几百毫微秒，磁芯存储器却需几百毫微秒到几微秒。存取周期越短，机器运行速度越高。

4. 运算速度

运算速度是反映运算器性能的主要技术指标，一般用每秒钟所能执行的指令条数来表示。但执行不同的指令所需要的时间不同，那么怎样表示运算速度呢，目前采用以下三种方法：

(1) 根据不同类型的指令在程序中出现的频率，取一个系数，各指令所需时间乘其系数，然后求和，得到加权平均值，这叫平均运算速度。

(2) 以执行时间最短的指令为标准表示运算速度。所说的微型机每秒50~200万次、小型机每秒200~500万次、中型机每秒600万次以上都是以作定点加法所需时间为标准表示的运算速度。

(3) 直接给出每条指令的实际执行时间和机器主频率。

5. 外设配备

计算机允许配备外设的多少，有无远程通讯接口，也是衡量机器性能的重要指标之一。另外，从整体看，打印机的速度也直接影响着计算机系统性能的发挥。

6. 软件丰富程度

衡量计算机系统的性能，并不完全取决于硬件，软件也是决定计算机系统性能的极其重要的因素。事实上，机器性能在很大程度上取决于软件的强弱。例如，同一个BASIC程序在同型号的不同机器上运行，所花费的时间可能相差许多倍，这是因为各机的BASIC解释程序的水平相差悬殊所致。因此，计算机系统的操作系统功能、语言处理程序水平、应用程序库的丰富程度也都是衡量计算机系统性能的主要技术指标。

五、软件发展简介

在计算机问世的初期，人们把主要力量放在硬件的研制和改进上，而很少考虑软件的研制问题。当时，用户直接用机器指令 (“0”和“1”) 编写程序，通过控制台上的开关、按钮，送入纸带，启动程序运行，显然这是一种十分繁琐的事，而且极易出错。

后来人们改用符号来表示各种机器码，使得程序比机器指令方便和清晰得多，这就是所谓的“汇编语言”。但是汇编语言所使用的符号，机器是不认识的，需要有另外一个程

序对它进行“翻译”。这种“翻译程序”不必由用户编写，而是作为产品连同机器一起提供给用户的，这就是最早引入计算机的软件。叫做“汇编程序”。

汇编语言比机器语言有了明显的进步，但与人们的日常语言仍有极大的距离，编程序仍然是一件十分困难的事情，程序读起来也不太容易。为此在1956年和1958年先后开发了FORTRAN和ALGOL两种高级语言。高级语言比较接近人们的日常语言，用高级语言写出的程序直观清晰，可读性好，而且高级语言已经不依赖于具体机器，成为独立的语言工具，同一个程序可以直接或稍作改动在其它的机器上运行。显然，这对于学习、交流和移植程序是十分有利的，而且为编写通用程序奠定了基础。

与汇编语言一样，计算机也不能直接执行用高级语言编写的程序，仍然需要有一个从事翻译的软件，这就是所谓的“编译程序”。而不同的高级语言其编译程序是不相同的，计算机只有装入了相应的翻译程序，才能运行相应的高级语言源程序，满足用户的需要。

随着技术的发展，计算机的运行速度得到很大提高，与此相比，输入输出的速度远远无法跟上主机的运转速度，为充分发挥主机的优势，在60年代初出现“脱机批处理”和“中断处理”。与之相配合，就出现了“中断处理程序”和“输入输出控制程序”。由于这两个程序几乎要被所有的程序用到，所以要把它们固定地常留在内存中，这些常驻内存的程序统称为“驻留管理程序”。管理程序和编译程序是第二代软件中最重要最有意义的两部分。

60年代中期，计算机硬件进入了第三代，相应地，软件也有了更新的发展，最突出的成就是在管理程序的基础上发展成的“操作系统”。这是计算机系统最重要的软件。它是计算机硬件、软件的“总管家”，正因为有了操作系统，计算机的效率才得到了更充分的发挥。例如，UNIX, CP/M, DOS以及它的改进型CCDOS等都是微机上目前较为著名的操作系统。

除了操作系统之外，各种各样的高级语言也大量涌现，据73年统计，在美国已被实现了的语言就有200多种。特别是BASIC一类会话式程序设计语言，进一步缩短了程序的编制周期。

软件的进一步发展，又出现了数据库管理系统，为数据共享提供了更为有利的途径。70年代初期，美国数据系统语言委员会的数据库任务组（DBTG）和IBM公司先后提出了DBTG数据库模型和IMS模型，为以后的网状模型和树状模型奠定了基础。到了70年代中期，IBM公司的E·F·Codd提出了关系型数据库的理论，PDP-11上的INGRES库和IBM-370上的System-R都是关系数据库的代表。目前，我国广为流行的dBASE-Ⅱ和dBASE-Ⅲ也属于这种关系型数据库。

80年代以后，计算机开始向第五代迈进，软件则面向人工智能方向发展。自然语言的处理问题引起了较多人的关注，模式识别也有所突破，各种智能库和专家系统先后出现，这一切都为计算机的拟人化创造了极为有利的条件，人们期待着会读、会说、会听、会思考的“智能机器人”的最终实现。

§ 1.2 信息与数据

一、信息与数据的区别和联系

什么是信息？信息是客观世界中各种事物的变化和特征的最新反映，是客观事物之间

联系的表征，也是客观事物状态经过传递后的再现。

信息作为客观事物特征和变化的反映，总是在不断地生成着和传递着。银河系的星群，微观世界的基本粒子，天空中的风雨雷电，市场上各种商品供求的动态变化，社会生活中人与人之间的关系和变化等，都是通过各种各样的信息反映出来。

信息是客观事物之间相互作用、相互联系的表征。例如，“山雨欲来风满楼”，“风满楼”作为一种“山雨欲来”的信息，正是“山雨欲来”的表征，这一信息就揭示了“风”和“雨”这两种自然现象之间的相互联系。

信息是要经过传递的，任何信息只有经过传递才能被人接收和利用。信息的范围极其广泛，语言文字是社会信息，湖光山色是自然信息，遗传密码是生物信息。总之，信息的概念是太大了，不论怎样描述和定义，总是带有一定的片面性。

什么是数据呢？数据是反映客观现实情况的数字或文字以及声音、颜色和图象等，是那些被处理后能产生信息的东西。因此，数据又可以被认为是任何的基本事实，利用一组基本事实去产生信息的过程就是所谓的数据处理。

由于数据是反映客观现实情况的，其中蕴含着有关现实情况的信息，因此数据是载荷信息的媒介。举例来说，当我们得到生产产量数据，就可以了解生产的情况，也即获得了有关生产情况的信息；当我们得到库存数据，就可以了解库存情况，也即获得了有关库存情况的信息等。其中产量数据一般是用数字描述的，库存数据中除了数字之外，可能还有一些名称，订货单位之类用文字表示的一些内容，它们也属于数据的范畴。本书我们只讨论这些由文字和数字构成的数据的处理。

数据和信息是两个密不可分的概念。数据是为信息传递服务的，它是信息得以传递的依据。但是信息在数据中的蕴含方式却是不同的，有些信息从数据的表面就可以得到，而有些信息却不是直接了当的。例如，当你看到一个数据“砌墙共用砖三万块”，你马上可以得到一个信息——砌墙一共用了三万块砖。但是究竟是否符合定额标准，这个信息就无法获得，只有和另外两个数据“所砌墙的体积”和“单位墙体用砖定额”相联系，经过运算和比较，才能得到有关“是否符合定额”的信息。又如军事上经常采用的密码和口令，其数据和信息既是相联的，又是分离的。这就是说，数据所带的信息分两类，一类可以使接收者直接得到信息，另一类只有经解释和加工之后才能产生信息。因此有人说信息是对数据的解释，是加工了的数据。

在一个企业管理信息系统中，一般来说，管理者所处的地位越高，他们所要求的信息就越需要加工和处理。例如，一个工段的管理者要了解各班组的生情况，只需把各班组的生记录拿来一看，便可基本达到目的。而工区的领导者并不大关心班组的具体情况，而是各工段的情况，这就需要在班组数据的基础上加以归纳合并，送上一张供工区领导使用的报表。类似的，公司级和局级领导也需要他们所属的各层管理人员，在原有数据基础上加工整理，产生出适合自己获取信息需要的数据内容，诸如报表、总结、分析、意见等等。

如上所述，一个企业内部的信息就是这样从最下层的数据，逐步加工提炼，一直到达各层管理人员和需要获得该信息的部门。每个信息管理部门的职责就是把蕴含着不很直接的信息的那些数据加工成可以直接向上级管理者或其它有关部门反映他们所关心的情况的

信息。严格说来,这种处理后所产生的内容仍然属于数据,但为了与前一种数据相区别,人们往往把它们叫做信息。这也就是许多有关数据处理的书上所常说的“数据处理的任务就是把数据加工成信息”中“信息”一词的含义。

无论如何,直观地看数据加工处理之后所产生的东西仍然是数据,充其量是“高一级水平”的数据。只有它被有关的人得到,并与他们关心的问题联系起来,产生某种新的认识的时候,才能转化为信息。但是为了适应大多数人的习惯,我们不妨把信息的两种概念混合使用,不再区分它究竟指的是“真正的信息”还是“高一级水平的数据”。在本书中,也不再严格区分数据和信息两个词的使用场合,因为在一般情况下,它们并没有什么明显的界限。

二、信息的性质与需求

信息的性质可以从许多不同的角度来描述,一般说来,它包括:准确性、表达形式、频度、宽度、来源、时间趋势、针对性、完整性、及时性和经济性。

准确性是指信息的真实程度。显然,只有准确的信息才能做出正确的决策。收集数据、处理数据、传递数据都有可能把不准确的内容掺杂到信息中去。例如,编制预算定额和产生单位估价表都涉及到大量的数据收集和处理工作,下发定额、修改定额也都有数据的传输问题。施工单位的预算工作也有类似的问题。因此,如何摒弃或减少各种不准确因素,保证信息的准确性,是数据和信息管理人员首先要考虑的问题。

信息的描述形式是指描述信息的方式和传递信息的媒介。例如,描述定额可以用表格形式,也可以用文字形式,下达定额可以用书面印刷作媒介,也可以用磁盘或磁带作媒介。究竟什么样的信息用什么样的形式和媒介,不同的情况有不同的要求,管理人员应妥善地选择形式和媒介,可保证信息的正确和用户的使用方便。

信息的频度是指使用该信息或变动该信息的经常程度如何。有些信息属于固定的或半固定的。例如,人事档案、设备档案等。而另一些信息是变化较频繁的。例如设备的运转记录、财务的收支帐目等。对于定额和预算工作来说,定额一般是属于固定的,而预算则是随着新项目的出现而更动的。研究信息的变化和使用频率,有益于考虑它们的组织形式和存取方式。

信息的宽度是信息所反映的事件有多大范围。例如,有关定额的信息可以是全国通用的,可以是部门通用的,也可以只适用于某一地区。一般说来,信息所适用的范围越宽,收集处理起来就越困难。当然信息宽度大,更适合于标准化和统一管理,这其中有一个折中和平衡的问题。

信息的来源可以是组织内部,也可以是组织外部,一般说来,内部信息的获取和处理都较为容易。但是随着我国经济的开放和搞活,越来越多的外部信息将对本地区的各个方面产生影响。如何适当地获取和利用外部信息,已经成为我们必须考虑的重要问题。计算机网络为迅速和准确地获得外部信息提供了有力工具,但这牵涉到成本、通讯、组织和控制等许多方面的问题,必须予以通盘考虑。

信息的时间趋势有三种:着眼于过去,着眼于目前和着眼于未来。定额工作一般是基于过去的资料,但随着新技术、新工艺的出现,必须随时有新的定额出现,以保证施工单位使用。预算工作就更应有着眼于未来的能力。由于我国许多原材料实行了市场浮动价

格,在进行预算时一定要对未来价格的变化有较充分的估计。这一点在当前的预算工作中尚不被重视或无力作到,但随着计算机的使用,应对此有所考虑。

针对性是区别信息与数据的主要因素。只有有针对性的数据才能转化为信息。例如,你把轮船设计图纸中的数据告诉给冰箱生产厂家,就决不会对它产生任何信息。定额工作也是如此,各个部门都有与自己业务有关的定额,如果把与它无关的定额下达给它,就是缺乏针对性。

完整性是指信息应能满足使用者的需要。显然,子目不全的定额是很难被施工单位使用的。同样各种材料的规格、型号、价格也都应尽可能完善。完整性和针对性有时会发生矛盾。过于追求完整性往往会使信息过于庞大,针对性不强;过于强调针对性有时会使一些不甚重要或不经常被使用的内容忽略掉。

及时性有两个含义。一是要求信息能够随用随有;另一个是要求信息永远是最新的。在手工处理中,及时性往往不易保证。例如某项定额的更动,往往要有一定的滞后时间才能真正贯彻使用。在这方面计算机具有较大的优势。

最后,信息的经济性是获得信息所花费的代价。要衡量代价,首先要了解信息本身的价值,信息的价值是什么呢?信息的价值不是经济学中那种使用价值和价值,而是以信息对人的各种有用程度来度量其大小,有用程度大则价值高,反之则低。可见信息的价值与其有用程度是一种正比关系。目前,在我国人们对信息的价值还没有引起足够的重视,所以开发计算机信息处理系统的经济代价是必须考虑的问题。

以上我们简单地介绍了信息的各种性质。研究信息性质的目的在于合理地予以考虑,达到各方面的优化安排。应该看到,不同性质的管理任务对于信息的各方面性质侧重不同。有些侧重于准确性,另一些则侧重于及时性,有些也许侧重外部信息,另一些则主要是处理内部信息。因此,必须针对自己所面临的具体问题来决定侧重点,以便最佳地满足对信息的各方面要求。

实践表明,对于预算系统来说,输出信息的准确性、及时性和格式规范化是必须侧重考虑的几个方面。

§ 1.3 数据的编码、组织和存储

一、数据的编码

编码是为了方便数据处理给数据编设的代码。所谓代码,指的是代表事物名称、属性和状态的符号与记号。在很多情况下,计算机是通过事物的代码来识别和处理事物的,所以为了记录、通讯、处理和检索数据,计算机处理必须代码化。

代码的主要作用有两个:一是代码可以为数据提供一个简单的明确而不含糊的代号,由于代码比数据本身要短得多,无论是记录还是记忆都较容易,同时也可以大大节省存储空间和处理时间。二是它可以提高数据处理的效益和精度,因为一旦编码以后,输入、查找、排序、累加和统计等操作都变得十分方便。

一套编码,如果有较多的使用者,码的编制必须标准化,尤其是使用计算机的数据处理系统,合理的编码结构已经成为数据处理系统是否具有生命力的一个重要因素,因此编