

21 世纪计算机专业大专系列教材

李大友 主编

数据结构

彭 波 编著



清华大学出版社

<http://www.tup.tsinghua.edu.cn>



21 世纪计算机专业大专系列教材

数 据 结 构

李大友 主编

彭 波 编著

清 华 大 学 出 版 社

(京)新登字 158 号

内 容 简 介

本教材是《21 世纪计算机专业大专系列教材》之一。全书共分 9 章,第 1 章综述数据、数据结构、算法描述、算法分析,以及数据结构与其他课程之间的关系等。第 2 章至第 7 章介绍了基本的数据结构,如线性表、栈、队列、串、数组、广义表、树、二叉树及图等,分别讨论了数据的逻辑结构和存储结构,以及相应运算的算法。第 8 章和第 9 章为查找和排序,介绍了常用的几种查找方法和内部排序方法。

教材中使用类 C 语言作为算法描述语言,且所有算法都可以在任何一种 C 语言的开发环境中实现。在随书的配套光盘中可以看到这些算法的 C 语言程序。

本书中所介绍的数据结构概念清楚,内容丰富。为了有助于学生加深对基础理论知识的理解,培养实际应用的能力,各章(除第 1 章外)都配有与该章内容相关的操作应用举例,且配有大量习题。

本书可作为高等院校计算机专业大专数据结构课程的教材,也可作为非计算机专业本科生的教材。

版权所有,翻印必究。

本书封面贴有清华大学出版社激光防伪标签,无标签者不得销售。

书 名: 数据结构

作 者: 彭 波 编著

出版者: 清华大学出版社(北京清华大学学研大厦,邮编 100084)

[http:// www. tup. tsinghua. edu. cn](http://www.tup.tsinghua.edu.cn)

印刷者: 北京顺义振华印刷厂

发行者: 新华书店总店北京发行所

开 本: 787×1092 1/16 **印张:** 17.25 **字数:** 390 千字

版 次: 2002 年 3 月第 1 版 2002 年 3 月第 1 次印刷

书 号: ISBN 7-900641-67-X

印 数: 0001~8000

定 价: 25.00 元(配光盘)

《21 世纪计算机专业大专系列教材》

编辑委员会名单

主 编 李大友

编 委 (排名不分先后)

刘乐善 (华中理工大学)

刘惠珍 (北京工业大学)

陈 明 (石油大学)

邵学才 (北京工业大学)

蒋本珊 (北京理工大学)

匙彦斌 (天津大学)

葛本修 (北京航空航天大学)

彭 波 (中国农业大学)

策划编辑 范素珍

序

这套教材为 21 世纪高等学校计算机专业大专系列教材。

我们从 1995 年开始组织《计算机专业大专系列教材》。当时根据中国计算机学会教育委员会与全国高等学校计算机教育研究会联合推荐的《计算机学科教学计划 1993》的要求,组织了《计算机组成原理》等 13 本教材,并由清华大学出版社出版。这套教材出版后,受到了高等学校师生的广泛欢迎和好评。

在组织上述教材的时候,主要是按《计算机学科教学计划 1993》的要求进行的。而 1993 教学计划主要是参照美国 IEEE 和 ACM《计算机学科教学计划 1991》并结合我国高等教育当时的实际情况制定的,反映的是 20 世纪 80 年代末计算机学科的发展状况。

计算机学科是一个飞速发展的新兴学科,发展速度之快可谓一日千里。近 10 年来,计算机学科已发展成为一个独立学科,计算机本身向高度集成化、网络化和多媒体化迅速发展。但从另一个方面来看,高等学校的计算机教育一直滞后于计算机学科的发展,特别是教材建设,由于受时间和软硬条件的限制,更是落后于现实需要,而大专层次的教材建设问题尤其严重。为了改变这种状况,高等学校的教育工作者和专家教授们应当仁不让地投入必要的时间和精力来完成这一历史使命。

为组织好这套教材,我们认真地研究了全国高等学校计算机专业教学指导委员会和中国计算机学会教育委员会联合推荐的《计算机学科教学计划 2000》和美国 IEEE 和 ACM 两个学会最新公布的《计算机学科教学计划 2001》。这两个教学计划都是在总结了从《计算机学科教学计划 1991》到现在计算机学科十年来发展的主要成果的基础上诞生的。它们所提供的指导思想和学科所涵盖的内容,不仅适合于大学本科,也适合大学专科的需求,关键在于要对其内容的取舍进行认真的研究。

在我国的《计算机学科教学计划 1993》和美国 IEEE 和 ACM 两个学会提出的《计算机学科教学计划 1991》中,根据当时的情况,只提出了 9 个主科目。而在《计算机学科教学计划 2001》中,根据学科的最新发展状况,提出了 14 个主科目,其中 13 个主科目又为核心主科目。这 14 个主科目是:算法与分析(AL)、体系结构(AR)、离散结构(DS)、计算科学(CN)、图形学与可视化计算(GV)、网络计算(NC)、人机交互(HC)、信息管理(IM)、智能系统(IS)、操作系统(OS)、程序设计基础(PF)、程序设计语言(PL)、软件工程(SE)、社会、道德、法律和专业问题(SP),其中除 CN 为非核心主科目外,其他 13 个主科目均为核心主科目。

将美国 IEEE 和 ACM 的教学计划 2001 与 1991 计划进行比较可看出:在 1991 计划中,离散结构只是作为数学基础提出,未被列为主科目;而在 2001 计划中,不但列为主科

目,而且为核心主科目。可见,已将离散结构提升为本学科的基础。

在 1991 计划中,未提及网络计算,而在 2001 计划中,不但提出,而且被列为核心主科目,以适应网络技术飞速发展的需求。

图形学与可视化计算也是为适应发展需求新增的内容,并且列为主科目。

除此之外,2001 计划在下述 5 个方面做了增加或调整:

- 将程序设计语言引论调整为程序设计基础和程序设计语言两个核心主科目,显然,加强了对程序设计的要求。

- 将人-机通信调整为人机交互,反映了人-机通信的实质是人机交互。在图形界面迅速发展的今天,人机交互理论和方法的研究和应用变得十分重要。

- 将人工智能与机器人学调整为智能系统,拓宽了对智能系统的要求。

- 将数据库与信息检索调整为信息管理,因为后者不仅概括了前者,而且反映了数据库与信息检索的实质是信息管理。

- 将数值与符号计算调整为计算科学,更具有概括性。

总之,上述变化不仅更好地反映了计算机学科的发展现状,而且使 2001 教学计划具有更强的科学性和实用性。

由于这套系列教材主要面向的对象是计算机专业三年制大专(高职)学生,其培养目标也应属于高级技术人才的层次。他们既要有一定的理论基础(较本科弱),又要更强调实用性,要有明确的应用方向。我们将应用方向定位在信息管理和计算机网络两个方向。这两个应用方向占计算机应用总计的 90%以上。

在系列教材的内容取舍上,2001 教学计划的 14 门主科目中,我们概括了除智能系统、计算科学和社会、道德、法律和专业问题之外的其他 11 个主科目。在每个主科目中,我们都以其中的基本概念、基本理论和基本方法作为主线组织教材,使学生既能掌握基本的基础理论和方法,又能为他们进一步深造打下必要的基础;在信息管理和计算机网络技术两个应用方向上,他们的应用能力将得到加强。

根据上述指导思想,初步确定组织 20 本左右的教材供各高校选用。这些教材包括:《离散数学》、《计算机应用基础》、《计算机组织与结构》、《微机系统与接口技术》、《计算机网络与通信》、《网络管理技术基础》、《计算机网络系统集成技术》、《数据结构》、《操作系统原理》、《实用软件工程基础》、《数据库原理与应用》、《管理信息系统原理与应用》、《办公自动化实用技术》、《多媒体技术及其应用》、《Internet 技术及其应用》、《计算机维护技术》、《C 语言程序设计》、《Java 语言程序设计》、《C++ 语言程序设计》、《VB 语言程序设计》、《计算机英语》等。

系列教材并不是教学计划,由于各高校情况不同,培养方向的侧重面也不一样,因此教学计划也不会雷同。教材按系列组织,力图能够反映计算机学科大专层次的总体要求,同时采用大拼盘结构,各校可根据自身情况选择使用。例如,语言类教材,我们就准备了多本,各校可选择其中的一本或两本,其他依此类推。

这套教材均由高等学校具有丰富教学实践经验的老师编写。所编教材体系结构严谨、层次清晰、概念准确、理论联系实际、深入浅出、通俗易懂。相信一定能够得到专科院校计算机专业师生的欢迎。

全国高等学校计算机教育研究会副理事长
课程与教材建设委员会主任

李大友

2001.6

前 言

本书是按照美国电气电子工程师学会(Institute of Electrical and Electronic Engineers, IEEE)和计算机协会(Association for Computing Machinery, ACM)的《计算机学科教学计划 2001》的要求编写的。

“数据结构”在计算机科学中是一门综合性的专业基础课。数据结构的研究不仅涉及到计算机硬件(特别是编码理论、存储装置和存取方法等)的研究范围,而且和计算机软件的研究有着更密切的关系,无论是编译程序,还是操作系统,都涉及到数据元素在存储器中的分配问题。在研究信息检索时也必须考虑如何组织数据,以便查找和存取数据元素更为方便。可以认为数据结构是介于数学、计算机硬件和计算机软件三者之间的一门核心课程,是从事计算机科学及其应用的科技工作者必须掌握的重要课程。

本教材作为《21 世纪计算机专业大专系列教材》之一。教材在内容组织和编排上力求体现“先理论,后应用,理论与应用相结合”的原则,强调对理论知识的理解和运用。通过对本课程的学习,掌握各种数据结构的基本概念、逻辑特性和物理表示法,以及相应运算的算法;灵活运用各种数据结构解决实际问题,并且为学习后继专业课程打下良好的基础。

全书共分为 9 章。

第 1 章综述数据、数据结构、算法描述、算法分析,以及数据结构与其他课程之间的关系等。

第 2 章至第 7 章介绍了基本的数据结构,如线性表、栈、队列、串、数组、广义表、树、二叉树及图等,分别讨论了数据的逻辑结构和存储结构,以及相应运算的算法。

第 8 章和第 9 章为查找和排序,介绍了常用的几种查找方法和内部排序方法。

本教材有以下主要特点:

(1) 基础理论知识的阐述由浅入深、通俗易懂、逻辑性强。在内容的组织与编排上,略去了一些理论上的推导和繁琐的数学证明。

(2) 为了有助于学生加深对基础理论知识的理解,培养实际应用的能力,各章(除第 1 章外)都配有与该章内容相关的操作应用举例,且配有大量习题。

(3) 教材中使用类 C 语言作为算法描述语言,且所有算法都可以在任何一种 C 语言的开发环境中实现。在随书的配套光盘中可以看到这些算法的 C 语言程序。

(4) 附录中汇总了本教材各章中介绍的各类数据结构时用到的数据存储结构类型说明,供学生上机实验时参考使用。

本教材讲课时数为 60 学时左右,上机实验时数在 20 学时以上。教师可以根据学时数和学生的实际情况选讲操作应用举例中的例子。

本书可作为高等院校计算机专业大专学生数据结构课的教材,也可作为非计算机专业本科生的教材。

由于作者水平有限,在教材中难免有错误,恳请读者谅解。如果读者有问题需要与作者联系,请发送电子邮件到:pb_bau@263.net。

本教材由李大友教授主编和审定,彭波副教授编写。在上机实现教材中所有算法的过程中得到了曾立、许振文等同志的帮助,在此表示感谢。

编著者

2001.6

目 录

第 1 章 绪论	1
1.1 数据结构概述	1
1.2 数据结构的发展概况	2
1.3 数据结构与其他课程的关系	4
1.4 基本概念	5
1.5 算法描述及分析	7
1.5.1 算法的重要特性	7
1.5.2 算法的描述方法	8
1.5.3 算法的设计要求	10
1.5.4 算法效率的度量	10
1.5.5 算法的空间需求	12
习题	12
第 2 章 线性表	15
2.1 线性表的逻辑结构	15
2.1.1 线性表的定义	15
2.1.2 线性表的基本操作	16
2.2 线性表的顺序存储结构	17
2.2.1 线性表的顺序存储表示	17
2.2.2 基本操作在顺序表上的实现	18
2.2.3 线性表顺序存储结构小结	22
2.3 线性表的链式存储结构	23
2.3.1 线性表的链式存储表示	24
2.3.2 基本操作在单链表上的实现	24
2.3.3 循环链表	28
2.3.4 双向链表	28
2.3.5 线性表链式存储结构小结	31
2.4 线性表的两种存储结构比较	31
2.5 线性表操作应用举例	32
习题	37
第 3 章 栈和队列	40
3.1 栈	40

3.1.1	栈的逻辑结构	40
3.1.2	栈的顺序存储结构	41
3.1.3	栈的链式存储结构	44
3.2	队列	46
3.2.1	队列的逻辑结构	46
3.2.2	队列的顺序存储结构	47
3.2.3	队列的链式存储结构	50
3.3	栈和队列操作应用举例	53
习题		58
第4章	串	60
4.1	串的逻辑结构	60
4.1.1	串的定义	60
4.1.2	串的基本操作	61
4.2	串的存储结构	62
4.2.1	定长顺序存储结构	62
4.2.2	堆分配存储结构	65
4.2.3	块链存储结构	67
4.3	串操作应用举例	69
习题		76
第5章	数组与广义表	78
5.1	数组的逻辑结构	78
5.1.1	数组的定义	78
5.1.2	数组的基本操作	79
5.2	数组的顺序存储结构	79
5.3	矩阵的压缩存储	83
5.3.1	特殊矩阵的压缩存储	83
5.3.2	稀疏矩阵的逻辑结构	85
5.3.3	稀疏矩阵的存储结构	86
5.4	广义表	91
5.4.1	广义表的逻辑结构	91
5.4.2	广义表的存储结构	93
5.5	数组与广义表操作应用举例	94
习题		96
第6章	树与二叉树	98
6.1	树	98

6.1.1	树的逻辑结构	98
6.1.2	树的存储结构	101
6.2	二叉树	104
6.2.1	二叉树的逻辑结构	104
6.2.2	二叉树的基本性质	107
6.2.3	二叉树的存储结构	109
6.3	遍历二叉树	113
6.3.1	遍历二叉树的操作定义	113
6.3.2	遍历二叉树的递归算法	114
6.3.3	遍历二叉树的非递归算法	115
6.3.4	建立二叉树的算法	121
6.4	二叉线索树	122
6.4.1	二叉线索树的引出	122
6.4.2	二叉线索树的定义	123
6.4.3	二叉线索树的存储结构	124
6.4.4	二叉线索树的操作	125
6.5	树和森林与二叉树的转换	129
6.5.1	树与二叉树的转换	129
6.5.2	森林与二叉树的转换	132
6.5.3	树和森林的遍历	133
6.6	赫夫曼树及其应用	135
6.6.1	基本概念	136
6.6.2	赫夫曼算法	137
6.6.3	赫夫曼编码	137
6.6.4	赫夫曼树和赫夫曼编码的存储表示	139
6.6.5	赫夫曼编码的算法	139
6.6.6	示例	140
6.7	树与二叉树操作应用举例	142
	习题	146

第7章 图

7.1	图的逻辑结构	149
7.1.1	图的定义	149
7.1.2	图的基本操作	149
7.1.3	图的基本概念	151
7.2	图的存储结构	154
7.2.1	邻接矩阵表示法	154
7.2.2	邻接表表示法	157

7.2.3	十字链表表示法	160
7.2.4	邻接多重表表示法	162
7.3	图的遍历	164
7.3.1	深度优先搜索	164
7.3.2	广度优先搜索	165
7.4	最小生成树	167
7.4.1	生成树	167
7.4.2	最小生成树	168
7.5	最短路径	173
7.5.1	求某个源点到其他顶点的最短路径	174
7.5.2	求每一对顶点之间的最短路径	177
7.6	拓扑排序	179
7.6.1	AOV 网	179
7.6.2	拓扑排序	180
7.7	关键路径	183
7.7.1	AOE 网	183
7.7.2	关键路径的概念	184
7.7.3	关键路径的算法	184
7.8	图操作应用举例	187
	习题	191

第 8 章	查找	194
8.1	基本概念	194
8.2	静态查找	195
8.2.1	静态查找的基本操作	195
8.2.2	静态查找表的顺序存储结构	196
8.2.3	顺序查找	196
8.2.4	折半查找	197
8.2.5	分块查找	199
8.3	动态查找	200
8.3.1	动态查找的基本操作	200
8.3.2	动态查找表的二叉链表存储结构	201
8.3.3	二叉排序树	201
8.3.4	二叉平衡树	206
8.3.5	B 树	210
8.4	散列表	211
8.4.1	散列表的概念	211
8.4.2	散列函数的构造方法	213

8.4.3 处理冲突的方法.....	216
8.4.4 散列表的查找和分析.....	218
8.5 查找操作应用举例	219
习题.....	221
第9章 排序.....	223
9.1 基本概念	223
9.2 插入排序法	224
9.2.1 直接插入排序.....	224
9.2.2 希尔排序.....	225
9.3 交换排序法	227
9.3.1 冒泡排序.....	227
9.3.2 快速排序.....	228
9.4 选择排序法	230
9.4.1 直接选择排序.....	231
9.4.2 堆排序.....	232
9.5 归并排序法	237
9.5.1 两个有序序列的归并.....	238
9.5.2 一趟归并排序.....	238
9.6 基数排序法	239
9.6.1 多关键字排序.....	239
9.6.2 链式基数排序.....	240
9.7 各种内部排序法的比较	244
9.8 排序操作应用举例	245
习题.....	247
附录 数据存储结构综合.....	249
参考文献.....	257

第 1 章 绪 论

自从世界上第一台电子计算机问世以来,计算机科学和计算机软件、硬件技术得到飞速发展;计算机的应用领域也从最初的科学计算逐步发展到人类社会的各个领域。计算机加工处理的对象由简单的数值、字符发展到文字、图像、声音等各种复杂的带有不同类型及相互有不同关系的数据。为了编制“好”的程序,必须要分析程序处理的数据的特性及数据之间的关系,这就是“数据结构”这门学科形成和发展的背景。

1.1 数据结构概述

众所周知,计算机的程序是对数据进行加工处理。在大多数情况下,这些数据并不是无组织的,数据之间往往具有重要的结构关系,这就是数据结构的重要内容。那么,什么是数据结构呢?先不妨举例说明,然后给出明确定义。

例 1.1 一个大学的学生健康情况管理。

表 1.1 中的学生健康情况登记表就是一个数据结构。表中每个学生的情况为一个记录,它由姓名、学号、性别、年龄、班级和健康状况等 6 个数据项组成。计算机学生健康情况管理的主要功能包括:查询、浏览、插入、修改、删除和统计等。

在这种数据结构中,计算机处理的数据之间存在的是一种“一个对一个”的简单线性关系,称为线性数据结构。

表 1.1 学生健康情况登记表

姓名	学号	性别	年龄	班级	健康状况
孙 晨	980631	男	18	计 98	健康
高成甜	980632	女	17	计 98	一般
钱嘉羽	980633	男	19	计 98	近视
黄佳佳	980634	女	18	计 98	健康
⋮	⋮	⋮	⋮	⋮	⋮

例 1.2 一个大学的人事档案管理。

图 1.1 中的数据结构像一棵树根在上枝干在下的倒挂的树,清晰地描述了教师和学生所在的系和专业。在这种结构中可以从树根沿着某系某专业很快找到某个教师或学生的数据,查找的过程就是从树根沿树枝到某个叶子的过程。

在这种数据结构中,计算机处理的数据之间存在的是一种“一个对多个”的层次关系,称为树型数据结构。

例 1.3 在 n 个城市之间建立通信网络,要求在其中任意两个城市之间都有直接的

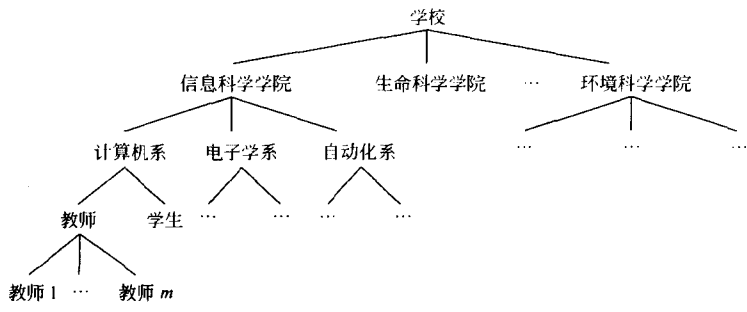


图 1.1 教师学生名册

或间接的通信线路,在已知某些城市之间直接通信线路预算造价的情况下,使网络的造价最低。

当 n 很大时,这样的问题只能用计算机来求解。用图 1.2(a)来描述 7 个城市之间的通信线路,其中:图中圆圈表示一个城市,两个圆圈之间的连线表示对应城市之间的通信线路,连线上的数值表示该通信线路的造价。这一描述的结构为图状结构,利用计算机可以求出满足要求的通信网络,如图 1.2(b)所示。

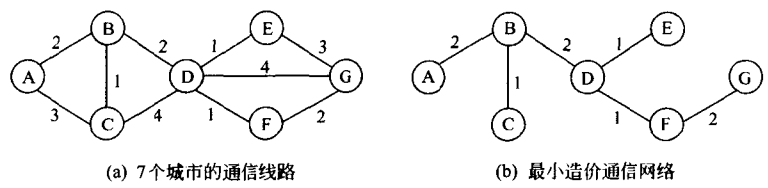


图 1.2 7 个城市的通信线路及最小造价通信网络

通过上面 3 个例子可以看出:数据结构中数据元素之间存在着逻辑关系。数据结构就是解决如何分析数据元素之间的关系,如何确立合适的逻辑结构和如何存储这些数据,并对为完成数据操作所设计的算法做出时间和空间的分析。

简单来说,数据结构是研究非数值计算的程序设计问题中数据以及它们之间的逻辑关系和对数据的操作的一门课程。重点分析数据之间的抽象的相互关系,而不涉及数据的具体内容。

1.2 数据结构的发展概况

回顾一下数据结构课程的发展与形成过程对于理解数据结构的内容和重要性是很有意义的。

(1) 20 世纪 60 年代初期 国外还没有专门的“数据结构”课程,但是数据结构的有关内容已经散见于《编译原理》和《操作系统》之书中。

(2) 20 世纪 60 年代中期 有些国家的大学在这期间开始设立有关课程,但是当时课程的名称并不叫“数据结构”,而命名为“表处理语言”。它的主要内容是研究当时已经出

现的几个表处理系统。比如, J. Weizenbaum(美国)在 20 世纪 50 年代初设计的 SLIP 系统(简单表处理语言);在 1954—1959 年期间由 A. Newell(美国)等人设计的 IPL-V 系统(信息处理系统);在 1959—1960 年由 J. McCarthy(美国)设计的 LISP 系统(表处理语言);以及在 1962 年由 D. Farber(美国)等人设计的 SNOBOL 系统(串处理语言)。它们的共同特点是数据对象的结构形式或者是表结构,或者是树结构。像 LISP 的数据结构就是二叉树结构,SNOBOL 的数据结构是表和树结构。

上述的几种语言是以数据为中心,为处理非数值问题设计的。比如 LISP 就是为解决人工智能问题而设计的表处理语言。而我们所熟知的 FORTRAN、ALGOL 等算法语言则是为解决数值问题而设计的,它们侧重于以程序为中心。比如 ALGOL-60 分程序和过程就是以程序为中心的例子。

以程序为中心的观点侧重于建立程序,只是当数据成为程序加工的对象时,才考虑到数据。这种观点适合于数值计算问题,这类问题属于在简单数据结构上进行复杂函数变换的问题。以数据为中心的观点是把数据结构作为问题的中心部分(如数据库),而把程序看成是围绕着数据结构缓慢爬行的小虫,它时而询问,时而修改或扩充当前驻留在内存中的数据。这种观点适合于航空订票系统、信息管理系统、情报检索系统等非数值问题的解决,它们都要求采用复杂的数据结构描述系统的状态,它们的运算是实现对于数据结构的访问或改变等。某些科学家曾断言,程序设计以设计为中心的观点,将对未来程序设计语言的设计产生重大影响。

(3) 20 世纪 60 年代后期 1968 年在美国一些大学计算机科学系的教学计划中曾明确规定“数据结构”为一门课程,但是对该课程的内容范围并没有做具体的限定。其后发表的有些文章和出版的书籍中,对“数据结构”这个术语有各种不同的理解和解释。最初,数据结构几乎和图论是同义语,特别是表和树的理论(这些理论是描述分层数据的有力武器)。随后这个概念又扩充到包括网络、代数、集合论、关系等现在称之为“离散数学结构”的那些内容。它与现在称为“数据结构”的某些内容混合在一起,总称为“数据结构”。

由于数据必须在计算机中进行处理,因此不能局限于数据本身的数学概念的研究,还必须考虑数据的物理结构,即数据在存储器中的表示问题,这就进一步扩大了数据结构的内容。

特别是 1968 年,美国著名计算机科学教授 D. E. Knuth(唐·欧·克努特)所著的《The Art of Computer Programming》Volume 1(《计算机程序设计技巧》第一卷)1968 年出版,对计算机科学的发展作出了重大贡献。该书作者论证了任何语言都可采用像表处理语言那样的技术。这部著作比较全面地、系统地讨论了几种数据结构(Knuth 称为信息结构),定义了运算,并用英文加上汇编语言描述算法,比较详细地分析了算法的效率。尔后,逐渐将数据的有关数学概念独立出来,形成了现在的“离散数学结构”,而把数据的逻辑结构、物理结构及对每种结构所定义的运算形成了“数据结构”的主要内容。

(4) 20 世纪 70 年代初期 此期间出现了大型程序,软件也相对独立,结构化程序设计成为程序设计方法学的主要内容,人们越来越重视数据结构,认为程序设计的实质是对确定的问题选择一种好的结构,加上设计一种好的算法,也就是人们所说的“程序 = 数据结构 + 算法”。