

高等教育自学考试自学辅导丛书

数据组织与管理辅导与练习

计算机信息管理专业

王 珊 主编



13

12

全国高等教育自学考试指导委员会办公室自学指导服务中心组织编写



经济科学出版社

1-1000.13
1-1000.13
WC/2

高等教育自学考试自学辅导丛书

数据组织与管理 辅导与练习

计算机信息管理专业

全国高等教育自学考试指导委员会办公室
自学指导服务中心组织编写

王 珊 主编



经济科学出版社

1354216

责任编辑：孙怡红

责任校对：徐领弟

版式设计：周国强

技术编辑：王世伟

数据组织与管理辅导与练习

计算机信息管理专业

王 珊 主编

社址：北京海淀区万泉河路 66 号 邮编：100086

网址：www.esp.com.cn

电子邮件：esp@public2.east.net.cn

(版权所有 翻印必究)

出版部电话：62630591 发行部电话：62568485

经济科学出版社出版、发行 新华书店经销

中国铁道出版社印刷厂印刷

河北省三河市华丰装订厂装订

850×1168 毫米 32 开 7 印张 250000 字

1999 年 11 月第一版 1999 年 11 月第一次印刷

印数：0001—3000 册

ISBN 7-5058-1701-9/G·354 定价：10.50 元

(图书出现印装问题，本社负责调换)

图书在版编目 (CIP) 数据

数据组织与管理辅导与练习/王珊主编. —北京: 经济科学出版社, 1999. 11

(高等教育自学考试自学辅导丛书)

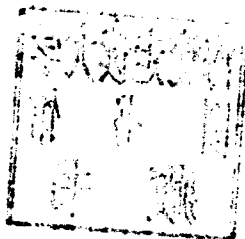
计算机信息管理专业

ISBN 7-5058-1701-9

I. 数… II. 王… III. 数据库系统-高等教育-自学考试-自学参考资料 IV. TP311. 13

中国版本图书馆 CIP 数据核字 (1999) 第 39096 号

JS338/29



前 言

为了完善高等教育自学考试教育形式,弥补考试的局限性,促进高等教育自学考试的发展,根据全国考办的工作部署,我们组织编印了《高等教育自学考试自学辅导丛书》。

该套丛书以全国考委公布的课程自学考试大纲为依据,以全国统编教材为蓝本,相关专家、学者担任各册辅导书的主编和主审,旨在帮助自学者达到学习目标,顺利通过国家考试。

自学辅导书是高等教育自学考试教育媒体的重要组成部分,我们将在全国考办和各专业委员会的指导帮助下,根据专业的开考情况和考生的实际需要,陆续组织编写、出版文字、音像和计算机多媒体自学辅导资料,由此构成与大纲、教材相配套的、完整的学习体系。

全国高等教育自学考试指导委员会办公室

自学指导服务中心

1997 年 1 月

愿天下有志者皆成人才

——《自学辅导丛书》总序

杨学为

中国独创的高等教育自学考试，既是一种国家考试制度，又是一种个人自学、社会助学、国家考试相结合的教育形式。自 20 世纪 80 年代初创立以来，已使许多自学者获得了大专、本科文凭。这一所投资省、适应面广、质量高的没有围墙的“大学”已受到了社会各界的欢迎，引起了世界的瞩目。为了进一步完善它，帮助更多的公民实现求学的理想与成才的追求，我们将不遗余力地把“学校”送到每个自学者的面前，把成才之路铺在每一个自学者的脚下。给自考者提供高质量、有实效的自学辅导材料就是这项工作的重要组成部分。基于这样的认识，我们邀请有关专家学者编写了这套《自学辅导丛书》（以下简称《丛书》）。

《丛书》的编写者、审订者们不仅有很深的学术造诣和丰硕的研究成果，而且在长期的教书育人事业中积累了丰富的经验。在繁重的教学、研究的同时，他们还以极大的热情投入到了高等教育自学考试这一崇高的事业中。对于高等教育自学考试的基本规律，自考者学习过程中的基本特点，加强自学指导的重要意义等等，他们都有深刻的认识和独到的见解。他们从内心深处愿意为每一位有勇气踏上自学之路的人奉献出全部的智慧 and 力量。所有这些，就是我们这套丛书的基石。

这套《丛书》既然是以帮助考生为根本宗旨，在编写过程中我们就力图体现下列特征：

1. 理清脉络，建立结构。掌握一门学科，最关键的是弄清其

独特的知识体系与结构，从总体上有一个明晰的框架。在此基础上，再装入基本事实、基本理论，就可以学得通，记得住，用得活。《丛书》在理清学科的脉络，帮助读者理解知识结构与体系方面想了不少办法。在编写每一部分时，都力图做到先总体后部分再回到总体。

2. 突出重点，突破难点。人们最初接触一门学科时，往往抓不住重点，找不出难点，平均使用力量，结果是费了不少力还不得要领。《丛书》根据各学科特点，把重点明确告诉读者，围绕重点把相关学科知识组织起来，把难点明确地提出来，简要地分析了其成为难点的原因及攻破难点的方法。

3. 学练结合，联系实际。学任何一门学科都必须做一定量的习题，都必须用学科知识解决现实生活、生产中的问题。《丛书》精选了那些有代表性的、能举一反三的问题并做了适当的分析，使读者能在做习题的过程中巩固已学的知识，加深对有关知识的理解。为帮助考生联系实际，《丛书》也选择了一些实例，通过这些实例，读者能够学习到理论联系实际的具体方法。

4. 纵横联系，指导方法。每门学科内部各章节之间，它与相关学科之间都有内在的联系。只有把握了这种纵横联系，才能加深理解，融会贯通，使各章节、各学科的学习相互补充、相得益彰。《丛书》充分考虑了读者在这一方面的困难，把应揭示的联系都予以揭示，但又把握了分寸，不至于使读者不知云里雾里。每门学科都有自己独特的学习、研究方法，只有掌握了这些方法，才算找到了打开该学科知识宝库的钥匙，才能收到既掌握知识又培养能力的实效，才能提高学习的效果与效率。《丛书》在指导方法上，密切结合学科内容，力求简明易懂，便于操作。

为了使考生读得懂，喜欢读，见成效，《丛书》在文字上力求简明扼要，通俗易懂；在行文上力求生动流畅，不绕弯子；在形式上力求灵活变化，适合自学者的情趣。

通过这些努力，我们期望达到的目的是：

1. 减缓坡度。在实现一个知识点转到另一个知识点、在从这一章转到下一章的转折时，借助各种中介而不至于使考生感到太吃力、太困难。

2. 减小难度。在学习新知识时有旧知识的铺垫，有相关的背景知识做向导，有深入浅出的分析，有一定的实例。这样，学习者所遇到的困难和压力就相应地得到缓解。

3. 拓展深度。在掌握一门学科时，不至于只知道一些表皮的东西，对一些基本理论、基本概念，要既知其然，又知其所以然；既知其一，又知其二。同时，对一些前沿问题也有所涉猎。

4. 实现高度。既能通过国家考试，获得文凭；又学到了知识，培养了能力，实现了个人素质的提高，这才是我们理解的高等教育自学考试，这才是我们助学的终极目标。

我们尽了绵薄之力来体现自己的宗旨，但能否如愿，应由广大考生去评定。我们诚恳地欢迎每一位考生提出意见和建议，从而进一步改进我们的工作，使每一个考生都能得到更切合实际，更有成效的指导与帮助。

作为高等教育自学考试的一名工作者，我有义务不厌其烦地告诫参加自考的朋友们：一定要在钻研大纲、教材的基础上使用《丛书》。那种平时不在大纲、教材上下功夫，只寄希望于突击背诵辅导材料以应付考试的办法是不足取的，它已使不少人走了弯路。“以大纲为纲，以教材为本”是我们自学者应遵循的基本原则。

人类的知识是无穷无尽的，自学之路也因之曲折而漫长。愿我们的工作能助自学者一臂之力，愿天下有志者皆成人才。

一九九六年冬

目 录

第一篇 数据结构

绪 论	1
第一章 数据组织概述	4
1.1 数据的逻辑结构	4
1.2 数据的物理结构	5
1.3 数据的运算及算法的“大 O”表示	8
第二章 线性结构	9
2.1 顺序存储的线性表	10
2.2 链接存储的线性表	13
第三章 树形结构	44
3.1 二叉树	45
3.2 二叉查找树	53
第四章 排序	57
4.1 几种简单的内排序方法	57
4.2 快排序和堆排序	63
4.3 合并排序与外排序	68
第五章 查找	75
5.1 线性表的查找	75
5.2 哈希表的查找	78
第六章 文件组织	84
6.1 无索引的文件组织方式	84
6.2 有索引的文件组织方式	86

附 1 综合练习题	92
附 2 参考答案	103

第二篇 数据库系统基础

第七章 数据库系统介绍	119
7.1 数据库系统概述	119
7.2 数据模型	122
第八章 关系数据库	124
8.1 关系模型	124
8.2 关系数据语言	126
8.3 数据库系统的三级模式与数据独立性	148
第九章 数据库保护	149
9.1 安全性	149
9.2 完整性	152
9.3 并发控制	154
9.4 恢复	157
第十章 数据库设计	161
10.1 数据库设计概述	161
10.2 需求分析	161
10.3 概念结构设计	162
10.4 逻辑结构设计	163
10.5 物理结构设计	164
10.6 数据库的实施和维护	165
 附 1 综合练习题	 169
附 2 参考答案	175

第三篇 关系数据库管理系统实例

第十一章	INFORMIX 概述	182
11.1	INFORMIX 公司简介	182
11.2	INFORMIX 产品构成	182
第十二章	关系数据库管理工具 ISQL	187
12.1	安装 ISQL、EasyBase 数据库初始化	188
12.2	启动 EasyBase 数据库和 ISQL	189
12.3	ISQL 窗口和菜单	189
第十三章	数据库应用开发举例	194
13.1	问题的背景、需求分析	194
13.2	系统设计	195
13.3	建立数据库	195
13.4	编写应用程序	198
第十四章	INFORMIX-OnLine 的基本功能	199
附 1	教材习题答案	200
附 2	综合练习题及答案	208

第一篇 数据结构

绪 论

一、《数据结构》是什么？

《数据结构》是研究非数值计算的程序设计问题中，计算机的操作对象以及它们之间的关系和操作的学科。通常，用计算机解决一个具体问题时，首先要分析问题，从具体问题中抽象出一个处理模型来，然后设计一组求解问题的算法，最后编码、调试，直到得到最终结果。这个处理模型就是所谓数据的逻辑结构，它包括操作对象及对象间的关系。在利用这个逻辑结构进行数据处理之前，又要先把它按一定方式存储到计算机中，这就是存储结构或物理结构。选定存储结构之后，我们就可以根据问题的要求来设计算法了。同一个逻辑结构，往往可以采用不同的存储方式，相应解决问题的算法也随之不同，因此，还应研究对不同的处理要求，选用什么样的存储结构更合理，以及如何评价不同的存储结构中，各种处理方法的时间和空间复杂度等问题。在本课程中，我们不研究如何从具体应用中抽象出处理模型的问题，而着重讨论对已定义好的逻辑结构，如何进行物理存储和进行有关操作的算法及对算法的粗略评价。

二、算法、算法描述及实现

学习《数据结构》的一个重要内容，就是要学会设计在每种数据结构中进行有关操作的算法。所谓算法，是对特定问题求解步骤的一种描述，它是指令的有限序列，这些指令是有穷的、确定的、可行的，并且是可获得最终输出结果的。为了描述算法，需要选用一种描述语言，它能清楚地反映算法的设计思想和处理逻辑，而且易读、易懂、易交流，又不拘泥于某种具体语言的语法细节。因为 PASCAL 是本课程的先修课，所以，本书采用类 PASCAL 作算法描述语言，它既能满足写算法的要求，又能方便地转换为具体的 PASCAL 语言或其他高级语言（如 C，C++ 等），以便算法的上机实现。

三、《数据结构》是计算机系的综合性的基础课。正如著名计算机科学家沃思（Niklaus Wirth）提出的《数据结构+算法=程序》，显然算法和数据结构是计算机上所处理问题的两个基本组成部分，其重要性不言而喻，它不仅在本书中是后两篇的基础，而且也是实现编译程序、操作系统和应用程序（特别是非数值计算的程序设计）的基础。

四、《数据结构》的内容安排和学习方法

教材第一篇共分六章，第一章概述对数据进行组织的基本内容和研究方法；第二、第三章是本篇的重点，着重介绍两种数据结构——线性表和树；第四章、第五章主要是对第二章中讨论的基本操作的延伸，专门讨论线性表的排序和查找（包括对哈希表的查找）；第六章主要是研究数据结构在外存中的组织。

从基本内容来说，本篇介绍的数据结构就是线性表（及其特例栈和队列）和树以及数据在外存的组织——文件。对每种数据结构，我们都要讨论三方面的内容，即逻辑结构、物理结构及有关操作的算法。逻辑结构可以说是数据结构的定义，它指结点的有限集合及结点间关系的有限集合，这是容易理解和掌握的。而物理结构（即存储结构），需要借助于高级语言的数据类型来描述，读者对高级语言中的数据类型是熟悉的，所以，也不存在困难。除了在外存各种

文件组织方式相对复杂一些外，在内存，我们主要使用顺序存储和链接存储，这也是易于理解和掌握的。因此，只有在各种数据结构中，设计进行有关操作的算法，才是相对困难的部分。

我们可把基本内容分为两大部分来掌握，即一部分是：基本理论、基本知识，属于理解、记忆的部分，而属于基本技术的另一部分则是算法设计和上机实践。

第一章、第五章的大部分内容及第六章基本不涉及写算法，着重掌握有关的基本内容即可。读者可先阅读教材，按课本上所给的习题及“自学指导”^①第三期的模拟训练题进行自测，再阅读参考答案，对未掌握的部分，继续钻研教材，补漏拾遗，不断反复，终究是不难掌握的。

对于第二章、第三章、第四章，除一些基本理论、基本知识外，主要精力要用在练习写算法上，可由浅入深地，先掌握最简单的、最基本的算法，然后再选更难一些的题目进行练习。

通常，在你动手写算法前，要先钻研教材，如对书上的例子，你可画出相应的示意图（若可能的话）或用一组实例数据来“手工”地执行一遍算法，看是否能得到预期的结果，如果还弄不明白，可以把它转化成 PASCAL 语言上机调试，并追踪程序的执行或在你不太理解的地方加上一些调试语句（如输出中间结果）。总之，写算法必须要真正理解，死记硬背是不行的。当你对例子没有疑问了，就可以做一些与例子相似的题目，或对例子加一些改动。如何检验自己写的算法是否正确，对自学者来说是比较困难的，除了可参阅《自学指导》第四期的参考答案外，惟一的办法就是上机调试了。通过上机调试，查出错误时，要认真总结错误原因，积累经验，这对培养自己的编程和调试能力都是很重要的。

下面是对原教材的分章辅导。

^① “自学指导”：全国电子信息应用教育中心编，全国高等教育自学考试计算机信息管理专业“自学指导”（本科段），第三期、第四期。后文中的“自学指导”均同此。

第一章 数据组织概述

通过本章的学习，应明确《数据结构》主要包括哪三方面的内容。

1.1 数据的逻辑结构

为了对数据进行组织，我们首先定义了什么是数据？数据是客观事物的符号表示，是能输入到计算机内并能被计算机处理的符号的总称。

数据有一定的层次，从低到高的划分是：项（基本项，又称域或字段，它是数据的具有独立含义的不可分割的最小单位）、组项、记录、文件。和 PASCAL 语言中的划分类似，PASCAL 的记录中包含若干域，这个域如果是简单类型就是这里所谓的基本项，如果是构造类型，它又可以是数组或记录等，我们这里可以称之为组项，意即它是由基本项组合而成的，是可以分解的。例如，一个学生记录中，学号、姓名、性别是基本项，不可再分，而“成绩”则可看成是组项，例如，它可以包含“PASCAL”、“数据结构”、“数据库”等三科成绩，在 PASCAL 中，可以把“成绩”这一组项定义成记录或数组，当然，也可以把它定义成三个基本项。同类型记录值的集合就是文件，和 PASCAL 的定义是完全一致的。

为了便于定义数据的逻辑结构，我们又引入了结点的概念，用它来代表实际问题中作为整体来进行考虑和处理的数据单位。结

点的描述可以和 PASCAL 的记录类型定义相对应,只是我们可以把它抽象成只含一个数据项,并可抽象地用 NODETYPE 来表示任意结点类型,这就给数据结构的研究带来极大方便。因为只有抽象的结点才可能用来代表任何我们所关心的处理对象中的一个成员(或一个元素),我们才可以在不考虑它的内部细节的情况下,对各种操作的算法进行讨论,而且不失一般性。当我们需要对实际问题进行处理时,又可以把一个结点用一个具体的记录来代替,算法却无需改变。例如,我们要对教材 P. 30 表 1—1 的工资表按实发工资从小到大排序,则我们不妨把表中的一行抽象成只有实发工资一项(在我们的课程中,称之为排序码)来讨论算法,显然不会影响到对算法本身的研究。

有了结点的概念,定义数据的逻辑结构就十分容易了。我们把“结点的有限集合和结点间关系的有限集合”定义为“数据结构”,这就是说,一个“数据结构”,首先要有有限个结点(即我们所要处理的对象中的成员集合),结点间又存在一定的关系(也称为“结构”),根据结点间的关系,可以把“数据结构”分为线性结构和非线性结构两大类,非线性结构又分为树结构和图结构,在我们的课程中,只讨论线性结构和树形结构这两类。

结点间的关系,无论是线性关系还是非线性关系,我们都可以用序偶或有向线段(图示法)来表示,而后者更为直观。例如,你只要看到《数据组织与管理》教材中,P. 32 的图 1—2 和 P. 33 的图 1—3,你的脑海中就有了清晰的线性表和树的形象,知道下一步的工作就是要把它们存到计算机中去,并讨论其若干操作的算法。

1.2 数据的物理结构

由于要把数据结构存放在计算机中,才能对它进行有关操作,

因此要考虑对数据结构的存储，称数据的物理存储。

数据结构的存储结构分两个方面，即对结点的存储和对结点间关系的存储。除本篇第六章专门讨论数据结构在外存的存储（文件组织）外，其余均考虑在内存的存储。

由于讨论算法时，首先就要选定存储结构，所以，物理结构可以在讨论算法时再详细考虑，本章只需了解我们要采用的存储结构有哪些种类就行了。

1.2.1 内存中所采用的存储结构

(1) 顺序存储

我们利用高级语言的一维数组来表示顺序存储，而数组元素的类型，即结点的类型，可用 PASCAL 的记录类型定义（有时也可以把结点抽象成一个数据项来处理）。顺序存储只能表示结点间的一个顺序关系，即利用存储顺序（数组的下标顺序）表示结点间的逻辑顺序。例如，我们有一张学生情况表，表的一行是一个学生的情况，表中的各行是按学号升序存到数组中，则下标的顺序既是物理顺序，也代表了逻辑顺序，因而不需另外存储结点间的逻辑关系了。显然，如果结点间有多于一种关系，则顺序存储最多只能反应其中一种关系，例如按学号的升序存储，某科成绩就不可能也是升序了，因此，其他关系就必须通过链接表示，比如可按某科成绩的升序将记录链接起来。若结点间的关系是非线性关系（例如树），则除特殊情况外，都应使用链接存储。

(2) 链接存储

链接存储可用来表示结点间的线性关系和非线性关系。对线性关系来说，当物理顺序和逻辑顺序不一致时，需要用链接存储来对结点间的逻辑关系进行存储。链接存储也有各种形式，最基本的是在每个结点中，增加一个能指出它在某种关系中的后继结点的地址的数据项，我们称之为指针。当然，若在某种关系中，结点有若干个后继，就需按后继的个数设指针项的数目。链接存储