



附：数据结构导论自学考试大纲

数据结构导论

组编 / 全国高等教育自学考试指导委员会
主编 / 陈小平

全国高等教育自学考试指定教材

计算机及应用专业 (专科)

11.12

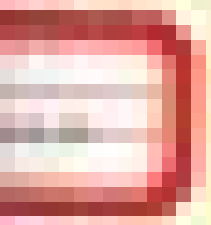
版社



新編增補

卷之...

...



全国高等教育自学考试指定教材

计算机及应用专业（专科）

数据结构导论

（附：数据结构导论自学考试大纲）

全国高等教育自学考试指导委员会组编

陈小平 主编

经济科学出版社

图书在版编目 (CIP) 数据

数据结构导论/陈小平主编. - 北京: 经济科学出版社, 2000.3

ISBN 7-5058-2046-X

I. 数… II. 陈… III. 数据结构-高等教育-自学考试-教材 IV. TP311.12

中国版本图书馆 CIP 数据核字 (1000) 号

数据结构导论

(附数据结构导论自学考试大纲)

全国高等教育自学考试指导委员会组编

陈小平 主编

经济科学出版社出版

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

北京第二外国语学院印刷厂印刷

787×1092 16 开 13.25 印张 280000 字

2000 年 3 月第一版 2000 年 3 月第一次印刷

印数: 001-20100

ISBN 7-5058-2046-x/G·435 定价: 18.00 元

(图书出现印装问题, 请与当地教材供应部门调换)

(版权所有 翻印必究)

内 容 提 要

本书系统地介绍了各种常用的数据结构,对基本概念、基本原理和基本方法做了深入浅出的介绍,对有关的算法设计做了详细和通俗的讲解,并对有关背景做了适当交待。每章附带小结和适量的习题。上述特点使本书特别适合于自学。

本书可供计算机及其应用专业自学考试学生及专业人员使用,也可作为高等院校数据结构课程的教科书或参考书。

组编前言

当您开始阅读本书时,人类已经迈入了二十一世纪。

这是一个变幻难测的世纪,这是一个催人奋进的时代。科学技术飞速发展,知识更替日新月异。希望、困惑、机遇、挑战,随时随地都有可能出现在每一个社会成员的生活之中。抓住机遇,寻求发展,迎接挑战,适应变化的制胜法宝就是学习——依靠自己学习、终生学习。

作为我国高等教育组成部分的自学考试,其职责就是在高等教育这个水平上倡导自学、鼓励自学、帮助自学、推动自学,为每一个自学者铺就成才之路。组织编写供读者学习的教材就是履行这个职责的重要环节。毫无疑问,这种教材应当适合自学,应当有利于学习者掌握、了解新知识、新信息,有利于学习者增强创新意识、培养实践能力、形成自学能力,也有利于学习者学以致用、解决实际工作中所遇到的问题。具有如此特点的书,我们虽然沿用了“教材”这个概念,但它与那种仅供教师讲、学生听,教师不讲、学生不懂,以“教”为中心的教科书相比,已经在内容安排、形式体例、行文风格等方面都大不相同了。希望读者对此有所了解,以便从一开始就树立起依靠自己学习的坚定信念,不断探索适合自己的学习方法,充分利用已有的知识基础和实际工作经验,最大限度地发挥自己的潜能,以达到学习的目标。

欢迎读者提出意见和建议。

祝每一位读者自学成功。

全国高等教育自学考试指导委员会

1995.5

编者的话

数据结构是计算机及其应用专业的一门重要的专业基础课,它为计算机专业人员提供重要的专业基础知识和必要的技能训练。

本书是根据全国高等教育自学考试指导委员会的要求,专门为自学考试计算机及其应用专业(专科)的考生而写的教材。为了满足自学的需要,在符合数据结构导论自学考试大纲的前提下,本书对有关内容做了精选,力求重点突出、叙述清楚,并对相关背景做了适当交待。对自学中比较困难的部分,特别是算法设计,做了较详细的、尽可能通俗的解释。因此,本书不仅适合于计算机及其应用专业自学考试的考生使用,也可作为普通高等院校数据结构课程的教材或教学参考书。

郑诚编写了本书的第三章、第五章和第七章;陈小平负责编写其余各章及全书的统稿。安徽大学计算机系倪志伟同志、中国科大计算机系胡国平同学和姜磊同学也参加了本书的部分工作。

南京大学计算机系金志权教授、中国科大计算机系赵保华教授和安徽大学计算机系谢荣传教授仔细审阅书稿和自考大纲,并提出了宝贵的意见和建议。作者还得到了中国科大计算机系陈国良教授和上海交通大学电子信息学院陈敏逊教授的帮助。本书的责任编辑崔岱远同志亦为本书付出辛勤的劳动。谨向以上诸位表示衷心的感谢。

受编者水平和时间方面的限制,本书难免存在各种缺陷,恳请读者批评指正。

主编

1999年11月17日

目 录

数据结构导论

第 1 章 概论	(1)
1.1 引言.....	(1)
1.2 数据、逻辑结构和运算.....	(3)
1.2.1 数据元素和数据项.....	(3)
1.2.2 数据的逻辑结构.....	(4)
1.2.3 运算和基本运算.....	(5)
1.3 存储实现和运算实现.....	(7)
1.3.1 存储实现.....	(7)
1.3.2 运算实现.....	(8)
1.4 算法分析.....	(9)
1.5 数据结构及其评价和选择.....	(11)
1.5.1 数据结构.....	(11)
* 1.5.2 数据结构的评价和选择.....	(12)
小结.....	(13)
习题.....	(14)
第 2 章 线性表	(15)
2.1 线性表的基本概念.....	(15)
2.1.1 线性结构.....	(15)
2.1.2 线性表.....	(16)
2.2 线性表的顺序实现.....	(16)
2.2.1 顺序表.....	(16)
2.2.2 基本运算在顺序表上的实现.....	(17)
2.2.3 顺序实现的算法分析.....	(20)
2.3 线性表的链接实现.....	(21)
2.3.1 单链表.....	(21)
2.3.2 单链表的简单操作.....	(23)
2.3.3 基本运算在单链表上的实现.....	(26)
2.4 其它运算在单链表上的实现.....	(28)

2.4.1	建表	(28)
2.4.2	清除重复结点	(30)
2.5	其它链表	(32)
2.5.1	循环链表	(32)
2.5.2	双链表	(33)
2.6	顺序实现与链接实现的比较	(35)
2.6.1	空间性能的比较	(35)
2.6.2	时间性能的比较	(35)
2.7	串	(35)
2.7.1	串的基本概念	(36)
2.7.2	串的基本运算	(36)
2.7.3	串的存储	(37)
小结		(39)
习题		(40)
第3章	栈、队列和数组	(42)
3.1	栈	(42)
3.1.1	栈的基本概念	(42)
3.1.2	栈的顺序实现	(43)
3.1.3	栈的链接实现	(45)
3.1.4	举例	(47)
3.1.5	栈的简单应用和递归	(48)
3.2	队列	(51)
3.2.1	队列的基本概念	(51)
3.2.2	队列的顺序实现	(51)
3.2.3	队列的链接实现	(56)
3.2.4	举例	(58)
3.3	数组	(59)
3.3.1	数组的逻辑结构和运算	(59)
3.3.2	数组的存储结构	(60)
3.3.3	矩阵的压缩存储	(61)
3.4	综合应用示例	(66)
小结		(71)
习题		(71)
第4章	树	(74)
4.1	树的基本概念	(74)
4.2	二叉树	(76)
4.2.1	二叉树的基本概念	(76)
4.2.2	二叉树的性质	(77)
4.3	二叉树的存储结构	(79)
4.3.1	二叉树的链式存储结构	(79)
4.3.2	二叉树的顺序存储结构	(81)
4.4	二叉树的遍历	(82)
4.5	递归消除	(86)

4.5.1	简单递归消除	(86)
4.5.2	基于栈的递归消除	(89)
4.6	树和林	(92)
4.6.1	树的存储结构	(92)
4.6.2	树的遍历	(96)
4.6.3	树、林与二叉树的关系	(96)
4.7	判定树和哈夫曼树	(98)
4.7.1	分类与判定树	(98)
4.7.2	哈夫曼树与哈夫曼算法	(100)
小结		(103)
习题		(104)
第5章	图	(106)
5.1	图的基本概念	(106)
5.1.1	图的实际背景	(106)
5.1.2	图的定义和术语	(107)
5.2	图的存储结构	(109)
5.2.1	邻接矩阵	(109)
5.2.2	邻接表	(111)
5.3	图的遍历	(113)
5.3.1	连通图的深度优先搜索	(113)
5.3.2	连通图的广度优先搜索	(115)
5.3.3	图的连通分量计算	(116)
5.4	最小生成树	(117)
5.5	拓扑排序	(121)
小结		(124)
习题		(124)
第6章	查找表	(128)
6.1	基本概念	(128)
6.1.1	集合的基本概念	(128)
6.1.2	查找表的基本概念	(129)
6.2	静态查找表的实现	(130)
6.2.1	顺序表上的查找	(130)
6.2.2	有序表上的查找	(132)
6.2.3	索引顺序表上的查找	(133)
6.3	树表	(134)
6.3.1	二叉排序树	(135)
6.3.2	平衡二叉排序树	(139)
6.4	散列表	(143)
6.4.1	散列函数的构造法	(144)
6.4.2	动态查找表在开散列表上的实现	(145)
6.4.3	动态查找表在闭散列表上的实现	(147)
6.4.4	开散列表与闭散列表的比较	(149)
小结		(150)

习题	(151)
第7章 文件	(152)
7.1 基本概念	(152)
7.1.1 文件结构	(152)
7.1.2 外存储器简介	(154)
7.2 顺序文件	(156)
7.3 索引文件	(156)
7.4 ISAM 文件	(157)
7.5 VSAM 文件	(159)
7.6 散列文件	(160)
7.7 多关键字文件	(161)
7.7.1 多重表文件	(162)
7.7.2 倒排文件	(162)
小结	(163)
习题	(163)
第8章 排序	(164)
8.1 概述	(164)
8.2 插入排序	(165)
8.3 交换排序	(166)
8.3.1 冒泡排序	(167)
8.3.2 快速排序	(168)
8.4 选择排序	(170)
8.4.1 直接选择排序	(170)
8.4.2 堆排序	(171)
8.5 归并排序	(177)
8.5.1 有序序列的合并	(177)
8.5.2 二路归并排序	(178)
8.6 外排简介	(179)
小结	(180)
习题	(181)
参考文献	(182)

数据结构导论自学考试大纲

出版前言	(185)
一、课程的性质与设置目的	(187)
二、课程内容与考核目标	(188)
第1章 概论	(188)
第2章 线性表	(189)
第3章 栈、队列和数组	(190)
第4章 树	(191)
第5章 图	(192)

第6章 查找表	(193)
第7章 文件	(195)
第8章 排序	(196)
三、有关说明与实施要求	(197)
附录 题型举例	(199)
后记	(200)

第 1 章 概 论

用数字式计算机解决任何问题都离不开程序设计。程序设计的实质是数据表示和数据处理,而这种表示和处理应通过一个渐进的过程逐步完成。数据结构课程主要讨论这个过程中的一些基本问题。本章将概括地介绍有关的基本概念、基本思想、基本原理及实际背景。

1.1 引 言

今日世界,能被计算机解决的问题种类繁多,确实达到了令人眼花缭乱的地步。然而,说到底,用数字式计算机解决问题的实质是对数据的加工处理。“数据”是计算机加工处理的对象,没有数据,计算机解题就变成“无米之炊”。但是,数据要能被计算机加工处理,首先必须能够存储在机器中,成为能被机器直接操作的对象。数据在计算机存储器中的这种存在形式称为机内表示。显然,数据的机内表示与数据在现实生活和实际问题中的表现形式(姑且称为“机外表示”)是不同的。因此,为了让计算机去加工处理数据,必须首先将数据从机外表示转化为机内表示。这项任务称为数据表示。另外,一个实际问题通常不仅包括数据,还包括处理要求。因此仅仅把数据转化为机内表示并不能完全解决问题,还要用适当的可执行语句编制程序,以便让计算机去执行对数据机内表示的各种操作,从而实现处理要求,即得到所需的结果。这项工作称为数据处理。因此,对计算机专业人员来说,无论面对的具体问题是什么,必须完成的两项基本任务是:数据表示和数据处理。下面用一个简单的例子来说明这两项任务。

【例 1-1】 考虑某单位职工档案管理问题。为了简单起见,假定每个职工的档案只包括以下五个项目:工作证号码、姓名、性别、出生日期和职称。一般地说,档案管理人员很可能将这些档案组织成表格形式,如图 1-1 所示。表中每一行反映了一个职工五个方面的情况,在本例的假设下构成一个职工的档案。由所有职工的档案组成的这张表格就是本问题中的数据。

0001	刘建国	男	19491001	工程师
0002	黄 红	女	19650506	助 工
0005	张 华	女	19461118	高 工
0007	王 军	男	19600110	技术员
⋮	⋮	⋮	⋮	⋮

图 1-1 数据示例

本问题中的处理要求包括该单位所需的所有能用计算机完成的档案管理工作,其中至少应包含以下功能:①查找,需要在表格中找出某人的档案;②读取,阅读通过查找找到的档

案;③插入,调入新职工时,将该职工的档案加到表格中;④删除,某职工调离时,从表格中撤销其档案;⑤更新,某职工情况变化(如提职称)时,修改其档案的有关内容。上述每一项功能又可称为一个“运算”(详见以下几节)。

假如现在要用计算机解决上述档案管理问题,即实现档案自动(计算机)管理,设计人员必须完成两项基本任务:第一,将如图 1-1 所示的表格形式的档案转化为某种适当的机内表示,使之成为能被计算机直接处理的对象。第二,用计算机程序代替档案管理人员完成档案管理工作,也就是说,要编写出实现全部处理要求的程序。这里,第一项任务就是数据表示,第二项任务就是数据处理。如果这两项任务都落实了,整个问题也就解决了。

明确了基本任务之后,下一个关键问题是:计算机专业人员应怎样完成他们的基本任务?也就是说,应该用什么样的方法、步骤去完成上述基本任务?根据程序设计理论和软件工程实践,为求解实际问题而进行的程序设计应是一个渐进的过程。从大范围来看,软件工程学认为,软件系统的生存期可分为软件计划、需求分析、软件设计、软件编码、软件测试和软件维护等六个阶段(参见图 1-2)。一般认为,程序设计应包括前五个阶段。因此,程序设计(包括数据表示和数据处理两个方面)是一个复杂的过程,需要经过一些不同的阶段,每个阶段完成不同的工作,编写程序只是其中的一个阶段。

阶段	任 务	工作结果
计划	理解工作范围	计划任务书
需求分析	定义用户要求	需求规格说明书
设计	建立软件结构	设计说明书
编码	编写程序	程 序
测试	发现和排除错误	可运行的程序系统
维护	运行和管理	改进的程序系统

图 1-2 软件开发阶段

数据结构课程并不研究由上述六个阶段构成的程序开发全过程,而是集中讨论以其中的设计阶段为核心、同时涉及编码阶段和分析阶段的一个小范围内的若干基本问题。概括地说,本课程的主要内容包括:数据的逻辑结构、定义在逻辑结构上的基本运算、数据的存储结构和运算的实现。其中,数据的逻辑结构是数据的组织形式,基本运算规定了数据的基本操作方式。由一种逻辑结构和一组基本运算构成的整体是实际问题的一种数学模型,这种数学模型的建立、选择和实现是数据结构的核心理论问题。存储结构是逻辑结构的存储实现,即数据按逻辑结构规定的形式在计算机存储器中的存放方式。运算实现是完成运算功能的算法,或这些算法的设计。图 1-3 表明了上述主要内容之间的关系。此外,本课程还要考虑各种结构和实现方法的评价和选择。

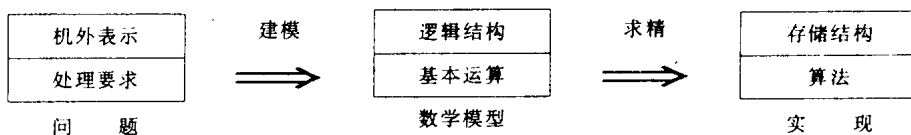


图 1-3 数据结构的主要内容

以下几节将对这些内容做概括性介绍。在进入详细讨论之前,首先必须明确一个基本观

点:程序设计是一个渐进的过程;这个过程在数据结构课程(研究范围)中的“表现”如图 1-3 所示。具体地说,需注意以下三点:

①数据表示任务是逐步完成的,即数据表示形式的变化过程是:机外表示——→逻辑结构——→存储结构。

②数据处理任务也是逐步完成的,即有转化过程:处理要求——→基本运算和运算——→算法。

③数据表示与数据处理是密切相关的,数据处理方式总是与数据的某种相应的表示形式相联系,反之亦然。

1.2 数据、逻辑结构和运算

从数据结构的观点看,通常所说的“数据”应分成三个不同的层次,即数据、数据元素和数据项。在这一区分的基础上,可以进一步建立数据表示和数据处理的数学模型——数据的逻辑结构和定义在它上面的运算。

1.2.1 数据元素和数据项

随着计算机科学的发展和计算机的应用普及,计算机加工处理的对象已从早期的数值、布尔值等扩展到字符串、表格、图像甚至语音等等。凡能被计算机存储、加工的对象通称为数据。在数据结构这门课程中,我们经常遇到的另外两个基本概念是数据元素和数据项。

数据元素是数据的基本单位,在程序中作为一个整体而加以考虑和处理。换句话说,数据元素被当作运算的基本单位,并且通常具有完整确定的实际意义。在后继内容中,根据需要,数据元素又被称为元素、结点、顶点或记录。

在很多情况下,数据元素又是由数据项组成的,但数据项通常不具有完整确定的实际意义,或不被当作一个整体对待。在有些场合下,数据项又称为字段或域。它是数据的不可分割的最小标识单位。

例 1-1 中的表格是一个数据,它是由各个职工的档案构成的。每一个职工的档案是本问题中的数据元素,这是因为每个档案在此问题中被当作运算的基本单位。例如,删除运算不是同时作用于整个表格或多个档案,也不是作用于一个档案中的个别项目,而是作用于一个职工的整个档案。这就是说,档案是删除运算的基本单位,因此,每个职工的档案是本问题中的一个数据元素。另一方面,档案中的各个项目是本问题中的数据项。显然,每个项目是作为档案的一个成分出现的,但单独的项目没有完整确定的实际意义。比如,将表格中第一条档案的各个项目拆开分别地看,则“0001”、“刘建国”、“男”、“19491001”和“工程师”等数据分别表示一个编号、一个人名、一种性别、一个日期和一个职称,它们各自在档案管理这个实际问题中并无完整的实际意义(但它们合在一起却构成职工刘建国的档案,因而具有完整的实际意义)。更重要的是,各个项目在本问题中不能当作运算的基本单位(否则,删除运算可以只删除一个档案中的一个项目,比如人名,那将造成严重的后果)。所以,档案中的项目是数据项而不是数据元素。

有些数据元素不能再分解为数据项的组合,我们认为这种数据元素是由一个数据项组成的,这个数据项就是该数据元素自身。

从某种意义上说,数据、数据元素和数据项实际上反映了数据组织的三个层次,数据可由若干个数据元素构成,而数据元素又可由若干个数据项构成。因此,尽管习惯上数据元素和数据项也可以看成是数据,本书以下不再把数据元素和数据项称为数据,以便明确区别数据的不同层次。

1.2.2 数据的逻辑结构

在科学研究中,为了把握一个难以直接把握的复杂对象,往往采取以下两个步骤:①将这个复杂对象分解成一些较简单的成分或要素;②考察这些成分或要素之间的关联方式。对于计算机工作者来说,数据是一种复杂对象。为了把握这种复杂对象,以便完成数据表示任务,我们接下来要继续研究数据元素之间,以及同一个数据元素中的数据项之间的关联方式。为了研究的方便,我们假定数据项之间的关联方式是简单的(即认为数据元素是数据项的集合)和次要的,集中精力研究数据元素之间的关联方式(注意,可将这种研究的结果应用于数据项之间的关系,如果数据项之间的关联方式也比较复杂的话)。

在任何问题中,数据元素之间可以存在多种关系。例如,对于图 1-1 所示的表格,若两个职工档案中“性别”项的值相同,则可以说这两条档案有“同性”关系。但这种关系对我们现在的问题(即如何组织数据)来说并不重要。从数据结构的观点看,重要的是数据元素之间的逻辑关系。所谓逻辑关系是指数据元素之间的关联方式或称“邻接关系”。例如,在图 1-1 所示的表格中,第一条档案与第二条档案是邻接的,即相互关联的,因此这两个数据元素之间有逻辑关系。但是,第一条档案与第三条档案不相邻接,即相互不直接关联,因此它们之间没有逻辑关系。数据元素之间逻辑关系的整体称为逻辑结构。数据的逻辑结构就是数据的组织形式。图 1-1 中数据的逻辑结构如图 1-4(b)所示。图 1-4 中的小圆圈称为结点。一个结点代表一个数据元素(有时也把结点和数据元素当作同义词)。结点之间的连线代表逻辑关系,即相应

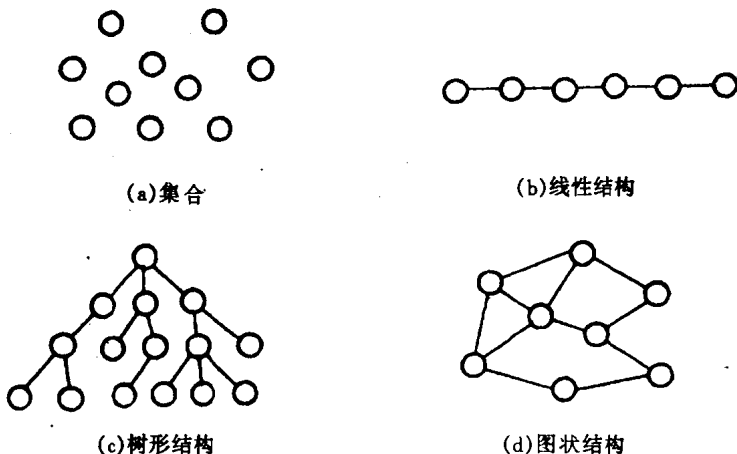


图 1-4 四类基本逻辑结构示意图

数据元素之间的邻接关系。图 1-4(b)中的逻辑结构反映了例 1-1 中表格作为一个数据的组织形式,这种组织形式就是数据元素(档案)“一个接一个地排列”。

在任何问题中,数据元素都不是孤立存在的,它们之间总是存在着某种关系(集合可以看成是没有关系的关系),我们称其为结构。根据数据元素之间关系的不同特性,通常有图 1-4 所示四类基本逻辑结构,它们反映了四类基本的数据组织形式。**集合**中任何两个结点之间都没有逻辑关系,组织形式松散。**线性结构**中结点按逻辑关系依次排列形成一条“锁链”。**树形结构**具有分支、层次特性,其形态有点像自然界中的树。**图状结构**最复杂,其中的各个结点按逻辑关系互相缠绕,任何两个结点都可以邻接。

关于逻辑结构,有以下几点需特别注意:

(1)逻辑结构与数据元素本身的形式、内容无关。例如,在图 1-1 所示表格中的每个档案里增加一个项目“工资额”,就得到另一个数据。但由于所有档案仍是“一个接一个排列”的,新数据的逻辑结构与原来的相同。

(2)逻辑结构与数据元素的相对位置无关。例如,将图 1-1 所示表格中的所有档案按照年龄由大到小的顺序重新排列,得到另一个表格。由于新表格中的所有档案仍是“一个接一个排列”的,其逻辑结构与原表格的逻辑结构相同,还是线性结构。

(3)逻辑结构与所含结点数无关。比如在图 1-1 所示的表格中增加或撤销若干档案,所得表格仍为线性结构。

由此可见,一些表面上很不相同的数据可以有相同的逻辑结构,因此,逻辑结构是数据组织的某种“本质性”的东西。事实上,逻辑结构是数据组织的主要方面。抓住这种本质,不仅有利于分析数据机外表示(如图 1-1 所示的表格)的组织形式,更重要的是使程序设计人员可以根据解题需要重新组织数据,即把数据中的所有数据元素按所需的逻辑结构重新加以安排。例如,对例 1-1 中的数据——表格,既可以照原样组织成线性结构,也可以重新组织成集合或树形结构,具体的选择应根据解题的需要而定。

1.2.3 运算和基本运算

为了有效地完成数据表示任务,我们引入了逻辑结构概念。相应地,为了有效地完成数据处理任务,需要引入“运算”概念。

【例 1-1(续)】 假设现采用线形结构(本例中记为 S)作为此问题的数据组织形式。下面考虑问题求解的数据处理方面。为简单起见,这里只考虑“更新”这一特殊的处理要求,即当该单位进行了一次职称评定之后,修改有关人员档案中“职称”项的值。

可以按下述基本思路来解决此问题:对每一位职称变动的职员,输入其工作证号码 a 和提升后的职称 b,在 S 中找出工作证号码为 a 的结点,然后将该结点“职称”项的值修改为 b。假定职变动的职员共有 n 人,根据上述思路可写出实现“更新功能”的求解过程框架如下:

```
1:   for (i = 1; i <= n; i++)
2:       { 输入(a,b);
3:           在 S 中找出“工作证号码”项的值等于 a 的结点 X;
4:           将结点 X“职称”项的值修改为 b;
5:       }
```