21世纪高等院校计算机教材系列

数据结构教程

朱明方 吴 及 编著

购书可获得增值回报

提供教学用电子教案

机械工业出版社
CHINA MACHINE PRESS

21世纪高等院校计算机教材系列

数据结构教程

朱明方　吴及　编著

机械工业出版社

随着计算机及其应用技术的发展，“数据结构”已成为许多非计算机专业的重要技术基础课程，同时以面向对象的观点来讨论数据结构已成为必然的要求。

本书针对非计算机专业的特点，从应用的角度出发，讲解了线性表、二叉树、图、查找、排序等常用的数据结构及基本运算。书中各部分内容力求少而实用，对各种数据结构和处理算法的讲解深入浅出，并与实际问题相结合，从而使读者很容易理解和掌握书中的知识点。

本书可以作为大专院校“数据结构”课程的教材，也可以作为从事计算机应用开发的科技人员的参考书。

图书在版编目（CIP）数据

数据结构教程 / 朱明方，吴及编著．—北京：机械工业出版社，2007.1

（21 世纪高等院校计算机教材系列）

ISBN 7-111-20364-X

Ⅰ．数… Ⅱ．①朱… ②吴… Ⅲ．数据结构—高等学校—教材
Ⅳ．TP311.12

中国版本图书馆 CIP 数据核字（2006）第 137887 号

机械工业出版社（北京市百万庄大街 22 号　邮政编码 100037）
策　　划：胡毓坚
责任编辑：罗子超
责任印制：李　妍
高等教育出版社印刷厂印刷

2007 年 1 月第 1 版 · 第 1 次印刷
184mm × 260mm · 20 印张 · 495 千字
0001 – 5000 册
定价：28.00 元

出版说明

信息技术是当今世界发展最快、渗透性最强、应用最广的关键技术，是推动经济增长和知识传播的重要引擎。在我国，随着国家信息化发展战略的贯彻实施，信息化建设已进入了全方位、多层次推进应用的新阶段。现在，掌握计算机技术已成为21世纪人才应具备的基本素质之一。

为了进一步推动计算机技术的发展，满足计算机学科教育的需求，机械工业出版社聘请了全国多所高等院校的一线教师，进行了充分的调研和讨论，针对计算机相关课程的特点，总结教学中的实践经验，组织出版了这套“21世纪高等院校计算机教材系列”。

本套教材具有以下特点：

(1) 反映计算机技术领域的新发展和新应用。

(2) 注重立体化教材的建设，多数教材配有电子教案、习题与上机指导或多媒体光盘等。

(3) 针对多数学生的学习特点，采用通俗易懂的方法讲解知识，逻辑性强、层次分明、叙述准确而精炼、图文并茂，使学生可以快速掌握，学以致用。

(4) 符合高等院校各专业人才的培养目标及课程体系的设置，注重培养学生的应用能力，强调知识、能力与素质的综合训练。

(5) 适合各类高等院校、高等职业学校及相关院校的教学，也可作为各类培训班和自学用书。

机械工业出版社

前　言

随着计算机及其应用技术的发展，数据的组织和处理方法对计算机处理效率的影响也日渐增大。“数据结构”问题已成为计算机应用中经常要遇到的问题。因此，“数据结构”课程也成了非计算机专业，特别是理工科专业的技术基础课。另一方面，面向对象的程序设计技术的发展，促进了“数据结构”学科的发展，在面向对象的程序设计方法日益流行的今天，以面向对象的观点来讨论数据结构已成为必然的要求和发展趋势。

目前已出版的数据结构教材，大多是面向计算机专业的，为非计算机专业而写的较少。实际上，由于知识背景不同，学时安排的差异，应用情况也不一样，非计算机专业对数据结构知识的要求和学习投入，都与计算机专业有所不同。少而精、注重实用是非计算机专业对该课程的要求。因此，突出重点、易学和实用应该是针对非计算机专业的数据结构教材的主要特点。

笔者在多年教学实践的基础上，结合理工科非计算机专业学生的知识基础和应用要求编写了这本教材。教材编写的指导思想是：从应用的角度、用面向对象的观点来介绍和讨论常用的数据结构和有关算法。对数据结构各部分内容的介绍力求少而实用，对各种数据结构和算法的讨论与实际问题相结合，从而更好地体现教材的实用性和针对性。为了便于非计算机专业仅有 C 语言基础的学生学习，教材中对算法的描述采用 C/C++语言，并对 C++语言的主要特点作了必要的介绍。

讲授本教材全部内容，建议授课时间不少于 48 学时；标有“*”的章节，教师可以视条件酌情选择。若不讲授这些内容，则建议授课时间为 32～36 学时。课程中应该有足够的上机大作业和实验学时相配合。

本教材第 1 章至第 4 章由朱明方编写，第 5 章至第 7 章由吴及编写，全书由朱明方统稿。

在编写过程中，施迎难老师提供和整理了部分资料，付出了辛勤的劳动。在此表示感谢。

书中的疏漏和错误之处，请读者批评指正。

为配合教师的教学，本书配有教学课件，可上网下载，网址：http://www.cmpbook.com。

编　者

目 录

第1章　绪　　论

从软件设计的角度来看，可以把计算机上处理的问题归结为数值计算和数据处理两类。如今的计算机应用多偏向于数据处理方面。因此，人们对计算机处理的数据的组织形式和相互关系的研究也越来越多、越来越深入，这就促进了“数据结构”这门学科的发展。

数据结构问题的提出，来源于程序设计技术的发展，对它的研究是为了提高程序运行的效率（时间效率、空间效率和复用程度等），数据结构主要包括：数据集合中数据元素之间的逻辑关系、它们在计算机存储器中的表示和基本运算的实现。

本章作为后续章节的预备知识，首先引出二元关系的定义，然后介绍有关数据结构的基本概念，最后讨论算法的描述与评价问题。

1.1　二元关系

1.1.1　二元关系的定义

二元关系是一个数学概念，它定义于集合的基本运算——笛卡尔积（Cartesian Product）的基础上。为了说明二元关系的概念，先回顾一下集合的笛卡尔积的定义。

1．集合的笛卡尔积

设有集合 A 和 B，则集合 A 对集合 B 的笛卡尔积，记作 A×B，定义为

$$A\times B=\{<a, b> \mid a\in A \text{ 且 } b\in B\}$$

根据定义，若有 $A=\{a_1, a_2\}$，$B=\{b_1, b_2\}$，则有

$$A\times B=\{<a_1, b_1>, <a_1, b_2>, <a_2, b_1>, <a_2, b_2>\}$$

其中，$<a_1, b_1>$等都是有序对（序偶），a_1是有序对的第一分量，b_1是有序对的第二分量，必须注意它的有序性。根据这个特点，显然有

$$A\times B\neq B\times A$$

同样地，有

$$(A\times B)\times C\neq A\times(B\times C)$$

当然，笛卡尔积与集合的并、集合的交运算结合起来，还可以得到其他性质。

只要理解了两个集合的笛卡尔积的定义，也就很容易理解 n 个集合的笛卡尔积的概念了。

2．二元关系

（1）二元关系的数学定义

二元关系是一个数学概念，它的定义为：

设有集合 M、N，则 M×N 的任意一个子集 R 称为 M 到 N 的一个二元关系。

如果 M=N，则称 R 为 M 上的一个二元关系，简称关系。此时，它表示的是 M 中元素

之间的某种关联。

例如，设集合 $M=\{m_1, m_2, m_3, m_4\}$，$N=\{n_1, n_2, n_3\}$，则

$$R_1=\{<m_1, n_1>, <m_1, n_2>, <m_2, n_2>, <m_3, n_2>\}$$

$$R_2=\{<m_1, n_1>, <m_2, n_2>, <m_3, n_1>, <m_2, n_1>, <m_4, n_2>\}$$

是集合 M 到 N 的两个二元关系。

$$R_3=\{<m_1, m_2>, <m_1, m_3>, <m_2, m_3>, <m_3, m_4>\}$$

是集合 M 上的一个二元关系。

$M\times N$ 可以有若干个子集。也就是说，集合 M 到 N 可以定义若干个不同的关系，其数目由集合 M 和 N 的元素个数决定。

为了描述关系中元素之间的关联情况，对于关系 R 中的任意一个有序对 $<a, b>$（可记作 $<a, b>\in R$ 或记作 aRb）定义：

a 是 b 的关于 R 的前件（直接前驱）。

b 是 a 的关于 R 的后件（直接后继）。

例如，在上面的关系 R_1 中，因为包含了有序对 $<m_1, n_2>$，因此，m_1 是 n_2 的关于 R_1 的前件。同样，n_2 是 m_1 的关于 R_1 的后件。

在上述关系 R_3 中，有序对 $<m_1, m_2>$ 表明 m_1 是 m_2 的关于 R_3 的前件。同样，m_2 是 m_1 的关于 R_3 的后件。

（2）二元关系的意义

二元关系的数学定义是在笛卡尔积的基础上给出的。从实际意义上来说，二元关系表示了集合中两个元素之间的某种相关性。这两个元素可以来自两个不同的集合，也可以同属于一个集合。下面通过两个简单例子说明二元关系表示的实际意义。

【例 1-1】 甲、乙、丙、丁和戊 5 个人参加某公司招聘面试，有 A、B、C、D 和 E 5 份试题，每个人以抽签方式抽取问题，即抽签结果应是每人对应一份题。对此抽签结果，可以用一个二元关系简单、明了地表示出来。即

$$R_4=\{<\text{甲}, B>, <\text{乙}, E>, <\text{丙}, A>, <\text{丁}, C>, <\text{戊}, D>\}$$

这里 R_4 是一个二元关系，其中每一个有序对表示每个人和他所抽到的题，有序对的第一分量是参加面试的某个人，第二分量是该人所抽到的题。这个二元关系清晰地表示了“参加面试人”集合｛甲、乙、丙、丁、戊｝中的元素“参加面试人”和“试题”集合｛A、B、C、D、E｝中的元素“试题”之间的关联。

【例 1-2】 有编号为 1、2、3、4 的 4 位选手进行象棋循环赛，比赛的结果为：1 号胜 2 号，3 号胜 2 号，3 号胜 4 号，3 号胜 1 号，1 号胜 4 号，4 号胜 2 号。我们可以用一个二元关系很清楚地表示上述比赛结果。即

$$R_5=\{<1, 2>, <3, 2>, <3, 4>, <3, 1>, <1, 4>, <4, 2>\}$$

这里，R_5 也是一个二元关系。其中每个有序对表示一场比赛的结果，每个有序对的第一分量是该场比赛的胜者，第二分量是该场比赛的负者。这个二元关系表示了同一个集合{1, 2, 3, 4}中的元素“选手”之间的比赛胜负关系。

从上述例子中可以看出，利用二元关系可以简洁、明了、确切地表示出集合中两个元素之间的某种相关性。

1.1.2 二元关系的基本性质和几种重要关系

1. 二元关系的基本性质

了解二元关系的基本性质，对分析数据元素之间的关系是有意义的。下面给出二元关系的几个主要的性质。

设 R 是集合 M 上的一个关系，则

1）若对于每一个 $a \in M$，都有 $<a, a> \in R$，那么称 R 是自反关系（自反性）。

2）若对于任何 $a \in M$，都有 $<a, a> \notin R$，那么称 R 是反自反关系（反自反性）。

3）如果 $<a, b> \in R$，必有 $<b, a> \in R$，那么称 R 是对称的（对称性）。

4）如果 $<a, b> \in R$ 且 $a \neq b$ 时，必有 $<b, a> \notin R$，那么称 R 是反对称关系（反对称性）。

5）如果 $<a, b> \in R$ 且 $<b, c> \in R$ 时，必有 $<a, c> \in R$，那么称 R 是传递关系（传递性）。

假设有集合 M={1, 2, 3, 4, 5}，M 上的一个关系为

R={<1, 2>, <2, 3>, <3, 4>, <3, 5>, <1, 3>, <1, 4>, <1, 5>, <2, 4>, <2, 5>}。

根据上述性质定义，很容易看出该关系 R 是反自反的、反对称的和传递的。

在实际问题中，利用二元关系的基本性质，可以更方便、更确切地分析和描述数据集合中数据元素之间的关联规律和特点，从而有利于对它们的处理。

2. 等价关系与等价类

等价关系是一类重要的二元关系，在处理问题时经常会用到它。等价关系的定义是：

如果非空集合 S 上的关系 R 是自反的、对称的和传递的，则称 R 为 S 上的一个等价关系。

如果 R 是 S 上的等价关系，a, b 是 S 的任意元素，若有 $<a, b> \in R$，则称 a 等价于 b，记作 $a \sim b$。

等价关系在现实世界中广泛存在。例如，假设有学生集合 S={a, b, c, d, e, f }，S 上的“同班”关系 R 就是等价关系。因为

1） 任何一个人都和自己同一个班，即具有自反性。

2） 若 a 和 b 是同一个班的，那么，b 也就和 a 同在一个班。即具有对称性。

3） 如果 a 和 b 在一个班，b 与 c 是同班的，那么，a 和 c 一定是同班的。即具有传递性。

由此可以看出，这个“同班关系”是一个等价关系。

假设 a、b、c 是同班的，d、e、f 是同班的，则同班关系为

R= {<a, a>, <b, b>, <c, c>, <a, b>, <b, a>, <b, c>, <c, b>, <a, c>,<c, a>, <d, d>, <e, e>,
<f, f>, <d, e>, <e, d>, <e, f>, <f, e>, <d, f>, <f, d>}

由等价关系可以进一步给出等价类的概念。假设 R 是非空集合 S 上的等价关系，则 S 上互相等价的元素构成了 S 的若干不相交的子集，这些子集称为 S 的 R 等价类。等价类的一般定义是：设 R 是非空集合 S 上的等价关系，对任意的 $x \in S$，由 $[x]_R=\{y \mid y \in S \wedge xRy\}$ 给出的集合 $[x]_R$，称为 x 关于 R 的等价类，简称为 x 的等价类，简记为 [x]。

例如，对于上述学生集合的同班关系，根据“同班”关系得到的不相交的学生子集：{a, b, c}和{d, e, f}分别是学生 a 和 d 的等价类。当然，它们也是学生 b、c 和学生 e、f 的等价

类。即

$$[a] = [b] = [c] = \{a, b, c\}，\qquad [d] = [e] = [f] = \{d, e, f\}；$$

在实际问题中，当需要根据某种关系来划分集合中的元素时，划分等价类是常用的方法。在后面的章节中，我们将结合具体应用讨论等价类的划分方法。

3．偏序关系和全序关系

偏序关系和全序关系也是重要的关系，它们提供了比较集合中元素的工具。

偏序关系的定义是：如果集合 S 上的关系 R 是自反的、反对称的和传递的，则称 R 是集合 S 上的偏序关系，称序偶<S, R>为偏序集合，也记作<S, ≤>。

例如，集合 S={2, 4, 6, 8}和其上的整倍数关系 R={<2, 2>, <4, 4>, <6, 6>, <8, 8>, <4, 2>, <6, 2>, <8, 2>, <8, 4>}，则构成偏序集合 <S, R>。

对于偏序集<S, ≤>中的 x, y∈S，如果有 x≤y 或 y≤x 成立，我们说 x 和 y 是可比较的。由偏序集合的定义可知，偏序集合中的元素不一定都是可比较的。也就是说，它们在偏序中不一定都有确定的位置上的先后关系。若要使集合中的元素之间都可比较，则需要构造下面定义的全序关系。

全序关系的定义是：在偏序集合<S, ≤>中，如果对于任意的 x, y∈S，x 和 y 都是可比较的，则称此偏序关系≤为 S 上的全序关系，称序偶<S, ≤>为全序集合。

例如，集合 S={1, 2, 3, 4}上的小于等于关系：R_1={<1, 1>, <2, 2>, <3, 3>, <4, 4>, <1, 2>, <1, 3>, <1, 4>, <2, 3>, <2, 4>, <3, 4>}，即是全序关系，<S, R_1>为全序集合。

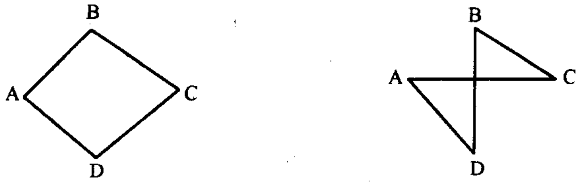
由全序集合的定义可以看出，全序集合中所有元素之间都是可以比较的，这为我们处理实际问题时从给定的一些“前题条件”（即偏序关系）出发，确定集合中各成员之间的先后位置关系提供了办法。因为根据偏序集合与全序集合的关系，能够以非空有限偏序集合<S, ≤>为基础，在集合 S 的元素上构造一个包含了原来的偏序关系的全序，从而使得集合中的全体成员在满足给定的偏序关系的前题下都可比较，这就是拓扑排序（也叫拓扑分类）的问题。

1.2 数据结构

1.2.1 数据结构的引出

前面已经提到，数据结构学科的形成及发展是与程序设计技术的发展、计算机应用的日益广泛联系在一起的。对于数据处理问题，在计算机中的处理效率和处理结果与数据的组织形式是密切相关的。请看以下例子：

【例 1-3】 给定 A、B、C、D 四个点的坐标，若对它们进行不同的组织——给出不同的边，则可以达到不同的处理目的。如果确定一组边 AB、BC、CD 和 DA，则可以画出一个四边形，如图 1-1a 所示；若确定的一组边是 AC、CB、BD 和 DA，则可以画出两个对接的三角形，如图 1-1b 所示。虽然是相同的四个顶点，由于对它们组织不一样（给出不同的顶点顺序），处理的结果完全不同。这就是说，对于不同的问题、不同的要求，对数据的组织是不一样的。



a)　　　　b)

图 1-1　顶点 A、B、C、D 所对应的不同图形

a) A、B、C、D 对应的四边形　b) A、B、C、D 对应的对接三角形

另一方面，为了得到可能高的处理效率，对于同一个问题，针对不同的数据组织形式应该采用不同的处理方法，也就是说要设计不同的算法。

【例 1-4】 一个单位的电话号码本，其中包含了该单位所属成员的电话号码和住址信息，对电话号码的排列可以采用不同的方式，因而形成了不同的电话号码本。人们在查找电话号码时，应该根据所用的电话号码本的特点采用不同的查找方法，才能快速地查到所要的电话号码。如果电话号码如图 1-2a 所示是按姓名的拼音顺序排列的，则根据所找人的姓名拼音可以较快查到其电话号码；如果电话号码如图 1-2b 所示是按部门来分的，那么在知道所查找人的工作部门的情况下，可以很快查找到其电话号码，但对于不知道所查找人的具体部门的时候，查找就会比较费时。这说明，电话号码的组织方式直接决定了查找电话号码的方法。

姓　名	电话号码	住　址
安　乐	76345	西区 7-3-501
⋮	⋮	⋮
蔡玉芹	77276	东区 2-4-702
⋮	⋮	⋮
郭　兴	75832	西南 6-2-402
⋮	⋮	⋮
王海涛	74964	东南 4-1-201
⋮	⋮	⋮
徐新阳	72521	西北 8-3-602
⋮	⋮	⋮

a)

部　门	姓　名	电　话	住　址
校机关 ⋮	王海涛 ⋮	74964 ⋮	东南 4-1-201 ⋮
机械学院 ⋮	蔡玉芹 ⋮	77267 ⋮	东区 2-4-702 ⋮
理学院 ⋮	徐新阳 ⋮	72521 ⋮	西北 8-3-601 ⋮
水土学院 ⋮	郭　兴 ⋮	75832 ⋮	西南 6-2-402 ⋮
信息学院 ⋮	安　乐 ⋮	76345 ⋮	西区 7-3-501 ⋮
⋮	⋮	⋮	⋮

b)

图 1-2　电话号码的不同组织方式

a) 以姓名拼音顺序排列的电话号码　b) 以部门区分的电话号码

以上两个例子说明，在数据处理问题中，数据的组织和表示是至关重要的，因为它直接决定了处理时所采用的方法，从而直接影响了数据处理的效率。

数据结构课程所要讨论的内容正是围绕不同情况下数据的组织和表示展开的。当然，在确定了数据的组织和表示的基础上，如何进行有效的处理是必然要讨论的问题，因为它是我们要达到的根本目的。

1.2.2　数据的逻辑结构和存储结构

1. 数据的逻辑结构

数据结构（Data Structure）讨论的是数据（Data）以及数据元素（Data Element）之间的

关系，实际上它表示的是客观事物及其相互之间的联系。计算机中的数据是一个广泛的概念，它包括所有能够被输入到计算机中，并且可以进行存储、处理和输出的信息。例如，字母、数字、声音、图像等都可能是计算机中的数据。整体数据中相对独立的单位称为数据元素，是数据处理的基本单位。例如，“字符串”数据中的每个字符，“数值”数据中的每个数等都是其所属数据的数据元素。有时候，一个数据元素可以由若干数据项组成。例如，记录是“文件”的数据元素，而一个记录可以包含若干数据项，此时的数据元素就不是一个简单的数据项。

客观事物之间的联系是各种各样的，反映到数据上来，数据元素之间的联系也各不相同。

例如，一年有二十四个节气（立春、雨水……、小寒、大寒），它们构成了一个简单的农历节气关系。

又如，一个年级的学生按成绩总分由高到低排列，每个学生的信息包括学号、姓名、班级、成绩四项信息，则学生信息表中各信息项之间构成了表 1-1 的学生成绩表。

表 1-1　学生信息表

班　级	学　号	姓　名	成　绩
W21	10120	张旭	95
W24	10048	李力坤	92
W22	10851	王仿固	90
W21	10252	徐畅	88
W23	10469	林致意	86
⋮	⋮	⋮	⋮

还有，操作系统管理下的文件系统，其关系是根目录包含多个子目录，每个子目录又可以包含下一层的多个子目录或文件，各层子目录之间的层次关系，如图 1-3 所示。

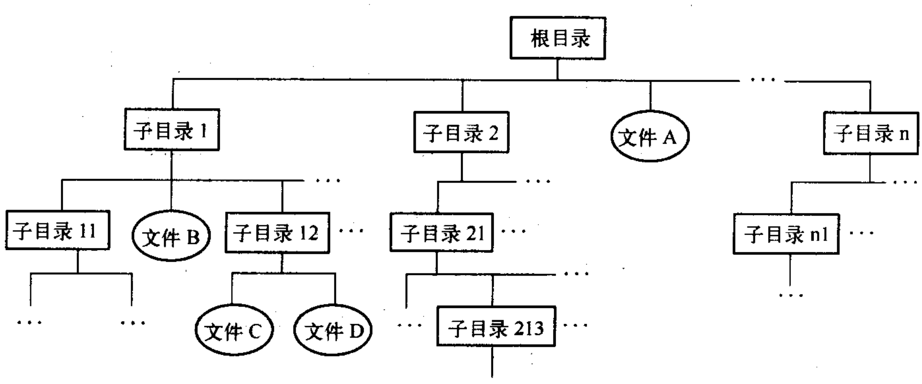


图 1-3　文件目录结构

数据元素之间的关系，有的是客观存在的，比如节气表中各节气之间的先后关系；有的则是因为解决问题的需要人为组织的，如学生成绩表中各学生信息之间的关系，文件目录之间的关系等。无论是客观存在的或是人为组织的，它们都表示数据元素之间逻辑上的关联，称为数据的逻辑结构，简称为数据结构。也就是说，数据结构反映的是数据元素以及它们之