新世纪高职高专实用规划教材

● 计算机系列

数据结构

SHUJU JIEGOU

王 钢 徐 红 主 编
杨德芳 李 国 副主编

清华大学出版社

新世纪高职高专实用规划教材　计算机系列

数　据　结　构

王　钢　徐　红　主编
杨德芳　李　国　副主编

清华大学出版社
北　京

内 容 简 介

本书系统介绍了最常用的数据结构，包括线性表、栈、队列、数组、矩阵的压缩存储、树与二叉树、图以及查找和排序的算法等。阐述各种数据结构的逻辑关系，分析讨论各种数据结构在计算机内的存储表示，以及在这些数据结构下的算法实现，并对各种算法的时间和空间性能作简要分析。

本书既注重原理又注重实践，对基本的算法均给出相应的C语言程序的描述，并加以较详细的注释。全书配有大量的图表，每章后都附有习题，内容丰富，概念讲解清楚，逻辑性强。在本书的最后给出实验内容的附录。

本书可作为高等院校计算机相关专业的教材，亦适合于计算机爱好者自学，还可供广大从事计算机应用和开发的技术人员参考。

图书在版编目(CIP)数据

数据结构/王钢，徐红主编；杨德芳，李国副主编. —北京：清华大学出版社，2005.2
(新世纪高职高专实用规划教材　计算机系列)
ISBN 7-302-10134-5

Ⅰ.数…　Ⅱ.①王…②徐…③杨…④李…　Ⅲ. 数据结构—高等学校：技术学校—教材　Ⅳ.TP311.12

中国版本图书馆CIP数据核字(2004)第133096号

出 版 者：清华大学出版社　　**地　　址**：北京清华大学学研大厦
http://www.tup.com.cn　　**邮　　编**：100084
社 总 机：010-62770175　　**客户服务**：010-62776969
组稿编辑：彭　欣
文稿编辑：李春明
封面设计：陈刘源
印 刷 者：北京市清华园胶印厂
装 订 者：三河市金元印装有限公司
发 行 者：新华书店总店北京发行所
开　　本：185×260　**印张**：15.25　**字数**：358千字
版　　次：2005年2月第1版　2006年7月第3次印刷
书　　号：ISBN 7-302-10134-5/TP·6931
印　　数：7001～9500
定　　价：20.00元

《新世纪高职高专实用规划教材》序

编写目的

目前，随着教育改革的不断深入，高等职业教育发展迅速，进入到一个新的历史阶段。学校规模之大，数量之众，专业设置之广，办学条件之好和招生人数之多，都大大超过了历史上任何一个时期。然而，作为高职院校核心建设项目之一的教材建设，却远远滞后于高等职业教育发展的步伐，以至于许多高职院校的学生缺乏适用的教材，这势必影响高职院校的教育质量，也不利于高职教育的进一步发展。

目前，高职教材建设面临着新的契机和挑战：

(1) 高等职业教育发展迅猛，相应教材在编写、出版等环节需要在保证质量的前提下加快步伐，跟上节奏。

(2) 新型人才的需求，对教材提出了更高的要求，即教材要充分体现科学性、先进性和实用性。

(3) 高职高专教育自身的特点是强调学生的实践能力和动手能力，教材的取材和内容设置必须满足不断发展的教学需求，突出理论和实践的紧密结合。

有鉴于此，清华大学出版社在相关主管部门的大力支持下，组织部分高等职业技术学院的优秀教师以及相关行业的工程师，推出了一系列切合当前教育改革需要的高质量的面向就业的职业技术实用型教材。

系列教材

本系列教材主要涵盖以下领域：

- 计算机基础及其应用
- 计算机网络
- 计算机图形图像处理与多媒体
- 电子商务
- 计算机编程
- 电子电工
- 机械
- 数控技术及模具设计
- 土木建筑
- 经济与管理
- 金融与保险

另外，系列教材还包括大学英语、大学语文、高等数学、大学物理、大学生心理健康等基础教材。所有教材都有相关的配套用书，如实训教材、辅导教材、习题集等。

教材特点

为了完善高等职业技术教育的教材体系，全面提高学生的动手能力、实践能力和职业技术素质，特意聘请有实践经验的高级工程师参与系列教材的编写，采用了一线工程技术人员与在校教师联合编写的模式，使课堂教学与实际操作紧密结合。本系列丛书的特点如下：

(1) 打破以往教科书的编写套路，在兼顾基础知识的同时，强调实用性和可操作性。

(2) 突出概念和应用，相关课程配有上机指导及习题，帮助读者对所学内容进行总结和提高。

(3) 设计了“注意”、“提示”、“技巧”等带有醒目标记的特色段落，使读者更容易得到有益的提示与应用技巧。

(4) 增加了全新的、实用的内容和知识点，并采取由浅入深、循序渐进、层次清楚、步骤详尽的写作方式，突出实践技能和动手能力。

读者定位

本系列教材针对职业教育，主要面向高职高专院校，同时也适用于同等学历的职业教育和继续教育。本丛书以三年制高职为主，同时也适用于两年制高职。

本系列教材的编写和出版是高职教育办学体制和运作机制改革的产物，在后期的推广使用过程中将紧紧跟随职业技术教育发展的步伐，不断吸取新型办学模式、课程改革的思路和方法，为促进职业培训和继续教育的社会需求奉献我们的力量。

我们希望，通过本系列教材的编写和推广应用，不仅有利于提高职业技术教育的整体水平，而且有助于加快改进职业技术教育的办学模式、课程体系和教学培训方法，形成具有特色的职业技术教育的新体系。

教材编委会

新世纪高职高专实用规划教材
计算机系列编委会

前　言

随着社会经济的高速发展，我国的高等教育已进入从精英教育走向大众化教育的发展阶段。高职高专教育近年来虽然得到了飞速的发展，但还仍处于探索阶段，教材改革已成为教育改革的重要方面。“数据结构”是计算机程序设计的重要理论基础，是计算机及相关专业的一门专业基础课，计算机的系统软件和应用软件都要用到各种不同的数据结构，因此学好“数据结构”对于学习计算机专业的其他课程，如操作系统、数据库管理系统、软件工程等都是十分有益的。本书在编写过程中充分考虑到高职高专学生的特点，本着注重基础、易于实践、通俗易懂的原则，旨在为高校教改教学提供实践教材。该书的编写符合现代教育技术和教学规律，以培养实际开发和应用能力为主要教学目的，同时也可供从事计算机应用等工作的技术人员参考。

C 语言以其灵活的编程方式，强大的指针处理功能，现已成为描述数据结构的最佳语言之一。考虑到计算机专业的学生都先学过 C 语言，本书完全用 C 语言描述，读者可根据具体情况轻松地升级到 C++或 Visual C++。考虑到本书面向的读者主要是专科生，“数据结构”又是较难学习的一门课程，所以在编写教材时，总结了笔者多年来在教学上的实践和体会，并参考了国内外较新的有关文献编写而成，所涉及到的数据结构都有较完整的数据类型定义，算法结构完整，可上机调试实现。

全书包含 10 章内容和 1 个附录：第 1 章介绍数据结构的主要内容、基本概念、算法的评价标准和评价方法。第 2 章至第 6 章介绍几种常用的线性结构，包括线性表、栈、队列、串以及数组、特殊矩阵和广义表。第 7 章介绍具有广泛应用价值的树形结构——树和二叉树及其重要应用。第 8 章介绍复杂的数据结构——图及其应用。第 9 章和第 10 章分别介绍数据处理中广泛应用的查找和排序技术。每一种数据结构都有较完整的数据类型定义，并详细地讨论各种数据结构在计算机内的存储表示及算法的实现，并对 C 语言描述的算法作相应的注释和简要的分析。本书最后的附录给出一些较典型的实验，使读者通过实验来加深对数据结构这门课程的理解和掌握。

本书可作为高等院校计算机相关专业的教材，亦适合于计算机爱好者自学，还可供广大从事计算机应用和开发的技术人员参考。

本书的第 1、2、3、7 章和附录部分由王钢编写，第 4、5、6 章由杨德芳编写，第 8、9、10 章由李国编写，最后由王钢和徐红统改定稿。

本书在编写过程中，参考了大量的文献资料和国内外较新的优秀教材，并得到许多老师和专家的大力支持，在此表示诚挚的谢意。

由于时间仓促及作者的水平有限，书中缺点和错误在所难免，恳请广大读者批评指正。

编　者

2004 年 7 月

目　录

第 1 章　数据结构概论

本章要点

- 数据结构的基本概念和术语
- 数据的逻辑结构和物理结构
- 数据类型和抽象数据类型
- 算法的时间复杂度和空间复杂度分析

本章难点

- 数据的逻辑结构和物理结构，算法的时间复杂度分析。

计算机在发展的初期，其应用的范围是数值计算，所处理的数据都是整型、实型、布尔型等简单数据，以此为对象的程序设计称为数值型程序设计。随着计算机技术的飞速发展，计算机已逐渐深入到社会生活的各个方面，广泛地应用于数据处理和过程控制等领域。计算机能处理的数据也不再是简单的数值，而是字符串、表格、图像、声音等复杂的数据，这些复杂的数据不仅量大，而且具有一定的结构。例如一名学生的信息是由若干字段组成的记录，描述一个单位的组织结构需要使用树形结构。此外，函数调用和递归过程中使用的栈，操作系统用到的队列、磁盘文件目录树等，都是有一定结构的数据。数据结构研究的就是如何处理这些有结构的数据，因此，数据结构的知识不论对开发系统软件还是应用软件都是非常重要的，是学习程序设计和提高软件设计水平的重要理论基础。

凡是能够输入计算机并能被计算机所处理的信息都被称为数据，计算机科学是一门研究数据表示和数据处理的科学。据统计，百分之九十以上的计算机用于数据处理，而处理的这些数据大部分都是有一定结构的，为了编写出高质量的程序，必须研究待处理的这些数据的特性、数据之间存在的关系及对应的存储表示，并利用这些特性和关系设计出相应的算法和程序。

1.1　数据结构的概念

1.1.1　什么是数据结构

在计算机发展的初期，人们使用计算机的目的主要是处理数值计算问题。为了更好的说明什么是数据结构，我们先来举一个简单数值计算的例子。求一个圆的面积。设圆的半径为 R，面积为 A，利用 C 语言编写程序处理此问题的主要语句如下：

```
double A , R=50.0
A=3.1415926*R*R
```

在这个简单的问题中实际上已经用到了数据结构的知识，使用了两个实型变量 A 和 R，

根据C语言的知识，我们知道这两个变量只能进行+、-、*、/等运算，在内存中各占4个字节，它们在内存中的关系是离散的，利用圆的面积公式可以求出面积。

实际上这些内容都是与数据结构有关的，只是由于数值计算所涉及到的运算对象是简单的整型、实型或字符型数据，所以程序设计者的主要精力是集中于程序设计的技巧和运算公式上，而无须重视数据结构。

从这个问题中我们还看到，利用计算机来解决一个具体问题时，一般需要经过下列几个步骤：首先从该具体问题抽象出一个适当的数学模型，并将此数学模型所用到的数据在计算机中表示出来，然后设计一个解此数学模型的算法，最后编写程序进行调试、测试，直至得到最终的解答。

随着计算机应用领域的不断扩大，非数值计算问题越来越显得重要，这类数据的处理涉及到的数据结构更为复杂，数据之间的关系一般无法用数学方程式或公式加以表达，相互间的运算也不再局限于+、-、*、/等基本运算。解决这类问题的关键不再是数学分析和计算方法，而是要设计出合适的数据结构，才能有效地解决问题，下面列举一些属于这一类的问题。

例1.1 学生信息检索系统。如表1.1所示，每一条学生记录对应一个数据，并且是有一定结构的数据，即每一个数据都包含有学号、姓名、专业等若干数据项，数据之间的关系是一一对应的，对这些数据进行+、-、*、/等算术运算已经没有任何意义，对这样一个二维数据表只能进行查找、插入、修改、删除等基本操作。当需要查找某个学生的有关情况或是查询某个专业或年级的学生有关情况时，只要建立了相关的数据结构，按相应的数据结构将这些记录存储在计算机中，并按照某种算法编写出相关程序，就可以实现计算机的自动检索。还可以分别按姓名、专业、年级顺序建立索引表，如表1.2、1.3、1.4所示。由这四张表所构成的文件就是学生信息检索的数学模型，计算机的主要操作便是按照某个具体要求实现对学生信息文件的自动检索。

表1.1 学生信息检索系统的数据结构

学 号	姓 名	性 别	专 业	年 级
20000101	孙祥林	男	计算机应用与维护	2000
20000208	王书香	女	应用电子技术	2000
20010316	李明玉	女	通信工程	2001
20010116	刘文慧	女	计算机应用与维护	2001
20010320	崔建国	男	通信工程	2001
20020114	赵文东	男	计算机应用与维护	2002
20020321	杨 威	男	通信工程	2002
20030226	刘文慧	女	应用电子技术	2003
20030230	宋鲁生	男	应用电子技术	2003
20030315	周晓阳	男	通信工程	2003

表 1.2　姓名索引表

姓　名	索　引
崔建国	5
李明玉	3
刘文慧	4，8
宋鲁生	9
孙祥林	1
王书香	2
杨　威	7
赵文东	6
周晓阳	10

表 1.3　专业索引表

专　业	索　引
计算机应用与维护	1，4，6
应用电子技术	2，8，9
通信工程	3，5，7，10

表 1.4　年级索引表

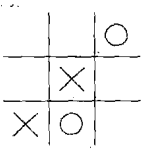
年　级	索　引
2000 级	1，2
2001 级	3，4，5
2002 级	6，7
2003 级	8，9，10

诸如此类的还有电话查号系统自动化、仓库账目管理等。在这类文件管理的数学模型中，计算机处理的对象之间通常存在着的是一种最简单的线性关系，这类数学模型称为线性数据结构。

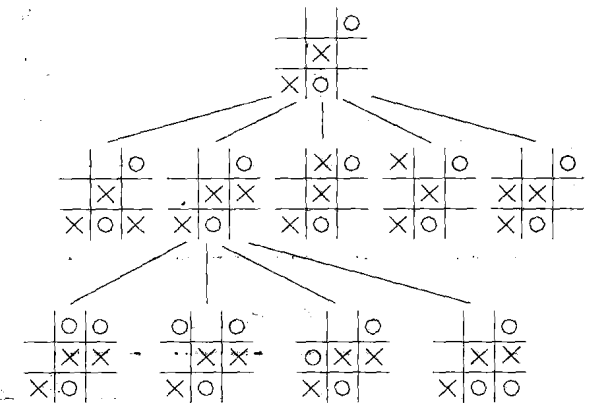
例 1.2　计算机和人对弈问题。计算机之所以能和人对弈，是因为有人将对弈的策略事先存入计算机。由于对弈的过程是在一定规则下随机进行的，为使计算机能灵活对弈，必须对对弈过程中可能发生的各种情况以及相应的对策都考虑周全。因此，在对弈过程中，计算机操作的对象是对弈过程中可能出现的各种棋盘状态，这种状态称为格局。

例如图 1.1(a)所示为井字棋的一个格局，井字棋是一个 3×3 的方格，由两人对弈，当一方的三个棋子率先占同一行、同一列或同一对角线时便获胜。从一个格局可以派生出几个格局，例如从图 1.1(a)所示的格局可以派生出五个格局，而从每一个新的格局有可能派生出四个可能出现的格局。因此，若将从对弈开始到结束的过程中所有可能出现的格局都画在一张图上，则可得到一棵倒长的“树”。“树根”是对弈开始之前的棋盘格局，而所有

的“叶子”就是可能出现的结局，对弈的过程就是从树根沿树杈到某个叶子的过程。



(a) 棋盘格局示例



(b) 对弈树的局部

图 1.1 井字棋对弈“树”

此外，对于描述单位的组织结构、操作系统的文件目录等，都要用到“树”这种处理非数值计算的数学模型，树形结构也是一种基本的数据结构。

例 1.3 城市间的通信布线问题。一个地区由许多城市组成，为了实现城市间高速大容量数据通信的要求，需要在这些城市间铺设通信线路，以达到在任意两个城市间实现通信的要求。为此，只要在任意两个城市间都铺设一条通信线路，就能实现上述目的，但这显然是很不经济而且有时是无法实现的。经过考察和预算，在 A 市到 F 市之间可铺设的通信布线图如图 1.2 所示。其中每个顶点代表一个城市，顶点间的连线代表两个城市间可以铺设通信线路，而线上的数值表示铺设该通信线路的费用(单位：万元)。在图 1.2 中，必然存在一个既能保证各个城市之间都能实现通信，而投资最小的设计方案。此外，还有教学计划安排问题，城市交通图设计问题，这类问题的研究需要建立新的数学模型，即建立一种称为“图”的数据结构。

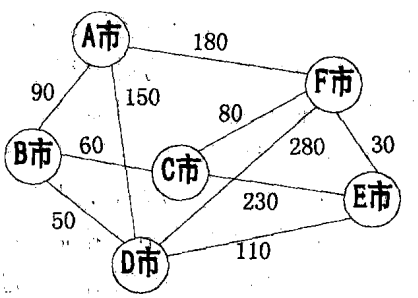


图 1.2 城市通信布线图

综合以上三个例子可以看出，研究这类非数值计算问题的数学模型不再是公式和数学方程，而是诸如表、树、图之类的数据结构。因此，可以说数据结构是一门研究非数值计算的程序设计问题中出现的计算机操作对象以及它们之间的关系和操作的学科。

数据结构是计算机科学与技术专业的专业基础课，是十分重要的核心课程。数据结构不仅涉及到计算机的硬件(数据的存储方式、存储装置和编码理论)，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存取分配问题，所有的计算机系统软件和应用软件都要用到各种类型的数据结构。因此，要想更好地运用计算机来解决各种实际问题，仅掌握几种程序设计语言是远远不够的。要想更好地使用计算机、充分发挥计算机的性能，必须学习和掌握数据结构的有关知识。在计算机科学中，数据结构不仅是一般程序设计的基础，而且是学习、设计和实现操作系统、编译程序、数据库管理系统、软件工程及其他系统程序和大型应用程序的重要基础。

1.1.2　基本概念和术语

在系统的学习数据结构的知识之前，先对一些概念和术语赋以确定的含义，这些概念和术语将在以后的章节中多次出现。

数据(data)：是对客观事物的符号表示，在计算机科学中是指所有能输入计算机并能被计算机处理的符号的总称。它是计算机程序加工处理的“原料”，它可以是数值数据，也可以是非数值数据。数值数据是一些整数、实数或布尔型数据，主要用于工程计算、科学计算等；非数值数据包括字符、文字、图像、声音等。因此，对计算机科学而言，数据的含义极为广泛。

数据元素(data element)：是数据的基本单位，在计算机程序中通常作为一个整体进行考虑和处理。例如学生信息检索系统中学生信息表中的一个记录，井字棋对弈过程树中的一个棋盘格局，城市通信布线图中的一个城市顶点等都被称为一个数据元素。有时一个数据元素可以由若干数据项(data item，又称字段)组成，如学生信息表中的一个记录为一个数据元素，而一个学生记录又由姓名、专业、年级等若干数据项组成，数据项是数据处理中不可分割的最小单位。

数据对象(data object)：是具有相同性质的数据元素的集合。在某个具体问题中，数据元素都具有相同的性质(数据元素的值不一定相等)，属于同一数据对象。例如在学生信息表中，所有学生的记录是一个数据对象；在城市通信布线图中，所有的城市顶点组成一个数据对象。

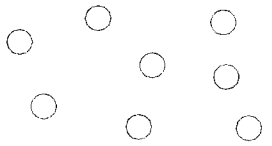
数据结构(data structure)：是相互之间存在一种或多种特定关系的数据元素的集合。从 1.1.1 节中的三个例子可以看到，在任何问题中，数据元素都不是孤立存在的，而是在它们之间存在着某种关系，这种数据元素相互之间的关系称为结构(structure)。根据数据元素之间的不同特性，通常有下列四类基本结构，如图 1.3 所示。

(1) **集合**：集合结构中的所有元素都“属于同一集合”，即只要满足同属于一个集合就是集合结构，这是一种极为松散的结构。

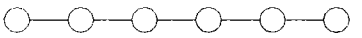
(2) **线性结构**：该结构的数据元素之间存在着一对一的关系。

(3) **树形结构**：该结构的数据元素之间存在着一对多的关系。

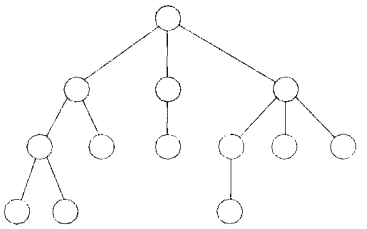
(4) **图形结构**：该结构的数据元素之间存在着多对多的关系，图形结构也称为网状结构。



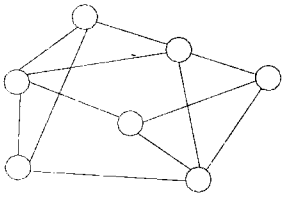
(a)集合结构



(b)线性结构



(c)树形结构



(d)图形结构

图 1.3 四类基本结构关系图

由于集合是数据元素之间关系极为松散的结构，因此也可用其他结构来表示它。

从数据结构的概念中可以知道，一个数据结构由两个要素组成：一个是数据元素的集合；另一个是关系的集合。在形式上，数据结构可以用一个二元组表示。数据结构的形式定义为：数据结构是一个二元组。

```
Data_structure=(D,R)
```

其中，D 是数据元素的有限集，R 是 D 上关系的有限集。

数据结构包括数据的**逻辑结构**和数据的**物理结构**。上述四种数据结构是从实际问题中抽象出来的数学模型，结构中的“关系”描述的是数据元素之间的逻辑关系，因此又称为数据的逻辑结构。我们研究数据结构的目的是为了在计算机中实现对它的操作，为此还需要研究如何在计算机中表示一个数据结构。数据结构在计算机中的表示(又称映像)称为数据的物理结构，或称**存储结构**。它所研究的是数据结构在计算机中的实现方法，包括数据元素的表示和相互之间关系的表示。

数据元素之间的关系在计算机中有两种不同的表示方法：**顺序存储结构**和**链式存储结构**。顺序存储结构是把逻辑上相邻的元素存储在物理位置相邻的存储单元中，由此得到的存储表示称为顺序存储结构。顺序存储结构是一种最基本的存储表示方法，通常借助于程序设计语言中的数组来实现。链式存储结构对逻辑上相邻的元素不要求其物理位置相邻，元素之间的关系通过附设的指针段来表示，即一个元素位置除了存储其自身的值以外，还存储下一个元素的地址，由此得到的存储表示称为链式存储结构，链式存储结构通常借助于程序设计语言中的指针来实现。

数据的逻辑结构和物理结构是密切相关的两个方面，任何一个算法的设计取决于选定的数据的逻辑结构，而算法的实现依赖于所采用的物理结构。

“数据结构”作为一门独立的课程在国外是从 1968 年才开始设立的。在这之前其某些

内容曾在其他课程(如表处理语言中)有所阐述。1968 年美国的唐·欧·克努特教授开创了数据结构的最初体系，他所著的《计算机程序设计技巧》第一卷《基本算法》是第一本较系统地阐述数据的逻辑结构和存储结构及其操作的著作。从 20 世纪 60 年代末到 70 年代初，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，人们就越来越重视数据结构，认为程序设计的实质是对确定问题选择一种好的结构，加上一种好的算法。从 20 世纪 70 年代中期到 80 年代初，各种版本的数据结构著作就相继出现。目前在我国，“数据结构”也已经不仅仅是计算机专业教学计划中的核心课程之一，而且也是其他非计算机专业的主要选修课程之一。

1.1.3　数据结构课程的内容和任务

数据结构是与数学、计算机硬件和软件有十分密切关系的学问，是介于数学、计算机硬件和软件之间的一门计算机科学与技术专业的核心课程，是高级程序设计语言、编译原理、操作系统、数据库、软件工程等课程的基础。同时，数据结构技术也广泛应用于信息科学、系统工程以及各种工程技术领域。

数据结构课程集中讨论软件开发过程中的设计阶段，同时涉及编码和分析阶段的若干问题。可以说，数据结构贯穿大型程序设计的始终。数据结构课程的内容体系可以从以下三个方面加以认识。

第一，抽象。从实际问题中抽象出逻辑结构，即分析解决该问题需要用到什么样的数据，数据元素之间的关系如何，以及基于这种逻辑结构上的基本运算。

第二，实现。将逻辑结构用适当的物理结构在计算机内表示出来，同时编写算法实现各种基本运算。

第三，评价。对不同的数据结构进行比较和算法分析。

数据结构的核心技术是分解与抽象。通过对问题的抽象，舍弃数据元素的具体内容，得到数据的逻辑结构。通过分解将问题处理的要求划分成各种功能，再通过抽象舍弃实现细节，只分析基本运算的定义。这是一个从具体问题抽象为解决问题的数据结构的过程，这类似于数学中将一个实际问题抽象成一个数学公式或方程式。然后，通过增加对实现细节的考虑进一步得到存储结构和实现运算。

总之，用计算机来处理问题，所面临的信息量往往是非常巨大的。面对成千上万的数据，首先面临的是这些数据如何在计算机中进行存放的问题，不同的存放方式其基本运算必然是不同的。**数据结构课程的任务**就是将实际问题抽象成数据的逻辑结构，再将逻辑结构用适当的存储结构在计算机中表示出来，并实现基本运算，为编写程序及实现用计算机处理实际问题打下坚实的基础。这同时也体现了数据结构这门课程的重要性。因此，熟练地掌握数据结构这门课程是学习计算机及相关专业的前提。

1.2　数据类型、抽象数据类型和参数传递

实际问题的千差万别，往往会涉及到许许多多的因素。而计算机是用来处理数据的，如果用计算机来处理这些实际问题，需要将这些实际问题抽象成相应的数学模型，即首先