普通高等教育“十二五”规划教材

数据结构（Java版）

吴仁群　编著

内容讲述由浅入深，符合初学者学习习惯
辅以图形或实例来帮助读者理解知识点

所有算法均使用Java语言实现

中国水利水电出版社
www.waterpub.com.cn

等教育“十二五”规划教材

数据结构（Java版）

吴仁群　编著

中国水利水电出版社
www.waterpub.com.cn

内容提要

本书是针对数据结构初学者编写的基础教程，书中不仅讲解了数据结构常用的基本理论知识，而且提供了大量应用实例，以帮助初学者对知识的理解。全书共分 8 章：绪论、线性表、栈和队列、串和数组、树和二叉树、图、查找、排序等。

本书内容实用，结构清晰，实例丰富，可操作性强，可作为高等学校数据结构的教材，也可作为计算机相关专业的培训和自学教材。

图书在版编目（CIP）数据

数据结构 : Java版 / 吴仁群编著. -- 北京 : 中国水利水电出版社, 2013.1
普通高等教育“十二五”规划教材
ISBN 978-7-5170-0401-1

Ⅰ. ①数… Ⅱ. ①吴… Ⅲ. ①数据结构－高等学校－教材②JAVA语言－程序设计－高等学校－教材 Ⅳ. ①TP311.12②TP312

中国版本图书馆CIP数据核字(2012)第285650号

书　　名	普通高等教育“十二五”规划教材 **数据结构（Java 版）**
作　　者	吴仁群　编著
出版发行	中国水利水电出版社 （北京市海淀区玉渊潭南路 1 号 D 座　100038） 网址：www.waterpub.com.cn E-mail：sales@waterpub.com.cn 电话：（010）68367658（发行部）
经　　售	北京科水图书销售中心（零售） 电话：（010）88383994、63202643、68545874 全国各地新华书店和相关出版物销售网点
排　　版	北京零视点图文设计有限公司
印　　刷	北京嘉恒彩色印刷有限责任公司
规　　格	184mm×260mm　16 开本　14.5 印张　399 千字
版　　次	2013 年 1 月第 1 版　2013 年 1 月第 1 次印刷
印　　数	0001—3000 册
定　　价	32.00 元

前　言

本书是计算机相关专业中一门重要的专业基础课程。当用计算机来解决实际问题时，就要涉及数据及数据之间关系的表示及处理，而数据及数据之间关系的表示及处理正是数据结构主要研究的对象。通过数据结构的学习可以为后续课程尤其是软件方面的课程打下了厚实的知识基础。因此，数据结构在计算机相关专业中具有举足轻重的作用。

作为一本关于数据结构的基础教材，本书具有以下特点。

（1）本书内容的讲述由浅入深，符合初学者的计算机语言学习习惯。

（2）本书在讲述每个知识点时，都辅以图形或具体实例，使读者能够从具体应用中掌握知识，能够很容易地将所学的知识应用于实践。

（3）本书每章后面附有习题，读者可通过做习题，巩固并掌握所学知识。

（4）本书所有算法均采用当今流行的 Java 语言编写，便于读者学以致用。

本书共有 8 章。第 1 章讲述数据、数据结构、数据类型、抽象数据类型等基本概念，算法和算法描述、算法的性能分析等有关知识。第 2 章介绍了线性表的含义及 ADT 描述、线性表的两种顺序存储和链式存储、不同存储方式下基本操作的实现及应用。第 3 章介绍了栈和队列的定义及 ADT 描述、栈和队列的存储结构、不同存储结构下基本操作实现及应用。第 4 章介绍了串和数组的定义及 ADT 描述、串的存储结构及应用、数组的存储方式、压缩存储及应用。第 5 章讲述了树和二叉树的概念及 ADT 描述、树和二叉树的存储方式、树和二叉树及森林的遍历及应用、树和二叉树及森林的转换、哈夫曼树及应用。第 6 章讲述了图的概念及 ADT 描述、图的存储方式、图的遍历、最小生成树、有限无环图及应用等。第 7 章介绍了查找的基本概念、静态查找和动态查找的基本方法、哈希表的概念及查找方法等。第 8 章讲述了排序的基本概念、插入排序、交换排序、选择排序、归并排序和基数排序等排序的主要方法。为便于学习，本书作者可以提供每章上机调试的源程序。

本书由吴仁群编写。在编写过程中，编者参考了本书参考文献所列举的图书，并得到了中国水利水电出版社的大力支持，在此对参考文献中图书的作者及中国水利水电出版社表示深深的感谢。

由于时间仓促，书中难免存在一些不足之处，敬请读者批评指正。

编　者

2012 年 11 月于北京印刷学院

目　录

第 1 章　绪论

- 了解数据结构的含义及有关概念。
- 了解数据结构的逻辑结构和物理结构。
- 了解算法的含义、描述方法及重要特性。
- 掌握估算算法的时间复杂度和空间复杂度的方法。

1.1　基本概念

数据结构是计算机科学与技术领域中广泛被使用的术语。它用来反映一个数据的内部构成，即一个数据由哪些成分构成，这些成分的构成是什么样，呈现什么结构。数据结构作为一门学科主要研究数据的各种逻辑结构和存储结构，以及对数据的各种操作。因此，数据结构主要有三个方面的内容：数据的逻辑结构、数据的物理存储结构、对数据的操作（或算法）。一般来说，算法的设计取决于数据的逻辑结构，算法的实现取决于数据的物理存储结构。

1.1.1　数据和数据结构

1. 数据

数据是对客观事物的符号表示，是所有能够输入到计算机中并被计算机程序处理的符号（比如，0 和 1，a、b 等）的集合。包括字符、文字、表格、图像等，都可称为数据。例如，学生信息管理系统所要处理的数据可能是一张如表 1-1 所示的表格。

表 1-1　学生信息表

姓名	性别	籍贯	出生年月	政治面貌	联系方式
张三	男	湖北	1969-10-01	中共党员	29291230
李四	男	湖南	1969-5-01	群众	29293456
⋮	⋮	⋮	⋮	⋮	⋮

2. 数据元素

数据元素是数据集合中的一个实体，是计算机程序中加工处理的基本单位。数据元素按其组成可分为简单型数据元素和复杂型数据元素。简单型数据元素由一个数据项组成，所谓数据项就是数据中不可再分割的最小单位；复杂型数据元素由多个数据项组成，它通常包含着一个概念的多方面信息。

例如，一个在校大学生的基本信息组成如下：

姓名	性别	籍贯	出生年月	政治面貌	联系方式

在上述信息中，除“出生年月”外，其他都是简单型数据元素。“出生年月”可分为年、月、日三部分，属于复杂型数据元素。

出生年月：	年	月	日

3. *数据结构*

简单地说，数据结构就是相互之间存在一种或多种特定关系的数据元素的集合。数据结构有逻辑上的数据结构（即逻辑结构）和物理上的数据结构（即物理结构）之分。

数据的逻辑结构是指数据元素之间的逻辑关系。常见的逻辑结构有集合结构、线性结构、树形结构和图状结构。

集合结构：数据元素之间的关系是“属于同一集合”，如图 1-1（a）所示。

线性结构：数据元素之间存在一对一的关系，如图 1-1（b）所示。

树形结构：数据元素之间存在一对多的关系，如图 1-1（c）所示。

图状结构：数据元素之间存在多对多的关系，如图 1-1（d）所示。

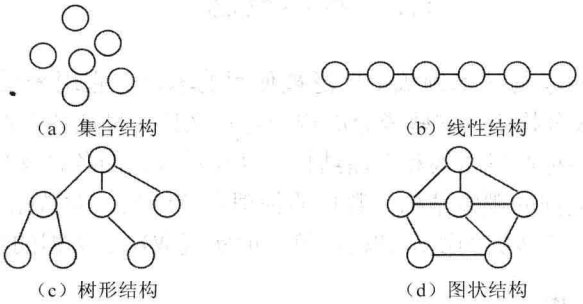


图 1-1　基本逻辑结构

一般来说，一个数据结构 DS（Data Struture）可以表示为一个二元组：

$$DS=(D, S)$$

式中：D 为数据元素的集合；S 为定义在 D（或其他集合）上的关系的集合。

例如 1，复数是一个数据结构：

$$Complex=(C, R)\quad C=\{c1, c2\}\quad R=\{<c1, c2>\}$$

数据的物理结构，也称存储结构，是指数据结构在计算机存储器中的具体实现，是逻辑结构的存储映像（Image）。常见的存储结构有顺序存储结构和链式存储结构。前者是借助于数据元素的相对存储位置来表示数据元素之间的逻辑结构，如图 1-2（a）所示；后者是借助于指示数据元素地址的指针表示数据元素之间的逻辑结构，如图 1-2（b）所示。

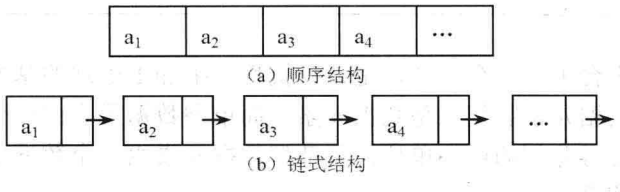


图 1-2　基本存储结构

为了叙述方便和避免产生混淆，通常把数据的逻辑结构统称为数据结构，把数据的物理结构统称为存储结构。

1.1.2　数据类型

在用高级程序语言编写的程序中，必须对程序中出现的每个变量、常量或表达式，明确说明它们所属的数据类型。数据类型是指数据的取值范围及其上可进行的操作的总称。例如，Java 语言中提供的基本数据类型有整型、浮点型、双精度型、逻辑型、字符型等。

高级程序设计语言中的数据类型可分为简单数据类型和复合数据类型。简单数据类型中的值是不可分的，例如整型、实型等。复合数据类型的值是由若干成分按某种结构组成的，是可分解的，如 Java 语言中的类是一种复合数据类型。例如：

```
class  NODE {
    int   age;
    char  name[20];
    float  score;
}
NODE  p=new NODE() ;
```

1.1.3　抽象数据类型

抽象数据类型（Abstract Data Type，ADT）是一个数学模型以及定义在该模型上的一组操作。抽象数据类型的定义仅取决于它的一组逻辑特性，而与其在计算机内部如何表示和实现无关，即不论内部结构如何变化，只要它的数学特性不变，都不影响其外部使用。因此抽象数据类型可实现信息隐蔽和数据封装，以及使用与实现相分离。

抽象数据类型可以表示为一个三元组：

$$\text{ADT}=(D,\ S,\ P)$$

这里，D 是数据对象集合，S 是 D 上的关系集合，P 是对 D 的基本操作。

实际中，可以按照如下结构来描述抽象数据类型：

```
ADT 抽象数据类型名{
    数据对象：〈数据对象的定义〉
    数据关系：〈数据关系的定义〉
    基本操作：〈基本操作的定义〉
}ADT 抽象数据类型名
```

数据对象和数据关系的定义用伪码表示，基本操作定义格式如下：

```
    基本操作名（参数表）
    初始条件：〈初始条件描述〉
    操作结果：〈操作结果描述〉
ADT Triplet{
    数据对象：D={e1,e2,e3|e1,e2,e3∈ElemSet}
    数据关系：R1={<e1,e2>,<e2,e3>}
    基本操作：
    InitTriplet(&T,v1,v2,v3)
    操作结果：构造了三元组 T，元素 e1、e2、e3 分别被赋予参数 v1、v2、v3 的值。
```

```
    DestroyTriplet(&T)
    操作结果：三元组被销毁。
    Put(&T,i,e)
    初始条件：三元组已存在，i∈[1,3]
    操作结果：改变 T 中第 i 个元素的值为 e。
    ……
}ADT Triplet
```

1.1.4　数据结构的符号描述举例

1. 集合结构

【实例 1-1】小组成员组成的数据结构。

Set＝（D，R）

D={张三，李四，王五，吴一，陈二}

R＝{<张三，李四>,<张三，王五>,<张三，吴一>,<张三，陈二>,<李四，王五>,<李四，吴一>,<李四，陈二>,<王五，吴一>,<王五，陈二>,<吴一，陈二>}

这里关系<a，b>表示 a 和 b 属于同一小组。

如图 1-3（a）所示。

2. 线性结构

【实例 1-2】排队购买车票成员组成的数据结构。

List＝（D，R）

D={张三，李四，王五，吴一，陈二}

R＝{<张三，李四>,<李四，王五>,<王五，吴一>,<吴一，陈二>}

这里关系<a，b>表示在队列中 a 是 b 的直接前驱。

如图 1-3（b）所示。

3. 树形结构

【实例 1-3】家庭成员组成的数据结构。

T＝（D，R）

D={祖父，姑姑，叔叔，父亲，儿子，孙子}

R＝{<祖父，姑姑>,<祖父，叔叔>,<祖父，父亲>,<父亲，儿子>,<儿子，孙子>}

这里关系<a，b>表示在队列中 a 是 b 的父辈。

如图 1-3（c）所示。

4. 图状结构

【实例 1-4】四个直辖城市航空网络的数据结构。

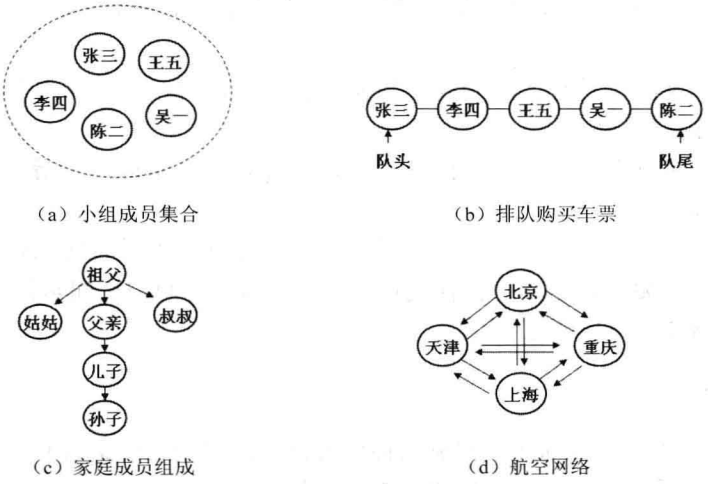
T＝（D，R）

D={北京，上海，天津，重庆}

R＝{<北京，上海>,<北京，天津>,<北京，重庆>,<上海，北京>,<上海，天津>,<上海，重庆>,<天津，北京>,<天津，上海>,<天津，重庆>,<重庆，北京>,<重庆，天津>,<重庆，天津}

这里关系<a，b>表示 a 有直达航班到 b。

如图 1-3（d）所示。



（a）小组成员集合　　（b）排队购买车票

（c）家庭成员组成　　（d）航空网络

图 1-3　数据结构应用举例

1.2　算法和算法描述

1.2.1　概念和特性

1. 算法概念

算法是在有限步骤内求解特定问题所使用的一组定义明确的规则。通俗点说，就是计算机求解特定问题的步骤的描述。特定的问题可以是数值的，也可以是非数值的。解决数值问题的算法称为数值算法，科学和工程计算方面的算法都属于数值算法，如求解数值积分，求解线性方程组、求解代数方程、求解微分方程等。解决非数值问题的算法称为非数值算法，数据处理方面的算法都属于非数值算法。例如各种排序算法、查找算法、插入算法、删除算法、遍历算法等。数值算法和非数值算法并没有严格的区别。

2. 算法的特征

一个算法应该具有以下五个重要的特征：有穷性、确定性、可行性、输入和输出。

（1）有穷性是指一个算法必须保证执行有限步之后结束。

（2）确定性是指算法的每一步骤必须有确切的定义，没有二义性。

（3）可行性是指算法中描述的每一步操作都可以通过已有的基本操作执行有限次实现。

（4）输入是指一个算法有零个或多个输入，以描述运算对象的初始情况，所谓零个输入是指算法本身定出了初始条件。

（5）输出是指一个算法有一个或多个输出，以反映对输入数据加工后的结果。没有输出的算法是毫无意义的。

1.2.2　算法设计要求

评价一个好的算法有以下几个标准：

（1）正确性。算法对于一切合法的输入数据都能产生满足规格说明的结果。

（2）可读性。算法应该好读，易于理解，一般在满足正确性的前提下，算法越简单越好。

（3）健状性。算法应具有容错处理。当输入非法数据时，算法应对其作出反应，而不是产生莫名其妙的输出结果。

（4）效率与存储量需求。效率是指算法执行的时间；存储量需求指算法执行过程中所需要的最大存储空间。

在保证满足标准（1）、（2）、（3）情况下，希望算法执行所需时间比较短，所占用存储空间比较小。实际中，要满足这两点往往是很困难的，因为时间和空间是彼此冲突的。因此，应该具体情况来权衡时间和空间。

1.2.3　算法描述

算法采取多种方式来进行描述。常见的描述方式有：采用自然语言来描述、采用程序流程图的形式来描述、采用某种具体程序语言来描述等。

1. 采用自然语言来描述

这种方式是使用自然语言来描述问题的求解过程。下面举例说明。

问题 P：判断正整数 N 是否为素数。

使用自然语言来描述的算法如下：

Step1：令 i=2。

Step2：判断 i 是否小于等于 $N/2$，若是，转到 Step3；否则，转到 Step4。

Step3：判断 N 除以 i 的余数 R 是否等于零？

若 R 等于零，转到 Step5。

否则，i 加 1，转到 Step2。

Step4：输出 N 为素数。

Step5：算法结束。

2. 采用程序流程图的形式来描述

这种方式是使用流程图符号来描述问题的求解过程。求解问题 P 所对应的流程如图 1-4 所示。

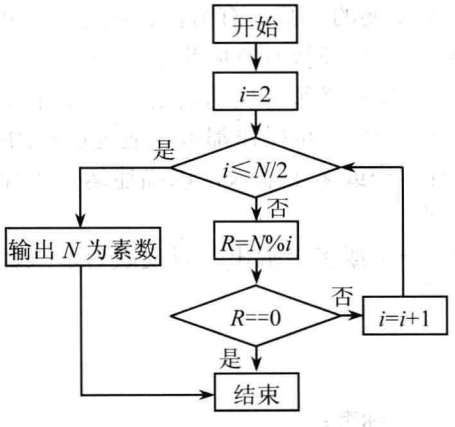


图 1-4　程序流程图

3. 采用某种具体程序语言来描述

这种方式是使用某种具体语言（如 Java 语言）来描述问题的求解过程。求解问题 P 所对应的 Java 程序如下：

```
boolean  isPrimeNumber(int N )
{
    int  i,j;
    for(i=2;i<=N/2;i++)  if  (N%i==0)  break;
    if(i<=N/2) return false;
    return true;
}
```

除了上述三种方式外，还有利用类语言（如类 C 语言、类 Pascal 等）来描述问题的求解过程，感兴趣者可以参看有关文献。不管采用哪种方式描述算法，都必须能够正确描述求解过程。本书中所有算法均采用 Java 语言来描述。

4. 设计算法的基本过程

一般来说，针对某个具体设计求解算法的基本过程包括以下四个阶段：

（1）通过对问题进行详细分析，抽象出相应的数学模型。

（2）确定使用的数据结构，并在此基础上设计对此数据结构实施各种操作的算法。

（3）选用某种语言将算法转换成程序。

（4）调试并运行这些程序。

1.3 算法的性能分析

算法的性能分析是指对算法质量优劣的评价。除了正确性、可读性和健壮性等特性外，重点要分析算法的时间复杂度和空间复杂度。

1.3.1 时间复杂度

算法的时间复杂度是依据该算法编制的程序在计算机上执行所消耗的时间来度量。这种度量可采用事后统计和事前估计两种方式。

1. 事后统计

事后统计就是利用计算机内计时功能，不同算法的程序可以用一组或多组相同的统计数据区分。这种方式缺点在于：必须先运行依据算法编制的程序，所得时间统计量依赖于硬件、软件等环境因素，掩盖算法本身的优劣。

2. 事前估计

一个高级语言程序在计算机上运行所消耗的时间取决于：

（1）依据的算法选用何种策略。

（2）问题的规模。

（3）程序语言。

（4）编译程序产生机器代码质量。

（5）机器执行指令速度。

同一个算法用不同的语言、不同的编译程序、在不同的计算机上运行，效率均不同，所

以使用绝对时间单位衡量算法效率不合适。

实际中，可以撇开那些与计算机硬件、软件有关的因素，可以认为一个特定算法的“运行工作量”的大小，只依赖于问题的规模（通常用整数量表示），或者说，它是问题规模的函数。

任何一个算法都是由控制结构和若干基本操作组成。一般情况下，算法中基本操作重复执行的次数是问题规模的 n 的函数，记为 $T(n)$。以下以一个矩阵相乘算法来说明如何计算一个算法中语句执行的次数。

语句	执行次数
`maxtrixMultiply( intA[], intB[][],int C[] ,int n)`	
`//int A[n][n],B[n][n],C[n][n]`	
`{`	
`for(int i=0;i<n;i++)`	$n+1$
`{`	
`for(int j=0;j<n;j++)`	$n(n+1)$
`{`	
`C[i][ j]=0;`	n^2
`for(int k=0;k<n;k++)`	$n^2\times(n+1)$
`C[i][j]=C[i][j]+A[i][k]*B[k][j];`	n^3
`}`	
`}`	
`}`	

总执行次数 $T(n)$为

$$T(n)=n+1+n(n+1)+n^2+n^2\times(n+1)+n^3=2n^3+3n^2+2n+1$$

定义：如果存在一个 $g(n)$，当 $n\to\infty$时，有

$$T(n)/g(n)=\text{常数}\neq 0,$$

则称函数 $T(n)$与 $g(n)$同阶，或者说，$T(n)$与 $g(n)$同一个数量级，记作

$$T(n)=O(g(n))$$

称上式为算法的时间复杂度，或称该算法的时间复杂度为 $O(g(n))$。其中，n 为问题的规模的量度。

基于高等数学中极限知识可知，当一个算法的执行次数可以表达为如下形式：

$$T(n)=\alpha_m n^{\beta_m}+\alpha_{m-1}n^{\beta_{m-1}}+\cdots+\alpha_0,\beta_i>\beta_{i-1},i=1,2,\cdots,m$$

则该算法的时间复杂度为 $O(n^{\beta_m})$。

例如

$$\lim_{n\to\infty}\frac{2n^3+3n^2+2n+1}{n^3}=2$$

算法 maxtrixMulti 的时间复杂度为 $O(n^3)$。

由上面时间复杂度的定义可以看出，在计算一个算法的时间复杂度时，往往只需分析影响算法运行时间的主要部分，也即只需分析循环嵌套数最多的嵌套循环中最内层语句执行的次数。

很多算法的时间复杂度不仅与问题的规模有关，而且还与它所处理的数据集的状态有关。因此，在分析一个算法的复杂度时往往根据数据集中可能出现的最坏情况和最好情况分别估计出算法的最坏时间复杂度和最好时间复杂度。有时，要对数据集的分布做出某种假设，估算在这种分布下算法的平均时间复杂度。

1.3.2　空间复杂度

类似于算法的时间复杂度，本书中以空间复杂度作为算法所需存储空间的量度，记作：

$$S(n)=O(f(n))$$

式中：n 为问题的规模（或大小）。

存储空间包括固定部分和可变两个部分。前者是指程序指令代码的空间，常数、简单变量、定长成分（如数组元素、结构成分、对象的数据成员等）变量所占的空间；后者大小与实例特性有关的成分变量所占空间、引用变量所占空间、递归栈所用的空间、通过 new 和 delete 命令动态使用的空间。

如果输入数据所占空间只取决于问题本身，与算法无关，那么只需分析除输入和程序之外的额外空间，否则应同时考虑输入本身所需空间。

1.3.3　分析算法时间复杂度举例

【实例 1-5】求 M 和 N 的最大值。

算法为

```
int maxMN(int M,int N)
{
    int max=0;
    Max=N;
    if M>N then  max=M;
    return max;
}
```

分析：本算法包含三个基本语句，执行次数均为 1，因此算法的执行时间是一个与问题规模 n 无关的常数，即算法的时间复杂度 $T(n)=O(1)$。

【实例 1-6】计算自然数 1 至 N 的和。

算法为

```
long sum(int N)
{
    int i;
    long sum=0;
    for(i=1;i<=N;i++)  sum=sum+i;
    return sum;
}
```

分析：本算法的执行次数 $T(N)=aN+b(a>0)$，是问题规模 N 的线性函数，因此算法的时间复杂度 $T(n)=O(N)$。

【实例 1-7】输出乘法表。

算法为

```
void output(int N)
{
    int i,j;
    for(i=1;i<=N;i++)
    {
        for(j=1;j<=i;i++) printf("%2d*%2d=%2d",i,j,i*j);
        printf("\n")
    }
}
```

分析：本算法的执行次数 $T(N)=aN^2+bN+c(a>0)$，是问题规模 N^2 的线性函数，因此算法的时间复杂度 $T(n)=O(N^2)$。

1.4　习题

1. 名词解释

简述下列术语：数据、数据项、数据元素、数据对象、数据逻辑结构、数据物理结构、数据结构、数据类型、算法。

2. 填空题

（1）数据的物理结构包括_____的表示和_____的表示。

（2）对于给定的 n 个元素，可以构造出的逻辑结构有_____、_____、_____、_____四种。

（3）数据的逻辑结构是指_____。

（4）一个数据结构在计算机中_____称为存储结构。

（5）抽象数据类型的定义仅取决于它的一组_____，而与_____无关，即不论其内部结构如何变化，只要它的_____不变，都不影响其外部使用。

（6）数据结构中评价算法的两个重要指标是_____。

（7）一个算法具有 5 个特性：_____、_____、_____，有零个或多个输入、有一个或多个输出。

3. 简答题

（1）数据结构是一门研究什么内容的学科？

（2）数据元素之间的关系在计算机中有几种表示方法？各有什么特点？

（3）抽象数据类型的主要特点是什么？使用抽象数据类型的主要好处是什么？

4. 计算时间复杂度

求下列程序段的时间复杂度：

（1）

```
for(i=1;i<=n;i++)
for(j=1;j<=n;j++)
s++;
```

时间复杂度为_____。

（2）

```
i=0;
s=0;
while(s<n) {
i++;
s=s+i;
}
```

时间复杂度为_____。

（3）

```
i=1;
do {
    j=1;
    do {
        System.out.printf("%d\n",i*j);
        j++;
    }while(j>n)
    i++;
}while(i>n);
```