

数据结构与算法

To do

□ 班级通讯录制作

- 包括：学号，姓名，出生日期，电话，手机， email
 - 其中学号、姓名最重要。bbs账号、ftp目录、登记成绩等都与此有关
 - 下周班长交给助教
-

课程目标

- 学会如何有效地组织信息，以便支持高效的数据处理
 - 掌握常用的基本数据结构及其应用
 - 学会合理地组织数据，有效地表示数据，有效地处理数据
 - 基本掌握算法的设计与分析技术
 - 提高程序设计能力与程序的质量
-

课程考核内容

- 1. 平时(书面作业、课堂测试): 20%
 - 2. 上机实习(+实习报告): 15%
 - 3. 期中考试: 20%
 - 4. 期末考试: 40%
 - 5. 考勤和态度: 5%
-

要求：1 诚信

- 端正学习态度、调动学习兴趣
 - 提倡讨论，但严禁抄袭
 - 可以讨论思路，请同学查看算法的逻辑问题和效率问题
 - 但要亲自动手实现
 - 发现抄袭，则抄袭者和被抄袭者本次作业或上机题计双倍倒扣分，即得-20分，且以后的作业题会得到重点检查。严重的期评将给予不及格处理
 - 数据结构教学计划和要求（参见课程网站）
-

要求：2 按时提交作业，严禁抄袭

- 所有书面作业和上机作业都必须在指定的期限内完成并提交
 - 一般周三交书面作业，除非有不可抗拒的客观原因。请严格按提交时间完成书面作业和上机作业。例如，一个满分为**10**分的作业题，记分标准为：
 - (1) 准时提交，满分可达**10**分（除个别加分）；
 - (2) 延迟**3**天之内提交，满分可达**7**分；
 - (3) 延迟**7**天之内提交，满分可达**3**分；
 - (4) **7**天之后提交或不交，得分 **-5**分；
 - (5) 抄袭得 **-20**分。
-

要求：3 书面作业提交格式

- ❑ 写学号、名字
 - ❑ 写算法分析、注释
 - ❑ 算法中直接使用的函数、过程需写**ADT**，并说明函数功能、入口参数、出口参数
 - ❑ 注意算法格式（层次嵌套、不同功能块之间留空）
-

要求：4 上机题提交格式

- ❑ 上机作业打成一个**zip**包后提交，建议命名方式： 学号+姓名+作业次数
e.g. , ” 00308096张宁1.zip”

包中须含有：

1. **readme.txt**文件，把你的程序运行环境、编译运行步骤、程序功能等等简单说明一下
 2. 附加了诚实代码保证和足够注释的源程序以及相关的项目和资源文件。例如，VC++中的**.dsw**, **.dsp**文件，**rc**目录中的图像资源文件；Jbuilder中的**.jpr**或**.jpx**文件，特殊的**Java**包等等
 3. 上机实习总结报告
-

要求：5 上机题编程风格

- ❑ 诚实代码保证
- ❑ 内部文档要求
- ❑ 过程代码要求
- ❑ 面向对象的代码要求

具体参见课程网站

教材

□ 张铭，王腾蛟，赵海燕，《数据结构与算法》，高等教育出版社，**2008年 6月**。——普通高等教育“十一五”国家级规划教材

□ 张铭、赵海燕、王腾蛟，《数据结构与算法——学习指导与习题解析》，高等教育出版社，**2005年 10月**。——“十五”国家级规划教材配套参考书



教材 & 参考书

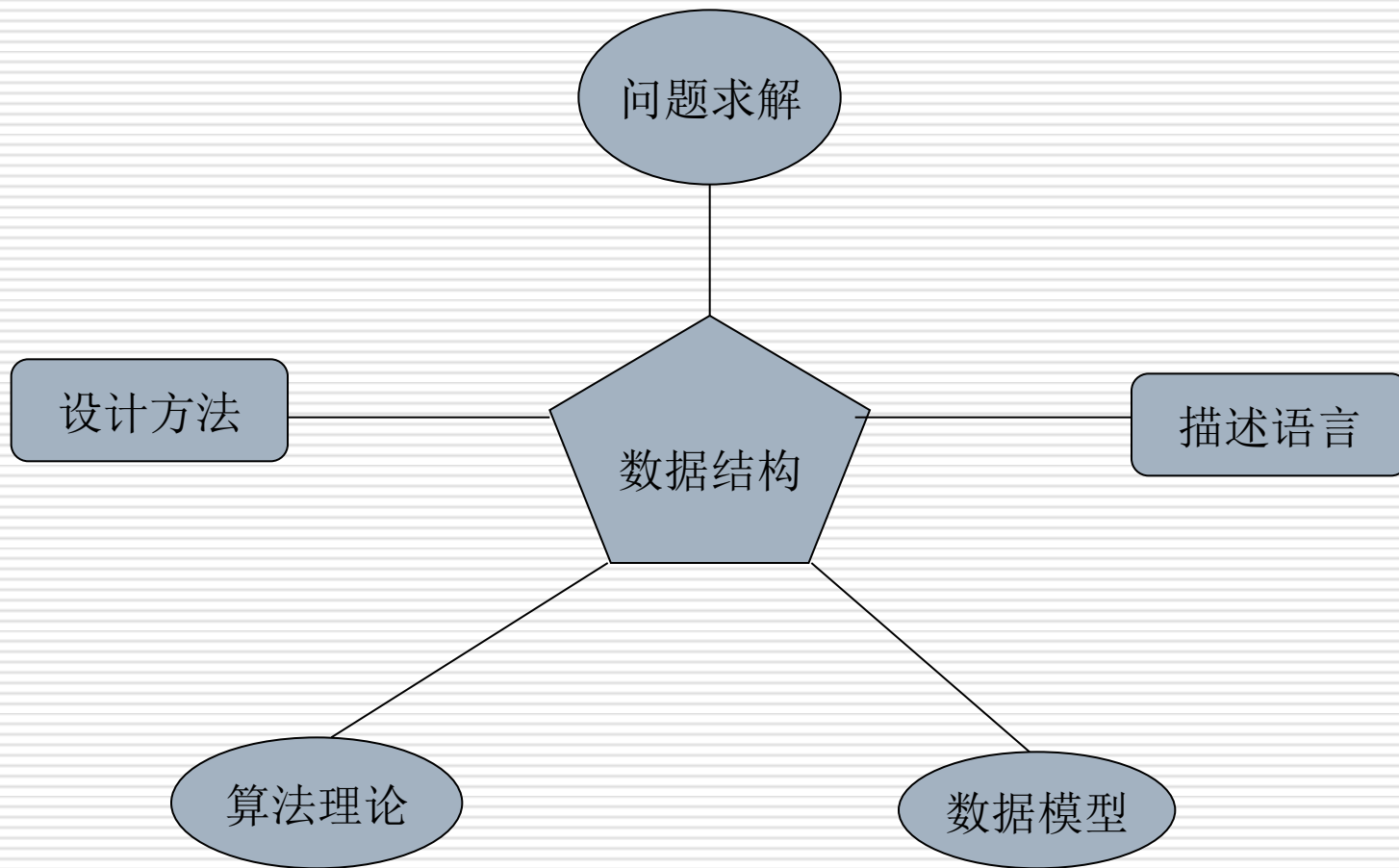
□ 参考书目

- 张铭、刘晓丹译,《数据结构与算法分析》(C++第二版), 电子工业出版社, 2002年。(用1998年的第一版也可以, 也可以用英文原版)
 - 张铭、刘晓丹译,《数据结构与算法分析——Java版》, 电子工业出版社, 2001年
 - 许卓群,《数据结构》, 高等教育出版社, 1988年
 - 严蔚敏,《数据结构题集》, 清华大学出版社
 - Donald E. Knuth, The Art of Computer Programming, Addison Wesley. Vol. 1, Vol 3. 国防工业出版社影印。(苏运霖译)
 - Thomas H.Cormen, Charles E.Leiserson, Ronald L. Rivest, Clifford Stein, Inroduction to Algorithms, MIT Press. 高等教育出版社影印。
 - William Ford,《Data Structure with C++》, 清华大学出版社
-

第一章 概论

- 1.1 为什么要学习数据结构
 - 1.2 什么是数据结构
 - 1.3 抽象数据类型
 - 1.4 算法的特性及分类
 - 1.5 算法的效率度量
 - 1.6 数据结构的选择和评价
-

1. 为什么学习数据结构？



为什么学习数据结构？

□ 求解问题

- 计算机科学就是“一种关于信息结构转换的科学” (Wegnor); (数据结构也称“信息结构”)
- 计算机科学是“算法的学问”，算法是精确定义的一系列规则，指出怎样从给定的输入信息经过有限步骤产生所求的输出信息 (Knuth)
- 其实数据结构与算法两者互为存在（数据结构离不开施于其上的操作，同时算法也必然离不开作为其处理对象和结果的数据）

□ 例子：

- 已知一组人的身高，从中找出最高、最矮的，再找出身材最适中的（诸如101个人中，找出高度第51位的那个）；
 - **Tower of Hanio:** 给出3个柱子和 n 个圆盘，起初所有盘子均放在最左边的柱子上，按大盘在下的顺序堆放；如何把所有盘子移到最右的柱子上？要求任何盘子都不能放到比它小的圆盘上面
-

为什么学习数据结构？

□ 核心基础课程

- 任何问题都离不开数据
- 数据是计算机化的信息（对现实世界的事物采用计算机能够识别、存储和处理的形式所进行的描述）
- 后续专业课程学习的必要知识与技能准备
 - 编译技术要使用栈、散列表及语法树
 - 操作系统中用队列、存储管理表及目录树
 - 数据库系统运用线性表、多链表、及索引树
 - etc.

□ 提高读者求解复杂问题的能力

Web信息处理

队列、图、字符、
矩阵散列、排序、
索引、检索

人工智能

广义表、集合、搜
索树及各种有向图

图形图像

队列、栈、图、矩阵、
空间索引树、检索

数据库概论

线性表、多链表、
排序及B+索引树

操作系统

队列、存储管理表、
排序及目录树

编译原理

字符串、栈、散
列表及语法树

算法分析与设计

数据结构与算法实习

数据结构与算法

程序设计实习

概率统计

计算概论

集合论与图论

1.2 什么是数据结构

□ 数据结构(data structure)涉及如下三个方面

- 数据的逻辑结构
- 数据的存储结构
- 数据的运算

简言之，一类按照一定逻辑关系组织起来的数据的表示及其相关操作

数据结构的定义

□ 数据的**逻辑结构**

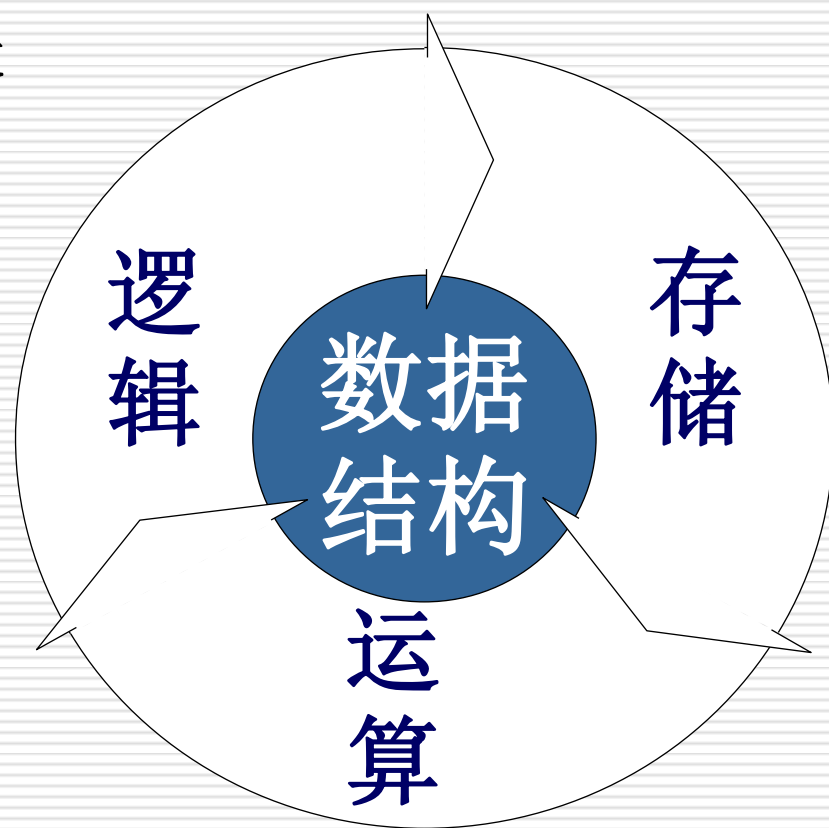
- 图 \supset 树 \supset 二叉树 \supset 线性表

□ 数据的**存储结构**

- 顺序方法、链接方法
- 索引方法、散列方法

□ 数据的**运算**

- 增、删、查、改
- 排序、检索



1.2.1 数据的逻辑结构

□ 二元组 $B=(K, R)$

- K ---- 结点(初等或组合类型)的有限集合
- R ---- K 上的有穷关系的集合 (一组二元关系)

■ K 是由有限个结点组成的集合，每一个结点都代表一个数据或一组有明确结构的数据

■ 关系集 R 是定义在集合 K 上的一组关系，其中每个关系 (relation) r ($r \in R$) 都是 $K \times K$ 上的二元关系，用它描述结点数据之间的逻辑关系

– 例如, $r = \{ \langle k_{i-1}, k_i \rangle \mid k_i \in K, 1 < i < n \}$

数据的逻辑结构：示例

□ 家族人员

- 把每个成员个体的属性描述作为数据结点，而全部人员组成结点集 K
 - 家族的各类亲属关系就是一组关系 R , 其中如母系血缘关系 r 、远亲关系 r^* 、和非血缘的亲情关系 r' 等等，每一个关系要给出具体人员的关系元组
 - 例如：母子关系（王爱莲，张选）
兄弟关系（张选，张立）
妯娌关系（王爱莲，李美英）
-

结点类型：基本数据类型

- ❑ 整数类型(integer): 该类型规定了所能表示的整数范围，在计算机中一般使用1个字节到4个字节来存储整数
 - ❑ 实数类型(real): 计算机的浮点数据类型所能表示的数值范围和精度是有限的。机器一般使用4个字节到8个字节来存储浮点数
 - ❑ 布尔类型(boolean): 取值为真(true)和假(false)，在C++语言中一般使用整数0表示false，用非0表示true
-

结点类型：基本数据类型

- 字符类型(char): 用单个字节（8bit，最高位bit为0）表示ASCII字符集中的字符
 - 汉字符号需要使用2个字节（每个字节的最高位bit为1）的编码，单个字节对于汉字是没有独立含义的
 - 在C++中把双字节表示中文符号的字节类型称为w_char类型（wide character）。
 - 目前国际上已经采用了统一的扩展字符集合标准UNICODE，这一标准允许英、日、韩、阿拉伯语等文字的混合文字处理
-

结点类型：基本数据类型

- 指针类型(pointer)：用于表示机器内存地址，指针表示指向某一内存单元的地址
 - 由于机器的指令系统一般采用32 bit或64bit的地址长度，所以指针类型也相应地用4个字节或8个字节来表示一个指针
 - 指针值的存储和指针值的运算方式，在形式上与正整数相似
 - 但指针的运算一般仅限于两个指针地址的比较，加减，或对一个指针增减一个整数量等
-

结点类型：复合数据类型

- 复合类型是由基本数据类型组合而成的数据结构类型。
 - 例如：在程序语言中常用的数组类型，结构（记录）类型等皆属复合数据类型
 - 复合数据类型本身，又可以参与定义结构更为复杂的结点类型
 - 结点的类型不限于基本数据类型，可以根据应用的需要来灵活定义
-

结构的分类

- 讨论逻辑结构 (K, R) 的分类，一般把讨论重点放在关系集 R 上。用 R 的性质来刻画数据结构的特点，并对数据结构进行分类：
 - 线性结构 (linear structure)
 - 树型结构 (tree structure)
 - 图结构 (graph structure)
-

线性结构

在
程
序
设
计
中
应
用
最
多

- 关系 r 是一种线性关系，或称为‘前后关系’，有时也称为‘大小关系’。关系 r 是有向的，且满足全序性和单索性等约束条件
 - 全序性是指，线性结构的全部结点两两皆可以比较前后（关系 r ）
 - 单索性是指，每一个结点 x 都存在唯一的一个直接后继结点 y 。如果其他结点 z 在 y 之前，则这个 z 也一定在 x 之前，不会在 x ， y 之间

树型结构

- 树型结构简称树结构，或称为层次结构。其关系 r 称为层次关系，或称‘父子关系’、‘上下级关系’等
 - 每一个结点可以有多个的‘直接下级’，但是它只能有唯一的‘直接上级’。
 - 树型结构的最高层次的结点称为根（root）结点。只有它没有父结点
 - 从数学上看，树型结构和图结构的基本区别就是“每个结点是否仅仅从属一个直接上级”。而线性结构和树型结构的基本区别是“每个结点是否仅仅有一个直接后继”
 - 树型结构存在着很多变种，如二叉树结构，堆结构等，它们都有着各自独特的有效应用
-

图结构

- 图结构有时称为结点互联的**网络结构**，因特网的网页链接关系就是一个非常复杂的图结构
 - 对于图结构的关系 **r** 没有加任何约束。这样也就无法象线性结构及树结构那样，利用关系 **r** 的约束来设计图结构的存储结构
 - 在日常应用中图结构往往只是层次结构的一种扩展——允许结点具有多个‘直接上级结点’，关系 **r** 表现为树型结构约束的放松
-

结点和结构

- 对于数据结构 (K, R) ，结点数据类型不限于基本数据类型，可以根据应用需要来灵活设计结点的数据类型
 - 可以认为数据结构的设计是一层一层地进行的
 - 先明确数据结点，及其主要关系 r
 - 在分析关系 r 的同时，也要分析其数据结点的数据类型
 - 如果数据结点的逻辑结构比较复杂，那么把它作为下一个层次，再分析下一层次的逻辑结构
 - 这是一种自顶向下的分析设计方法
-

1.2.2 数据的存储结构

□ 计算机的主存储器的特性

- 其存储空间提供了一种具有非负整数地址编码的，相邻单元的集合，其基本的存储单元是字节
- 计算机的指令具有按地址随机访问存储空间内任意单元的能力，访问不同地址所需的访问时间基本相同

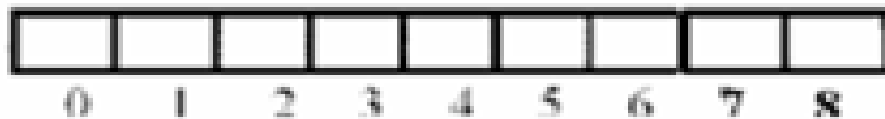
数据的存储结构

- 用数学上的映射来表示，数据的存储结构是建立一种映射，对于数据逻辑结构 (K, r) ，其中 $r \in R$
 - 对它的结点集合 K 建立一个从 K 到存储器 M 的单元的映射： $K \rightarrow M$ ，对于每一个结点 $j \in K$ 都对应一个唯一的连续存储区域 $c \in M$
 - 每一个关系元组 $(j_1, j_2) \in r$ （其中 $j_1, j_2 \in K$ 是结点），亦即 j_1, j_2 的逻辑后继关系应映射为存储单元的地址顺序关系（或指针的地址指向关系）
 - 四种基本存储映射方法：顺序、链接、索引、散列
-

顺序（sequential）方法

- 用一块无空隙的存储区域存储数据称为顺序存储
- 顺序存储把一组结点存储在按地址相邻的顺序存储单元里，结点间的逻辑后继关系用存储单元的自然顺序关系来表达
- 顺序存储法为使用整数编码来访问数据结点提供了便利

S:



顺序方法

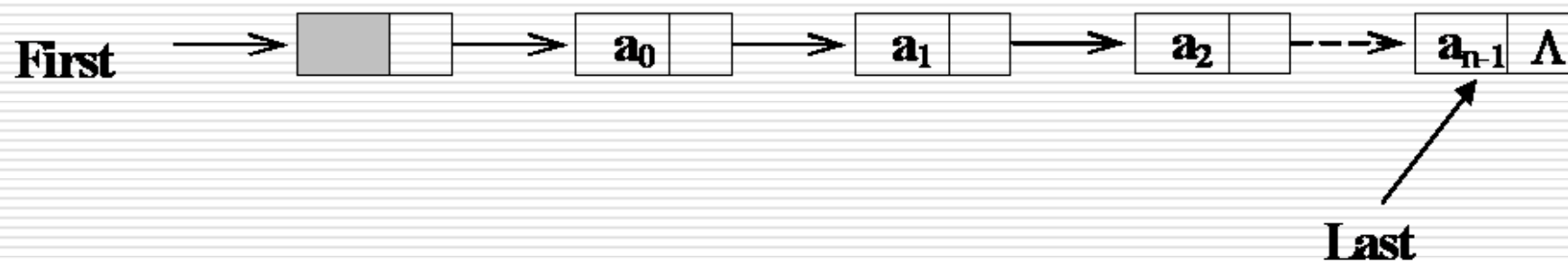
- 顺序存储结构称为**紧凑存储结构**，其紧凑性是指它的存储空间除了存储有用数据外，没有用于存储其他附加的信息
 - 紧凑性可以用“**存储密度**”来度量：它是一个存储结构所存储的“有用数据”和该结构（包括附加信息）整个存储空间大小之比
 - 有时为了“**用空间换取时间**”，在存储结构中存储一些附加信息还是很必要的
 - 譬如用于提高算法的执行速度，或者让算法实现更为简便等
-

链接（linked）方法

- 利用指针，在结点的存储结构中附加指针字段称为**链接法**。两个结点的逻辑后继关系可以用**指针的指向**来表达
 - 任意的逻辑关系 r ，也可以使用这种指针地址来表达。一般的做法是将数据结点分为两部分：
 - 一部分存放结点本身的数据，称为**数据字段**
 - 另一部分存放指针，称**指针字段**，链接到某个后继结点，指向它的存储单元的开始地址。多个相关结点的依次链接就会形成**链索**
-

链接方法

单链表



链接方法

- 对于经常增删结点的复杂数据结构，顺序存储往往会遇到困难，链接方法结合new动态存储为这些复杂问题提供了解决方法
 - 缺陷：
 - 为了访问结点集 K 中某个结点，必须用该结点的存储指针
 - 当不知道结点指针时，为了在结点集 K 中寻找某个符合条件的结点，就要沿着链接结点的链索，一个个结点比较搜索。所需花费的时间是相当客观的
-

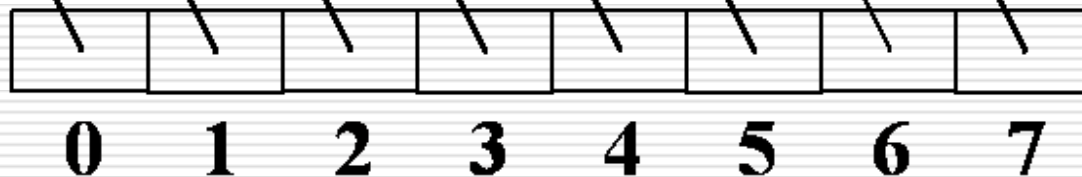
索引（indexing）方法

- 索引法是顺序存储法的一种推广，它也使用整数编码来访问数据结点位置
- 索引方法是要建造一个由整数域 Z 映射到存储地址域 D 的函数 $Y: Z \rightarrow D$ ，把结点的整数索引值 $z \in Z$ 映射到结点的存储地址 $d \in D$ 。Y称为索引函数
 - 一般而言索引函数并不象数组那样，是简单的线性函数
 - 当数据结点长度不等的情况下，索引函数就无法用线性表达式给出

索引方法

数据结点的存储区域

索引表



索引方法

- 为了构造任意的索引函数，可以为索引函数提供附加的存储空间，称为索引表S
- 索引表中每一元素是指向数据结点的指针。因为索引表 S 由等长元素（指针）组成，所以可以进行线性的索引计算：

始址(元素S[i])

$$= \text{始址(元素S[0])} + i \bullet (\text{指针尺寸})$$

通过上述公式，由索引号 I 可以计算出索引表中的单元S[i]

的始址，再通过读出S[i]元素的内容（是指针），访问真正需要访问的数据结点

索引方法

- 索引方法也付出了存储开销，其数据结点要附加用于指针的存储空间
 - 索引方法在程序设计中是一种经常使用的方法，其主要原因是对于非顺序的存储结构来说，使用索引表是快速地由整数索引值找到其对应数据结点的**唯一**方法
-

散列（hashing）方法

- ❑ 散列方法是索引方法的一种延伸和扩展
 - ❑ 利用一种称为散列函数(hash functions)的机制来进行索引值的计算，然后通过索引表求出结点的指针地址
 - ❑ 散列函数是将字符串 s 映射到非负整数 z 的一类函数 $h: S \rightarrow Z$,
对任意的 $s \in S$ ，散列函数 $h(s)=z$ ， $z \in Z$
-

散列方法

- 散列函数 $h(s)$ 除了它取非负整数值外，关键的问题包括：
 - 如何恰当地选择散列函数
 - 如何建造散列表
 - 如何在构建散列表时解决“碰撞”问题
-

抽象数据类型 (abstract data structure)

- 抽象数据类型是描述数据结构的一种理论工具
 - 特点是把数据结构作为独立于应用程序的一种抽象代数结构来描述，以在更大程度上帮助人们独立于程序的实现细节来理解数据结构的性质和约束条件
- 抽象数据类型不同于具体的数据结构，前者所描述的是一种模板以及模板的结构和性质。而模板的类型参数 T （元素的数据类型）必须用具体的数据类型所代入，才能成为具体的数据类型

抽象数据类型 (ADT)

□ 抽象数据结构 Λ 由

 <取值空间, 访问方式, 运算集>

三个不可分割的部分组成:

- Λ 的取值空间
 - Λ 的访问方式
 - Λ 的运算集
-

抽象数据类型： Λ 的取值空间

- Λ 的取值空间大小取决于它的元素类型 T 。
具体的数据结构是 Λ 和 T 联合确定的
- 对于取值空间，进一步说明它的大小是否可以动态改变是很重要的

抽象数据类型： Λ 的访问方式

- 在程序中读写数据结构及其组成部分的访问，一般是使用变量名访问，或使用组合几个变量名的方式
 - Λ 的访问方式可以归并到下面的 Λ 的运算集讨论，因为使用变量名来访问也可以看作为一种基本的运算
-

抽象数据类型： Λ 的运算集

- 一般使用函数来定义运算，参与运算的对象成为函数的输入参数，运算结果是函数值
 - 为适合不同应用领域以及它的元素类型 T 的不同，同一个结构的含义会有所不同，也就需要适当地添加一些新的运算种类
 - 运算集的选择也会受到该数据结构的存储方案的影响
-

ADT示例：向量vector容器

```
Template <class ELEM>
```

```
class vector {
```

```
// 向量 类模板vector，模板参数ELEM。当程序使用此vector模
```

```
// 板时，应该在前面附加 #include <vector>
```

```
    // 1. 向量的取值类型：
```

```
    // 它的元素类型为ELEM，k个元素的顺序存储成为一个向量
```

```
    // vector<double> v(100); 用于声明元素为浮点数的一个向量
```

```
    // v，程序中创建一个向量就是用这种格式的语句。其中
```

```
    // double替换了模板参数ELEM，向量长度k确定为100;
```

```
    // 如果创建整型向量，那就用<int>替换模板参数
```


ADT示例：向量vector容器

// 2. 用名字访问，即使用变量访问向量及其元素的方法

// 用整型变量 i 表示向量的下标，用 $v[i]$ 访问向量的第 i 个元素，

// 用语句 $v = w$ ；可以将同型的向量 w 拷贝到向量 v

// 用相等算子 $v == w$ 可以比较两个向量的值是否全同

// 3. 私有变量和 运算集：

private:

// 私有元素

ELEM *V[];

// 私有变量，存储向量元素

int maxLength;

// 向量最大长度

int currenSize;

// 向量当前长度

ADT示例：向量vector容器

```
public:          // 公有成员
    // 构造函数有三个：
    vector<ELEM> v;           // 创建一个长度为0的向量
    vector<ELEM> v(k);        // 创建一个长度为k的向量
    vector<ELEM> v(k,value);  // 创建一个长度为k的向量
                                // 其元素的初值全部为value
    int size();              // 函数v.size()将返回整数，向量的当前长度
    // 还可以定义其他类似size的函数，如两个向量的加法、乘法等
}
```

ADT示例：二维矩阵容器matrix

```
template<class ELEM>
```

```
class matrix          // 二维矩阵类模板matrix ，模板参数ELEM
```

```
// 当程序使用此matrix模板时，应该在前面附加 #include
```

```
// <vector> 以及 #include <matrix>
```

```
{
```

```
    // 1. 二维矩阵的取值类型：
```

```
    // 它的k个元素组成行向量，然后用k个行向量组成二维矩阵
```

```
    // 用matrix <double> M(100);声明一个浮点的k×k的矩阵
```

```
    // M. 程序中创建一个二维矩阵就可以用这样的语句其中double
```

```
    // 替换了模板参数ELEM，矩阵的各维的长度k = 100；
```

ADT示例：二维矩阵容器matrix

// 2. 用名字访问，即使用变量访问矩阵及其元素的方法

// 用整型变量 i 表示下标，用 M[i] 访问矩阵的第 i 行，

// 用 语句 M[i] = M[j]；可以将矩阵的第 j 行拷贝到第 i 行

// 3. 私有变量和 运算集：

typedef vector<ELEM> Mrow;

typedef vector< Mrow> Mmatrix;

private:

Mmatrix amatrix; //私有变量，存储一个二维矩阵

int maxLength; //存储各维的尺度

ADT示例：二维矩阵容器matrix

public:

// 创建函数有三个：

Mmatrix <ELEM> m; // 创建一个二维矩阵，尺寸待定

Mmatrix <ELEM> m(k); // 创建一个 $k \times k$ 的二维矩阵

// 创建 $k \times k$ 的二维矩阵

Mmatrix <ELEM> m(k,Mrow(k,value);

// 其元素的初值全部为value

int v.size() // 返回函数值为矩阵的当前长度

// 可以补充定义其他函数，如两个矩阵的加法、乘法等

}

1.4 算法及其特性

- 算法(**algorithms**)是为了求解问题而给出的指令序列
 - 程序是算法的一种实现，计算机按照程序逐步执行算法，实现对问题的求解
 - 一个**求解问题**通常用该问题的输入数据类型和该问题所要求解的结果（算法的输出数据）所应遵循的性质来描述
-

算法及其特性

□ 算法应该具有如下性质：

■ 一定的通用性

- 对于那些符合输入类型的任意输入数据，都能根据算法进行问题求解，并保证计算结果的正确性

■ 算法的有效性

- 算法是有限条指令组成的指令序列，其中每一条指令都必须是被能够被确切执行的，被人或机器所执行。指令的类型应该明确规定，仅限于若干明确无误的指令动作，是一个有限的指令集。其结果应具有确定的数据类型，是能够预期的

算法及其特性

■ 算法的确定性

- 算法每执行一步之后，关于它的下一步，应该有明确的指示。下一步动作可以是条件判断、分支指令、或顺序执行一条指令、或者指示整个算法的结束等。算法的确定性要保证每一步之后都有关于下一步动作的指令，不能缺乏下一步指令（被锁住）或仅仅含有模糊不清的指令

■ 算法的有穷性

- 算法的执行必须在有限步内结束。换句话说，算法不能含有死循环
-

计算复杂性和算法的效率

- 根据计算理论(theory of computation), 存在着
一类问题虽然能够被准确定义, 但却不存在能够解
决该问题的算法。称为不可解问题
 - 计算复杂性理论(computational complexity
theory)指出, 理论上存在一大类难解问题, 它们
虽然存在着求解算法, 但是在算法的计算时间上,
都是组合爆炸型的求解算法
-

计算复杂性和算法的效率

- 所谓组合爆炸型，是指随着问题的规模 n 的增大，算法的时间开销不能约束在 n 的 k 阶多项式数量范围内。
(其中 k 是任意不依赖于 n 的常数)
 - 根据计算复杂性理论，**难解问题**的定义就是它的求解算法均无法在多项式时间 n^k 数量级内解决（其中 k 是任意正整数）。比较常见的难解问题有：图论中的求最优巡游路径问题，判定命题逻辑公式是否为恒真等
-

1.5 算法的执行效率及其度量

- 解决同一个问题总是存在着多种算法，而算法设计者在所花费的时间和所使用的空间资源往往要两者之间采取折中，采用某种以空间资源换取时间资源的策略
 - 算法设计者可以通过算法分析，判断所提出的算法是否现实，设计出更好的算法
-

算法的渐进分析(asymptotic analysis)

- 算法的渐进分析，简称**算法分析**。算法在计算机上实际执行时，需要消耗时间资源（归结为**CPU**执行指令的总数），和使用空间资源（归结为所需占用的存储单元数量，字节数）
 - 由于算法分析和它所求解的问题规模直接有关，因此通常将**问题规模 n** 作为一个参照量，求算法的时空开销和 n 的关系
-

算法的渐进分析

- 算法的渐进分析就是要估计，当数据规模 n 逐步增大时，资源开销 $T(n)$ 的增长趋势
 - 从数量级大小的比较来考虑，当 n 增大到一定值以后，资源开销的计算公式中影响最大的就是 n 的幂次最高的项，其他的常数项和低幂次项都是可以忽略的
-

算法的渐进分析

- 算法的渐进分析就是要得到一个大O渐进表达式，简写为：

$$\text{rate}_{n \rightarrow \infty} T(n) = O(F(n))$$

其中O是数学分析常用符号‘大O’，而 $F(n)$ 是自变量为 n 的某个具体的整函数表达式，例如

$$F(n) = n^2$$

大O表示法的运算规则

□ 单位时间

- 简单布尔或算术运算
- 简单**I/O**
- 函数返回

□ 加法规则： $f_1(n) + f_2(n) = O(\max(f_1(n), f_2(n)))$

- 顺序结构，**If**结构，**switch**结构

□ 乘法规则： $f_1(n) \cdot f_2(n) = O(f_1(n) \cdot f_2(n))$

- **for, while, do-while**结构
-

算法的渐进分析示例：

矩阵求和（matrix_addition）

```
void matrix_addition(double **M1, double **M2, int k) {  
    // 矩阵M1,M2求和，即两两元素求和，结果存回M1  
    // k×k是矩阵的规模  
    for (int i=0 ; i<k; i++)  
        for (int j=0 ; j<k; j++) //对应位置的矩阵元素相加  
            M1[i][j] = M1[i][j] +M2[i][j] ;  
}
```

算法的渐进分析示例：

矩阵求和（`matrix_addition`）

- 此问题的规模由输入数据规模决定，主要是浮点矩阵 $M1, M2$ 的尺寸 $n=2*k^2$
 - 在渐进分析时，通常我们把浮点运算和数组元素的读写操作的执行时间都看作在同一个数量级内，忽略其执行时间的细微差别
 - 矩阵加法的时间开销是和问题规模 n 成正比的，或称该算法（的时间开销）是线性增长率的。简写为 rate
 $\lim_{n \rightarrow \infty} T(n) = O(n)$ ，即 $F(n)=n$
-

算法的渐进分析

- 用于渐进分析的常见的 $F(n)$ 还有以下若干种：
 - $F(n) = 1$, 常数函数, 不依赖于 n
 - $F(n) = \log n$, 对数函数, 它比线性函数 n 增长慢
 - $F(n) = n$, 线性增长, 随着问题规模 n 而增长
例如, $\text{rate } n \rightarrow \infty$ (n 个 a 相加)
 $\quad = \text{rate } n \rightarrow \infty \quad (n \bullet a)$
 $\quad = O(n)$
 - $F(n) = n^2$, 二阶增长
-

算法的渐进分析

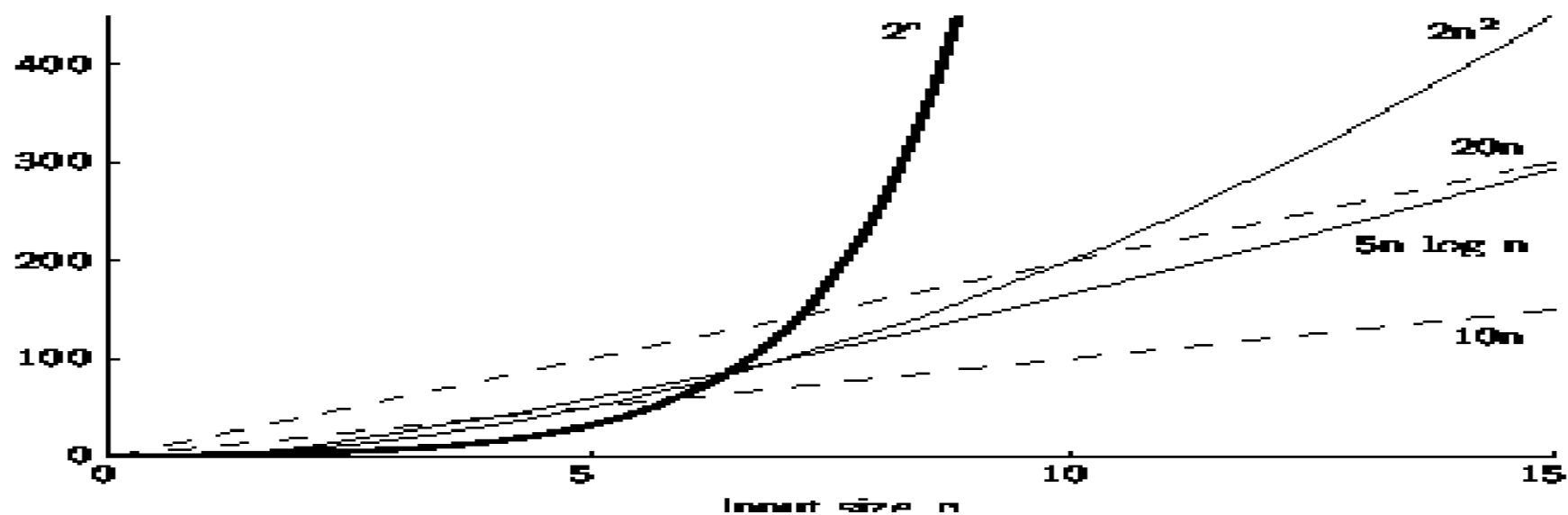
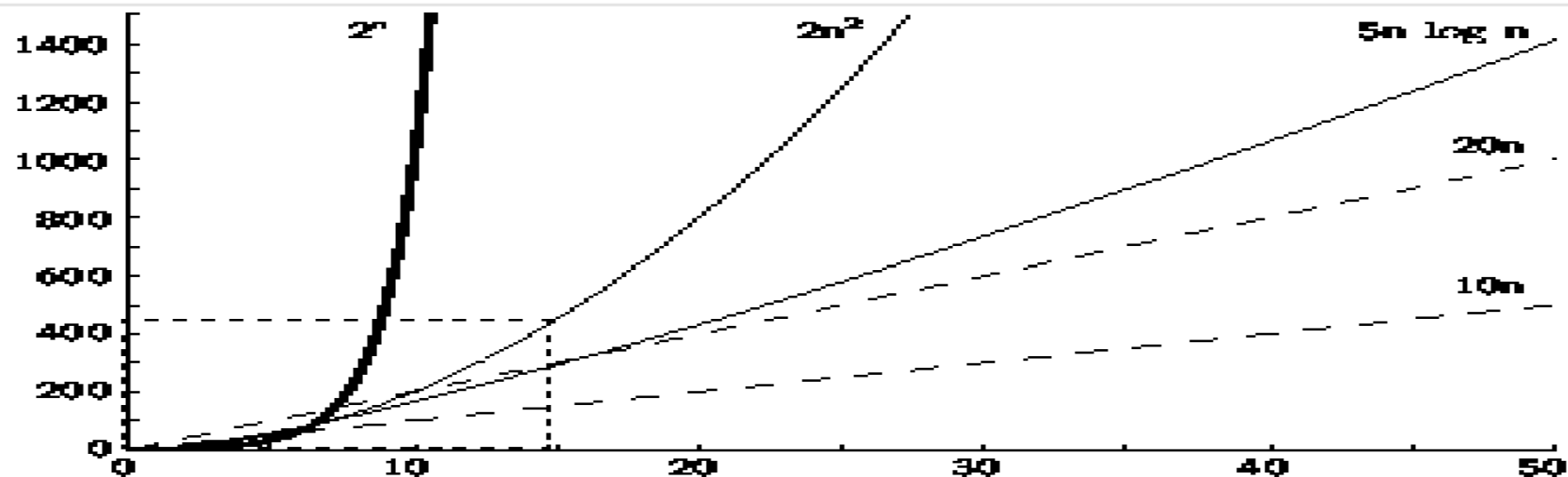
□ 用于渐进分析的常见的 $F(n)$ 还有以下若干种：

- $F(n) = n \cdot \log(n)$ ，其增长率的阶数低于二阶，但高于一阶线性。例如，

$$\text{rate } n \rightarrow \infty \quad \sum_{i=1}^{\log n} \sum_{j=i}^n (a) = O(n \cdot \log(n))$$

- $F(n) = a^n$ 指数增长，随问题规模 n 而增长极快
这种指数型的渐进函数往往由递归定义的函数产生。例如函数 $H(n)$ ，令 $H(1)=1$ ，且 $H(n)=2 \cdot H(n-1)$ ，那么通过递推可以计算出 $H(n) = 2^{(n-1)}$

增长率函数曲线



Orders of Magnitude in Speed

Meters per Sec.	English Equivalence	Example
10^{-11}	1.2 in./century	Stalactites growing
10^{-10}	1.2 in./decade	Slow continent drifting
10^{-9}	1.2 in./year	Fingernails growing
10^{-8}	1 ft/year	Hair growing
10^{-7}	1 ft /month	Weeds growing
10^{-6}	3.4in./day	Glacier
10^{-5}	1.4 in/hour	Minute hand of a watch
10^{-4}	1.2 ft/hour	Gastro-intestinal tract
10^{-3}	2 in./min	Snail
10^{-2}	2 ft/min	Ant
10^{-1}	20 ft./min	Giant tortoise
1	2.2 mi/hour	Human walk
10^1	22 mi/hour	Human sprint
10^2	220 mi /hour	Propeller airplane
10^3	37 mi /min	Fastest jet plane
10^4	370 mi/min	Space shuttle
10^5	3700 mi/min	Meteor impacting earth
10^6	620 mi/sec	Earth in galactic orbit
10^7	6200 mi/sec	LA to satellite to NY
10^8	62,000 mi /sec	One-third speed of light

最坏、最好、和平均情况

- ❑ 在具体进行算法增长率估计时，往往会由于算法中的条件分支而遇到困难
 - ❑ 由于算法实际执行的操作往往依赖于分支条件的走向，而输入数据的取值又影响这些分支走向，因此很多算法都无法得出独立于输入数据的渐进估计
 - ❑ 针对这一情况，提出了最坏情况估计、平均情况估计、和Theta (Θ 希塔) 估计等三种方法
-

示例：求矩阵M中绝对值最大的元素

```
double abs_biggest(double **M, int k) {  
    // 求矩阵M中绝对值最大的元素， $k \times k$ 是矩阵的规模  
    double current_big = 0;    //临时存储单元，初始化为0  
    for (int i=0 ; i<k; i++)  
        for (int j=0 ; j<k; j++) {                // 对每一个矩阵元素进行大小比较  
            // fabs是浮点数求绝对值函数  
            double temp = fabs(M[i][j]);  
            if ( temp > current_big)  
                current_big = temp;                // current_big存储当前最大者  
        }  
    return current_big; //函数值返回  
}
```

示例：求矩阵M中绝对值最大的元素

- ❑ 输入参数为浮点矩阵M，它的大小是影响算法时空开销的主要因素。令 $n=k \times k$ 为这个问题的规模
 - ❑ 在最坏的情况下，每一次判断(`temp`) > `current_big`) 都为真，赋值语句执行n次
 - ❑ 在最好的情况下（在第一次对`current_big`赋值后，程序再没有对它作赋值操作），则基本没有时间开销
 - ❑ 平均而言，在假定浮点矩阵的主元素（绝对值最大）位置随机分布的情况下，可以说，其平均时间开销正比于 $n/2$
-

最坏、最好、和平均情况

- 对于时间开销，一般不注意算法的“最好估计”。特别是处理应急事件，计算机系统必须在规定的响应时间内做完紧急事件处理。这时，**最坏估计**是唯一的选择
 - 对于多数算法而言，最坏情况和平均情况估计两者，它们的时间开销的公式虽然不同，但是往往只是常数因子大小的区别，或者常数项的大小区别。因此不会影响渐进分析的增长率函数估计
-

时间和空间资源开销

- 对于空间开销，也可以实行类似的渐进分析方法
 - 很多算法使用的数据结构是静态的存储结构，即存储空间在算法执行过程中并不发生变化
 - 对于静态数据结构，空间开销的估计往往是容易的，它们或者与所涉及的问题规模成正比（空间开销为线性增长），或者不随问题的规模而增大，空间开销为常数
-

时间和空间资源开销

- 使用动态数据结构算法的存储空间是变化的，在算法运行过程中有时会有数量级地增大或缩小。对于这种情况，空间开销的分析和估计是十分必要的

时空资源的折中原理

- 对于同一个问题求解，一般会存在多种算法。而这些算法在时空开销上的优劣往往表现出“**时空折中**”（**trade-off**）的性质
 - “时空折中”，是指为了改善一个算法的时间开销，往往可以通过增大空间开销为代价，而设计出一个新算法来
 - 有时也可以为了缩小算法的空间开销，而牺牲计算机的运行时间，通过增大时间开销来换取存储空间的节省
-

大 Θ 表示法及其分析规则

- 大 Θ （希塔）表示法是算法分析的一种渐进估计，简称 Θ 希塔估计。它是前面讨论的大 O 表示的扩展
- 对于任意的资源开销函数 $T(n)$ ，以及对 $T(n)$ 的一个渐进估计式 $F(n)$ ，称 $F(n)$ 为 $T(n)$ 的一个渐进 Θ 希塔估计，当且仅当满足如下数学性质：

存在正常数 C_1, C_2 ，以及正整数 n_0 ，使得对于任意的正整数 $n > n_0$ 有下列两不等式同时成立：

$$|T(n)| > C_1 \cdot F(n) \text{ 和 } |T(n)| < C_2 \cdot F(n)$$

大 Θ 表示法示例

假定 $T(n) = 100 \cdot n^2 + 5 \cdot n + 500$,

令 $F(n) = n^2$, 这时存在正常数

$C_1 = 100$, $C_2 = 105$, 以及 $n_0 = 10$

使得当 $n > n_0$ 时,

$$|T(n)| > C_1 \cdot F(n) \text{ 和 } |T(n)| < C_2 \cdot F(n)$$

同时成立

上述不等式可以用算术计算验证, 由此说明 $T(n)$ 的希塔估计式等于 $F(n) = n^2$

大 Θ 表示法及其分析规则

- 对于算法的资源开销，可以给出它的下界估计和上界估计两种估计式
 - 如果一个算法存在一个 $F(n)$ 渐进估计式，它可以同时充当它的上界和下界的渐进估计，那么就把这个 $F(n)$ 称作算法开销 $T(n)$ 的希塔估计，或称 Θ 估计
-

1.6 数据结构的选择和评价

- ❑ 仔细分析所要解决的问题，特别是求解问题所涉及的数据类型和数据间逻辑关系
 - ❑ 数据结构的初步设计往往在算法设计之先
 - ❑ 注意数据结构的可扩展性。包括考虑当输入数据的规模发生改变时，数据结构是否能够适应。同时，数据结构应该适应求解问题的演变和扩展
 - ❑ 数据结构的设计和选择也要比较算法的时空开销的优劣
-

算法的度量

□ 算法设计的两个目标:

- 易读、易编码和调试（软件工程）
- 充分利用计算机资源（算法和数据结构）

□ 评估方法

- 实验（运行程序）
 - 渐进分析
 - 关键资源
 - 影响运行时间的因素
 - 往往与问题的输入规模 n 有关
-

时间/空间权衡

□ 数据结构

- 一定的空间来存储它的每一个数据项
- 一定的时间来执行单个基本操作

□ 代价和效益

- 空间和时间的限制
 - 程序设计工作量
-

小结

- 数据结构的地位与重要意义
 - 数据结构的主要内容
 - 抽象数据类型的概念
 - 算法及其特点
 - 算法的有效性度量
 - 数据结构的选择
-