

数据结构与算法

计算机学院

朱晨阳 副教授

zhuchenyang07@nudt.edu.cn



教学团队

- 主讲教师：朱晨阳，计算机学院副教授
 - zhuchenyang07@nudt.edu.cn
 - 15580096235
- 教辅：张斯顿、于丰源
- 帮助大家解决上机和课后作业中的疑问
- Educoder课堂：邀请码 45WS8
 - <https://www.educoder.net/>



课程内容安排

- 第一章：数据结构概述
- 第二章：向量、栈和队列
- 第三章：链表
- 第四章：串
- 第五章：排序
- 第六章：查找
- 第七章：树、二叉树
- 第八章：树形结构的应用
- 第九章：图
- 第十章：算法



教材与参考书

- **教材：**

- 熊岳山，数据结构与算法（C语言描述，第2版），清华大学出版社，2016

- **参考书：**

- 严蔚敏, 吴伟民，数据结构（C语言版），清华大学出版社，2008
- Mark Weiss著，冯舜玺译，数据结构与算法分析：C语言描述（原书第2版），机械工业出版社，2014





成绩评定

- 期末笔试成绩：70%
- 课堂表现：5%
- 平时作业：10%
 - 平时作业通过Trustie平台进行提交
 - 拍照后上传
- 上机实训作业：15%
 - 分为代码验收以及实验报告两个部分



课程推荐软件

- 课程实训作业使用语言：C语言（验收编译器为gcc）
 - 不建议使用非教学提供的第三方函数库
 - 学习如何用Debug模式调试程序
- 推荐使用IDE：VSCode





课程概述

前言

基本概念

教学目标

前言

- 数据结构与算法这门课，和程序设计有什么不同？
- 为什么要学数据结构与算法？



工具

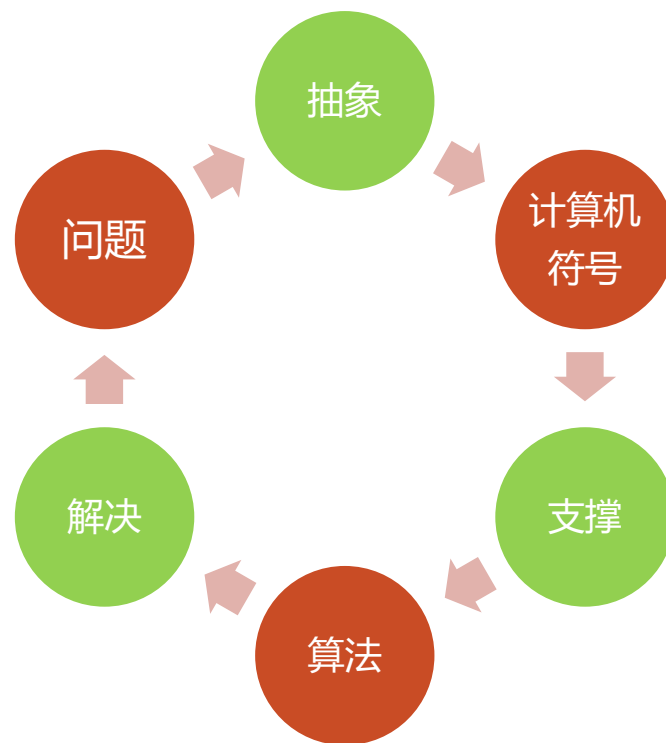


更高效地使用这些工具



基本概念

- 计算机程序和现实问题之间的关联往往并不直观。
- 程序=算法+数据
 - 数据结构是算法与计算机之间的桥梁
 - 算法实现依赖具体的数据结构





算法

- 例：计算10的平方根

初始点

10

操作序列

预期结果

3.16...

方法1（保留2位小数）

$$\because 0.01^2 \leq 10 \not\leq (0.01 + 0.01)^2$$

$$\therefore \sqrt{10} \neq 0.01 \dots$$

$$\because 0.02^2 \leq 10 \not\leq (0.02 + 0.01)^2$$

$$\therefore \sqrt{10} \neq 0.02 \dots$$

$$\because 0.03^2 \leq 10 \not\leq (0.03 + 0.01)^2$$

$$\therefore \sqrt{10} \neq 0.03 \dots$$

.....

$$\because 3.16^2 \leq 10 \leq (3.16 + 0.01)^2$$

$$\therefore \sqrt{10} = 3.16 \dots$$



算法

- 例：计算10的平方根

初始点

10

操作序列

预期结果

3.16...

方法2

$$\because 3^2 \leq 10 \leq (3 + 1)^2$$

$$\therefore \sqrt{10} = 3. \dots$$

$$\because 3.1^2 \leq 10 \leq (3.1 + 0.1)^2$$

$$\therefore \sqrt{10} = 3.1 \dots$$

$$\because 3.16^2 \leq 10 \leq (3.16 + 0.01)^2$$

$$\therefore \sqrt{10} = 3.16 \dots$$

.....

(直到满足精度要求)



算法

- 例：计算10的平方根

初始点

10

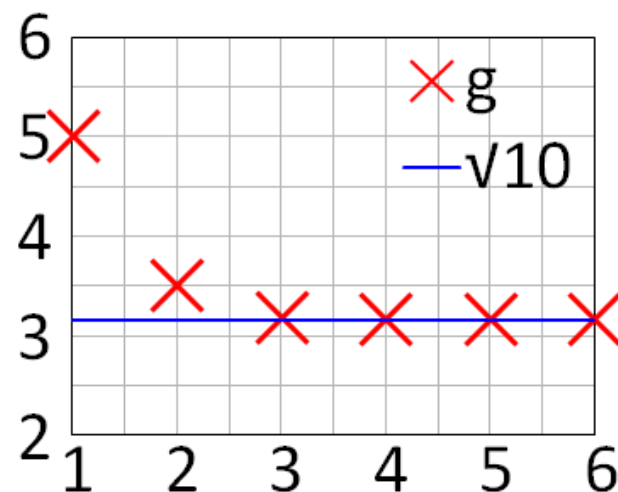
算法

预期结果

3.16...

方法3

- ①对正数 x ，猜测其平方根为 g ；
- ②若 g^2 足够接近 x ，计算完成；
- ③否则，将 $(g+x/g)/2$ 作为新的猜测值，仍记为 g ；
- ④重复步骤②和③。





数据结构与算法

- 实际问题并不完全都是数学题
- 如何建立算法与计算机之间桥梁?
- 形式化的数据结构

基本概念

- 数据是客观事物的**符号表示**，是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**进行描述的符号的集合。

表 1-1 学生情况表

[illegible]



基本概念

- 数据是客观事物的**符号表示**，是对现实世界的事务采用**计算机能够识别、存储和处理的形式**进行描述的符号的集合。

```
struct Information{  
    int id;  
    char* name;  
    char* gender;  
    int age;  
    char* hometown;  
    int class_id;  
    Grade grade;  
};
```

```
struct Grade{  
    int math;  
    int english;  
    int physic;  
    int chemistry;  
};
```



教学目标

素质：从具体到抽象，再从抽象到具体的计算机科学素养（计算思维）

能力：针对实际问题选择合适的数据结构、存储结构以及相应算法的能力，程序设计与编程实现的能力

知识：计算机数据表示、存储以及处理的过程，常用算法的设计方法



数据结构定义

- 数据项
 - 初等项，是数据的不可分割的最小单位，
 - 组合项，由若干数据项组成。
- 数据元素
 - 数据的基本单位，一般由多个数据项组成。
- 数据对象
 - 性质相同的数据元素的集合。



数据结构定义

数据项

表 1-1 学生情况表

学号	姓名	性别	年龄	籍贯	班别
6001	张三	女	19	北京	15
6002	李四	男	20	上海	15
6003	王五	男	18	湖南	15
⋮	⋮	⋮	⋮	⋮	⋮

初等项

数据项

成绩			
数学	物理	化学	外语
82	85	90	92
90	92	91	95
93	95	91	94
⋮	⋮	⋮	⋮

组合项



数据结构定义

表 1-1 学生情况表

学号	姓名	性别	年龄	籍贯	班别	成绩			
						数学	物理	化学	外语
6001	张三	女	19	北京	15	82	85	90	92
6002	李四	男	20	上海	15	90	92	91	95
6003	王五	男	18	湖南	15	93	95	91	94
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

数据元素



数据结构定义

表 1-1 学生情况表

学号	姓名	性别	年龄	籍贯	班别	成绩			
						数学	物理	化学	外语
6001	张三	女	19	北京	15	82	85	90	92
6002	李四	男	20	上海	15	90	92	91	95
6003	王五	男	18	湖南	15	93	95	91	94
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

数据对象



数据结构定义

- 什么是数据结构?
- 数据的结构由数据对象以及该对象中数据元素之间的(逻辑)关系组成。一个数据结构可以描述为一个二元组:

$$\text{Data Structure} = (D, R)$$

- **D**是数据对象
- **R**是该数据对象中数据元素之间关系的集合, 描述的是数据的逻辑结构。



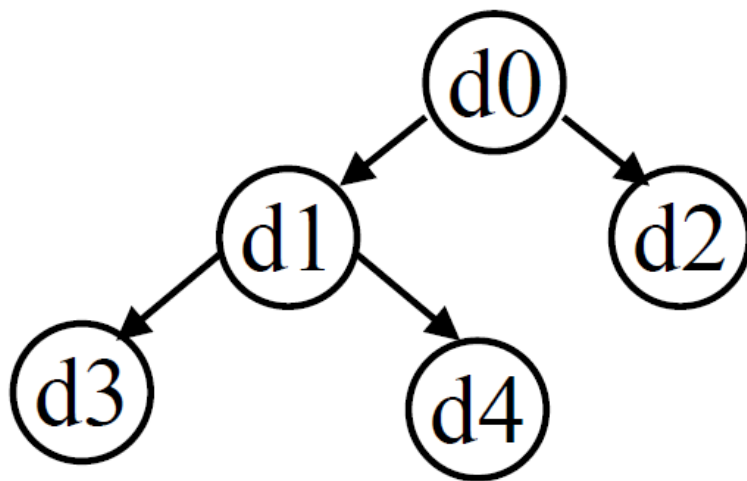
数据间的逻辑关系

- 用 $\langle d_i, d_j \rangle$ 表示数据元素 d_i 和 d_j 的先后关系, 称 d_i 为 d_j 的前驱, d_j 为 d_i 的后继, (D, R) 中的 R 即为类似 $\langle d_i, d_j \rangle$ 的二元关系的集合。
- 例1, Data Structure= (D, R)
 - $D=\{d_0, d_1, d_2, d_3, d_4\}$
 - $R=\{\langle d_0, d_1 \rangle, \langle d_1, d_2 \rangle, \langle d_2, d_3 \rangle, \langle d_3, d_4 \rangle\}$
- 例2, Data Structure= (D, R)
 - $D=\{d_0, d_1, d_2, d_3, d_4\}$
 - $R=\{\langle d_0, d_1 \rangle, \langle d_0, d_2 \rangle, \langle d_1, d_3 \rangle, \langle d_1, d_4 \rangle\}$



数据间的逻辑关系

- 数据的逻辑结构图：用结点表示数据元素，结点间的连线表示数据元素间的关系





数据结构定义

- 数据结构一般包含三方面的内容
 - 数据的逻辑结构
 - 数据的存储结构(或称物理结构)
 - 数据的运算

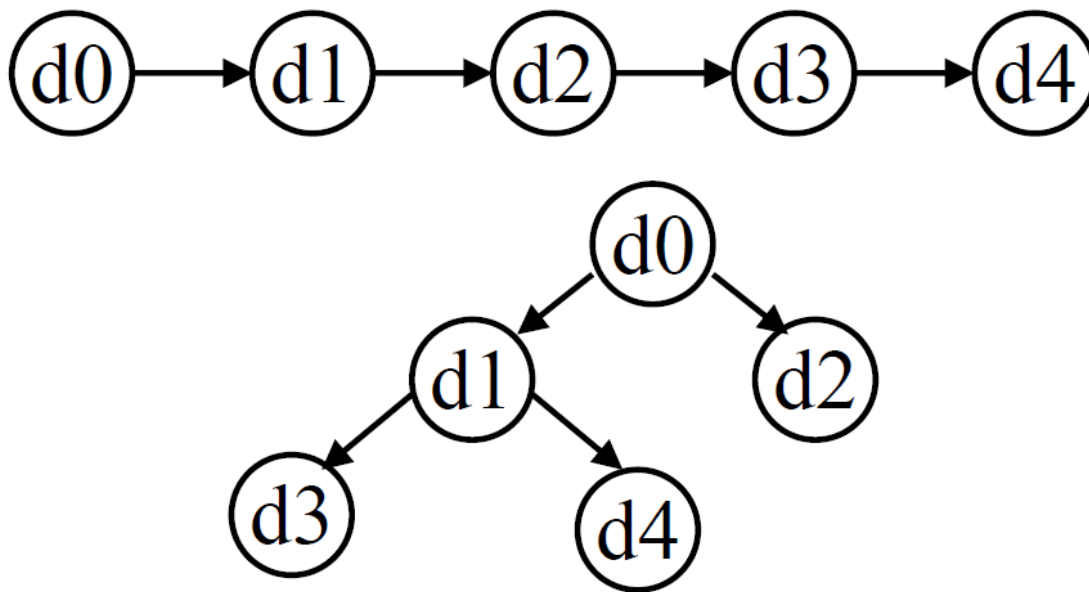


数据结构定义

- 数据的逻辑结构描述数据元素之间的逻辑关系，是面向问题的，是独立于计算机的。
- 数据的存储结构是数据对象和逻辑结构在计算机中的物理表示方式，根据处理速度和存储空间而建立
- 每种逻辑结构都有一个运算的集合，如检索、插入、排序等

逻辑结构的分类

- 数据的逻辑结构可分为两大类
 - 线性结构
 - 非线性结构



线性逻辑结构

- 线性结构只有一个开始结点和一个端结点，且所有结点最多只有一个前驱和一个后继。
 - 线性表是典型的线性结构。
 - 学生情况表是一个线性表。

表 1-1 学生情况表

[illegible]



逻辑结构优缺点

- 讨论：线性逻辑结构 vs. 非线性逻辑结构
- 优缺点各有什么？



存储结构

- 存储结构存储什么？
 - 存储数据对象，并体现逻辑关系
 - 存储数据结构(数据对象+逻辑关系)
- 怎么储存？
 - 线性逻辑结构？
 - 非线性逻辑结构？

存储结构

- 数据的存储结构包括四种基本的存储方法
 - 顺序存储
 - 链接存储
 - 索引存储
 - 散列存储

顺序存储

1001	k_1
1002	k_2
1003	k_3
1004	k_4
1005	k_5

(a)

链接存储

1001	k_2	1007
1002		
1003	k_1	1001
1004	k_5	\wedge
1005	k_4	1004
1006		
1007	k_3	1005

(b)

索引存储

2001	key ₁	1001
2002	key ₂	1005
2003	key ₃	1003
2004	key ₄	1002
2005	key ₅	1004

k_1	1001
k_4	1002
k_3	1003
k_5	1004
k_2	1005

(c)

散列存储

d_2	k_2
d_5	k_5
	\vdots
d_1	k_1
d_4	k_4
d_3	k_3

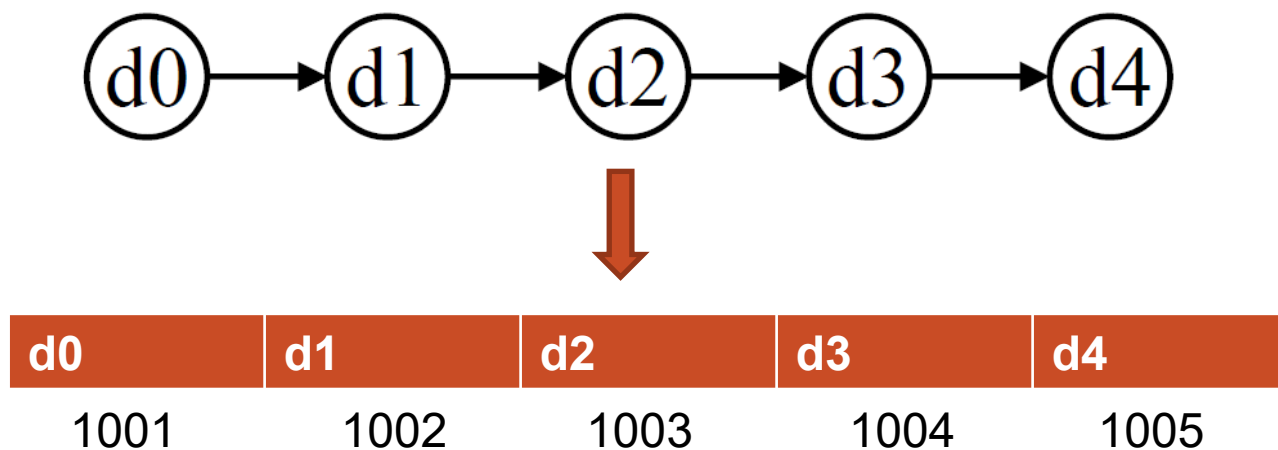
$d_i = h(\text{key}_i)$

(d)



顺序存储

- 顺序存储把逻辑上相邻的结点存储在物理位置相邻的存储单元。
- 结点间的逻辑关系用存储单元的邻接关系来体现。



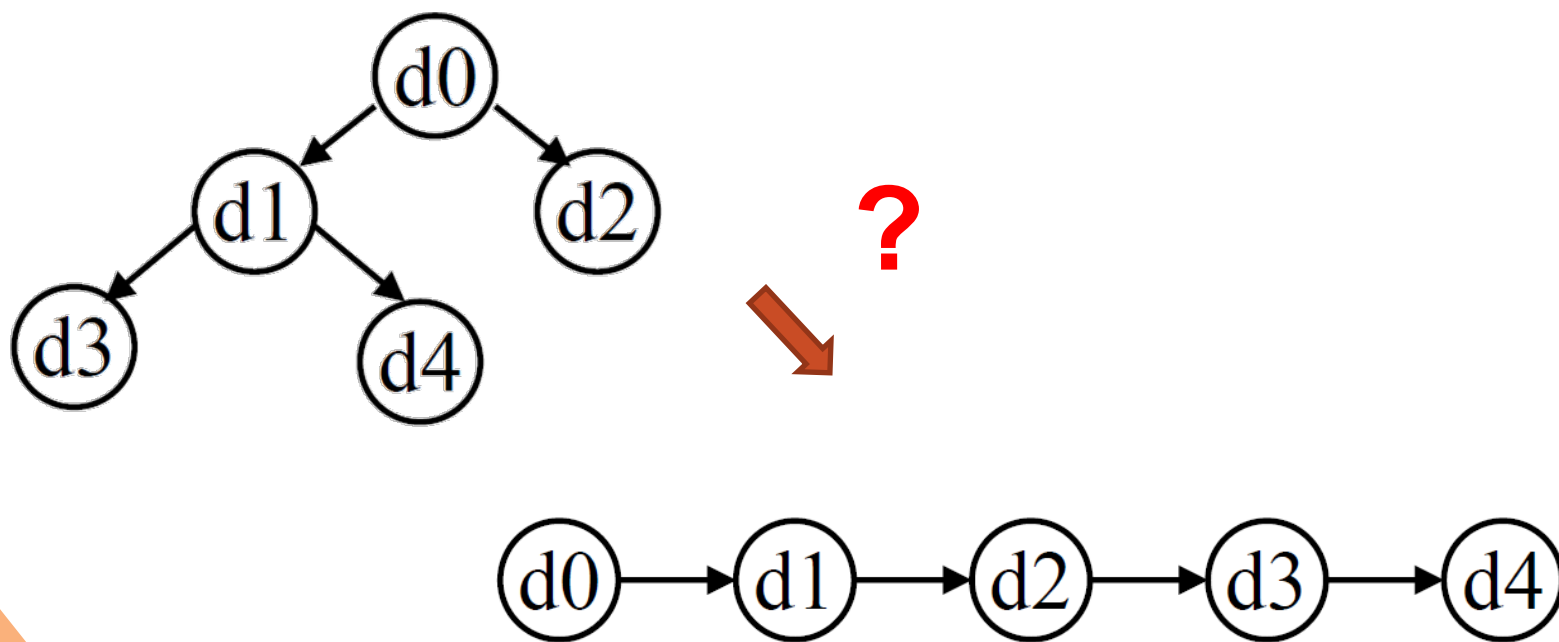


顺序存储

- 顺序存储可用什么C语言语句实现？
- 非线性逻辑结构可不可以用顺序存储？

顺序存储

- 非线性逻辑结构可不可以用顺序存储？
 - 非线性结构通过某种线性化方法也可实现顺序存储。

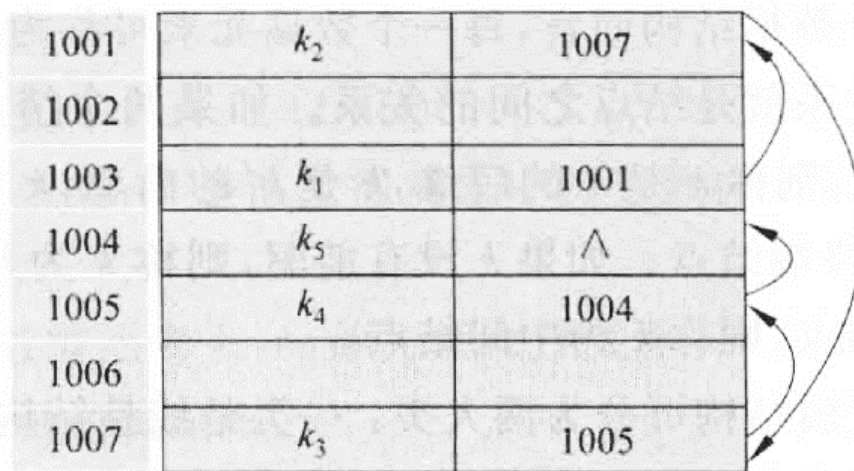




链接存储

- 链接存储对逻辑上相邻的结点不要求存储在相邻的物理位置上，结点间的逻辑关系由附加的指针表示。非线性结构常用链接存储，线性结构也可链接存储。链接存储通常用程序语言的指针来描述。

链接存储





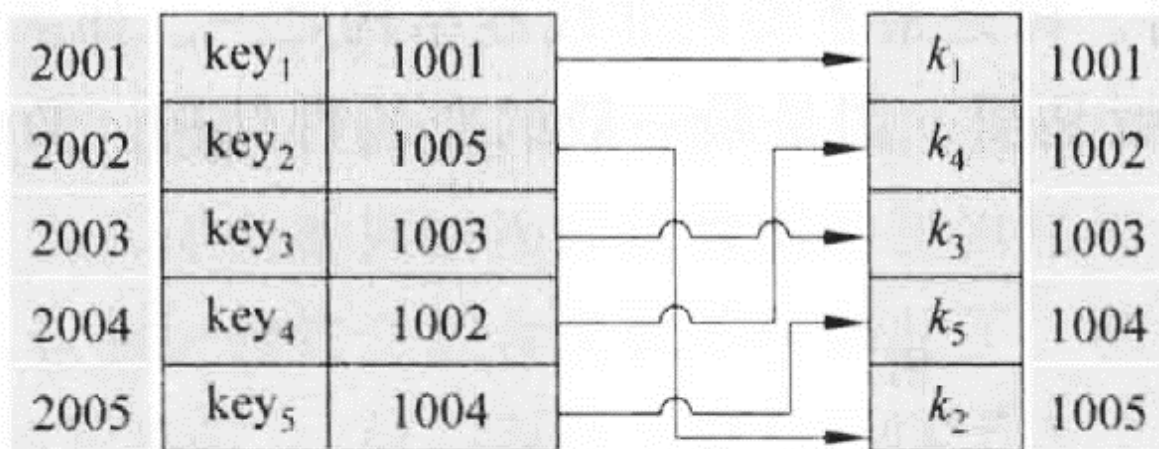
索引存储

- 索引存储方法在保存结点数据的同时，还建立附加的索引表。
- 索引表的每一项称为索引项，索引项一般由关键字和地址组成。
- 关键字是结点的一个字段和多个字段的组合，其值能唯一标识数据结构中的一个结点，地址指示该结点的存储位置。

索引存储

- 什么情况下适合采用索引存储？为什么需要索引表？

索引存储





散列存储

- 散列存储方法根据结点的关键字计算出该结点的存储地址，然后按该存储地址存放该关键字对应的数据元素。
- 数据元素 di 的存储地址为：
 - $ai = h(di.key)$
 - h 为散列函数
 - $di.key$ 为 di 的关键字

散列存储

d_2	k_2	$d_i = h(key_i)$
d_5	k_5	
	\vdots	
d_1	k_1	
d_4	k_4	
d_3	k_3	



散列存储

- 散列存储和索引存储有什么相同？
 - 都不能直接检索到数据
 - 都是通过key值来进行数据检索
- 散列存储和索引存储有什么不同？
 - 索引表是一一对应的
 - 散列函数的对应关系不是一一对应的



散列存储

- 散列存储的优势在哪里？
- 散列存储的局限性在哪里？



数据类型

- 高级程序设计语言中，数据结构是通过数据类型来描述的。
 - 数据类型是一组性质相同的值的集合以及定义在这个集合上的一组操作的总称。
- 数据类型用于刻画操作对象的特性，规定了可能取值的范围以及允许进行的操作。
 - **基本数据类型**，如整型，字符型
 - **组合数据类型**，由基本数据类型组合而成



基本数据类型

- C语言中有三种基本数据类型：
 - 整型数据
 - 浮点型数据
 - 字符型数据
- 我们可以拿这三种数据类型组合出更符合实际需求的组合型数据类型
 - 例如：学生信息



抽象数据类型

- 抽象是指从特定实例抽取共同的性质以形成一般化概念的过程。
- 数据抽象是一种对数据和操作数据的算法的抽象。
- 抽象数据类型(ADT)是指**抽象数据的组织和与之相关的操作**。可以看成是数据的**逻辑结构**以及在逻辑结构上**定义的操作**。



抽象数据类型

```
ADT ADT_Name
{
    Data:                                /* 数据说明 */
        数据元素之间逻辑关系的描述

    Operations:                          /* 操作说明 */
        Operation1
            Input:                        对输入数据的说明
            Preconditions:                执行操作前系统应满足的状态
            Process:                      对数据执行的操作
            Output:                      对返回数据的说明
            Postconditions:               执行操作后系统的状态

        Operation2
        :
} /* ADT */
```



抽象数据类型

- 练习：建立一个药店双黄连库存信息的ADT



抽象数据类型

• 练习：建立一个药店双黄连库存信息的ADT

ADT Shuanghuanglian

{

Data:

非负整型数, 表示双黄连药品的库存

Operations:

InitStock

Input: 初始的药品库存

Preconditions: 无

Process: 将库存数据设置为该函数输入的数据

Output: 无

Postconditions: 无

RefillStock

Input: 进货数量

Preconditions: 无

Process: 将库存数量设为原库存量加上进货数量

Output: 库存数量

Postconditions: 无

Sale

Input: 售出数量

Preconditions: 售出数量小于库存数量

Process: 将库存数量设为原库存量减去售出数量

Output: 库存数量

Postconditions: 无

}