

1.1 数据结构的基本概念

基本概念和术语

- 数据 数据是信息的载体, 是描述客观事物属性的数、字符及所有能输入到计算机中并被计算机程序识别和处理的符号的集合
- 数据元素 数据元素是数据的基本单位, 通常作为一个整体进行考虑和处理
- 数据对象 数据对象是具有相同性质的数据元素的集合, 是数据的一个子集
- 数据类型 数据类型是一个值的集合和定义在此集合上的一组操作的总称
 - 原子类型: 其值不可再分的数据类型
 - 分类
 - 结构类型: 其值可以再分解为若干成分(分量)的数据类型
 - 抽象数据类型: 抽象数据组织及与之相关的操作
- 数据结构 数据结构是相互之间存在一种或多种特定关系的数据元素的集合
 - 逻辑结构
 - 数据结构包括三方面
 - 存储结构
 - 数据的运算

概述

- 逻辑结构是指数据元素之间的逻辑关系, 即从逻辑关系上描述数据
- 与数据的存储无关, 是独立于计算机的

数据的逻辑结构

分类

- 线性表
 - 栈
 - 队列
 - 数组
- 非线性结构
 - 集合
 - 树
 - 图

概述

- 存储结构是指数据结构在计算机中的表示(又称映像), 也称物理结构
- 数据元素的表示和关系的表示
- 存储结构是用计算机语言实现的逻辑结构, 它依赖于计算机语言

数据的存储结构

分类

- 顺序存储
 - 把逻辑上相邻的元素存储在物理位置上也相邻的存储单元中, 元素之间的关系由存储单元的邻接关系来体现
 - 优点: 可以实现随机存取, 每个元素占用最少的存储空间
 - 缺点: 只能使用相邻的一整块存储单元, 因此可能产生较多的外部碎片
- 链式存储
 - 要求逻辑上相邻的元素在物理位置上也相邻, 借助指示元素存储地址的指针来表示元素之间的逻辑关系
 - 优点: 不会出现碎片现象, 能充分利用所有存储单元
 - 缺点: 每个元素因存储指针而占用额外的存储空间, 且只能实现顺序存取
- 索引存储
 - 在存储元素信息的同时, 还建立附加的索引表
 - 优点: 检索速度快
 - 缺点: 附加的索引表额外 占用存储空间
 - 增加和删除数据时也要修改索引表, 会花费较多的时间
- 散列存储
 - 根据元素的关键字直接计算出该元素的存储地址, 又称哈希 (Hash) 存储
 - 优点: 检索、增加和删除结点的操作都很快
 - 缺点: 若散列函数不好, 则可能出现元素存储单元的冲突, 而解决冲突会增加时间和空间开销

数据结构三要素

施加在数据上的运算包括运算的定义和实现。

数据的运算

- 运算的定义是针对逻辑结构的, 指出运算的功能
- 运算的实现是针对存储结构的, 指出运算的具体操作步骤

1.2算法和算法评价

算法的基本概念

算法是对特定问题求解步骤的一种描述，它是指令的有限序列，其中的每条指令表示一个或多个操作

重要特性

- 有穷性：一个算法必须执行有穷步之后结束，且每一步都可在有穷时间内完成
- 确定性：算法中每条指令必须有确切的含义，对于相同的输入只能得出相同的输出
- 可行性：算法中描述的操作都可以通过已经实现的基本运算执行有限次来实现
- 输入：一个算法有零个或多个输入，这些输入取自于某个特定的对象的集合
- 输出：一个算法有一个或多个输出，这些输出是与输入有着某种特定关系的量

优秀算法的标准

- 正确性：算法能够正确地解决求解问题。
- 可读性：算法应具有良好的可读性，以帮助人们理解
- 健壮性：输入非法数据时，算法能适当地做出反应或进行处理，而不会产生莫名其妙的输出结果
- 效率与低存储量需求：效率是指算法执行的时间，存储量需求是指算法执行过程中所需要的最大存储空间，这两者都与问题的规模有关

算法效率的度量

算法效率的度量是通过时间复杂度和空间复杂度来描述的

时间复杂度

- 一个语句的频度是指该语句在算法中被重复执行的次数
- 常见的渐近时间复杂度为

一般情况下，嵌套的循环次数是时间复杂度指数（存在特殊情况）

空间复杂度

- 算法的空间复杂度 $S(n)$ 定义为该算法所耗费的存储空间,它是问题规模 n 的函数。
- 一个程序在执行时除需要存储空间来存放本身所用的指令、常数、变量和输入数据外，还需要一些对数据进行操作的工作单元和存储一些为实现计算所需信息的辅助空间
- 算法原地工作是指算法所需的辅助空间为常量，即 $O(1)$

计算规则

- 加法规则： $O(f(n)) + O(g(n)) = O(\max(f(n), g(n)))$
- 乘法规则： $O(f(n)) \times O(g(n)) = O(f(n) \times g(n))$