

Lecture 1: 数据结构的基本概念

数据结构的存储结构

概念：顺序存储

- 定义
 - 把逻辑上相邻的元素存储在物理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体现。
- 特点
 - 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

概念：链式存储

- 定义
 - 不要求逻辑上相邻的元素在物理上也相邻，借助元素存储地址的指针来表示元素之间的逻辑关系。
- 特点
 - 优点
 - 不会产生碎片现象，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存储空间，且只能实现顺序存储；

概念：索引存储

- 定义
 - 在存储元素信息时，还建立附加索引表。
 - 索引表中的每项称为索引项，索引项的一般形式是(关键字，地址)
- 特点
 - 优点
 - 检索速度快
 - 缺点
 - 附加的索引表额外占用存储空间；
 - 另外，增加和删除数据时也要修改索引表，因而会花费较多的时间；

概念：散列存储

- 定义
 - 根据元素的关键字直接计算出该元素的存储地址，又称**哈希表(Hash)** 存储；
- 优点
 - 检索，增加和删除结点的操作很快

- 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

其他

面试题：用循环比递归的效率高吗？

- 循环和递归两者是可以互换的，不能决定性的说循环的效率比递归高；
- 递归的优点：代码简洁清晰，容易检查正确性；
- 缺点是：当递归调用的次数较多时，要增加额外的堆栈处理，有可能产生堆栈溢出的情况，对执行效率有一定的影响；
- 循环的优点：结构简单，速度快；
- 缺点是：它并不能解决全部问题，有的问题适合于用递归来解决，不适合用循环；

面试题：如何在一个数组中，找出两个数相加等于一个固定值的所有数对

- (1) **方法1：**使用暴力穷举法，用到了2层for循环，时间复杂度为 $O(n^2)$
- (2) **方法2：**先将数组排序，然后依次取一个数后，在数组中用二分查找，查找 $sum - val$ 是否存在，如果存在，则找到了一对二元组，它们的和为 sum 。排序需要 $O(n \log 2n)$ ，二分查找需要 $(\log 2n)$ ，查找 n 次，所以时间复杂度为 $O(n \log 2n)$
- (3) **方法3：**可以用hash表来存储数组中的元素，这样取得一个数后，去判断 $sum - val$ 在不在数组中，如果在数组中，
 - 则找到了一对二元组，它们的和为 sum ，不过该算法的缺点就是需要用到一个hash表，增加了空间复杂度。

面试题：什么是数据结构？C 语言、数据结构、算法以及程序之间的关系是什么？

- (1)数据结构是指所有数据元素以及数据元素之间的关系，可以看作是相互之间存在着某种特定关系的集合因此可以将数据结构看成是带结构的数据元素的集合。此外，在通常情况下，选择合适的数据结构可以带来提高运行效率或者存储效率。
- (2)数据结构、算法以及程序之间的关系是：数据结构+算法=程序。算法不一定要能够在计算机上直接运行但程序必须要求能在计算机中运行。C 语言只是对算法或者数据结构的描述，描述数据结构和算法不局限于 C 语言，也可以是 C++语言和其他的计算机语言甚至也可以用人的自然语言。

面试题：程序和软件的区别是什么？

- (1)软件是相对于硬件而言的。是一系列按照特定顺序组织的计算机数据和指令的集合。而程序是软件的一个组成部分，是实现软件功能的主要动力之一，它是软件的必要元素；
- (2)软件=程序+文档=数据结构+算法+文档