

数据结构与算法 (Python)

期末笔试复习

谢正茂 webg@PKU-Mail

北京大学计算机系

June 10, 2021

- 概述
- 算法分析
- 基本数据结构
- 递归与动态规划
- 排序与搜索
- 树及其算法
- 图及其算法

- 基于有穷观点的能行方法
 - 计算的基本概念
- 抽象计算模型-图灵机
- 计算复杂性及不可计算问题
 - 不同问题的计算是有不同复杂度的，有些问题是无法计算的。
- 数据结构和抽象数据类型
 - 抽象的概念，及什么是 ADT 和 DS(三.1)
- 算法及衡量算法
 - 有穷、确定、可行 (一.6)
 - 算法基本概念和衡量算法的方法

- 程序与算法的关系
- 算法分析的概念
- 计算资源及资源消耗指标, 运行时间检测方法
- 算法复杂度的衡量指标: 大 O 表示法
- 确定大 O 的方法及常见的大 O 数量级函数 (二.1)
 - $O(1)$, $O(\log(n))$, $O(n)$, $O(n * \log(n))$, $O(n^2)$, $O(2^n)$
- 对算法的实现代码进行分析, 以得到大 O 数量级 (二.2)
- 理解常见数据类型中操作的大 O 数量级

基本数据结构

- 线性数据结构的概念，理解 ADT 的不同实现方案及其复杂度分析
- 栈的概念、特性和 ADT Stack
- 队列的概念、特性和 ADT Queue(一.3)
- 双端队列的概念、特性和 ADT Deque(二.4)
- 列表的概念、特性和 ADT List、ADT OrderedList

递归与动态规划

- 递归的概念及初步例子
- 递归的“三定律”
- 用递归解决进制转换问题
- 递归调用的内部实现: 与栈相关
- 递归与自相似图形, 理解绘制自相似图形的递归算法
- 用递归解决河内塔问题和探索迷宫问题
- 动态规划算法策略
- 从兑换硬币问题对比递归算法和动态规划算法, 如何避免递归爆炸

- 顺序搜索算法，以及在无序表和有序表数据结构中的不同实现
- 二分搜索算法，分而治之的算法策略
 - 高效算法的额外开销问题，以及依据实际应用来选择算法
- 散列的概念，及散列冲突概念，完美散列函数
- 散列函数设计的几种方法 (一.7/8/9)
- 散列冲突解决方案，负载因子
 - 开放定址法: 线性探测
 - 数据链法
- 抽象数据类型 ADT Map 及实现的算法分析

排序算法

- 冒泡排序算法, 及性能改进
- 选择排序算法 (多趟比对, 但减少交换次数)
- 插入排序算法 (为“新项”寻找插入位置, 逐步扩大已排序子列表)
- 谢尔排序算法 (固定间隔的多个子列表进行插入排序, 减小间隔)
- 归并排序算法 (将列表持续分裂为两半后, 再合并完成排序)
- 快速排序算法 (以中值作为基准将列表分为小于和大于两部分)
- 根据数据特征来选择排序算法
 - 算法性能退化
 - 如何选择一个好的算法来解决问题

排序算法

- (一.10/11/12/13/14, 二.9) 重点
- 什么时候出现最好情况, 什么时候出现最差情况?
- 复杂度主要讨论比较次数
- 稳定性的定义是什么? 保证相同key的前后次序
- 适用场景有什么限制?

算法名称	最好情况	最差情况	平均情况	稳定性
冒泡排序	$O(n)$	$O(n^2)$	$O(n^2)$	稳定
选择排序	$O(n)$	$O(n^2)$	$O(n^2)$	不稳定
插入排序	$O(n)$	$O(n^2)$	$O(n^2)$	稳定
谢尔排序	$O(n)$	$O(n^2)$	$O(n^{\frac{3}{2}})$	不稳定
归并排序	$O(n\log(n))$	$O(n\log(n))$	$O(n\log(n))$	稳定
快速排序	$O(n\log(n))$	$O(n^2)$	$O(n\log(n))$	不稳定
分配排序	$O(d(m+n))$	$O(d(m+n))$	$O(d(m+n))$	稳定
堆排序	$O(n\log(n))$	$O(n\log(n))$	$O(n\log(n))$	不稳定

堆排序——二叉树建堆 $O(n)$ +取出元素 $\log n$

A、 $H(\text{key}) = \text{key} \% 17$

B、 $H(\text{key}) = \text{key} \% 18$

C、 $H(\text{key}) = \text{key} \% 19$

D、 $H(\text{key}) = \text{key} \% 20$

B: 只对小片段访问; C中值不访问; D错

10. 对于排序算法特性的叙述正确的是()

A、 冒泡排序不需要访问那些已排好序的记录

B、 谢尔排序中, 当对确定规模的这些小序列进行插入排序时, 要访问序列中的所有记录

C、 快速排序过程中, 递归树上根据深度划分的每个层次都要访问序列中的所有记录

D、 选择排序需要访问那些已排好序的记录

11. 下面的排序算法哪个是不稳定的 ()。

A、 插入排序

B、 归并排序

C、 冒泡排序

D、 快速排序

12. 在第一趟排序之后, 不能确保将数据表中某一个元素放在其最终位置上的排序算法 ()。

A. 选择排序

B. 冒泡排序

C. 快速排序

D. 归并排序

树及其算法

- 树的概念及例子，树的两种定义 高度和深度的辨别
 - 熟悉树相关的术语与定义
- 实现树的方法: 嵌套列表法、节点链接法
- 树的应用: 解析树 (语法树和表达式树)
- 表达式树的建立算法，利用表达式解析树求值
- 树的遍历: 前序、中序及后序遍历 (二.3/7) 后中推前
 - 在表达式生成和求值中的应用
- Huffman 树编码
- 优先队列的概念，完全二叉树与二叉堆 (一.5, 二.5/6/8)
- 二叉搜索树 BST 及平衡树 AVL 树的概念及实现 (一.18, 二.11, 四.2)
- 树 (林) 与二叉树的互相转化

二叉树的度是指非空子树的数量；满二叉树是指每一层都是满的

图及其算法

- 图的概念，用图来表示的网络
- 熟悉图的术语及定义，ADT Graph
- 图的实现方法: 邻接矩阵及邻接列表法
- 词梯问题及广度优先搜索 BFS
- 骑士周游问题及深度优先搜索 DFS
- 通用的深度优先搜索算法 (一.16)
- DFS 用于解决拓扑排序和强连通分支问题 (一.15/17, 三.2)
- 路由选择, 最短路径问题及 Dijkstra 算法 (一.19)
- 信息广播, 最小生成树问题及 Prim 算法 (二.10)

期末闭卷笔试

- 时间：6 月 22 日（周二）下午 1:00 点-3:00 点
- 地点：理教 406
- 百分制，占总评百分之四十