

上海科技大学 2023 年攻读硕士学位研究生

招生考试试题

科目代码：991

科目名称：数据结构与算法

考生须知：

1. 本试卷满分为 150 分，全部考试时间总计 180 分钟。
 2. 所有答案必须写在答题纸上，写在试题纸上或草稿纸上均无效。
 3. 考生可用中文或者英文作答。
-

一、 判断题(10 题，每题 2 分，共 20 分)

请在答题纸上写明题号后，对正确的命题则打“√”，错误的命题则打“×”。

1. 单链表 (singly linked list) 实现的有序表可使用二分查找法来提高查找速度。
2. 利用单链表实现栈和队列，出栈、入栈、出队和入队操作均可以在 $\Theta(1)$ 时间内完成。
3. 在设计哈希表 (hash table) 时，采用线性探查 (linear probing) 可以避免一次群集 (primary clustering) 现象。
4. 对一棵结点数目为 n ，高度为 h 的树进行广度优先搜索 (breadth-first search) 遍历，其复杂度为 $\Theta(h)$ 。
5. 高度为 h 的二叉树 (binary tree) 包含的结点数至多为 $2^{h+1}-1$ (注：空树的高度为-1)。
6. 对二叉搜索树 (binary search tree) 进行中序遍历 (in-order traversal)，遍历结果为有序序列。
7. 任何一个贪心 (greedy) 算法总是能给出最优的解。
8. 给定任何一个所有权重为正数的有权无向图，从图中任何一个顶点到所有其他点的最短路径所构成的树也是这个图的最小生成树 (minimum spanning tree) (假设图连通)。
9. 动态规划 (dynamic programming) 算法将一个问题拆解成多个完全独立的子问题进行求解。

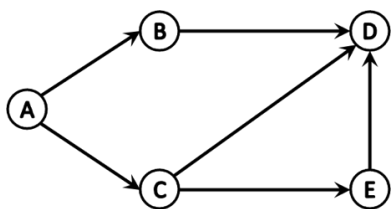
10. 如果一个问题 A 可以在多项式时间内求解, 则 A 属于 P 问题。

二、 单选题(15 题, 每题 2 分, 共 30 分)

每题只有一个正确选项。请在答题纸上写明题号后写下正确选项的序号。

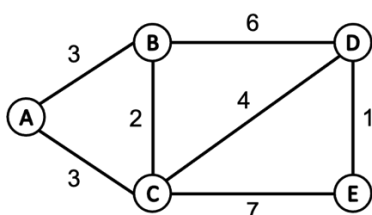
11. 请问下列关于线性表的叙述正确的是
- A. 线性表中每个元素都有一个直接前驱和一个直接后继
 - B. 线性表中至少要有 1 个元素
 - C. 线性表中所有元素的排列顺序必须是从小到大或者从大到小
 - D. 数组可以用来实现线性表
12. 假设元素入栈的先后顺序为 a、b、c, 并且元素进栈后可停留、可出栈, 直到所有的元素都出栈。那么, 所有可能的出栈序列个数为
- A. 6
 - B. 5
 - C. 4
 - D. 3
13. 假设有一个用一维数组 $A[1..n]$ 表示的循环队列, Front 为当前队头元素在数组中的索引 (index), Back 为队尾元素的索引加 1。若队列非空, 队列中元素个数为 (注: mod 为取模操作符)
- A. $(Back + Front + n) \bmod n$
 - B. $(Back - Front + n) \bmod n$
 - C. $(n - Back - Front) \bmod n$
 - D. $(n - Back + Front) \bmod n$
14. 请问下列关于哈希表的叙述错误的是
- A. 适当扩大哈希表容量, 可以避免负载因子 (load factor) 过大
 - B. 构造性能良好的哈希函数可以减少冲突 (collision)
 - C. 通常采用链接法 (chaining) 或者开放寻址法 (open addressing) 解决冲突
 - D. 哈希函数的构造一定越复杂越好
15. 在下列排序算法中, 不稳定的算法是
- A. 插入 (insertion) 排序
 - B. 快速 (quick) 排序
 - C. 归并 (merge) 排序
 - D. 冒泡 (bubble) 排序
16. 给定一个具有 n 个元素的最小堆 (min-heap), 查询最小值和最大值的时间复杂度分别为
- A. $O(1)$ 、 $O(\log n)$

- B. $O(1)$ 、 $O(n)$
C. $O(\log n)$ 、 $O(\log n)$
D. $O(n)$ 、 $O(\log n)$
17. 给定一个有序数组 S (包含 n 个元素), 基于 S 构建相应的二叉搜索树 T 。请问下列叙述正确的是
- A. 基于 S 构建二叉搜索树 T 的平均时间复杂度是 $\Theta(\log n)$
B. 在 S 中查询中位数要比在 T 中查询中位数的平均时间复杂度小
C. 给定元素 e , 在 S 中插入 e 要比在 T 中插入 e 的平均时间复杂度小
D. 给定元素 e , 在 S 中删除 e 要比在 T 中删除 e 的平均时间复杂度小
18. 给定一棵二叉树 (包含 n 个结点), 请问下列叙述错误的是
- A. 该树存在 n 条边
B. 根结点到二叉树中任意结点存在唯一路径
C. 若该树为完全二叉树 (complete binary tree), 其根结点的左子树和右子树高度差至多为1
D. 对该树进行广度优先遍历, 可计算出每个结点所在的深度
19. 对包含 n 个元素的无序序列进行堆排序 (heapsort), 请问下列叙述错误的是
- A. 采用最小堆或最大堆都可实现堆排序
B. 堆排序中建堆的平均时间复杂度可以是 $O(n)$
C. 堆排序的平均时间复杂度为 $O(n \log n)$, 最坏时间复杂度为 $O(n^2)$
D. 堆排序可原址 (in-place) 实现, 所以空间复杂度为 $O(1)$
20. 完全二叉树通常可采用数组高效实现。若考虑一般二叉树 (包含 n 个结点), 采用数组实现所需的数组长度在最坏情况下为
- A. $O(n)$
B. $O(n^2)$
C. $O(n^{\log n})$
D. $O(2^n)$
21. 下列关于 P, NP 和 NPC (NP 完备) 问题的描述中, 正确的是
- A. P 不属于 NP
B. 如果 X 属于 NPC 且 X 可以多项式求解, 则 $NP=P$
C. 如果 X 可以多项式时间转成 Y, 且 Y 属于 NPC, 则 X 属于 NPC
D. 如果 X 可以多项式时间转成 Y, 且 X 可以多项式求解, 则 Y 也可以多项式求解
22. 给定如下有向无环图, 其所有可能的拓扑排序结果数量为
- A. 1
B. 2
C. 3
D. 4



23. 给定如下带权无向图, 用 (X, Y) 表示连接顶点 X 和顶点 Y 的边, 请问下列边集能构成此图的最小生成树的是

- A. $(A, B), (A, C), (B, D), (C, E)$
- B. $(A, B), (B, C), (C, D), (C, E)$
- C. $(A, C), (B, C), (C, D), (D, E)$
- D. $(A, B), (B, C), (C, E), (D, E)$



24. 下列算法中采用了动态规划思想的是

- A. 归并排序
- B. 快速排序
- C. 克鲁斯克尔算法 (Kruskal's algorithm)
- D. 弗洛伊德算法 (Floyd-Warshall algorithm)

25. 给定 n 个任务, 每个任务 t 有严格的开始和结束时间, 分别记为 t_{start} 和 t_{end} 。对这 n 个任务运行如下算法:

算法: 把 n 个任务按照 t_{end} 升序排序, 从左到右 (从小到大) 对排序好的任务依次进行选择决策: 如果当前任务和已被选的所有任务没有时间重叠, 则选择当前任务, 否则不选择当前任务。(注: 运行算法开始时已被选任务集合为空)

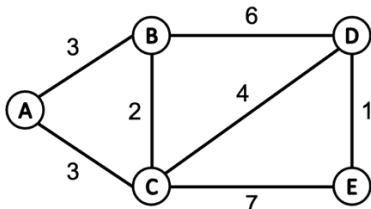
请问下列关于上述算法的表述正确的是

- A. 选出的任务集合是两两时间不交叉且任务数量最大的任务子集
- B. 是一个分治算法
- C. 是一个动态规划算法
- D. 运行时间复杂度为 $\Theta(n)$

三、多选题(6 题, 每题 5 分, 共 30 分)

每题有一个或者多个正确选项。请在答题纸上写明题号后写下所有正确选项的序号。每题少选可得 2 分, 多选或错选都不得分。

26. 假设使用两个队列实现一个栈, n 为栈中元素个数。请问下列叙述正确的是
- 单个元素入栈和出栈操作时间复杂度均为 $\Theta(1)$
 - 单个元素入栈操作时间复杂度为 $\Theta(n)$, 且出栈操作时间复杂度为 $\Theta(1)$
 - 单个元素入栈操作时间复杂度为 $\Theta(1)$, 且出栈操作时间复杂度为 $\Theta(n)$
 - 以上叙述均不正确
27. 假设待排序的元素个数为 n , 请问下列关于快速排序算法的叙述正确的是
- 快速排序算法的平均时间复杂度为 $\Theta(n)$
 - 快速排序算法的平均时间复杂度为 $\Theta(n \log n)$
 - 快速排序算法的最坏时间复杂度为 $\Theta(n^2)$
 - 快速排序算法是原址排序方法
28. 给定一段文本, 统计字符 (a, b, c, d, e) 出现的频率分别为 (0.1, 0.1, 0.2, 0.2, 0.4), 构建霍夫曼生成树 (二叉树结构) 并对字符 (a, b, c, d, e) 进行霍夫曼编码 (Huffman coding), 请问下列叙述正确的是
- 霍夫曼生成树中结点的度 (degree) 只能为0或者2
 - 字符 e 的编码一定不是字符 c 的编码的前缀 (prefix)
 - a 的编码位数为4, e 的编码位数为1
 - 所有字符的平均霍夫曼编码位数为2.2
29. 对于最小堆, 请问下列叙述正确的是
- 最小堆是实现优先队列的一种数据结构
 - 采用完全二叉树来实现最小堆, 该堆的最大值对应的结点一定是叶结点
 - 如果最小堆包含 n 个元素, 则一次 Pop 和一次 Push 操作复杂度都为 $O(\log n)$
 - 最小堆 (包含 n 个元素) 通常可采用数组高效实现, 空间复杂度为 $O(n)$
30. 给定如下带权无向图, 从顶点 D 开始用普里姆算法 (Prim's algorithm) 求解其最小生成树。用 (X, Y) 表示连接顶点 X 和顶点 Y 的边, 请问此算法运行过程中依次选边的序列 (从左到右) 是
- (D, E), (B, C), (A, B), (C, D)
 - (D, E), (C, D), (B, C), (A, B)
 - (D, E), (C, D), (B, C), (A, C)
 - (D, E), (B, C), (A, C), (C, D)



31. 下列叙述正确的是
- 动态规划算法通过增加空间复杂度来降低时间复杂度

- B. 在归并排序算法中通过存储子问题的解可以进一步降低该算法的时间复杂度
- C. 迪杰斯特拉算法 (Dijkstra's algorithm) 具备贪心的思想
- D. 0-1 背包问题 (knapsack problem) 可以用动态规划算法求解

四、 排序算法 (4 题, 每题 5 分, 共 20 分)

给定一个待排序的正整数序列 $\{2, 11, 10, 5, 4, 13\}$, 回答下列四个问题。

32. 利用插入排序算法对这个序列按照升序排序。从第三个元素开始, 请写出每个元素插入后的排序结果, 并写出每次插入过程中比较 (compare) 两个元素大小和交换 (swap) 两个元素位置的次数。

例如, 插入第二个元素 11 后排序结果为 $2, 11, 10, 5, 4, 13$ 。比较次数: 1, 交换次数: 0。

33. 给定一段正整数序列 $\{a_1, a_2, \dots, a_n\}$ 。假设升序为有序, 如果一对元素 $a_i > a_j$ 且 $i < j$, 那么 a_i 和 a_j 就组成了一对逆序对 (inversion)。例如, $\{9, 8, 7\}$ 中有 3 对逆序对: $\langle 9, 8 \rangle$ 、 $\langle 9, 7 \rangle$ 、 $\langle 8, 7 \rangle$ 。请写出 $\{2, 11, 10, 5, 4, 13\}$ 这个序列中所有的逆序对。如果一个正整数序列有 n 个元素, 那么这个序列最少有多少对逆序对? 最多有多少对逆序对?
34. 假设待排序的序列的元素个数为 n , 请写出插入排序的最坏时间复杂度 (须用渐近符号 Θ 、 n 表示) 并给出具体分析过程。
35. 假设某元素个数为 n 的序列有 d 对逆序对, 请写出对该序列运行插入排序的时间复杂度 (须用渐近符号 Θ 、 n 、 d 表示), 并给出具体分析过程。假设插入排序中执行一次比较操作的时间为 $\Theta(1)$, 执行一次交换操作的时间为 $\Theta(1)$ 。

五、 二叉搜索树和 AVL 树 (5 题, 每题 5 分, 共 25 分)

对于一棵二叉搜索树, 如果树中任意结点的左子树 (若存在) 和右子树 (若存在) 的高度差不超过 1, 则称为 AVL 树。在二叉搜索树中, 约定二叉搜索树的结点值大于该结点左子树 (若存在) 中所有结点的值且小于右子树 (若存在) 中所有结点的值。树中结点 x 的值、左孩子结点和右孩子结点分别表示为 $x.value$ 、 $x.left$ 和 $x.right$ 。

将如下元素依次 (从左到右) 插入初始为空的二叉搜索树

20、15、30、17、26、37、23、28

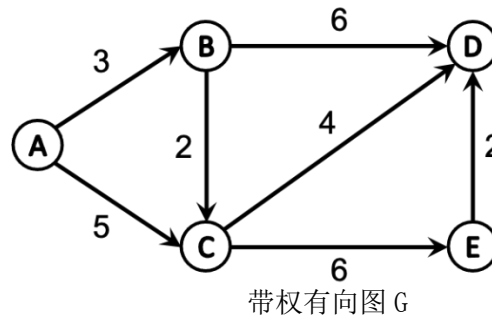
回答下列五个问题。注: 二叉搜索树的插入和删除须维持二叉搜索树的结构。

36. 画出插入所有元素后的二叉搜索树 T 。

37. 对 T 进行后序遍历 (post-order traversal), 写出遍历结果。
38. 若采用数组表示 T, 所得到的数组是? (空结点可用 NIL 表示)
39. 描述在 T 中删除结点 15 的过程, 判断删除之后的二叉搜索树是否为 AVL 树? 如果是, 请简述理由; 如果不是, 请画图示意调整成 AVL 树的过程。
40. 给定一个存储 n 个整数的二叉搜索树 (假设 n 为奇数), 定义任意非空结点 x 所对应的子树的结点数为 $x.size$, 即 $x.size = x.left.size + x.right.size + 1$ (空结点的 $size = 0$)。假设所有结点的 $size$ 已知, 设计一个查询该树中所有整数的中位数的算法, 且算法时间复杂度为 $O(\log n)$ 。用伪代码描述。

六、图 (5 题, 每题 5 分, 共 25 分)

给定如下带权有向图 G, 顶点 x 到 y 的有向边的权重用函数 $w(x, y)$ 表示, $w(x, y)$ 的实际数值标记在每条边上, 例如 $w(A, B) = 3$ 。请根据此图回答下列五个问题。



41. 给出从 A 点开始, 用广度优先搜索访问上图所有顶点的访问顺序 (每个顶点第一次被访问的顺序)。假设一个点的邻居按照字母顺序访问。
42. 假设我们要算出从 A 点到达所有点的最短距离。给定顶点 x, 我们用 $x.d$ 存储从 A 到 x 的已知最短距离, 用 $x.p$ 存储这个已知最短路径上到达 x 之前的一个顶点 (这个顶点可以为空, 空用 NIL 表示)。初始化所有点的 d 和 p 如下表。

A. d	B. d	C. d	D. d	E. d
0	∞	∞	∞	∞
A. p	B. p	C. p	D. p	E. p
NIL	NIL	NIL	NIL	NIL

在上述初始化的 d 和 p 数据下, 在广度优先搜索中每次访问一个顶点 u 时, 对其所有还没有被访问的邻居顶点 v 做如下 $relax(u, v, w)$ 操作。给出广度优先搜索结束时所有点的 d 和 p 的值, 用与初始化相同的表格作答 (注: 不需要中间结果)。

```
relax(u, v, w)
{
    if v.d > u.d + w(u, v)
    {
        v.d = u.d + w(u, v)
        v.p = u
    }
}
```

43. 如果把广度优先搜索每次选择搜索下一个顶点的条件改成“对还没有被搜索的所有顶点中 d 最小的顶点进行搜索”（如有多个顶点同时具有最小的 d 时，按照字母顺序选择最小的），在此改变下重新给出第 42 题中最后所有点的 d 和 p 值，作答要求同第 42 题。
44. 第 42 题和第 43 题中所描述的两算法，
- 哪一个能正确输出从 A 到所有点的最短距离？
 - 假设将顶点 E 到顶点 D 的边的权重从正 2 改为负 2，即 $w(E, D) = -2$ ，这两种算法是否能正确输出从 A 到所有点的最短距离？
45. 画出带权有向图 G 的邻接表表示 (adjacency-list representation) 结果。假设用 V 和 E 分别表示一个有向图的总顶点数和边数，第 41 题中广度优先搜索在图用邻接表输入的情况下的时间复杂度是多少（须用渐近符号 Θ 表示）？