上海科技大学 2020 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一律无效。

3. 每道题的英文部分均已翻译为中文，考生可在中英文中任选一种语言作答。

---

1. True or False (totally 20 points, 2 points each) 判断题（共20分，每题 2 分）

Please indicate in the answer sheet whether each statement is true or false. Write down "T" for being true and "F" for being false.

请在答题纸上写明下列每个命题的真假。真则打 "√"，假则打 "×"。

1. Given three functions $f(n)$, $g(n)$ and $h(n)$, if $f(n) \in \Omega(h(n))$ and $g(n) \in \Omega(h(n))$, then $f(n) \cdot g(n) \in \Omega([h(n)]^2)$. （　）
   给定 3 个函数 $f(n), g(n)$ 和 $h(n)$，如果 $f(n) \in \Omega(h(n))$ 而且 $g(n) \in \Omega(h(n))$，那么 $f(n) \cdot g(n) \in \Omega([h(n)]^2)$。（　）

2. If a problem is in NP, it cannot be solved by a deterministic algorithm in polynomial time. （　）
   如果一个问题属于 NP，它将不能被一个时间复杂度为多项式的确定性算法解决。（　）

3. The in-order depth-first traversal of a binary search tree (BST) produces elements in descending order. （　）
   二叉搜索树的中序深度优先遍历可以产生一个降序的有序序列。（　）

4. For a DFS graph traversal, we can use a queue to keep track of vertices. （　）
   在图的深度优先（DFS）遍历中，可以利用一个队列实现顶点跟踪。（　）

5. For a rooted binary tree, n_0 denotes the number of nodes with degree 0, n_1 denotes the number of nodes with degree 1 and n_2 denotes the number of nodes with degree 2. Then we have the property that $n_0 = n_2 + 1$. （　）
   在一棵有根二叉树中，n_0代表度为 0 的节点的个数，n_1代表度为 1 的节点的个数，n_2代表度为 2 的节点的个数，那么等式 $n_0 = n_2 + 1$ 成立。（　）

6. In a binary heap (a heap that is implemented using a complete binary tree), if a node A is the ancestor of a node B, then we can deduce that the insertion of node A is performed

before the insertion of node B.　（　　）

在一个二叉堆（用完全二叉树实现的堆）中，节点 A 是节点 B 的祖先，那么可以推断出节点 A 先于节点 B 插入这个二叉堆中。（　　）

7. The operation of insertion before the current position is not efficiently supported by a singly-linked list.　（　　）

   单链表不能有效地支持当前位置之前的插入操作。（　　）

8. Queue is typically used to implement recursion in a program.　（　　）

   队列通常用于在程序中实现递归。（　　）

9. With a poorly chosen hash function, it is possible to have a situation in which the search time in a hash table of N items goes to $\Theta(N^2)$.　（　　）

   如果使用的哈希函数选择不当，则可能会出现以下情况：存有 N 项元素的哈希表中的搜索时间会变为$\Theta(N^2)$。（　　）

10. For the linked list implementation of the queue, the enqueue is performed after the last element, dequeue is at the first element.　（　　）

    对于用链表实现的队列，入队在最后一个元素之后执行，而出队则在第一个元素上执行。（　　）

2. Multiple Choices – Select One (totally 20 points, 2 points each) 单选题（共 20 分，每题 2 分）

Each question has only one correct choice. Please indicate the correct choice in the answer sheet.

每题只有一个正确选项。请在答题纸上写下正确选项的序号。

1. Given an array that has been almost sorted in ascending order, to sort this array in ascending order, which of the following sorting algorithm would be the most efficient in running time?
   A. Insertion sort
   B. Merge sort
   C. Quicksort
   D. None of the above

   如果给定的一个数组已经基本上按升序排列，要把这个数组排为升序，下面哪一个排序算法在运行时间上最高效？
   A. 插入排序
   B. 归并排序
   C. 快速排序
   D. 以上都不正确

2. Which of the following list of functions is in the increasing order of asymptotic growth rate? If two (or more) functions are of the same asymptotic order, the order between them does not matter.

   下面哪一组的函数是按照渐进增长率的递增排序？如果两个（或多个）函数在同一个渐进阶，那么它们之间的次序无需考虑。

   A. $100n^2$, $n\log_2(10n)^2$, $\log_2 n^{20}$, 2^n, n^2+1000
   B. $n\log_2(10n)^2$, $\log_2 n^{20}$, $100n^2$, n^2+1000, 2^n
   C. $\log_2 n^{20}$, $n\log_2(10n)^2$, $100n^2$, 2^n, n^2+1000
   D. $\log_2 n^{20}$, $n\log_2(10n)^2$, $100n^2$, n^2+1000, 2^n

3. If array P = [5, 8, 3, 1, 10, 14, 12, 11] is the output of running the partition function in Quicksort, which of the following elements is the pivot at its final position?

   如果数组 P = [5, 8, 3, 1, 10, 14, 12, 11] 是快速排序算法里的分区函数的输出，下面哪个元素是已经被放在其最终位置的枢轴？

   A. 1
   B. 8
   C. 10
   D. 11

4. Consider a 4-degree tree with 20 4-degree nodes, 10 3-degree nodes, 1 2-degree node and 10 1-degree nodes, then how many leaf nodes are there in the tree?

   已知一个度为 4 的树中，有 20 个度为 4 的节点，10 个度为 3 的节点，1 个度为 2 的节点和 10 个度为 1 的节点。那么这棵树上有几个叶节点？

   A. 41
   B. 82
   C. 113
   D. 122

5. For the two binary trees in Figure 1 below, Tree 1 and Tree 2, which DFS (Depth First Search) traversal will produce the same sequence of nodes?
   A. Post-order of Tree 1, Post-order of Tree 2
   B. Post-order of Tree 1, In-order of Tree 2
   C. Pre-order of Tree 1, Post-order of Tree 2
   D. In-order of Tree 1, Pre-order of Tree 2

   对于图 1 中所示的两个二叉树 Tree 1 和 Tree2，下列哪两种深度优先遍历会产生一致的节点输出序列？
   A. Tree 1 后序遍历，Tree 2 后序遍历
   B. Tree 1 后序遍历，Tree 2 中序遍历
   C. Tree 1 先序遍历，Tree 2 后序遍历
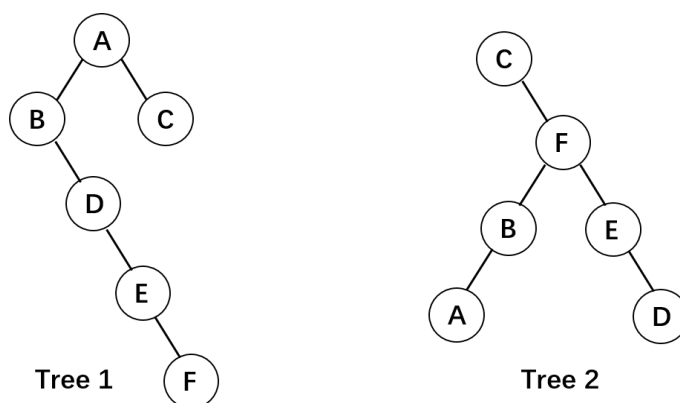   D. Tree 1 中序遍历，Tree 2 先序遍历

Figure 1. 图 1

6. If a complete binary tree has a number of 1880 nodes, what is the height of the tree?
   一个完全二叉树有 1880 个节点，那么这棵树的高度是：
   A. 9
   B. 10
   C. 11
   D. 12

7. A hash table has b buckets where each bucket can hold k keys. What is the minimum number of distinct keys that we must insert into the hash table such that overflow will surely occur?

   一个哈希表有 b 个存储桶，每个存储桶可容纳 k 个键。在这个哈希表中，至少需要插入多少个不同的键才会使溢出必然发生？
   A. 2
   B. $k + 1$
   C. $b + 1$
   D. $b*k + 1$

8. If we have a doubly linked list [a b c d e f] and we want to delete the element d and add a new element after f, how many pointers of the list must be changed or initialized?

   如果我们有一个双向链表[a b c d e f]，并且想要删除元素 d 并在 f 之后添加一个新元素，那么必须更改或初始化该链表的多少个指针？
   A. 1
   B. 2
   C. 3
   D. 4

9. Suppose we have two non-empty stacks stored in arrays: S1=[a, b, c] and S2=[f, e, d]. The left side of an array indicates the bottom of a stack. After the following sequence of operations:

假设我们有两个存储在数组中的非空堆栈：S1 = [a，b，c]和 S2 = [f，e，d]。数组的左端代表堆栈的底部。在以下操作序列之后：

Push(S1, Pop(S2))

Push(S2, c)

Push(S1, Pop(S2))

Then the two stacks become:

这两个堆栈变为：

A. S1=[c, f, a, b, c] and S2=[e, d]

B. S1=[a, b, c, d, c] and S2=[f, e]

C. S1=[a, b, c, d, c, e] and S2=[f]

D. S1=[a, b, c, d, e] and S2=[c, f]

10. The UNIX editor vi allows searching in both directions, with wrap-around if necessary. If the sequence of lines is stored as a linked list, which of the following is most reasonable in terms of time efficiency?

A. Singly linked list

B. Doubly linked list

C. Circular singly linked list

D. Circular doubly linked list

UNIX 编辑器 vi 允许双向搜索，必要时可以使用环绕式搜索。如果将行序列存储为链表，那么以下哪一项在时间效率上最合理？

A. 单链表

B. 双向链表

C. 循环单链表

D. 循环双向链表

3. Multiple Choices – Select One or More (totally 15 points, 3 points each) 多选题（共 15 分，每题 3 分）

Each question may have one or more correct choices. Please indicate all the correct choice(s) in the answer sheet.

每题有一个或多个正确选项。请在答题纸上写下所有正确选项。

1. In which of the following situations, using an adjacency matrix would be better than using an array of adjacency lists to represent a graph?

A. The graph has 1,000 vertices and 450,000 edges, and space efficiency is important.

B. The graph has 1,000 vertices and 2,000 edges, and it is important to save memory usage.

C. We need to determine whether two vertices in the graph are adjacent or not as fast as possible, whereas space efficiency is less important.

D. None of the above.

在下面哪种情况下，用一个邻接矩阵来表示图比用一组邻接表更合适？

A. 这个图有 1,000 个顶点和 450,000 条边，并且空间效率很重要。

B. 这个图有 1,000 个顶点和 2,000 条边，并且节省内存的使用很重要。

C. 我们需要快速判断这个图的两个顶点是否相邻，而空间效率相对次要。

D. 以上都不正确。

2. Consider a BST (Binary Search Tree) created by inserting the following keys in an initially empty BST in the given order: 4 2 6 5 1 3 7. Which of the following sequences would produce the same BST?

将一个整数序列 4 2 6 5 1 3 7 依次插入到一个初始状态为空的二叉搜索树中，建立一个二叉搜索树。下列序列中哪些（个）序列建立的二叉搜索树与上文所述的二叉搜索树一致？

A. 4 2 6 3 1 5 7

B. 4 6 5 7 1 2 3

C. 4 2 6 7 1 5 3

D. 4 6 2 1 5 3 7

3. Which of the following statements about graphs with N vertices is (are) true?

A. The maximum number of edges in an undirected graph is $O(N^2)$.

B. The maximum number of edges in a directed graph is $O(N^2)$.

C. The maximum degree of a vertex in an undirected graph is $O(N)$.

D. The maximum number of edges in an undirected acyclic graph with N vertices is $N(N-1)/2$.

对于一个包含 N 个顶点的图，下列选项中哪些表述是正确的？

A. 一个无向图的最大边数是一个 $O(N^2)$ 的函数。

B. 一个有向图的最大边数是一个 $O(N^2)$ 的函数。

C. 无向图中一个顶点的最大度数是 $O(N)$ 的函数。

D. 一个有 N 个顶点的无向无环图最多有$N(N-1)/2$条边。

4. Suppose you have a sorted list of numbers stored in one of the following data structures. Which of the following data structures should you use if you want to print the numbers in reverse order in linear time without using any auxiliary data structure except for a pointer?

A. A linked list

B. A circular linked list

C. A doubly linked list

D. An array

假设你有存储在以下数据结构之一中的数字排序列表。如果要以相反的顺序打印数字，为实现线性时间复杂度，且除一个指针外不允许使用任何辅助数据结构，应该使用哪种数据结构？
A. 链表
B. 循环链表
C. 双向链表
D. 数组

5. Which of the following is (are) solution(s) for a hashing collision:
   A. Replace the original hash function with a hash function without collision and re-hash all elements in the table
   B. Build a linked list of the hashed items for each bucket
   C. Find a different, empty bucket for each colliding item
   D. None of above

   以下哪项是对一个哈希冲突的解决方案：
   A. 替换为无冲突的哈希函数，然后重新建立含所有元素的哈希表
   B. 为每个存储桶的哈希项建立一个链表
   C. 为冲突的元素找到另一个空桶
   D. 以上都不是

4. Linked List (25 points) 链表（25 分）

Detect duplicates from an unsorted linked list. Given an unsorted linked list, [20,7, 12, 31, 25, 46, 52, 67, 85, 94, 73, 12, 46, 85, 12, 7, 46, 99], our goal is to detect all the keys with duplicates in the list: 7, 12, 46, 85.

从未排序的链表中检测重复项。给定一个未排序的链表，[20,7, 12, 31, 25, 46, 52, 67, 85, 94, 73, 12, 46, 85, 12, 7, 46, 99]，我们的目标是检测所有在链表中有重复的项：7, 12, 46, 85。

1) Using a hash table with separate chaining, 10 positions and hash function h(x) = x mod 10 for this task, describe your algorithm and show the resulting table after all the data are inserted. (5 points)
   使用具有单独链表和 10 个位置以及哈希函数为 h(x) = x mod 10 的哈希表，来描述你的算法，并在插入所有数据后显示最终的哈希表。 （5 分）

2) What is the average time and space complexity of the detection algorithm based on this type of hash table (with M positions) as in 1)? (5 points)
   基于如题 1）描述的此类哈希表（有 M 个位置）的检测算法的平均时间和空间复杂度是多少？ （5 分）

3) Use a sorting algorithm rather than hashing to implement the detection algorithm. (5

points)

使用排序而不是哈希算法来实现该检测算法。 （5 分）

4) Design an algorithm for duplication detection if no auxiliary data structure or sorting is allowed. (10 points)

如果不允许使用辅助数据结构或排序，请设计一种用于重复检测的算法。（10 分）

5. AVL tree (25 points) 二叉平衡树（25 分）

Given an initially empty AVL tree, insert the sequence of integers 15, 20, 23, 10, 13, 7, 30, 25 from left to right into the AVL tree. If there is an imbalance after you insert an integer, you should restore the balance of AVL tree before you insert the next node.

将一个的整数序列 15, 20, 23, 10, 13, 7, 30, 25 按从左到右的顺序依次插入到一个初始状态为空的二叉平衡树（AVL tree）。在插入一个整数之后，如果破坏了当前状态下树的平衡，在插入下一个节点之前需要将树重新调整至平衡。

1) Draw the final tree after all these insertions. (10 points)

画出将所有整数插入之后所建立的平衡树。（10 分）

2) Write the sequence of integers resulted from pre-order traversal of the final AVL tree built in 1). (5 points)

写出对题 1）中所建立的平衡树进行先序遍历所得到的整数序列。（5 分）

3) Continue with the final AVL tree in 1), erase the node with integer 23 from the AVL tree. Draw the resulting AVL tree after the deletion. (10 points)

将 1）中得到的平衡树中 23 所在位置的节点删除，画出删除节点后得到的新的平衡树。（10 分）

6. Minimum Spanning Tree (25 points) 最小生成树（25 分）

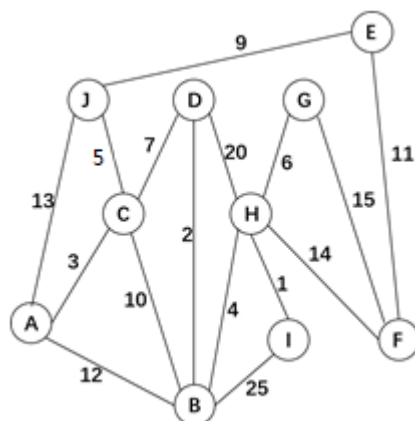Given an undirected weighted graph in Figure 2.

已知如图 2 中所示的无向有权图。

Figure 2 图 2

1) Draw the adjacent-matrix representation of the graph. (7 points)

   画出该图的邻接矩阵表示。（7 分）

2) Write down the edges of the MST (Minimum Spanning Tree) in the order in which Prim's algorithm adds them to the MST. Use the format (node1, node2) to denote an edge. For example, adding the edge between A and B into the MST is denoted as (A, B). (9 points)

   请写出用 Prim 算法计算最小生成树过程中添加边的顺序，用（顶点 1，顶点 2）的结构表达。例如，如果计算过程中需要添加顶点 A 和顶点 B 之间的边到最小生成树中，可以表示为（A, B）。（9 分）

3) For the same weighted graph as in Figure 2, write down the edges of the MST in the order in which Kruskal's algorithm adds them into the MST. Please use the same format as in 2). (9 points)

   采用 Kruskal 算法，求图 2 中的有权图的最小生成树。请采用题 2）中同样的方法表达计算过程。（9 分）

7. Sorting (20 points) 排序（20 分）

1) Write the pseudocode of the merge function of the merge sort algorithm, to merge two lists sorted in ascending order into one list sorted in ascending order, of totally n elements. Analyze the worst-case time complexity of the merge function. (5 points)

   写出归并排序算法里的归并函数的伪代码，把两个总共包含 n 个元素的已经排为升序的列表合并成一个升序的列表。分析这个归并函数在最坏情况下的时间复杂度。（5 分）

2) Suppose you are given k lists with a total of n distinct elements, where k > 2. Each list has been sorted in descending order. Design an algorithm using the strategy of Merge sort to merge the k lists into one list sorted in descending order, in running time $O(n\log_2 k)$. Write the pseudocode of your algorithm and analyze its worst-case time complexity. (7 points)

   假如给定 k 个总共包含 n 个不同元素的列表，这里 k > 2。每个列表已经排为降序。用归并排序算法的策略设计一个算法，将这 k 个列表在 $O(n\log_2 k)$ 时间内归并到一个降序排列的列表。写出你设计的算法的伪代码并分析它的最坏时间复杂度。（7 分）

3) Design an algorithm using a maximizing heap to solve the problem of merging k lists into one list in 2), in running time $O(n\log_2 k)$. Assume that the subroutines of Heapsort algorithm using a maximizing heap are already given. Write the pseudocode of your algorithm and analyze its worst-case time complexity. (8 points)

   设计一个算法，用一个最大化堆来解决题 2）提到的归并 k 个列表的问题，使得时间复杂度是 $O(n\log_2 k)$。假设使用最大化堆的堆排序算法的子程序都已经给定。写出你的算法的伪代码，并分析它的最坏时间复杂度。（8 分）