

# 北京大学信息科学技术学院考试试卷

考试科目: 数据结构与算法 A 姓名: \_\_\_\_\_ 学号: \_\_\_\_\_

考试时间: 2022 年 10 月 26 日 任课教师: \_\_\_\_\_

题号	一	二	三	四	五	总分
分数						
阅卷人						

## 北京大学考场纪律

- 1、考生进入考场后,按照监考老师安排隔位就座,将学生证放在桌面上。无学生证者不能参加考试;迟到超过 15 分钟不得入场。在考试开始 30 分钟后方可交卷出场。
  - 2、除必要的文具和主考教师允许的工具书、参考书、计算器以外,其它所有物品(包括空白纸张、手机等)不得带入座位,已经带入考场的必须放在监考人员指定的位置,并关闭手机等一切电子设备。
  - 3、考试使用的试题、答卷、草稿纸由监考人员统一发放,考试结束时收回,一律不准带出考场。若有试题印制问题请向监考教师提出,不得向其他考生询问。提前答完试卷,应举手示意请监考人员收卷后方可离开;交卷后不得在考场内逗留或在附近高声交谈。未交卷擅自离开考场,不得重新进入考场答卷。考试结束监考人员宣布收卷时,考生立即停止答卷,在座位上等待监考人员收卷清点后,方可离场。
  - 4、考生要严格遵守考场规则,在规定时间内独立完成答卷。不准旁窥、交头接耳、打暗号,不准携带与考试内容相关的材料参加考试,不准抄袭或者有意让他人抄袭答题内容,不准接传答案或者试卷等。凡有严重违纪或作弊者,一经发现,当场取消其考试资格,并根据《北京大学本科考试工作与学习纪律管理规定》及其他相关规定严肃处理。
  - 5、考生须确认自己填写的个人信息真实、准确,并承担信息填写错误带来的一切责任与后果。
- 学校倡议所有考生以北京大学学生的荣誉与诚信答卷,共同维护北京大学的学术声誉。

请把答案写在答题纸上。

得分

一、 选择填空(8 小题, 共 20 分)

1. (2 分) 数据结构的核心三要素不包含\_\_\_\_\_。

- A. 逻辑结构      B. 存储结构      C. 时间复杂度      D. 运算

2. (2 分) 以下算法的时间复杂度是\_\_\_\_\_。

(提示: 已知当  $n$  趋于无穷大时,  $1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n} - \ln(n) \rightarrow c$ , 其中  $c$  为欧拉常数, 约为 0.577215665)

```
int m = 1, i, j;
for (i = 1; i <= n; i++)
    for (j = 1; j <= n; j += i)
        m *= 2;
```

- A.  $\Theta(\log_2 n)$       B.  $\Theta(n \log_2 n)$       C.  $\Theta(n)$       D.  $\Theta(n^2)$

3. (2 分) 以下关于顺序表和链表的应用场合的说法, 不正确的是\_\_\_\_\_:

- A. 经常插入删除时, 不宜使用顺序表  
B. 无法确定线性表长度的最大值时, 不宜使用链表  
C. 当读操作比插入删除操作频率大时, 不应选择链表  
D. 当指针的存储开销, 和整个结点内容所占空间相比其比例较大时, 应该慎重选择链表

4. (2 分) 双循环链表中, 于结点  $p$  后插入新结点  $s$ , 以下做法不正确的是\_\_\_\_\_

- A.  $p \rightarrow next \rightarrow prev = s; s \rightarrow next = p \rightarrow next; p \rightarrow next = s; s \rightarrow prev = p;$   
B.  $p \rightarrow next \rightarrow prev = s; p \rightarrow next = s; s \rightarrow next = p \rightarrow next; s \rightarrow prev = p;$   
C.  $s \rightarrow next = p \rightarrow next; p \rightarrow next \rightarrow prev = s; s \rightarrow prev = p; p \rightarrow next = s;$

5. (3 分) 五个元素  $abcde$  顺序入栈, 下列选项中哪些不是可行的出栈序列? \_\_\_\_\_  
(可能不止一处, 若都可行则回答“无”)

- A.  $cbdae$       B.  $cabed$       C.  $decab$       D.  $dceab$

(3 分) 这些元素组成的所有可行出栈序列共\_\_\_\_\_种。

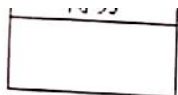
6. (2 分) 现有 4 个值为 1 到 4 的结点, 他们能组成\_\_\_\_\_种互不相同的二叉搜索树。

(2 分) 将序列  $\{12, 70, 33, 65, 24, 56, 48, 92, 86, 33\}$  按  $O(n)$  时间复杂度的建堆方法调整为最小值堆, 调整后的序列为\_\_\_\_\_。

7. (2 分) 设树  $T$  的度为 4, 其中度为 1、2、3、4 的结点个数分别为 4、2、1、1。则  $T$  中有\_\_\_\_\_个叶子结点:

- B. 4      B. 6      C. 8      D. 10





## 二、 辨析与简答(5 小题, 共 30 分)

- (3 分) 顺序栈和链式栈在底层分别通过定长数组和链表存储数据。请对比顺序栈和链式栈的时间效率和空间效率上的优劣。
- 对于字符串“1212123”:
  - (3 分) 按基于 ASCII 编码的字典序 (注意不算是数字串的数值大小序) 从小到大写出其所有不同非空子串。
  - (4 分) 写出该字符串优化后的 next 特征向量。
- (6 分) 一棵二叉树的前序、中序、后序序列如下所示, 其中一部分未标出, 请构造出该二叉树, 并填充尚未标出部分。要求画出该二叉树并写出他们的先序、中序、后序序列。  
先序: Z \_ C V U F \_ Y \_ W  
中序: C \_ U V Z \_ G F I \_  
后序: C \_ \_ B Y G W I \_ \_
- 某网络公司在通过网络传输数据之前使用压缩技术对消息进行编码, 假设消息包含以下字符及其频率:

Character	Frequency
a	5
b	9
c	12
d	13
e	16
f	45

- (3 分) 采用 Huffman 编码, 请画出对应的 Huffman 树;
  - (2 分) Huffman 编码平均长度还是多少?
  - (2 分) Huffman 编码比这 6 个字母等长编码能节省多少空间?
5. 并查集合并操作时, 我们可以使用重量权衡合并规则: 把子树结点少的并入子树结点多的 (结点多的那个作为新树根, 结点少的作为新树根的直接子结点)。为此, 我们除了维护每个结点的父结点下标, 还需要记录每颗子树的结点个数。一种简化的表示方式是, 我们只维护一个父结点下标数组: 对非根结点, 仍然记录其父结点下标; 但是对于根结点, 数组里的值为 -n, 其中 n 为这个根结点所在子树的结点个数。
- 现已知集合元素 0~8 所以对应的父结点下标数组值分别是 [1, -4, 3, 1, -3, 4, 4, 8, -2]。
- (4 分) 请画出所有元素构成的森林。
  - (3 分) 我们采用重量权衡规则与路径压缩, 对元素 2 与元素 7 所在集合合并。请写出合并后的数组。

得分

### 三、 算法填空(每空 2 分, 共 10 分)

给定以下树的结点定义, 其中 `getChildrenNum` 函数与 `getKthChild` 函数均已实现。现对于根结点 `root`, 已知 `root` 没有兄弟结点。请对以 `root` 为根的树实现非递归后根遍历。

```
class TreeNode {
public:
    int getChildrenNum(); // 返回子结点个数 n
    TreeNode* getKthChild (int k); // 对于  $0 \leq k \leq n-1$ , 返回从左到右第 k 个子结点; 对于其他 k, 返回空指针
}

class stackElement {
    TreeNode *node;
    int rd;
    stackElement(TreeNode * nd, int rd) {
        node = nd;
        rd = rd;
    }
};

void postOrderTraversal(Node *root) {
    stack<stackElement> st;
    st.push(stackElement(root, 0));

    while (____ /* 填空 1 */ ____ ) {
        stackElement topElem = st.top();
        ____ /* 填空 2 */ ____
        if (topElem.node->getChildrenNum() == 0 || ____ /* 填空 3 */ ____ ) {
            Visit(topElem.node);
        }
        else {
            st.push(stackElement(topElem.node, 1));
            for (int i = 0; i < topElem.node-> getChildrenNum (); ++i) {
                child = topElem.node->getKthChild( /* 填空 4 */ ____ );
                st.push(____ /* 填空 5 */ ____ );
            }
        }
    }
}
```

得分

#### 四、 算法设计与实现 (3 题, 共 30 分)

1. (10 分) 现有两个满足先进先出原则的队列 A 和 B, 在队列上可进行以下 4 个操作:

front(): 返回队首元素。

push\_back(T x): 把元素 x 插到队尾, 无返回值。

pop\_front(): 删除队首元素, 无返回值。

empty(): 返回布尔值, true 代表队列为空。

请用 A 和 B 模拟一个栈的 4 个操作:

T top();

void pop();

void push(T x);

bool empty();

请简述核心思路并给出这四个操作的伪代码 (必要之处给出注释等文字说明), 对时间效率无具体要求, 但要求使用不超过  $O(1)$  额外空间 (假设函数递归不占用空间), 不得使用除了以上四个提供的四个队列操作以外的操作。(pop() 和 top() 时假设 A、B 代表的栈不空, 不必考虑栈溢出问题)

2. (10 分) 请设计一个算法, 判断串 S 中是否有形如 AA 的子串, 如果有, 则返回该类子串第一次出现的起始下标。AA 串: 即可以看成由一个非空串重复两次构成的新串。

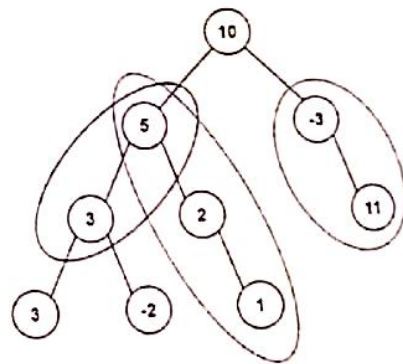
例如, S="dabcabc" 中, "abcabc" 就是 AA 子串, 其中 A="abc"。所以返回值应该是 1 (下标从 0 开始)。

请给出算法流程 (可给伪代码, 也可以用文字描述做法), 简单说明正确性, 并计算时间复杂度。请尽量使算法的时间复杂度小于等于  $O(n^2)$ , 否则会酌情扣分。

3. (10 分) 给定一个二叉树的根结点 root, 和一个整数 targetSum, 请设计一个算法, 求该二叉树里结点值之和等于 targetSum 的路径的数目, 并分析算法的时间复杂度。

注: 路径不需要从根结点开始, 也不需要从叶子结点结束, 但是路径方向必须是向下的 (只能从父结点到子结点)。(示例见右图)

例: 对于 targetSum=8, 一共有右图所示 3 条符合要求的路径。



得分

#### 五、 分析证明题 (10 分)

1. 如果一棵非空  $k(k \geq 2)$  叉树 T 中每个非叶子结点都有 k 个孩子。请回答下列问题并给出推导过程。

(1) 若 T 有 m 个非叶子结点, 则 T 中的叶子结点有多少个? (4 分)

(2) 若 T 的高度为 h (独根树高为 1), 则 T 中结点最多有几个? 最少有几个? (6 分)