

装
订
线
内

不
要
答
题

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：2019 年 11 月 13 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。

以下为试题和答题纸，试题共 4 页，请把答案写在答题纸。

扣

得分

一、 选择(每空 2 分, 共 24 分)

1. 现有三个复杂度分别表示为下述式子的算法

- A. $T_1(n) = 3n \lg n + \lg n$;
- B. $T_2(n) = 2^n + n^3 + 25$;
- C. $T_3(n, k) = k + n$ (且 $k \leq n$);

给出这些算法的复杂度 _____, 并将其从低到高排列 _____。

2. 带头结点的双循环链表 L 为空的条件是 _____。

- A. $L \rightarrow \text{prior} == L \ \&\& \ L \rightarrow \text{next} == \text{NULL}$
- B. $L \rightarrow \text{prior} == \text{NULL} \ \&\& \ L \rightarrow \text{next} == \text{NULL}$
- C. $L \rightarrow \text{prior} == \text{NULL} \ \&\& \ L \rightarrow \text{next} == L$
- D. $L \rightarrow \text{prior} == L \ \&\& \ L \rightarrow \text{next} == L$

3. 下列叙述中正确的是 _____。

- A. 结点中具有两个指针域的链表一定是非线性结构。
- B. 有序表可以存储在不连续的存储空间内。
- C. 循环链表的入队不会发生溢出。
- D. 循环链表是非线性结构。

4. 下面代码片段的运行结果应为 _____。

```
char c[6] = {'x','y','z','\0','m','\0'};
printf("%s", c);
```

5. 给定一个字符串数组 {"abc", "ced", "haha", "cbd", "hhaa"}, 其按字典顺序从小到大的排序序列为 _____。

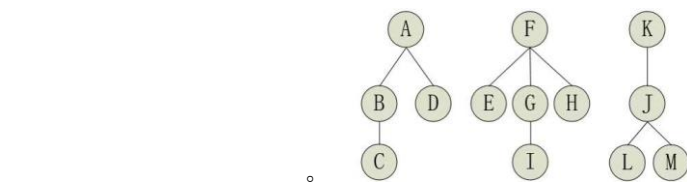
6. 现有字符串 $S_1 = \text{"ABCDEFG\&"}$, $S_2 = \text{"1234"}$, $S_3 = \text{"\#PKU\#"}$, $S_4 = \text{"THU"}$, 则 $\text{concat}(\text{replace}(S_1, \text{substr}(S_1, \text{length}(S_2), \text{length}(S_4)), S_4), \text{substr}(S_3, \text{index}(S_2, \text{"2"}), \text{length}(S_4)))$ 为 _____。

7. 对于入栈序列 1, 2, 3, ..., n, 按照每入栈 3 次后出栈 1 次, 无元素可入栈后, 栈内元素依次出栈。 1) 当 $n=10$ 时, 这样操作下得到的出栈序列为 _____; 2) 若 n 为 3 的倍数, 那么实现此过程所需的栈至少应可容纳 _____ 个元素。

8. 依顺序输入数组 {70, 65, 43, 28, 59, 36, 10, 19}, 利用筛选法建立最小值堆。进行两次 deleteMin 操作后, 得到的最小值堆的数组表示为 _____。

9. 一棵树按照先根次序遍历的结点序列为 SABEFGCDHIJ, 后根次序遍历该树为 BEGFACHJIDS, 那么 E 结点的兄弟结点有 _____。

10. 如图所示的森林 T 有三棵树, 该森林转化为二叉树后, 与结点 G 深度相同的结点有 _____。



得分

二、 辨析与简答(共 7 题，共 36 分)

- (2 分) 假设两种算法用来解决同一问题，其中一个运行时间为 $T_1(n) = 400n$ ，另一种为 $T_2(n) = n^2$ ，请分析 n 取何值时，两种算法哪个更为高效？
- (4 分) 请解释何为栈的“上溢”和“下溢”？何为队列的“假溢出”及如何避免？
- (9 分) 请分别计算模式 $p = \text{"aabaac"}$ 优化前和优化后的 next 数组，并按照优化后的 next 数组针对目标 $t = \text{"aabaabaabaac"}$ 进行 KMP 快速模式匹配，请画出匹配过程的示意图并计算匹配过程中的比较次数。
- (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。
 前序序列：_ _ C D E _ G H I _ K
 中序序列：C B _ _ F A _ J K I G
 后序序列：_ E F D B _ J I H _ A
- (5 分) 一棵哈夫曼树由字符 $abcde$ 组成，它们的出现频率满足 $a > b > c > d > e$ ，已知字符 e 的编码为 1001。请将电文 10001010100111 解码成字符串。
- (4 分) 试证：一个具有 $n(n > 1)$ 个结点的 m 叉树有 $n(m-1)+1$ 个空指针。
- (6 分) 亲戚关系具有自反性、对称性和传递性。下面给出了几组亲戚关系：
 $(1, 2), (3, 4), (3, 5), (1, 7), (3, 6), (8, 9), (3, 11), (3, 12), (3, 13),$
 $(14, 15), (16, 10), (14, 16), (1, 3)$
 (1) 请把上述 1-16 编号的人群归为若干个家族，使得家族内部都是亲戚，而家族之间不存在亲戚关系。
 (2) 假设一个小城镇有 10 万人口，已知亲戚关系有 30 万。请问你采用什么数据结构来辅助进行上述家族关系划分？需要采取什么措施来保持这个数据结构的高效。

得分

三、 算法填空(每空 2 分，共 10 分)

将下列算法补充完整，使其能将一个用带双标记的层次表示法表示的森林转换为左子结点/右兄弟表示。

// 注：代码中为了简洁将 top()和 pop()合并为了 Pop()，即读栈顶并出栈。

```
template <T>
struct DualTagTreeNode <T> {
    int ltag;          // 左标记
    T info;
    int rtag;          // 右标记
}; // 带双标记的层次表示法的结点类

struct TreeNode<T> {
public:
    T info;
    TreeNode* Child;    // 指向最左子结点的指针
    TreeNode* Sibling;  // 指向右兄弟的指针
}; // 左子/右兄弟结点中的结点类

TreeNode<T>* Convert (DualTagTreeNode* nodes, int size) {
// nodes 为带双标记的层次法表示的森林的序列，size 为序列长度。
    queue <TreeNode <T>*> aQueue;
    TreeNode<T> *pointer = new TreeNode<T>;
    root = pointer;          // 根结点
    for (int i=0; i<size-1; i++) {
        pointer->info = nodes[i].info;
        if(nodes[i].ltag == 1)
            _____ 填空 1 _____ ;
        else
            _____ 填空 2 _____ ;
        TreeNode<T>* temppointer = new TreeNode<T>;
        if(nodes[i].rtag == 0) {          // 有右兄弟，其后的结点即为其右兄弟
            pointer->Sibling = temppointer;
        }
        else {          // 否则，Sibling 为空并出队列
            pointer->Sibling = NULL;
            pointer = aQueue.Pop();
            _____ 填空 3 _____ ;
        }
        _____ 填空 4 _____ ;
    }
}
```

```

    }
    // 处理最后一个结点
    填空 5 ;
    pointer->Child = NULL;
    pointer->Sibling = NULL;
}

```

得分

四、 算法设计与实现（每题 10 分，共 20 分）

1. 请设计一个算法，判断输入的一个由 $1, 2, \dots, n$ 组成的排列是否可以通过入栈、出栈操作得到有序序列，并输出相应入栈出栈序列。请给出算法伪代码，并分析算法的时间复杂度。
例如：
输入序列 4 3 1 2；
输出 Yes; push push push pop push pop pop pop
2. 请设计一个算法对树序列化和反序列化（所谓序列化是将一棵树用一个字符串表示；而反序列化则为根据字符串恢复原本的树结构）。此处对序列化的格式没有限制，可以自定义。
可写伪代码，需要相应的注释；或者写出详细算法设计思路。

得分

五、 分析证明题（10 分）

设一棵度为 k 的非空树上的叶子结点数为 n_0 ，度为 i 的结点数为 $n_i (1 \leq i \leq k)$ 。试证明以下关系成立：

$$n_0 = 1 + \sum_{i=1}^k (i-1) * n_i$$