

59.7.11.20
四川大学计算机学院 2004 年研究生复试题

四川大学计算机学院 2004 年研究生复试题

考试科目: 数据结构

一、单项选择题 (在每小题的四个备选答案中, 选出一个正确的答案, 并将其号码填在题干后的括号内。每小题 2 分, 共 20 分)

1. 数据结构可形式地定义为 (D, S) , 其中 S 是 D 上 () 的有限集。

- A) 操作 B) 存储映象 C) 关系 D) 数据元素

2. 表达式 $x*(y-z)+u$ 的逆波兰式表示是 ()

- A) $xyzau++$ B) $xyz-u+$ C) $xyz*-u+$ D) $+-*xyzu$

3. 已知一个栈的入栈序列是 $a_1, a_2, a_3, \dots, a_n$, 其输出序列为 $a_{p_1}, a_{p_2}, a_{p_3}, \dots, a_{p_n}$

若 $p_1=n$, 则 p_i 为 ()。

- A) i B) $n-i$ C) $n-i+1$ D) 无法确定

4. 从一个具有头结点的单链表中查找数据元素值为 x 的结点时, 在查找成功的情况下,

平均比较次数是 ()

- A) n B) $n/2$ C) $(n-1)/2$ D) $(n+1)/2$

5. 在数组 A 中, 假设每个元素的大小为 3 字节, 行下标从 0 到 8, 列下标从 0 到 9, 从首地址为 1000 的位置开始以列序为主序连续存储于存储器内, 则元素 $A[3][6]$ 的起始地址是

- A) 1099 B) 1108 C) 1171 D) 1189

6. 广义表 $((a,b),(c,d))$ 的表尾是 ()。

- A) (c,d) B) $((c,d))$ C) (d) D) d

7. 设某二叉树前序为 $abdcfe$, 中序为 $dbaecf$, 则此二叉树的后序为 ()。

- A) $dbefca$ B) $debfea$ C) $dfebeca$ D) $dbfeca$

8. 采用邻接表存储的图按深度优先搜索方法进行遍历的算法类似于二叉树的 ()。

- A) 先序遍历 B) 中序遍历 C) 后序遍历 D) 层次遍历

9. 采用分块查找时, 如某线性表中共有 256 个元素, 查找每个元素的概率相同, 假设采用顺序查找来确定元素所在的块, 则每块包含 () 个结点时, 平均查找长度最小。

- A) 256 B) 15 C) 16 D) 18

10. 倒排文件的最大优点是 ()

- A) 便于进行文件的归并
B) 有利于文件的插入与删除
C) 能大大地提高主关键字的查找速度
D) 能大大地提高次关键字的查找速度

多重索引、以主关键字为索引

二、简答题 (每小题 5 分, 共 45 分)

1. 有如下数据结构的形式定义, 试画出此结构的图形表示。

$DS = \{D, S\}$, 其中

$D = \{1, 2, 3, 4\}$

$S = \{R\}$

$R = \{ \langle 1, 2 \rangle, \langle 1, 3 \rangle, \langle 2, 3 \rangle, \langle 2, 4 \rangle, \langle 3, 4 \rangle \}$

假设用如下两种结点来描述广义表的存储结构:

表结点: $\text{tag}=1 \quad \text{hlink} \quad \text{rlink}$

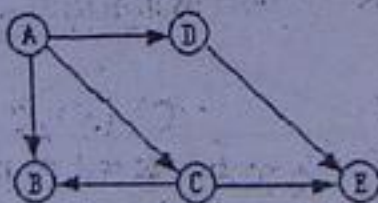
元子结点: $\text{tag}=0 \quad \text{data}$

画出如下广义表的存储结构:

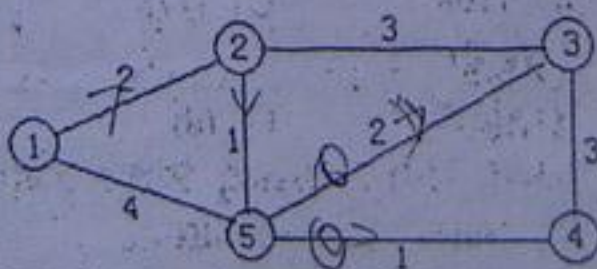
$(((), (a)), (b, c))$

3. 已知 Huffman 树的叶结点数为 m , 试写出此树的结点总数。

4. 给出下图的所有拓扑有序序列。

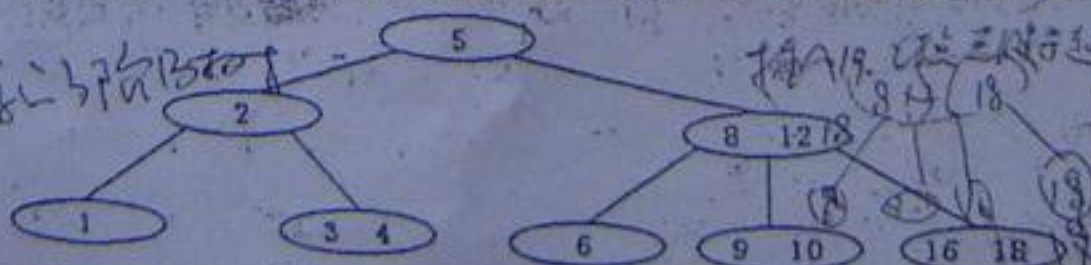


5. 对于下图, 用 Prim 算法从结点 1 出发构造出一棵最小生成树, 要求图示出每一步的变化情况。



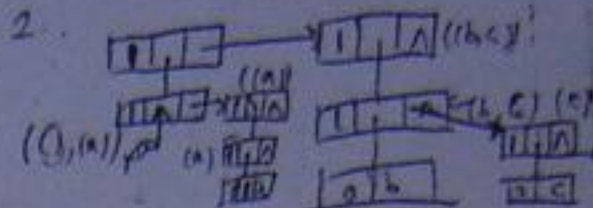
6. 设有 n 个不同的英文单词, 这些单词是无序的, 它们的长度都相同, 均为 d (并且 $d \leq 10$), 设 $n \gg d$, 试问使用哪种排序方法的时间复杂度最佳, 并说出理由。

7. 有下图所示的一棵 3 阶 B-树, 试画出插入关键字为 19 的结点之后的结果。



8. 对于下图从顶点 d 出发分别按深度优先搜索与广度优先搜索方法进行遍历, 写出相应的顶点序列。

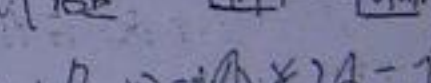
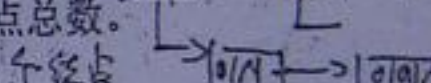
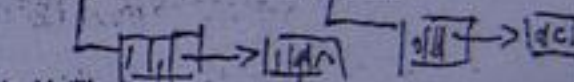
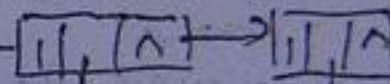
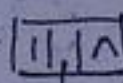
$(((), (a)), (b, c))$



有向与无向

(v_i, v_j)

$\langle v_i, v_j \rangle$



书本

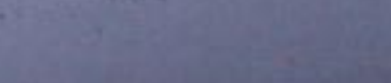
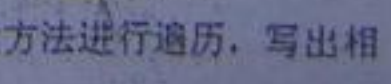
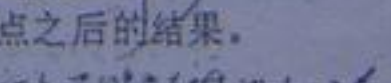
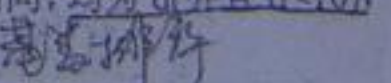
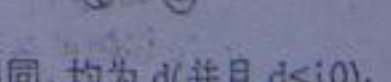
2n-1 个结点

依次输出所有前驱的结点

拓扑序列
5 个拓扑序列

图 4

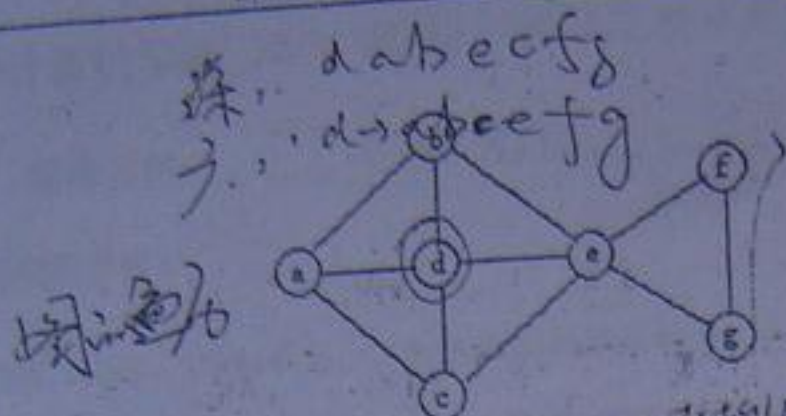
10-10 步



30/11/1:37AM

二叉树？

四川大学计算机学院 2004 年研究生复试题



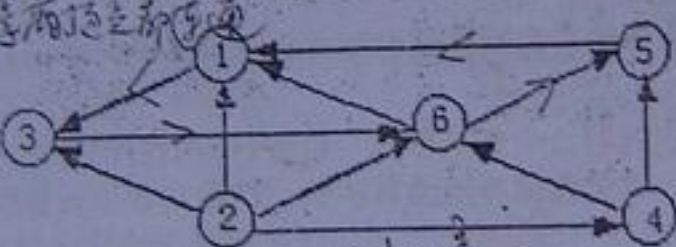
9. 已知稀疏矩阵如下图所示，试写出三元组表示

$$\begin{bmatrix} 15 & 0 & 0 & 0 & 16 \\ 0 & 6 & 0 & 0 & 0 \\ 0 & 0 & 8 & 0 & 0 \\ 9 & 0 & 0 & 18 & 0 \end{bmatrix}$$

a	1	2	3
0	4	5	6
1	1	1	15
2	1	5	16
3	2	2	6
4	3	3	8
5	4	1	9
6	4	4	18

三、应用及算法题 (共 35 分)

1. (9 分) 对下面所示的有向图 G 试给出:



(1) 各顶点的入/出度

(2) 拓扑排序

(3) 强连通分量

2. (8 分) 已知哈希表地址空间是 0-8，哈希函数是 $H(k) = k \text{ MOD } 7$ ，采用线性探测再散列处理冲突，将序列 (100, 20, 21, 35, 3, 78, 99, 45) 中的数据依次存入到此哈希表中，列出插入时的比较次数，并求出在等概率情况下的平均查找长度。

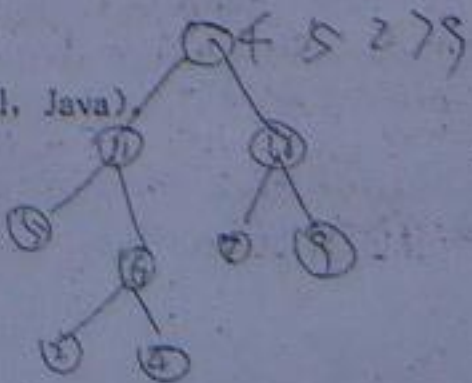
3. (9 分) 设有一份电文中共有 5 个字符: a, b, c, d, e，它们出现的次数分别是: 4, 5, 2, 7, 9，试画出对应的 Huffman 树 (要求任一结点的左孩子结点的权值小于等于右孩子结点的权值)，并写出每个字符对应的 Huffman 编码。

4. (9 分) 试写出按层次遍历二叉树的算法。

注: 算法可用任何语言实现 (如 C 语言、C++、Pascal、Java)。

$H(100) = 100 \text{ MOD } 7 = 2$
 $H(20) = 20 \text{ MOD } 7 = 6$
 $H(21) = 21 \text{ MOD } 7 = 0$
 $H(35) = 35 \text{ MOD } 7 = 0$
 $H(3) = 3 \text{ MOD } 7 = 3$
 $H(78) = 78 \text{ MOD } 7 = 1$
 $H(99) = 99 \text{ MOD } 7 = 1$
 $H(45) = 45 \text{ MOD } 7 = 3$

$H(100) = 100 \text{ MOD } 7 = 2$
 $H(20) = 20 \text{ MOD } 7 = 6$
 $H(21) = 21 \text{ MOD } 7 = 0$
 $H(35) = 35 \text{ MOD } 7 = 0$
 $H(3) = 3 \text{ MOD } 7 = 3$
 $H(78) = 78 \text{ MOD } 7 = 1$
 $H(99) = 99 \text{ MOD } 7 = 1$
 $H(45) = 45 \text{ MOD } 7 = 3$



30 11:37AM

四川大学计算机学院 2004 年硕士研究生复试题 (150 分钟)

姓名_____ 报考方向_____ 可调配方向_____ 总分_____

试卷与答卷一起交回存档

笔试部分 共7题, 60分。按知识点、, 分析和解题步骤判分, 尽力解答, 不要弃题。

题目: 分数	1题: 5分	2题: 10分	3题: 5分	4题: 10分	5题: 10分	6题: 10分	7题: 10分	总计 60分
得分	20分	10分	10分	10分	10分	10分	10分	

三. (5分) 谈谈你对报考研究生的认识和入学后的打算。

(10分) 用100—200字描述一项你自己的科研实践或科研作品或成果 (作过的软、硬件习
作, 发表的文章, 参加竞赛的作品、获奖成果, 以上各项满分为10分), 要求真实, 对于获
奖, 要求在面试时提供证明材料和证明人。如果没有已经作出的作品, 可以谈你准备作的科技
作品的创意、方案和措施 (满分5分)。 (本题考察平时科研实践和表达能力)

三. (5分)

(1) 你学习过哪些程序设计语言? 这些程序设计语言的共同的要素有哪些?

(2) 基于上述要素, 选择下列问题之一回答:

(a) 如果现在要自学一个新的程序设计语言, 你将采取什么样的方法来学习?

(b) 如果你是讲授程序设计语言的教师, 你怎样组织你的教学?

(考察归纳、领悟能力, 从学会到会学的能力)

四. (10分) 假定某高校要建立一个学生通过Web选课的系统, 以数据库为后台, 以Browser-
Server方式实现。请根据你的常识、分析能力和想像力, 回答下列问题 (400字左右):

(1) 分析该系统的特殊困难。 (3分)

(2) 给出解决特殊困难的方案 (3分)

(3) 列出你需要的硬件和软件平台, 大致的系统框图。 (4分)

(考察把握项目要点、方向以及写项目中请的能力)

五. (10分) (考察构思算法的能力) 设下列函数已经通过了调试:

bool Sort_Array(ArrayType * pInputArray, ArrayType* pOutArray);

该函数在内存中排序, 能把字节数最大为100M字节的ArrayType类型的数组排序。其中
ArrayType 是一个预定义的数组类型 (其细节无关紧要)。pInputArray和pOutArray分别是指向
排序前和排序后数组的指针。请用C语言风格的伪码设计一个算法, 它调用上面给出的函数, 完
成下列从输入到输出的任务:

输入: 排序前的大文件, 名称为 char * pInputFileName: 其内容为用分号分隔的ArrayType
的数组的元素, 可装满4个100M字节的数组。

输出: 排序后的大文件 char* pOutputFileName。

数据库常识 ↑

SQL

六 (10分) 研究生入学考试没有考察数据库课程的科目, 但下列最基本的数据库常识是计算机专业硕士生入学前应该具备的知识: 设学生选课数据库有以下关系模式: * 为注释符号

S(StudentNum, StudentName, Age, Department) *学生(学号, 姓名, 年龄, 所在系)

Enroll(studentNum, CourseNum, Score) *选课(学号, 课程编号, 成绩)

Corse (CourseNum, CourseName) *课程(课程编号, 课程名)

对下列每一个问题, 写出关系代数表达式和 SQL 查询程序:

(1) 查询学号为 SNum 的同学所选修课程的课程名及成绩: (3分)

(2) 查询选修 "CS415" 号课程的学生中成绩最高的学生的学号: (3分)

(3) 查询未选 "计算机网络" 课程的学生姓名及所在系: (4分)

七 (10分) 本题分3个小题, 各小题单独计分, 即使前面的小题没有作, 后面的小题仍然可以引用前面小题的结果。(本题考察分析问题、解决问题的能力 and 基本编程技能, 要求程序思路清楚、大框架正确、对文件操作和指针的概念清楚, 请加上必要汉字注释)。

题目背景 比较文件字节数是检查计算机病毒的一种方法, 其简化版只处理当前目录, 思路如下: 在当前目录下, 在系统没有病毒时获取该目录中未感染病毒文件的(名称, 字节数)数据, 保存在该目录下名称为 GoodSize.TXT 的文件中, 需要检查时, 获取待检查文件字节数, 与保存的字节数对比, 如果字节数变化, 报告疑似病例。

可利用函数: 设 void GetFileNameList(char *DirectoryName, char* pNameListFile) 是一个已经开发好的函数, 它能将名称为 DirectoryName 的目录中的所有文件名写在名称为 *pNameListFile 的文本文件中, 形如

ABC.EXE 111111 字节

CDE.COM 222222 字节

FGH.DLL 333333 字节

请完成下列工作:

用 C 语言或 C++ 写出下列函数(如用 C++, 函数形式可根据需要作适当调整)

(1) (2分) int GetFileSize (Char * pFileName);

返回值为文件 *pFileName 的当前实际字节数。

(2) (4分) CheckOneFile (Char * pFileName); 对比文件 *pFileName 的当前字节数和 GoodSize.TXT 中保存的该文件的字节数, 如果不一致, 报告疑似病例。

(3) (4分) void main();

它运行时, 生成如下菜单:

- ```
=====
1 在系统无病毒时, 保存当前目录中所有文件字节数于 GoodSize.TXT 中
2 怀疑有病毒时, 对指定文件, 通过比较字节数检查疑似病例。
3 退出
=====
```

请输入 (默认为 2):

要求在用户键入 1 或 2 时, 执行相应操作, 执行后又回到菜单界面, 只有键入 3 才退出, 按其它键按默认值 2 处理, 不允许 goto 语句。