**数据结构研讨课报告**

1. **二叉排序树**

**1.1 题目描述**

它或者是一棵空树;或者是具有下列性质的二叉树:

1. 若左子树不空，则左子树上所有结点的值均小于它的根结点的值
2. 右子树不空，则右子树上所有结点的值均大于它的根结点的值
3. 左、右子树也分别为二叉排序树

请思考:

(1)请给出一棵二叉排序树;

(2)如何利用该二叉排序树得到一个递升序列?

(3)如何利用该二叉排序树得到一个递减序列?

**1.2题目理解与分析**

1. 根据题目要求，可以构造出一棵简易二叉排序树如下面解答部分的图所示；
2. （3）分析时可以将问题简化，简化成只有三个”结点“的二叉树，总是左孩子<根节点<右孩子。由递归的思想可以得到，如果要得到一个递减的序列，就先访问左孩子，再访问根结点，最后访问右孩子，即中序遍历；递减序列则先访问右孩子，再访问根节点，最后访问左孩子，即逆中序遍历。

**另：经过小组讨论和课堂交流，递归虽然程序简单易写，但在时间复杂度上不太友好。如果需要同时得到递减和递升序列，没必要两次遍历。只需一次中序遍历即可，然后同时将输出的递减序列存储在栈中，然后从栈顶直接输出即可得到递升序列。**

**1.3问题解答**

1. **一棵二叉排序树可以是如下：**

7

9

6

4

8

5

**(2)&&(3)算法思想：**

1. 采用基本的中序遍历（递归)算法，输出该二叉排序树的递减序列；
2. 输出递减序列时，每输出一个数据，同时将其压入栈中；
3. 重复步骤2，直到遍历完毕；
4. 依次从栈顶弹出存储的值，即得到递升序列。

**逻辑结构：树和栈**

**存储结构：链式存储**

1. **树形结构在文件管理中的应用**

**2.1题目描述**

(1)怎样利用树形结构来管理文件目录，并能够将文件和文件夹加以区分;

(2)如何统计一个节点下文件夹和文件的数目;

(3)从目录树的管理上看，要实现文件夹或文件的删除、复制、移动，请描述算法实现的思路:

(4)地址路径和目录树结构怎么映射，给定一个地址路径(例如: D:\ProgramFiles\Microsoft Office\office14)，怎么实现定位。反之，给定一个节点，获得相应路径地址。请描述算法实现的思路。

**2.2题目理解与分析**

**不考虑空文件夹的情况下，**可以简单地认为--当一个结点没有孩子时，是文件；如果有孩子，则是文件夹。但是**如果考虑空文件夹**，以上想法则不能区分文件和空文件夹。本题接下来的分析解答是考虑了空文件夹的情况。

因此，在一般树结点的结构定义中，增设一个“类型域”，规定如果是1，代表是文件夹；如果是0，规定为文件。（当然，也可以规定其他值对应分类的情况。）

**2.3题目解答**

1. **算法思想：**
2. 树结点的定义中设置：data域、左孩子指针、右孩子指针、type域（类型域）。规定type为1时是文件夹；type为0时是文件。
3. 构建文件树。从根目录开始，递归地添加文件夹或文件结点作为子结点。

**逻辑结构：树**

**存储结构：链式存储**

1. **算法思想：**
2. 以指定的目标结点为根节点，递归遍历其所有子结点。
3. 遍历过程中访问type域，增设两个变量，分别记录文件夹和文件的个数。

**逻辑结构：树**

**存储结构：链式存储**

1. **算法思想：**
2. 删除：

·1遍历找到目标结点；

·2访问其所有子结点的type域，如果是1，则递归地删除其所有子节点；如果是0，直接删除即可。

1. 复制：

·1遍历找到目标结点；

·2创建一个新结点，复制目标结点所有的域值；

·3访问其type域，如果是1，则递归地新建-复制其所有子结点；如果是0，复制已完成。

1. 移动：

·1执行上述复制操作，复制到目标位置；

·2执行删除操作，删除原来的部分。

**逻辑结构：树**

**存储结构：链式存储**

1. **算法思想：**
2. 给定地址，定位结点：

·1对给定地址分层（如 D:\ProgramFiles\Microsoft Office\office14，划分为D：、ProgramFiles、Microsoft Office、office14）；

·2由分析可得，目录树的每一层对应一部分的地址。故从根结点开始，遍历各个孩子，找到该部分地址对应的孩子；

·3重复步骤1，直到找到对应的结点。

1. 给定结点，得到路径：

·1从目标结点开始，将目标结点的地址压入栈中，递归地向上回溯其父结点；

·2将其父节点的地址也压入栈中，重复步骤1，直到访问到根节点；

·3将栈中存储的地址以此弹出，则得到了所求的地址路径。

**逻辑结构：树和栈**

**存储结构：链式存储**

1. **问题三**

**3.1题目描述**

有一千万条短信，有重复，以文本文件(ASCII)形式保存，一行一条，请找出重复出现最多的前10条。

**3.2题目理解与分析**

由于数据量非常大，高达一千万条，不可能每一条都再和其他所有逐个去比对是否有重复，这样的话时间消耗太大。由于ASCII码由128个字符组成，所以完全可以通过字符结点的连接实现信息的存储。详细的解答如下。

**3.3问题解答**

1.先随机选取5条短信进行遍历。（当然这个遍历的短信数量是随机取的，不需要太大，只是起到一个提供初始集的作用。）假设此时构造的链如下：（之所以称为链是因为还没有组装成树）**（注意，每个链的最后一个结点要再跟一个数据域为空的结点，用于计数）**

@

9

1

··· ···

5

1

\*

5

1

9

8

0

7

1

（结点内的字符是随机写的，可以相同可以不同，均为ASCII码集里的字符就行）

2.接着遍历其他短信，如果有部分和已存在的链重复，则不需再重新构造，只需要在不同处重新开一个孩子结点即可。如右侧所示：

5

7

7

1

3.在构造树时，某条短信的链如果和已有的链完全重复，则其最后跟的计数器便+1，以此来记录重复的次数。

4.一次遍历完，即可构造出所有短信的链。此时创建一个新节点作为根节点，将每条信息的首字符结点作为根节点的孩子结点连接起来，即可完成树的构建。

5.构建好树以后，遍历整棵树，得到每个计数器的值；

6.构造一个辅助数组，将计数器的值放入数组中。当数组中的元素个数小于10时，计数器的值直接进入数组；当数组中的元素个数达到10时，再得到的计数器的值与这10个相比较，将比它小的移出数组，将大的放入数组；

7.最终数组内存放的十个值就是重复次数最多的前十条。

**逻辑结构：树**

**存储结构：链式存储**