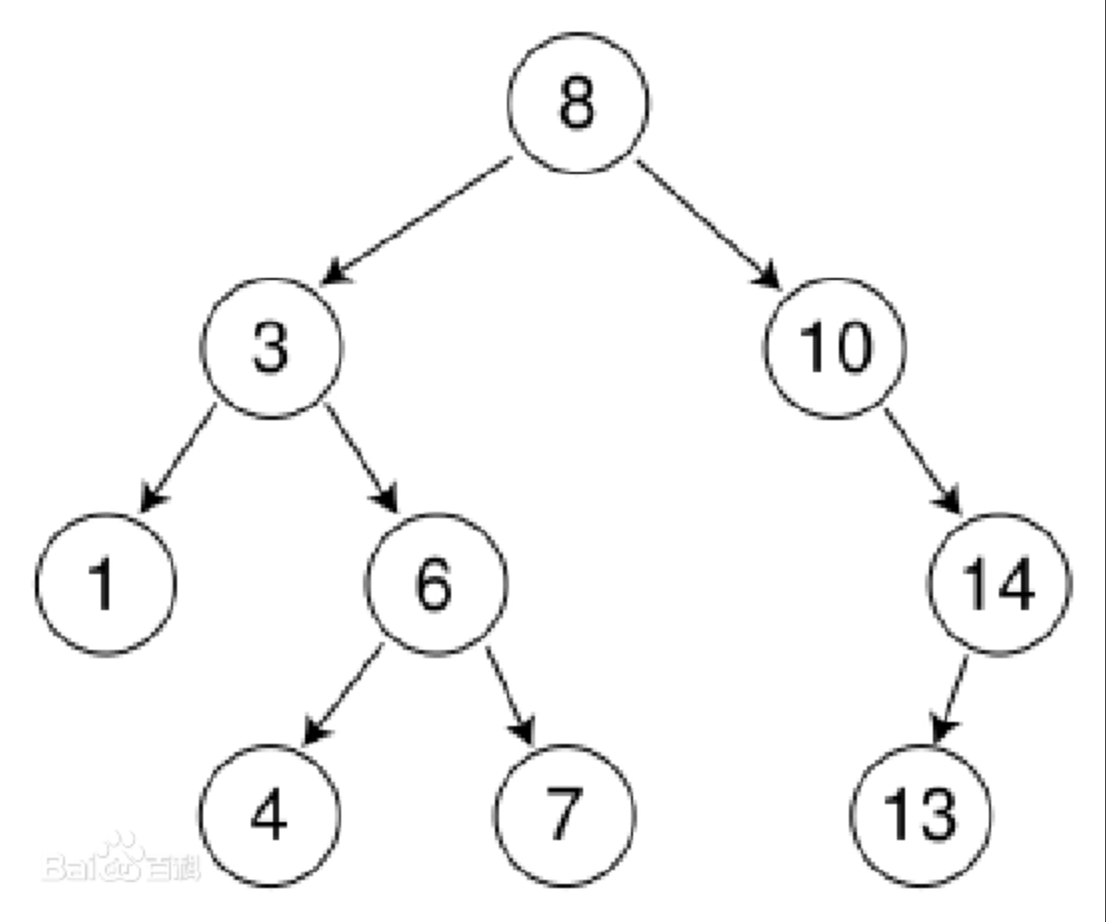
**第二次研讨实验报告**

姓名：陆明奇 学号：2050283 日期：2021 年 11 月 11 日

**1. 实验内容**

**1.1 二叉排序树**

1. **请给出一棵二叉排序树**



1. **如何利用该二叉排序树得到一个递升序列**

中序遍历该二叉排序树

1. **如何利用该二叉排序树得到一个递减序列**

先遍历二叉排序树的右子树，再遍历树的跟节点，最后遍历该二叉排序树的左子树，嵌套执行以上步骤便得到了一个递减序列（逆中序遍历）。由于逆中序遍历实现起来较为复杂，因此可以先用中序遍历，获得一个递升序列后通过栈输出的方式将其倒置成递减序列。

**1.2 树形结构在文件管理中的应用**

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**

在树的节点中加入一个bool型标记位来判断是文件还是文件夹。

1. **如何统计一个节点下文件夹和文件的数目**

遍历树的所有节点。对每个节点，通过判断其标记位来辨别是文件还是文件夹，从而对文件和文件夹进行计数。

1. **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**

如果要实现文件的删除，复制，移动就必须先在树中找到需要操作的节点，然后进行相应的操作，删除则是将该节点删除，复制则是在树的另一个节点上赋值该节点的信息，移动则是在另一个节点上赋值该节点的信息并删除该节点。若操作对象是文件夹的话，就必须将此文件夹的子树完全遍历一遍，再逐一操作。

1. **地址路径和目录树结构怎么映射，给定一个地址路径，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。**

定位时，按string类读入地址路径，从头开始扫描此路径，从一个‘\’到下一个‘\’之间的内容作为查找对象在当前根的孩子中遍历查找，将找到的孩子作为根重复之前的操作，就可以定位到目标文件。在给定一个节点时，从此节点往回查找其父亲节点，找到后把父亲节点的信息放入栈内，重复此操作直至找到根目录，将栈内的元素输出就可得到相应路径地址。

**1.3 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。**

（1）：哈希表

先将1千万条短信分成若干组，边扫描边建哈希表。第一次扫描，取关键字的首字符、尾字符、中间任意两字符的ASCII值之和用除留余数法得到其哈希地址，插入到哈希表中，并记录其地址、信息长度和重复次数。同哈希地址且等长就疑似相同，比较一下。相同记录只加1次进哈希表,并将重复次数加1。一次扫描以后，已经记录各自的重复次数。进行第二次哈希表的处理。可以用O(n)的复杂度完成前10 条的寻找。

（2）：按字数从小到大排序

根据经验，字数越少的短信出现重复的概率越高。可以从字数少的短信开始找起，比如一开始搜索一个字的短信，找出重复出现的top10 并分别记录出现次数。然后搜索两个字的，以此类推。对于相同字数的比较长的短信的搜索，除了用哈希表，还可以选择只抽取头、中和尾等几个位置的字符进行粗判。这样加快查找速度，但未必能得到真正期望的top10。因此还需要做标记，这样搜索一遍后，可以从各次top10结果中找到备选的top10。如果这次top10中有刚才做过标记的，则对其对应字数的所有短信进行精确搜索，以找到真正的top10并再次比较。

**2. 对教学和讨论课的建议**

目前教学多以老师讲解知识概念为主，但只理解概念本身并不能使我们很好地运用这些数据结构。希望老师上课时可以多讲一些例题，讲讲如何利用所学的数据结构解决这些题目，我想这样可以让我们对上课所讲的数据结构有更深入的理解。