

第三次作业

1. Dr.Stranger 的电脑染上了一种特殊的病毒，该病毒发作时会将字母用其他字母代替，但不会将单词顺序交换，也不会产生新的字母。在病毒发作前文档 D 中的单词顺序为字典排序（比如在字典序下单词 ab 会排在单词 ac 前），且文档 D 中所有的单词只由字母集合{a, b, c, d, e}中的字母组成，病毒发作后文档 D 中的所有单词顺序如下：{cebdbac, cac, ecd, dca, aba, bac}，求文档 D 在病毒发作前的内容，并说明解题思路和步骤。

2. Toole 教授提出了一种新的分治算法来计算最小生成树，该算法是这样的：给定一个图 $G=(V, E)$ ，将顶点集合 V 划分成两个集合 V_1 和 V_2 ，使得 $|V_1|$ 和 $|V_2|$ 至多差 1。设 E_1 为一个边集，其中的边都与 V_1 中的顶点关联， E_2 为另一个边集，其中的边都与 V_2 中的顶点关联。在两个子图 $G_1=(V_1, E_1)$ 和 $G_2=(V_2, E_2)$ 上，分别递归地解决最小生成树问题。最后，从 E 中选择一条通过割 (V_1, V_2) （两个端点分别在 V_1 和 V_2 上）的最小权边，并利用该边，将所得的两棵最小生成树合并成一棵完整的生成树。请论证该算法能正确地计算出 G 的最小生成树，或者给出一个使该算法不能正确工作的例子。

3. 给定数组 $A[1, \dots, n]$ ，在空间复杂度 $O(1)$ 的条件下实现归并排序（时间复杂度尽可能小）。给出算法流程，并分析时间复杂度。

4. 假设数组 $A[1, \dots, n]$ 中的各元素独立同分布，给定非严格单调递增函数 H ，将数组 $A[1, \dots, n]$ 的元素等概率地映射为 0 到 $m-1$ ($m \leq n$) 的 m 个整数。这等价于将数组 A 划分为 m 个桶，并且 $A[i]$ 落入各个桶的概率相等，即 $1/m$ 。然后分别对各个桶进行插入排序，最后按顺序扫描各个桶，将数组 $A[1, \dots, n]$ 递增输出。假设映射操作的时间复杂度均为 $O(1)$ ，1) 证明算法的正确性 2) 这个算法的最好、最坏和平均时间复杂度是多少？

5. 现在有 9 个长度不同的的顺串，其长度分别为 17, 40, 16, 55, 25, 11, 6, 21, 41。请用二路归并的方法对其进行归并。

(1) 构造最佳归并树。

(2) 根据最佳归并树计算每一趟和总的读记录次数。

(3) 在多路归并的时候，显然归并的路数 K 越大速度越快，那么是什么限制了 K 的大小？

6. 胜者树和败者树都是完全二叉树，是树形选择排序的一种变型。每个叶子结点相当于一个选手，每个中间结点相当于一场比赛，每一层相当于一轮比赛。不同的是，胜者树的中间结点记录的是胜者的标号；而败者树的中间结点记录的败者的标号。

胜者树与败者树可以在 $\log(n)$ 的时间内找到最值。任何一个叶子结点的值改变后，利用中间结点的信息，还是能够快速找到最值。在 k 路归并排序中经常用到。

举例说明为什么败者树的访问内存次数要比胜者树少，并分析是什么原因造成的。

7. 1. 将关键字序列 (7、8、30、11、18、9、14) 散列存储到散列表中。散列表是一个下标从 0 开始的一维数组，散列函数为： $H(\text{key}) = (\text{key} * 3) \text{MOD } 7$ ，处理冲突采用线性探测法，要求装填（载）因子为 0.7。

(1) 请画出所构造的散列表。

(2) 分别计算等概率情况下查找成功和查找不成功的平均查找长度。

8. 给一个整数集合 S , 定义 S 的子集 D 为连续子集当且仅当 D 中的整数构成连续的整数序列。求 S 的最大连续子集, 即包含连续整数最多的子集。如 $\{1, 3, 4, 100, 200, 2\}$ 的最大连续子集为 $\{1, 2, 3, 4\}$ 。

9. 若一个磁盘页块大小为 1024 ($=1K$) 字节, 存储每个记录需占用 16 字节, 其中关键码占 4 个字节。所有记录均已按关键码有序存储在磁盘文件中, 每个页块的第 1 个记录用于存放线性索引。另外在内存中开辟了 256K 字节的空间用于存放线性索引。试问:

(1) 若将线性索引常驻内存, 文件中最多可以存放多少个记录? (每个索引项 8 个字节, 其中关键码 4 字节, 地址 4 字节)

(2) 若使用二级索引, 第二级索引占用 1024 字节 (有 128 个索引项), 这时文件中最多可以存放多少个记录?

10. 将下列字符序列 EASYQUESTION 依次插入到初始为空的红黑树 (RB-tree) 中, 请画出最终得到的红黑树。(用圆圈表示红色结点, 方框表示黑色结点, 外部空叶结点可省略不画)。

11. 画出下列广义表的图形表示和他们的存储表示。

(1) $D: (A: (c), B: (e), C: (a, L: (b, c, d)))$

(2) $J1: (J2: (J1, a, J3: (J1)), J3)$

12. 3 种顺序适配方法各有其使用范围。现假定存储区中有大小为 900、1500、700 的三块存储块。请给出符合以下几种情况的存储请求的例子, 并描述三种适配法的分配过程。

(1) 首先适配法可满足请求, 另外两种方法不可满足请求。

(2) 最佳适配法可满足请求, 另外两种方法不可满足请求。

(3) 最差适配法可满足请求, 另外两种方法不可满足请求。