

# 吉林大学《数据结构》上机实验题



Time Limit: 50MS

Memory Limit: 64MB

编写一个哈夫曼编码译码程序。针对一段文本，根据文本中字符出现频率构造哈夫曼树，给出每个字符的哈夫曼编码，并进行译码，计算编码前后文本大小。为确保构建的哈夫曼树唯一，本题做如下限定：

- (1) 选择根结点权值最小的两棵二叉树时，选取权值较小者作为左子树。
- (2) 若多棵二叉树根结点权值相等，则先生成的作为左子树，后生成的作为右子树，具体来说：  
i. 对于单结点二叉树，优先选择根结点对应字母在文本中最先出现者，如文本为 cba，三个字母均出现 1 次，但 c 在文本中最先出现，b 第二出现，故则选择 c 作为左子树，b 作为右子树。  
ii. 对于非单结点二叉树，先生成的二叉树作为左子树，后生成的二叉树作为右子树。  
iii. 若单结点和非单结点二叉树根结点权值相等，优先选择单结点二叉树。
- (3) 生成哈夫曼编码时，哈夫曼树左分支标记为 0，右分支标记为 1。

## 输入格式：

输入为 3 行。第 1 行为一个字符串，包含不超过 5000 个字符，至少包含两个不同的字符，每个字符为 a-z 的小写字母。第 2、3 行为两个由 0、1 组成的字符串，表示待译码的哈夫曼编码。

## 输出格式：

输出第一行为用空格间隔的 2 个整数，分别为压缩前后文本大小，以字节为单位，一个字符占 1 字节，8 个二进制位占 1 字节，若压缩后文本不足 8 位，则按 1 字节算。输出从第二行开始，每行为 1 个字符的哈夫曼编码，按各字符在文本中出现次数递增顺序输出，若多个字符出现次数相同，则按其在文本出现先后排列。每行格式为“字母:编码”。最后两行为两行字符串，表示译码结果，若译码失败，则输出 INVALID。

## 样例：

输入	输出
cbaxyyzz 0100 011	8 3 c:100 b:101 a:110 x:111 y:00 z:01 zy INVALID