

作业 2 树型结构及其应用

作业题目：哈夫曼编码与译码方法

哈夫曼编码是一种以哈夫曼树（最优二叉树，带权路径长度最小的二叉树）为基础变长编码方式。其基本思想是：将使用次数多的代码转换成长度较短的编码，而使用次数少的采用较长的编码，并且保持编码的唯一可解性。在计算机信息处理中，经常应用于数据压缩。是一种一致性编码法（又称“熵编码法”），用于数据的无损压缩。要求实现一个完整的哈夫曼编码与译码系统。

要求：

1. 从文件中读入任意一篇英文文本文件，分别统计英文文本文件中各字符（包括标点符号和空格）的使用频率；
2. 根据已统计的字符使用频率构造哈夫曼编码树，并给出每个字符的哈夫曼编码（字符集的哈夫曼编码表）；
3. 将文本文件利用哈夫曼树进行编码，存储成压缩文件（哈夫曼编码文件）；
4. 计算哈夫曼编码文件的压缩率；
5. 将哈夫曼编码文件译码为文本文件，并与原文件进行比较。

以下可以不做，供思考，做了可以适当加分

6. 能否利用堆结构，优化的哈夫曼编码算法。
7. 上述 1-5 的编码和译码是基于字符的压缩，考虑基于单词的压缩，完成上述工作，讨论并比较压缩效果。
8. 上述 1-5 的编码是二进制的编码，可以采用 K 叉的哈夫曼树完成上述工作，实现“K 进制”的编码和译码，并与二进制的编码和译码进行比较。

作业说明：

1. 上传内容：（1）源程序代码；（2）实验数据和实验结果数据
2. 上传格式：（1）打包为 rar 或 zip 文件；（2）命名规则：学号-姓名-作业编号，如 1170300100-张岩-作业 2.rar
3. 上传方法和网址：用 Google 或 Firefox 浏览器；网址：10.160.3.21:8080
4. 上传截止时间：另行通知