

实验 2 树形结构及其应用

实验项目：树形结构及其应用

实验题目：哈夫曼编码与译码方法

实验内容：

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**方法。其**基本思想**是：将使用次数多的字符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**。本实验要求实现一个完整的哈夫曼**编码与译码**系统。

实验要求：

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. 将哈夫曼编码文件译码为文本文件，并与原文件进行比较。
5. 计算你的哈夫曼编码文件的平均编码长度和压缩率，并与实验结果比较验证；
6. 选做：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。
7. 选做：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。
8. 选做：利用堆结构，优化哈夫曼编码算法。

实验说明：

1. 提交内容：（1）源程序文件；（2）实验测试数据和实验结果数据；（3）实验报告（参见“实验报告参考模板-DSA2023”）。
2. 提交格式：（1）打包为 rar 或 zip 文件；（2）命名规则：学号姓名-实验编号，如 2022110924 王梓诺-实验 2.rar。注意：必须用**一个且仅一个压缩**文件。
3. 提交方法：

同时发送至**老师**和**助教**邮箱：1444528362@qq.com 和 1797990956@qq.com

4. 提交截止时间：2023 年 10 月 15 日（第 7 星期日）23:59:59 之前