Project 2

Huffman Code

GroupID：17

Programmer：范若曦

Report Writer：牟傲

Tester：黄蕴

Date: 2019-12-1

**Chapter 1: Introduction**

本实验实现的主要功能是两个：压缩和解压缩。

1、压缩：输入文本文件将其通过Huffman Tree编码转化成二进制文件存储；

2、解压缩：输入通过Huffman Tree压缩的二进制文件，将其转化成为文本文件存储。

其中涉及Huffman Tree的构造，堆的构造，文件的读写，字符串的处理等各项算法。

**Chapter 2: Algorithm Specification**

* 主要数据结构设计说明

该程序定义了两个头文件，一个名为“CreateHeap.h”，用于存储构造最小堆的函数，其中数据主要以线性表的形式进行存储；另外一个名为“CreateHuffman.h”，用于存储Huffman Tree构造的函数，其中调用了“CreateHeap.h”，将构造出的新树存入线性表，而后存入构造好的堆中。

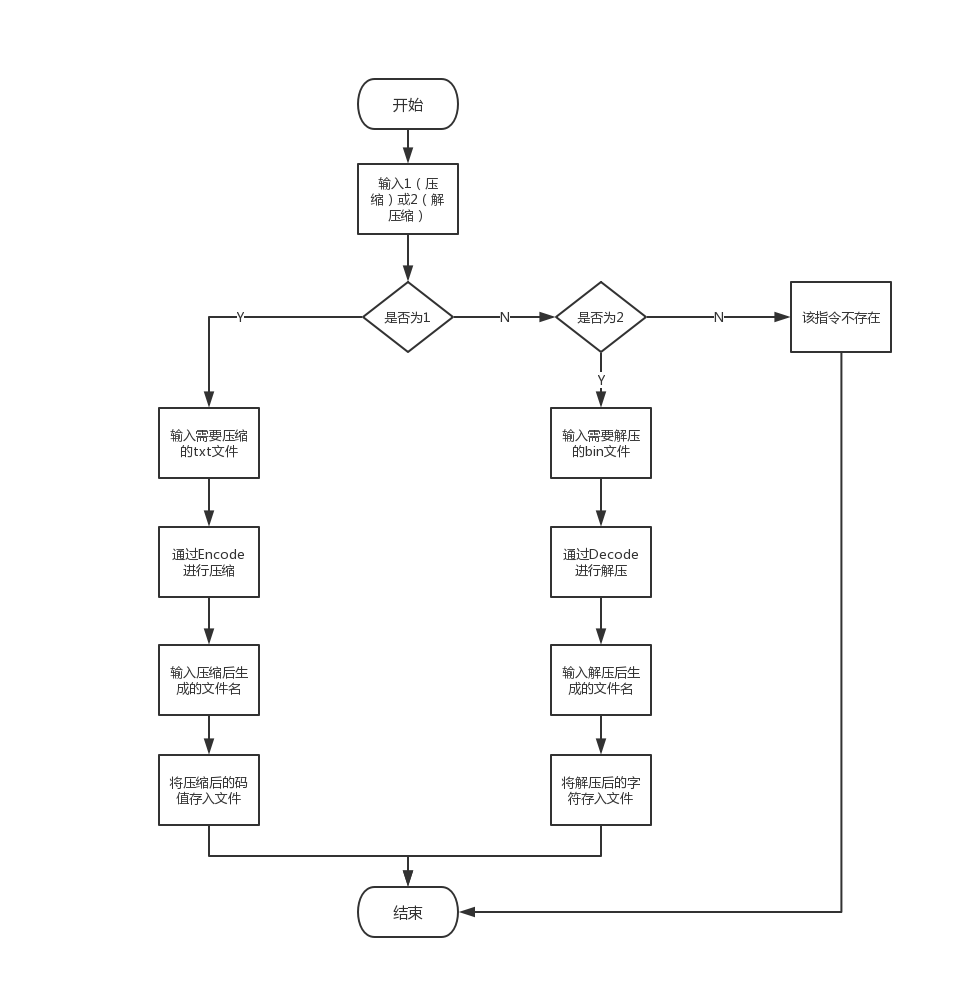
* 系统设计思想：

该程序系统主要分为两个大模块：Encode和Decode。

1. Encode：首先通过openFileAndEncode()来打开输入的txt文件，将文本读入数组，然后通过调用calWeight()计算文本中每一个字母的权重，通过build()进行Huffman Tree的构造，Encode()对消息进行编码，最后通过调用EncodeOutputFile()将编码以二进制文件的形式新建并存入输入的文件中，关闭文件。
2. Decode：首先通过openFileAndDecode()来打开输入的二进制文件，将读入的码值存入数组，然后调用findVal()来在堆中找到该码值对应的字母，并按顺序存入数组，最后调用DecodeOutputFile()将解码后的字符新建并存入输入的txt文件中，关闭文件。

每一个模块的函数都具有较高的可读性，结构清晰。

* 程序流程图

****

**Chapter 3: Testing Results**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **测试用例编号** | **测试的功能** | **系统初始状态** | **输入** | **期望的输出** | **实际输出** | **通过否？** |
| 1 | 实现编码压缩 | 等待输入待压缩文件名和压缩后文件名 | 待压缩文件：big.txt  压缩后文件：big.bin | 文件：big.bin | 文件：big.bin | Pass |
| 2 | 实现解码压缩 | 等待输入待压缩文件名和压缩后文件名 | 待压缩文件：hi.txt  压缩后文件：hi.bin | 文件：hi.bin | 文件：hi.bin | Pass |
| 3 | 实现编码压缩 | 等待输入待压缩文件名和压缩后文件名 | 待压缩文件：letter.txt  压缩后文件：letter.bin | 文件：letter.bin | 文件：letter.bin | Pass |
| 4 | 实现编码压缩 | 等待输入待压缩文件名和压缩后文件名 | 待压缩文件：letter2.txt  压缩后文件：letter2.bin | 文件：letter2.bin | 文件：letter2.bin | Pass |
| 5 | 实现编码压缩 | 等待输入待压缩文件名和压缩后文件名 | 待压缩文件：test1.txt  压缩后文件：test1.bin | 文件：test1.bin | 文件：test1.bin | Pass |
| 6 | 实现编码压缩 | 等待输入待压缩文件名和压缩后文件名 | 待压缩文件：test2.txt  压缩后文件：test2.bin | 文件：test2.bin | 文件：test2.bin | Pass |
| 7 | 实现编码压缩 | 等待输入待压缩文件名和压缩后文件名 | 待压缩文件：test3.txt  压缩后文件：test3.bin | 文件：test3.bin | 无输出 | Fail |
| 8 | 实现编码压缩 | 等待输入待压缩文件名和压缩后文件名 | 待压缩文件：test4.txt  压缩后文件：test4.bin | 文件：test4.bin | 文件：test4.bin | Pass |
| 9 | 实现编码压缩 | 等待输入待压缩文件名和压缩后文件名 | 待压缩文件：test5.txt  压缩后文件：test5.bin | 文件：test5.bin | 文件：test5.bin | Pass |
| 10 | 实现编码压缩 | 等待输入待压缩文件名和压缩后文件名 | 待压缩文件：test6.txt  压缩后文件：test6.bin | 文件：test6.bin | 文件：test6.bin | Pass |
| 11 | 实现编码压缩 | 等待输入待压缩文件名和压缩后文件名 | 待压缩文件：test7.txt  压缩后文件：test7.bin | 文件：test7.bin | 文件：test7.bin | Pass |
| 12 | 实现解码解压缩 | 等待输入待解压文件名和解压后文件名 | 待解压文件：big.bin  解压后文件：1big.txt | 与用例1的big.txt完全相同的1big.txt | 与用例1的big.txt完全相同的1big.txt | Pass |
| 13 | 实现解码解压缩 | 等待输入待解压文件名和解压后文件名 | 待解压文件：hi.bin  解压后文件：1hi.txt | 与用例2的hi.txt完全相同的1hi.txt | 与用例2的hi.txt完全相同的1hi.txt | Pass |
| 14 | 实现解码解压缩 | 等待输入待解压文件名和解压后文件名 | 待解压文件：letter.bin  解压后文件：1letter.txt | 与用例3的letter.txt完全相同的1letter.txt | 与用例3的letter.txt完全相同的1letter.txt | Pass |
| 15 | 实现解码解压缩 | 等待输入待解压文件名和解压后文件名 | 待解压文件：letter2.bin  解压后文件：1letter2.txt | 与用例4的letter2.txt完全相同的1letter2.txt | 与用例4的letter2.txt完全相同的1letter2.txt | Pass |
| 16 | 实现解码解压缩 | 等待输入待解压文件名和解压后文件名 | 待解压文件：test1.bin  解压后文件：1test1.txt | 与用例5的test1.txt完全相同的1test1.txt | 与用例5的test1.txt完全相同的1test1.txt | Pass |
| 17 | 实现解码解压缩 | 等待输入待解压文件名和解压后文件名 | 待解压文件：test2.bin  解压后文件：1test2.txt | 与用例6的test2.txt完全相同的1test2.txt | 与用例6的test2.txt完全相同的1test2.txt | Pass |
| 18 | 实现解码解压缩 | 等待输入待解压文件名和解压后文件名 | 待解压文件：test4.bin  解压后文件：1test4.txt | 与用例8的test4.txt完全相同的1test4.txt | 与用例8的test4.txt完全相同的1test4.txt | Pass |
| 19 | 实现解码解压缩 | 等待输入待解压文件名和解压后文件名 | 待解压文件：test5.bin  解压后文件：1test5.txt | 与用例9的test5.txt完全相同的1test5.txt | 与用例9的test5.txt完全相同的1test5.txt | Pass |
| 20 | 实现解码解压缩 | 等待输入待解压文件名和解压后文件名 | 待解压文件：test6.bin  解压后文件：1test6.txt | 与用例10的test6.txt完全相同的1test6.txt | 与用例10的test6.txt完全相同的1test6.txt | Pass |
| 21 | 实现解码解压缩 | 等待输入待解压文件名和解压后文件名 | 待解压文件：test7.bin  解压后文件：1test7.txt | 与用例11的test7.txt完全相同的1test7.txt | 与用例11的test7.txt完全相同的1test7.txt | Pass |

**Chapter 4: Analysis and Comments**

Analysis of the time and space complexities of the algorithms. Comments on further possible improvements.

* 算法分析

该算法具有正确性，能够满足对文件进行编码压缩和解码解压缩的需求；具有可读性，容易供人阅读和交流，方便理解和修改；算法具有健壮性，当输入错误数据时，能提示错误信息；算法具有一般性，对于一般的数据集合都成立。

* 算法特色

该算法是标准的huffman编码方式，计算了不同字母的权重并对其用数组进行储存，使用树构造了huffman tree；多使用while和for循环语句以及if选择分支语句实现函数的功能；通过ifstream和ofstream完成文件读写。算法排版清晰，多处注释，简洁易懂。

* 不足

编码和解码功能每实现一次就需要重新启动，使用的时候相对麻烦；使用数组存储字符，当字符种类过多的时候会有溢出的危险（比如对中文日文这些字符很多的文字进行解码和编码）；

**Declaration**

***We hereby declare that all the work done in this project titled "Huffman Code" is of our independent effort as a group.***

**Duty Assignments:**

**Programmer: 范若曦**

**Tester: 黄蕴**

**Report Writer: 牟傲**

Please keep in mind that these are the “minimum” requirements. Other requirements will be specified according to each project assignment.