石家庄铁道大学课程设计

压缩解压文件

**二○一九 届 信息科学与技术 学院**

**专 业 计算机科学与技术**

**学 号 20192163**

**学生姓名 崔金泽**

**指导教师 陈娜**

**完成日期 2021 年7 月7 日**

摘 要

我们都知道根据人类使用文字对应的每个字符都是有特定的频率的，比如说英文，一般来说字母a或者e的使用频率很高。如果我们能给出现频率最高的字符很短的编码，出现最少的字符最长的编码，而且保证每个编码都不是任意一个编码的前缀码，不会产生歧义，这样能保证根据压缩的编码来解压不会出现错误。如果我们用这样的编码来对一个文件进行压缩，那么出现最多的字符将会有最短的编码，可以节省很多的空间，将一个大的文件压缩成一个比较小的文件。而哈夫曼树就是一个可以用来生成这样一种编码的树，以树的叶子节点代表一个字符，从根节点到叶节点的路径形成一个编码，左子树表示比特1，右子树表示比特0，可以交换。这样每个字符对应一个唯一的编码而且任意一个编码都不是其他编码的前缀码。如果我们要压缩一个文件，需要先统计这个字符文件中每个字符出现的频率，然后根据频率来生成一颗哈夫曼树，使频率最高的字符对应于最短的编码。

**关键词**：哈夫曼编码 ;文件的解压缩;数据结构算法

目 录

[第一章 引言 1](#_Toc76452094)

[第二章 系统功能需求分析 1](#_Toc76452095)

[2.1 问题描述: 1](#_Toc76452096)

[2.2 需求分析 1](#_Toc76452097)

[3.1 抽象数据类型(ADT) 2](#_Toc76452098)

[第四章 详细设计 3](#_Toc76452099)

[4.1 重要程序段1 5](#_Toc76452100)

[4.2 重要程序段2 11](#_Toc76452101)

[4.3 重要程序段3 14](#_Toc76452102)

[4.4 重要程序段4 14](#_Toc76452103)

[第五章 调试分析 17](#_Toc76452104)

[第六章 运行结果 17](#_Toc76452105)

## 第一章 引言

数据结构是计算机存储、组织数据的方式。数据结构是指相互之间存在一种或多种特定关系的数据元素的集合。通常情况下，精心选择的数据结构可以带来更高的运行或者存储效率的算法。数据结构往往同高效的检索算法和索引技术有关。 (简要介绍数据结构课程)

本次课程设计，是利用哈夫曼树来实现压缩解压文件。

本文档从概要设计，详细设计，参考源程序代码以及显示程序运行的截屏等方面详尽的介绍了软件开发过程。

## 第二章 系统功能需求分析

2.1 问题描述:

实现一个基于哈夫曼树的文件压缩程序和文件解压程序。

基本要求：

(1) 要求压缩程序读入源文件，分析每种字符的频度，然后建立相应的哈夫曼树，

再求出相应哈夫曼编码，根据编码对源文件进行压缩，得到源文件对应的压缩文

件。

(2) 解压程序读入压缩文件，根据相应的哈夫曼编码解压还原，得到对应的源文件。

(3) 求出压缩率；

2.2 需求分析

压缩:对任意文件进行压缩。

解压缩:对以压缩的文件进行解压缩。

加密:对任意文件进行加密。(MD5加密)

解密:对以解密的文件进行解密。

第3章 概要设计

3.1 抽象数据类型(ADT)

**ADT Huffman**

{ **数据对象：D={ai|ai∈ElemTtpe,i=1,2,…,n,n≥0}**

**数据关系：R={<ai-1,ai>|ai-1,ai∈D，i=1，2，…,n}**

**基本操作：**

(1)Status creat(T a[], int size);

操作结果：创建哈夫曼树

(2) Status print();

初始条件：haffman树已存在

操作结果：输出哈夫曼树

(3) Status preOrder();

初始条件：haffman树已存在

操作结果：前序遍历哈夫曼树

(4) Status inOrder();

初始条件：haffman树已存在

操作结果：中序遍历哈夫曼树

(5) Status postOrder ();

初始条件：haffman树已存在

操作结果：后序遍历哈夫曼树

(6) Status destory();

初始条件：haffman树已存在

操作结果：删除哈夫曼树

}

3.2 存储结构

struct HuffmanTreeNode

{

HuffmanTreeNode<W>\* \_left;

HuffmanTreeNode<W>\* \_right;

HuffmanTreeNode<W>\* \_parent;

W \_w;//权值，也就是哈夫曼树的叶子节点的数据

HuffmanTreeNode(const W& w)

:\_left(NULL)

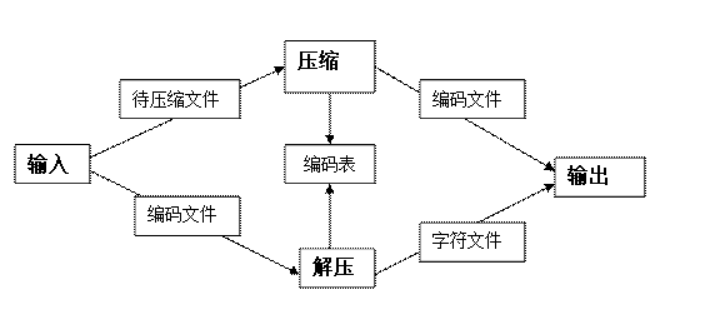
,\_right(NULL)

,\_parent(NULL)

,\_w(w)

{}

};



## 第四章 详细设计

项目主要思路：

1.统计：首先读取一个文件，统计出要压缩的文件中256个字符中各字符出现的次数，利用数据结构中的小堆来构建Huffman树，每次在堆顶选出两个出现次数较小的字符先进行构建，将它们相加的结果又放进堆里面，直到堆里面的数据被取完，这样字符出现次数多的离Huffman树的根节点就比较近，字符出现次数少的离Huffman树的根节点就比较远。

2.建树：根据构建好的Huffman树写出Huffman code，规定向左为“0”向右为“1”，这样字符出现次数多的Huffman code较短，字符出现次数少的Huffman code较长。

3.压缩：再次读取原文件，按照该字符对应的编码压缩文件；Huffman code刚好为“0”、“1”序列，所以可以每8位为一个字节写进压缩文件中，当余下的位数不够8位写不成一个字节时，需要在后面补0够一个字节再写入，这样就完成了文件的压缩，但要注意后面补0的位数为文件解压缩留下了隐患。

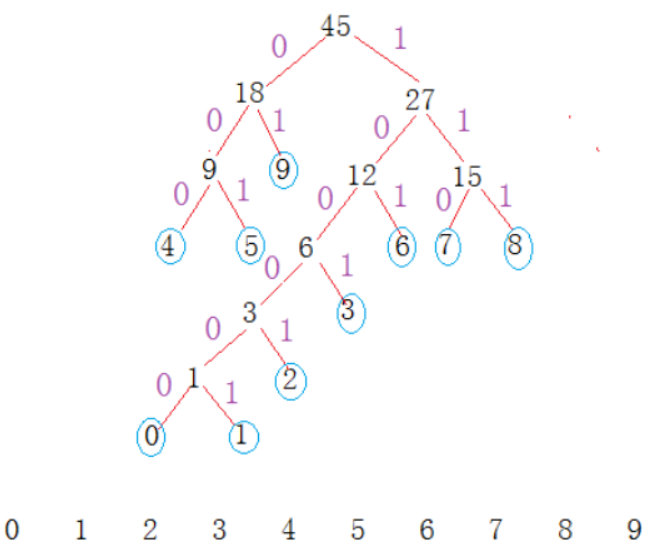
4.加工：将字符出现的总次数以及各个字符出现的次数写进配置文件，因为我们并不知道在进行压缩的时候在后面到底补了多少位0。

5.解压：利用压缩文件以及配置文件恢复原文件；

6.测试：首先对压缩和解压缩两个功能进行测试，看压缩后的文件是否比原文件小，以及解压缩文件与原文件的大小一致；在这个基础上再验证解压缩文件与原文件的内容是否保持一致，这里采用 beyond compare 软件来对文件内容进行验证，这样才验证了程序的大概功能是正确的。其次是压缩不同大小的文件计算出压缩所需要的时间来考量压缩的性能；

7.计算：经过多次压缩不同大小的文件不同类型的文件来计算文件的压缩率；

8.时间：使用 GetTickcount 函数计算完成压缩或者解压缩所需要的时间，GetTickcount 函数它返回从操作系统启动到当前所经过的毫秒数，常常用来判断某个方法执行的时间。



0->100000

1->100001

2->10001

3->1001

4->000

5->001

6->101

7->110

8->111

9->01

由图可以看到，权值越小的，它的Huffman编码越长，权值越大的，它的Huffman编码越短

假设一个文件，假设一个文件的内容是“abbcccdddd”，“a”出现的次数是1次，“b”出现的次数是2次，“c”出现的次数是3次，“d”出现的次数是4次，现在以各字符出现的次数构建一个Huffman树，并为各字符编码，则每个字符所对应的Huffman编码：

‘a’：100 ‘b’：101‘c’：11‘d’：0

根据每个字符的huffman编码对照原文件字符出现的顺序写出Huffman code：

10010110 11111110 000

Huffman code 一共19个比特位，每8位为一个字节写入压缩文件，当余下的不够一个字节时，在后面补0够一个字节再写入，这里需要补上5个0总共3个字节，而原文件是10个字节，少了7个字节，实现了压缩。

4.1 重要程序段1

#define \_CRT\_SECURE\_NO\_WARNINGS

#include <iostream>

#include <stdlib.h>

#include <stdio.h>

#include <string.h>

#include <ctype.h>

using namespace std;

struct head

{

int b,c; //字符,数

long count; //文件中该字符出现的次数

long parent, lch, rch; //make a tree

char bits[256]; //哈夫曼树编码

};

struct head header[512], tmp; //节点树

//函数：compress()

//作用：读取文件内容并加以压缩

//将压缩内容写入另一个文档

int compress(const char\* filename, const char\* outputfile)

{

char buf[512];

unsigned char c;

long i, j, m, n, f;

long min1, pt1, flength;

FILE\* ifp, \* ofp;

int per = 10;

ifp = fopen(filename, "rb"); //打开原始文件

if (ifp == NULL)

{

printf("打开文件失败:%s\n", filename);

return 0; //如果打开失败，则输出错误信息

}

ofp = fopen(outputfile, "wb"); //打开压缩后存储信息的文件

if (ofp == NULL)

{

printf("打开文件失败:%s\n", outputfile);

return 0;

}

flength = 0;

while (!feof(ifp))

{

fread(&c, 1, 1, ifp);

header[c].count++; //读文件，统计字符出现次数

cout << c;

flength++; //记录文件的字符总数

}

cout << endl;

flength--;

header[c].count--;

for (i = 0; i < 512; i++) //HUFFMAN算法中初始节点的设置

{

if (header[i].count != 0)

header[i].b = (unsigned char)i;

else

header[i].b = -1;

header[i].parent = -1;

header[i].lch = header[i].rch = -1;

}

for (i = 0; i < 256; i++) //将节点按出现次数排序

{

for (j = i + 1; j < 256; j++)

{

if (header[i].count < header[j].count)

{

tmp = header[i];

header[i] = header[j];

header[j] = tmp;

}

}

}

for (i = 0; i < 256; i++) //统计不同字符的数量

{

if (header[i].count == 0)

break;

}

n = i;

m = 2 \* n - 1;

for (i = n; i < m; i++)

{

min1 = 999999999;

for (j = 0; j < i; j++)

{

if (header[j].parent != -1) continue;

if (min1 > header[j].count)

{

pt1 = j;

min1 = header[j].count;

continue;

}

}

header[i].count = header[pt1].count;

header[pt1].parent = i;

header[i].lch = pt1;

min1 = 999999999;

for (j = 0; j < i; j++)

{

if (header[j].parent != -1) continue;

if (min1 > header[j].count)

{

pt1 = j;

min1 = header[j].count;

continue;

}

}

header[i].count += header[pt1].count;

header[i].rch = pt1;

header[pt1].parent = i;

}

for (i = 0; i < n; i++) //构造HUFFMAN树，设置字符的编码

{

f = i;

header[i].bits[0] = 0;

while (header[f].parent != -1)

{

j = f;

f = header[f].parent;

if (header[f].lch == j)

{

j = strlen(header[i].bits);

memmove(header[i].bits + 1, header[i].bits, j + 1);

header[i].bits[0] = '0';

}

else

{

j = strlen(header[i].bits);

memmove(header[i].bits + 1, header[i].bits, j + 1);

header[i].bits[0] = '1';

}

}

cout << header[i].b << " : " << header[i].bits << endl;

}

//下面的就是读原文件的每一个字符，按照设置好的编码替换文件中的字符

fseek(ifp, 0, SEEK\_SET); //将指针定在文件起始位置

fseek(ofp, 8, SEEK\_SET); //以8位二进制数为单位进行读取

buf[0] = 0;

f = 0;

pt1 = 8;

printf("当前文件有:%d字符\n", flength);

printf("正在压缩\n");

while (!feof(ifp))

{

c = fgetc(ifp);

f++;

for (i = 0; i < n; i++)

{

if (c == header[i].b) break;

}

strcat(buf, header[i].bits);

j = strlen(buf);

c = 0;

while (j >= 8) //当剩余字符数量不小于8个时

{

for (i = 0; i < 8; i++) //按照八位二进制数转化成十进制ASCII码写入文件一次进行压缩

{

if (buf[i] == '1') c = (c << 1) | 1;

else c = c << 1;

}

fwrite(&c, 1, 1, ofp);

pt1++;

strcpy(buf, buf + 8);

j = strlen(buf);

}

if (f == flength)

break;

}

if (j > 0) //当剩余字符数量少于8个时

{

strcat(buf, "00000000");

for (i = 0; i < 8; i++)

{

if (buf[i] == '1') c = (c << 1) | 1;

else c = c << 1; //对不足的位数进行补零

}

fwrite(&c, 1, 1, ofp);

pt1++;

}

fseek(ofp, 0, SEEK\_SET); //将编码信息写入存储文件

fwrite(&flength, 1, sizeof(flength), ofp);

fwrite(&pt1, sizeof(long), 1, ofp);

fseek(ofp, pt1, SEEK\_SET);

fwrite(&n, sizeof(long), 1, ofp);

for (i = 0; i < n; i++)

{

tmp = header[i];

fwrite(&(header[i].b), 1, 1, ofp);

pt1++;

c = strlen(header[i].bits);

fwrite(&c, 1, 1, ofp);

pt1++;

j = strlen(header[i].bits);

if (j % 8 != 0) //当位数不满8时，对该数进行补零操作

{

for (f = j % 8; f < 8; f++)

strcat(header[i].bits, "0");

}

while (header[i].bits[0] != 0)

{

c = 0;

for (j = 0; j < 8; j++)

{

if (header[i].bits[j] == '1') c = (c << 1) | 1;

else c = c << 1;

}

strcpy(header[i].bits, header[i].bits + 8);

fwrite(&c, 1, 1, ofp); //将所得的编码信息写入文件

pt1++;

}

header[i] = tmp;

}

fclose(ifp);

fclose(ofp); //关闭文件

cout << "压缩后文件有:" << pt1<<"位二进制字节\n";

cout <<"压缩后文件有"<<pt1/8+1<<"位字符\n" ;

float persent = (pt1 / 8 + 1) / 1.0\*flength;

cout << "压缩率是" << persent<<"%"<<endl;

return 1; //返回压缩成功信息

}

4.2 重要程序段2

//函数：uncompress()

//作用：解压缩文件，并将解压后的内容写入新文件

int uncompress(const char\* filename, const char\* outputfile)

{

char buf[255], bx[255];

unsigned char c;

char out\_filename[512];

long i, j, m, n, f, p, l;

long flength;

int per = 10;

int len = 0;

FILE\* ifp, \* ofp;

char c\_name[512] = { 0 };

ifp = fopen(filename, "rb"); //打开文件

if (ifp == NULL)

{

return 0; //若打开失败，则输出错误信息

}

//读取原文件长

if (outputfile)

strcpy(out\_filename, outputfile);

else

strcpy(out\_filename, c\_name);

ofp = fopen(out\_filename, "wb"); //打开文件

if (ofp == NULL)

{

return 0;

}

fseek(ifp, 0, SEEK\_END);

len = ftell(ifp);

fseek(ifp, 0, SEEK\_SET);

printf("当前文件有:%d字符\n", len);

printf("正在解压\n");

fread(&flength, sizeof(long), 1, ifp); //读取原文件长

fread(&f, sizeof(long), 1, ifp);

fseek(ifp, f, SEEK\_SET);

fread(&n, sizeof(long), 1, ifp); //读取原文件各参数

for (i = 0; i < n; i++) //读取压缩文件内容并转换成二进制码

{

fread(&header[i].b, 1, 1, ifp);

fread(&c, 1, 1, ifp);

p = (long)c;

header[i].count = p;

header[i].bits[0] = 0;

if (p % 8 > 0) m = p / 8 + 1;

else m = p / 8;

for (j = 0; j < m; j++)

{

fread(&c, 1, 1, ifp);

f = c;

\_itoa(f, buf, 2);

f = strlen(buf);

for (l = 8; l > f; l--)

{

strcat(header[i].bits, "0"); //位数不足，执行补零操作

}

strcat(header[i].bits, buf);

}

header[i].bits[p] = 0;

}

for (i = 0; i < n; i++)

{

for (j = i + 1; j < n; j++)

{

if (strlen(header[i].bits) > strlen(header[j].bits))

{

tmp = header[i];

header[i] = header[j];

header[j] = tmp;

}

}

}

p = strlen(header[n - 1].bits);

fseek(ifp, 8, SEEK\_SET);

m = 0;

bx[0] = 0;

while (1)

{

while (strlen(bx) < (unsigned int)p)

{

fread(&c, 1, 1, ifp);

f = c;

\_itoa(f, buf, 2);

f = strlen(buf);

for (l = 8; l > f; l--)

{

strcat(bx, "0");

}

strcat(bx, buf);

}

for (i = 0; i < n; i++)

{

if (memcmp(header[i].bits, bx, header[i].count) == 0) break;

}

strcpy(bx, bx + header[i].count);

c = header[i].b;

fwrite(&c, 1, 1, ofp);

m++;

cout << header[i].b << " : " << header[i].bits << endl;

if (m == flength) break;

}

fclose(ifp);

fclose(ofp);

cout << "解压后文件为" << out\_filename<<endl;

cout << "解压后文件有:" << flength << "位字符\n";

return 1; //输出成功信息

}

4.3 重要程序段3

void text2bin(const char\* filename, const char\* extractfilename2) {

int count = 0;

int ch, a;

char temp;

FILE\* fin = fopen(filename, "r");

FILE\* fout = fopen(extractfilename2, "w");

while (fscanf(fin, "%c", &temp) != EOF) {

ch = temp;

for (a = 7; a >= 0; a--) fprintf(fout, "%d", ch >> a & 1);

}

fclose(fin);

fclose(fout);

}

4.4 重要程序段4

int main(int argc, const char\* argv[])

{

int flag;

char c;

char filename[100], extractfilename[100], extractfilename2[100];

while (1)

{

printf("\t \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\n");

printf("\t| |\n");

printf("\t| C-压缩文件 |\n");

printf("\t| E-解压缩 |\n");

printf("\t| Q-退出 |\n");

printf("\t|\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_|\n");

printf("\n");

do

{

printf("\n\t请选择功能:");

scanf(" %c", &c);

c = toupper(c); //小写变大写

putchar('\n');

if ('C' != c && 'E' != c && 'Q' != c)

{

printf("\t选项错误,请重新输入!\n");

}

} while ('C' != c && 'E' != c && 'Q' != c);

if ('C' == c)

{

printf("\t请您输入需要压缩的文件:");

fflush(stdin); //fflush(stdin)刷新标准输入缓冲区，把输入缓冲区里的东西丢弃

cin>>filename;

putchar('\n');

printf("\t请您输入压缩后的文件:");

fflush(stdin);

cin >> extractfilename;

putchar('\n');

printf("\t请您输入要保存的压缩后的二进制文件名:");

fflush(stdin);

cin >> extractfilename2;

putchar('\n');

flag = compress(filename, extractfilename);

putchar('\n');

if (-1 == flag)

{

printf("\t文件打开失败!\n");

exit(1);

}

else

{

printf("\n\t压缩操作完成!\n\n");

text2bin(filename, extractfilename2);

}

}

else if ('E' == c)

{

printf("\t请您输入需要解压的文件:");

fflush(stdin);

cin >> extractfilename;

putchar('\n');

printf("\t请您输入解压缩后的文件:");

fflush(stdin);

cin >> filename;

putchar('\n');

flag = uncompress(extractfilename, filename);

putchar('\n');

if (-1 == flag)

{

printf("\t文件打开失败!\n");

exit(1);

}

else

{

printf("\n\t解压缩操作完成!\n\n");

}

}

if ('Q' == c) {

printf("\t感谢使用!\n");

exit(0);

}

}

return 0;

}

## 第五章 调试分析

调试程序分析是课程设计必不可少的环节，调试分析可以在程序运行之后发现原来没有出现的问题，以及做进一步的更正，是程序达到预期的目的。

问题一：如何验证音频和图片压缩是否正确？

beyond compare 对于图片文件和音频文件不识别，这里只能通过人工的方法进行验证，因为在进行解压缩的时候加了后缀信息，所以在进行验证的时候需要将后缀修改为和原文件类型相同的后缀再去打开，通过验证得出结论：图片文件和音频文件也能够正常压缩和解压缩。

## 第六章 运行结果

1、首先进入初始界面，如图1所示。

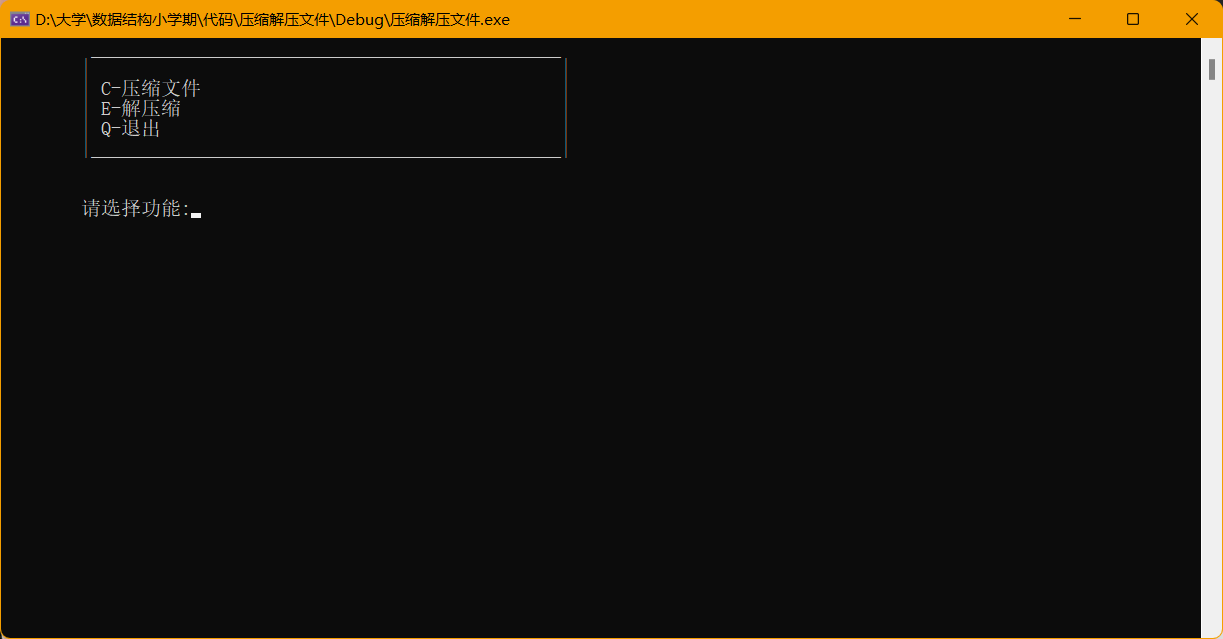


图1 系统初始界面

2、进入压缩功能，如图2所示

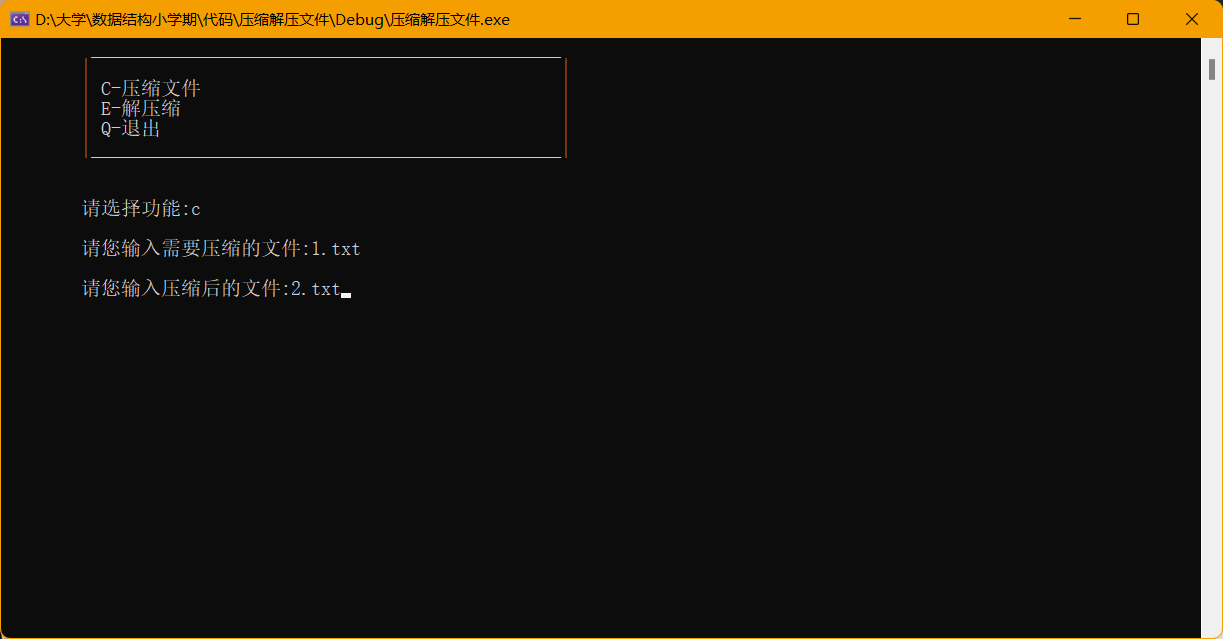


图2 压缩功能

3、压缩成功界面如图3所示。

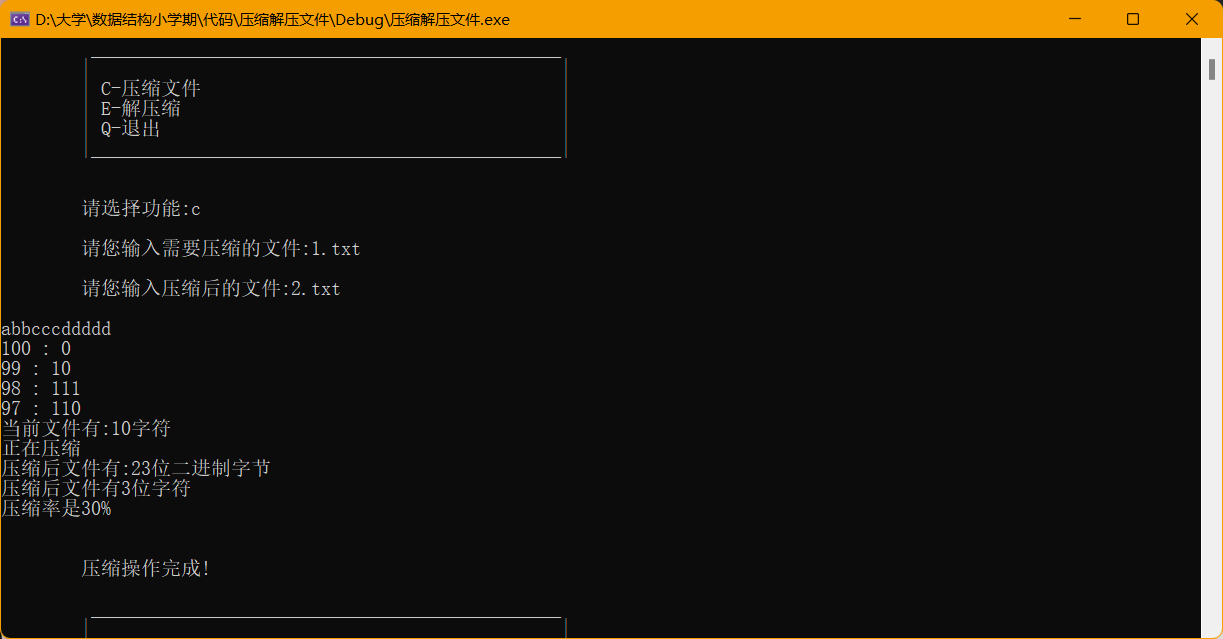


图3 压缩成功界面

4、选择解压缩功能。

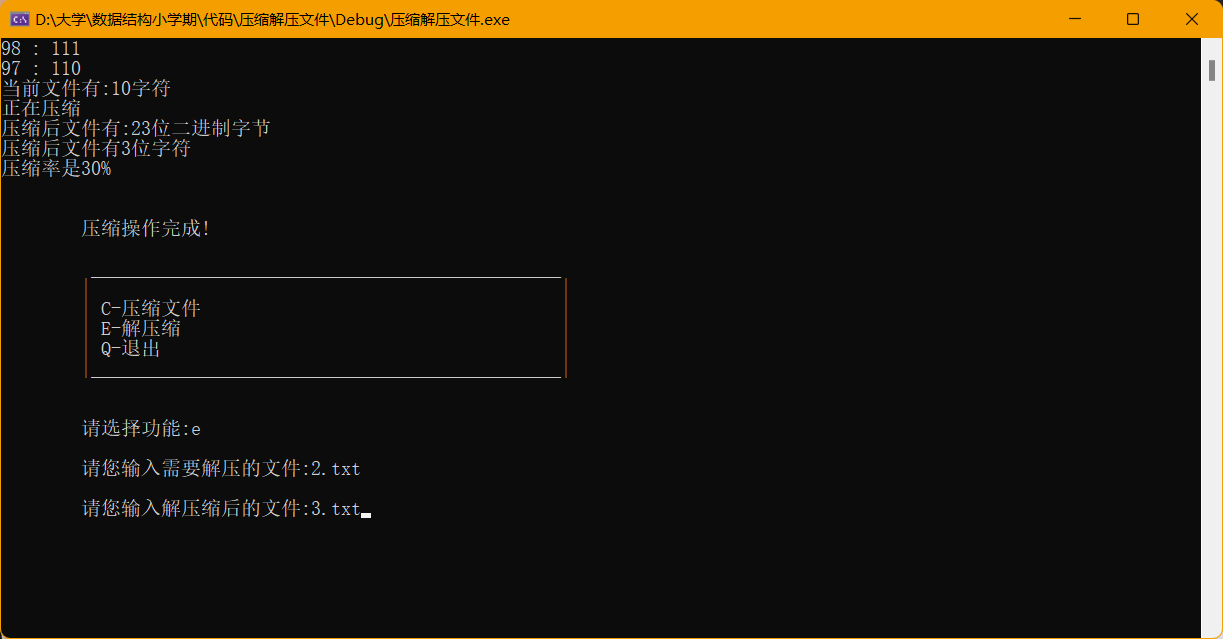


图4 解压缩功能

5、解压缩成功，如图5所示



图5查看记录

6、退出，如图6所示

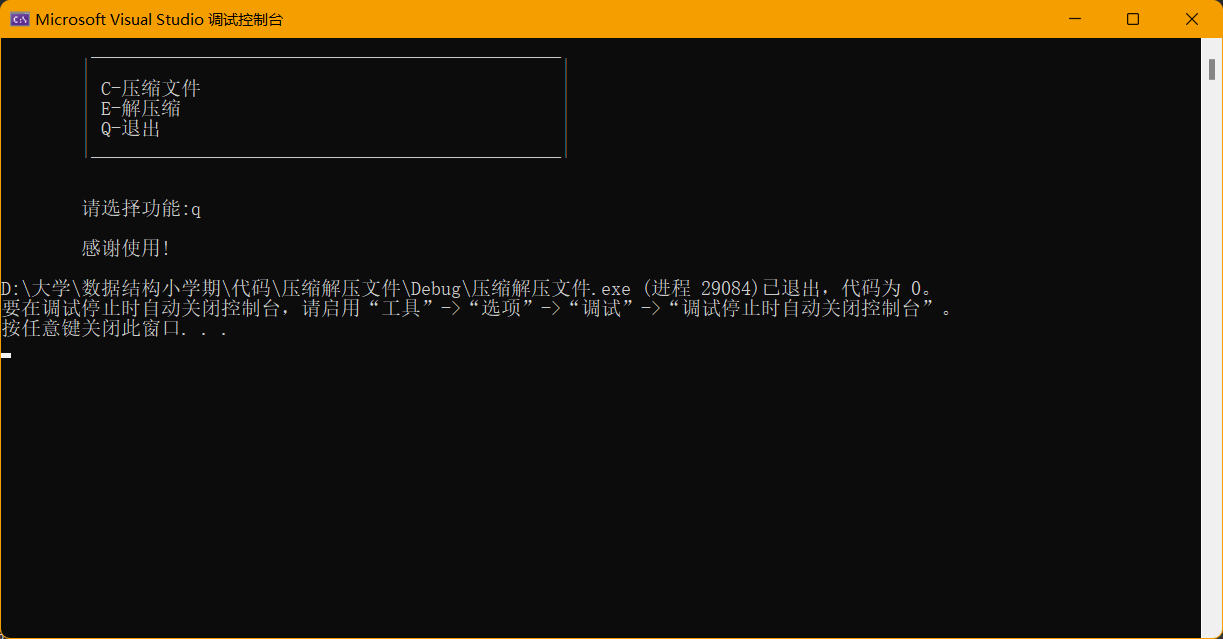
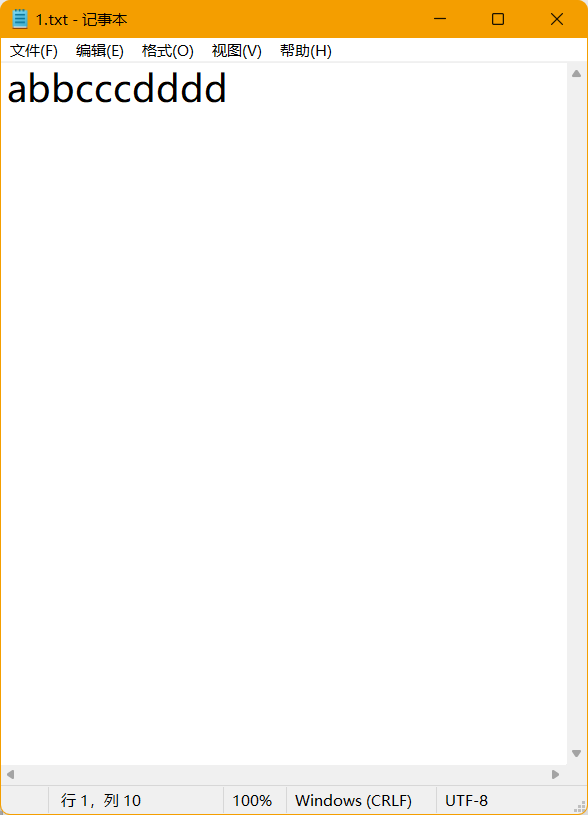
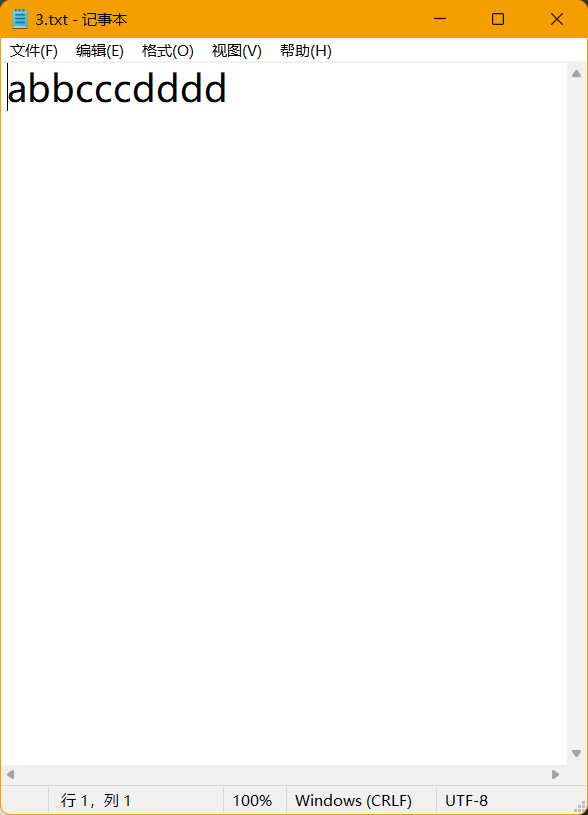
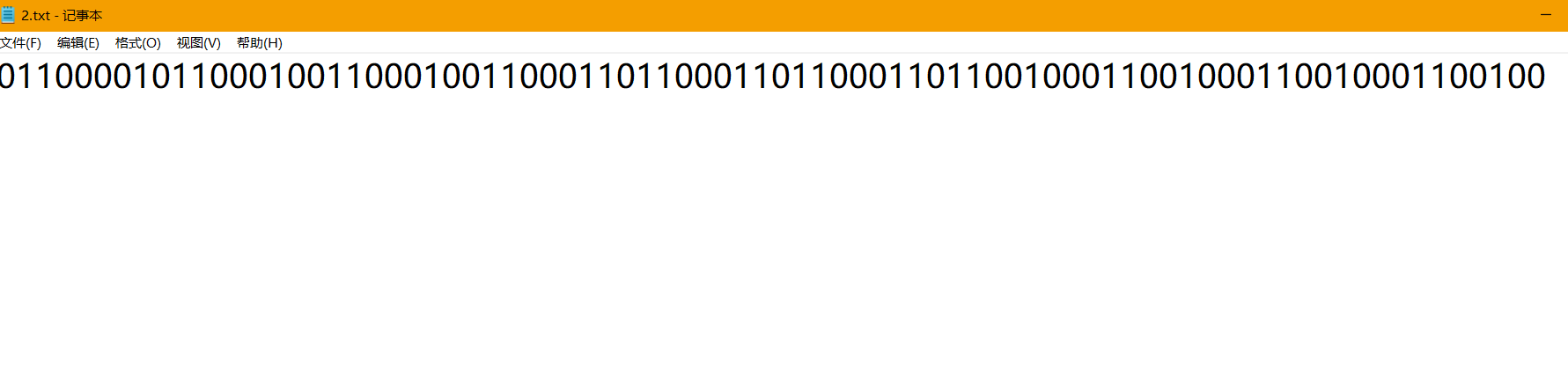


图6 退出

7.打开文件，验证是否正确





## 附录：完整代码

#define \_CRT\_SECURE\_NO\_WARNINGS

#include <iostream>

#include <stdlib.h>

#include <stdio.h>

#include <string.h>

#include <ctype.h>

using namespace std;

struct head

{

int b,c; //字符,数

long count; //文件中该字符出现的次数

long parent, lch, rch; //生成树

char bits[256]; //哈夫曼树编码

};

struct head header[512], tmp; //节点树

//函数：compress()

//作用：读取文件内容并加以压缩

//将压缩内容写入另一个文档

int compress(const char\* filename, const char\* outputfile)

{

char buf[512];

unsigned char c;

long i, j, m, n, f;

long min1, pt1, flength;

FILE\* ifp, \* ofp;

int per = 10;

ifp = fopen(filename, "rb"); //打开原始文件,rb 以读/写方式打开一个二进制文件，只允许读/写数据。

if (ifp == NULL)

{

printf("打开文件失败:%s\n", filename);

return 0; //如果打开失败，则输出错误信息

}

ofp = fopen(outputfile, "wb"); //打开压缩后存储信息的文件 wb:只写方式打开或新建一个二进制文件，只允许写数据。

if (ofp == NULL)

{

printf("打开文件失败:%s\n", outputfile);

return 0;

}

flength = 0;

while (!feof(ifp)) //功能是检测流上的文件结束符

{

fread(&c, 1, 1, ifp);

//fread( void \*buffer, size, count, FILE \*stream )

/\*buffer 指向要读取的数组中首个对象的指针 size 每个对象的大小（单位是字节）

count 要读取的对象个数 stream 输入流\*/

header[c].count++; //读文件，统计字符出现次数

cout << c;

flength++; //记录文件的字符总数

}

cout << endl;

flength--;

header[c].count--;

for (i = 0; i < 512; i++) //HUFFMAN算法中初始节点的设置

{

if (header[i].count != 0)

header[i].b = (unsigned char)i; //表示byte时，都用unsigned char

/\*将每个哈夫曼码值及其对应的ASCII码存放在一维数组header[i]中，

且编码表中的下标和ASCII码满足顺序存放关系\*/

else

header[i].b = -1; //哈夫曼初始化节点都是-1

header[i].parent = -1;

header[i].lch = header[i].rch = -1;

}

for (i = 0; i < 256; i++) //将节点按出现次数排序

{

for (j = i + 1; j < 256; j++)

{

if (header[i].count < header[j].count)

{

tmp = header[i];

header[i] = header[j];

header[j] = tmp;

}

}

}

for (i = 0; i < 256; i++) //统计不同字符的数量

{

if (header[i].count == 0)

break;

}

n = i; //外部叶子结点数为n个时，内部结点数为n-1，整个哈夫曼树的需要的结点数为2\*n - 1.

m = 2 \* n - 1;

for (i = n; i < m; i++)//构建哈夫曼树

{

min1 = 999999999;//结点出现的最大次数

for (j = 0; j < i; j++)

{

if (header[j].parent != -1) continue;

// parent != -1说明该结点已存在哈夫曼树中，跳出循环重新选择新结点

if (min1 > header[j].count)

{

pt1 = j;

min1 = header[j].count;

continue;

}

}

header[i].count = header[pt1].count;

header[pt1].parent = i;

// 依据parent域值（结点层数）确定树中结点之间的关系

header[i].lch = pt1;//计算左分支权值大小

min1 = 999999999;

for (j = 0; j < i; j++)

{

if (header[j].parent != -1) continue;

if (min1 > header[j].count)

{

pt1 = j;

min1 = header[j].count;

continue;

}

}

header[i].count += header[pt1].count;

header[i].rch = pt1;//计算右分支权值大小

header[pt1].parent = i;

}

for (i = 0; i < n; i++) //构造HUFFMAN树，设置字符的无重复前缀编码

{

f = i; //i=1，f=1

header[i].bits[0] = 0; //根节点编码为0

while (header[f].parent != -1)//1的母亲非空

{

j = f;//j=1

f = header[f].parent;//1=1的母亲结点0

if (header[f].lch == j)//如果左孩子是1，置左分支编码0

{

j = strlen(header[i].bits);//j=1

memmove(header[i].bits + 1, header[i].bits, j + 1);//memmove用于拷贝字节

header[i].bits[0] = '0';//左0右1，依次存储连接“0” “1”编码

}

else//置右分支编码1

{

j = strlen(header[i].bits);

memmove(header[i].bits + 1, header[i].bits, j + 1);

header[i].bits[0] = '1';

}

}

cout << header[i].b << " : " << header[i].bits << endl;

}

//读原文件的每一个字符，按照设置好的编码替换文件中的字符

fseek(ifp, 0, SEEK\_SET); //将指针定在文件起始位置

fseek(ofp, 8, SEEK\_SET); //以8位二进制数为单位进行读取

buf[0] = 0;//定义缓冲区

f = 0;

pt1 = 8;

/\*假设原文件第一个字符是"A"，8位2进制为01000001（65），编码后为0110识别编码第一个’0’，

将其左移一位，。下一个是'1'，应该|（或） 1，结果00000001

同理4位都做完，应该是00000110，由于字节中的8位并没有全部用完，们应该继续读下一个字符，

根据编码表继续拼完剩下的4位，如果字符的编码不足4位，还要继续读个字符，

如果字符编码超过4位，那么我们将把剩下的位信息拼接到一个新的字节里\*/

printf("当前文件有:%d字符\n", flength);

printf("正在压缩\n");

while (!feof(ifp))

{

c = fgetc(ifp);

f++;

for (i = 0; i < n; i++)

{

if (c == header[i].b) break;

}

strcat(buf, header[i].bits);

j = strlen(buf);

c = 0;

while (j >= 8) //当剩余字符数量不小于8个时，压缩存储

{

for (i = 0; i < 8; i++) //按照八位二进制数转化成十进制ASCII码写入文件一次进行压缩

{

if (buf[i] == '1') c = (c << 1) | 1;

else c = c << 1;

}

fwrite(&c, 1, 1, ofp);

pt1++;//统计压缩后文件的长度

strcpy(buf, buf + 8);//一个字节一个字节的拼接

j = strlen(buf);

}

if (f == flength)

break;

}

if (j > 0) //当剩余字符数量少于8个时

{

strcat(buf, "00000000");

for (i = 0; i < 8; i++)

{

if (buf[i] == '1') c = (c << 1) | 1;

else c = c << 1; //对不足的位数进行补零

}

fwrite(&c, 1, 1, ofp);

pt1++;

}

fseek(ofp, 0, SEEK\_SET); //将编码信息写入存储文件

fwrite(&flength, 1, sizeof(flength), ofp);

fwrite(&pt1, sizeof(long), 1, ofp);

fseek(ofp, pt1, SEEK\_SET);

fwrite(&n, sizeof(long), 1, ofp);

for (i = 0; i < n; i++)

{

tmp = header[i];

fwrite(&(header[i].b), 1, 1, ofp);

pt1++;

c = strlen(header[i].bits);

fwrite(&c, 1, 1, ofp);

pt1++;

j = strlen(header[i].bits);

if (j % 8 != 0) //当位数不满8时，对该数进行补零操作

{

for (f = j % 8; f < 8; f++)

strcat(header[i].bits, "0");

}

while (header[i].bits[0] != 0)

{

c = 0;

for (j = 0; j < 8; j++)//字符的有效存储不超过8位，则对有效位数位移实现两字符编码的连接

{

if (header[i].bits[j] == '1') c = (c << 1) | 1;//| 1不改变原位置的0,1值

else c = c << 1;

}

strcpy(header[i].bits, header[i].bits + 8);//把字符的编码按照原先的存储顺序进行连接

fwrite(&c, 1, 1, ofp); //将所得的编码信息写入文件

pt1++;

}

header[i] = tmp;

}

fclose(ifp);

fclose(ofp); //关闭文件

cout << "压缩后文件有:" << pt1<<"位二进制字节\n";

cout <<"压缩后文件有"<<pt1/8+1<<"位字符\n" ;

float persent = (pt1 / 8 + 1) / 1.0\*flength;

cout << "压缩率是" << persent<<"%"<<endl;

return 1; //返回压缩成功信息

}

//函数：uncompress()

//作用：解压缩文件，并将解压后的内容写入新文件

int uncompress(const char\* filename, const char\* outputfile)

{

char buf[255], bx[255];

unsigned char c;

char out\_filename[512];

long i, j, m, n, f, p, l;

long flength;

int per = 10;

int len = 0;

FILE\* ifp, \* ofp;

char c\_name[512] = { 0 };

ifp = fopen(filename, "rb"); //打开文件

if (ifp == NULL)

{

return 0; //若打开失败，则输出错误信息

}

//读取原文件长

if (outputfile)

strcpy(out\_filename, outputfile);

else

strcpy(out\_filename, c\_name);

ofp = fopen(out\_filename, "wb"); //打开文件

if (ofp == NULL)

{

return 0;

}

fseek(ifp, 0, SEEK\_END);

len = ftell(ifp);

fseek(ifp, 0, SEEK\_SET);

printf("当前文件有:%d字符\n", len);

printf("正在解压\n");

fread(&flength, sizeof(long), 1, ifp); //读取源文件长度，对文件进行定位 //读取原文件长

fread(&f, sizeof(long), 1, ifp);

fseek(ifp, f, SEEK\_SET);

fread(&n, sizeof(long), 1, ifp); //读取原文件各参数

for (i = 0; i < n; i++) //读取压缩文件内容并转换成二进制码

{

fread(&header[i].b, 1, 1, ifp);

fread(&c, 1, 1, ifp);

p = (long)c;//读取源文件字符的权值

header[i].count = p;

header[i].bits[0] = 0;

if (p % 8 > 0) m = p / 8 + 1;

else m = p / 8;

for (j = 0; j < m; j++)

{

fread(&c, 1, 1, ifp);

f = c;

\_itoa(f, buf, 2); //将f转换为二进制表示的字符串

f = strlen(buf);

for (l = 8; l > f; l--)

{

strcat(header[i].bits, "0"); //位数不足，执行补零操作

}

strcat(header[i].bits, buf);

}

header[i].bits[p] = 0;

}

for (i = 0; i < n; i++) //根据哈夫曼编码的长短，对结点进行排序

{

for (j = i + 1; j < n; j++)

{

if (strlen(header[i].bits) > strlen(header[j].bits))

{

tmp = header[i];

header[i] = header[j];

header[j] = tmp;

}

}

}

p = strlen(header[n - 1].bits);

fseek(ifp, 8, SEEK\_SET);

m = 0;

bx[0] = 0;

while (1) // 通过哈夫曼编码的长短，依次解码，从原来的位存储还到字节存储

{

while (strlen(bx) < (unsigned int)p)

{

fread(&c, 1, 1, ifp);

f = c;

\_itoa(f, buf, 2);

f = strlen(buf);

for (l = 8; l > f; l--)//在单字节内对相应位置补O

{

strcat(bx, "0");

}

strcat(bx, buf);

}

for (i = 0; i < n; i++)

{

if (memcmp(header[i].bits, bx, header[i].count) == 0) break;

}

strcpy(bx, bx + header[i].count);/\*从压缩文件中的按位存储还到按字节存储字符，

字符位置不改变\*/

c = header[i].b;

fwrite(&c, 1, 1, ofp);

m++;

cout << header[i].b << " : " << header[i].bits << endl;

if (m == flength) break;

}

fclose(ifp);

fclose(ofp);

cout << "解压后文件为" << out\_filename<<endl;

cout << "解压后文件有:" << flength << "位字符\n";

return 1; //输出成功信息

}

void text2bin(const char\* filename, const char\* extractfilename2) {

int count = 0;

int ch, a;

char temp;

FILE\* fin = fopen(filename, "r");

FILE\* fout = fopen(extractfilename2, "w");

while (fscanf(fin, "%c", &temp) != EOF) {

ch = temp;

for (a = 7; a >= 0; a--) fprintf(fout, "%d", ch >> a & 1);

}

fclose(fin);

fclose(fout);

}

int main(int argc, const char\* argv[])

{

int flag;

char c;

char filename[100], extractfilename[100], extractfilename2[100];

while (1)

{

printf("\t \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\n");

printf("\t| |\n");

printf("\t| C-压缩文件 |\n");

printf("\t| E-解压缩 |\n");

printf("\t| Q-退出 |\n");

printf("\t|\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_|\n");

printf("\n");

do

{

printf("\n\t请选择功能:");

scanf(" %c", &c);

c = toupper(c); //小写变大写

putchar('\n');

if ('C' != c && 'E' != c && 'Q' != c)

{

printf("\t选项错误,请重新输入!\n");

}

} while ('C' != c && 'E' != c && 'Q' != c);

if ('C' == c)

{

printf("\t请您输入需要压缩的文件:");

fflush(stdin); //fflush(stdin)刷新标准输入缓冲区，把输入缓冲区里的东西丢弃

cin>>filename;

putchar('\n');

printf("\t请您输入压缩后的文件:");

fflush(stdin);

cin >> extractfilename;

putchar('\n');

printf("\t请您输入要保存的压缩后的二进制文件名:");

fflush(stdin);

cin >> extractfilename2;

putchar('\n');

flag = compress(filename, extractfilename);

putchar('\n');

if (-1 == flag)

{

printf("\t文件打开失败!\n");

exit(1);

}

else

{

printf("\n\t压缩操作完成!\n\n");

text2bin(filename, extractfilename2);

}

}

else if ('E' == c)

{

printf("\t请您输入需要解压的文件:");

fflush(stdin);

cin >> extractfilename;

putchar('\n');

printf("\t请您输入解压缩后的文件:");

fflush(stdin);

cin >> filename;

putchar('\n');

flag = uncompress(extractfilename, filename);

putchar('\n');

if (-1 == flag)

{

printf("\t文件打开失败!\n");

exit(1);

}

else

{

printf("\n\t解压缩操作完成!\n\n");

}

}

if ('Q' == c) {

printf("\t感谢使用!\n");

exit(0);

}

}

return 0;

}