《数据结构》第四次作业

Huffman树编解码实验

# 读入文本：

输入一篇ASCII文件（不低于1k字符）

文件名： ascll.txt

文件内容：

ASCII Code: Its Importance and Applications

ASCII, the American Standard Code for Information Interchange, is a character encoding standard that has been instrumental in the development of modern electronic communication. It represents text in a way that computers can understand and process, paving the way for the digital revolution we experience today.

The ASCII code was first developed in the 1960s, and it remains an essential component of computer systems worldwide. Its simplicity and universality have made it a staple in the computing industry, despite the emergence of more complex character encoding systems like Unicode.

The ASCII code assigns a unique number, ranging from 0 to 127, to each character. These numbers, known as ASCII values, allow computers to store, retrieve, and manipulate text data efficiently. Whether it's a letter, a digit, or a punctuation mark, each character has a corresponding ASCII value that computers can use to represent it.

The importance of ASCII cannot be overstated. It is the foundation of many computer systems and programming languages, and it plays a crucial role in data transmission and storage. The internet, for example, relies heavily on ASCII to transmit text-based data between computers. Emails, web pages, and other forms of digital communication would not be possible without ASCII.

ASCII also facilitates cross-platform compatibility. Since ASCII is a universal standard, text encoded in ASCII can be read and interpreted by computers running different operating systems and using different programming languages. This interoperability has been crucial in enabling the widespread adoption of computers and the internet.

In addition to its practical applications, ASCII also has a significant cultural impact. It has allowed for the digitization of books, newspapers, and other forms of printed media, preserving our cultural heritage for future generations. ASCII-encoded text can be easily stored and accessed, making it possible to share information and ideas across vast distances and different cultures.

Moreover, ASCII has been a driving force in the development of computer science and software engineering. It has been used as a basis for creating more complex character encoding systems, such as Unicode, which support a wider range of characters and languages. The principles and concepts underlying ASCII have also influenced the design of many modern programming languages and computer systems.

In conclusion, ASCII code is a fundamental element of modern computing and communication. Its simplicity, universality, and interoperability have made it an indispensable tool in the digital age. From email to web pages to digital libraries, ASCII has enabled us to share information, connect with others, and preserve our cultural heritage in a way that was not possible before the advent of computers. As we continue to move into an increasingly digital world, the importance of ASCII and its role in shaping our technological landscape will continue to be felt.

# 统计字符和频度

统计文档中字符出现的频度，并根据频度对每个字符生成Huffman编码

character frequency

\10 16

\32 485

' 1

, 35

- 3

. 24

0 2

1 2

2 1

6 1

7 1

9 1

: 1

A 22

C 21

E 1

F 1

I 50

M 1

S 21

T 7

U 2

W 1

a 219

b 29

c 108

d 104

e 278

f 47

g 57

h 64

i 184

k 5

l 87

m 78

n 206

o 167

p 71

q 1

r 153

s 155

t 209

u 67

v 23

w 25

x 9

y 27

z 1

# 打印字符对应的Huffman编码

根据字符频度，生成出现的字符对应的Huffman编码（不可打印字符统一用类似‘\A0’表示，示例中“A0”表示该字符的ASCII码值）

character code

\10 00101101

\32 000

' 011111000101

, 101111

- 0111110110

. 0111111

0 01111100011

1 00101100100

2 001011001111

6 001011001110

7 01111101111

9 011111000100

: 011111000011

A 1001011

C 1011101

E 001011001101

F 001011001010

I 011110

M 001011001100

S 1011100

T 011111001

U 01111100000

W 01111101110

a 0011

b 0010111

c 01001

d 01110

e 111

f 100100

g 010000

h 001010

i 1000

k 011111010

l 10011

m 10110

n 0110

o 1010

p 001000

q 001011001011

r 1101

s 1100

t 0101

u 001001

v 1001010

w 0100011

x 001011000

y 0100010

z 011111000010

# 文件的Huffman编码

根据字符的Huffman编码，将原文档转换为Huffman编码

1001011101110010111010111100111100001011101101001110111011111000011000011110010111000000111101011000100010101101010100110110010011110000011011001110000100101100100000100010011100001001001101011000101001101100000000001011010000000010110110010111011100101110101111001111010111100001010010101110001001011101101111101100001001001101100001011100010100110110011100011110101110000101110110100111011100010010010101101000011110011010010010101101101100011010110001010011000001111001100101111110101001001010001101100100001111011110001000110000000110000100100101000111101001101001010111111010001110110010011010011101000011001000000011000101001101100111000111101011100000101001010001101010000010100011110000000101111111110110000100001101100010111010010011011011101100101001110011000100001100000101001010111000011101111001010111100111010001000101101110110010100010101001000001011010100111011111010110000111100111110100101011101101001101000010010000100110101011010110001001011010000100100110101100010100110011111100001111001010001101111001000110111111001110110010111000000101111001011000010100010000110000001100001000110011010001000001010010100011010100001001101010110001000001001010111111011100000010010011011000000100101100111011111011100010100110110011100000011011001110000001000110110100100111111001100101111000001000001110010101000011001000000001010010101110000100011001101000100001001001010110100001010010101110000111010000100001000010100111001100011011111001010101010011001001010110001010011000001000111110001110010110000010001111101100011101100100111100001011010011100011010001001111110000000010110100000000101101011111001001010111000100101110111001011101011110011110000010011010011101110000100011001111000001001001000110111000101000011101111001010111100111010001000111011100001000011000001010010101110000010110010001111100010000101100111001111100011110010111100000110110011100001000010100011011111011000111000011011000000011011000011111001100111011001011000001110011000010011010101100010001010011011101100101000101010010000001001101010110001000001001010111111010001100010001011000101111101101100000010001110101101100110111001000111000011101110111111000011110010111000001100100010110001000100111000010011000010101000100000011011001110000001001011010001001010111110111000011100111000010101000100000010100011100101011100010110001101110111000100001010000011000110001010011001000100111110001000011000001010010101110000100110101011000100000100101011000011001000000010000110011100010011100010111010100010101111000011101111100001000100001011110000101001010111000111101101111101010000111011001001111000101010010000010110101011011110000100110101011000100010011111001011000000010010010100011110100110100101011111101000111011001001101001110100001100100000001100010001011000101111101101100000100111000011111010111000011111000000110100001001101001110111011111100000000101101000000001011010111110010010101110001001011101110010111010111100111100000100110100111011100000111100110010000100000110110000000110000010010110100000101100101100100111100001100010011011000101111111101101111000110100110110010000100001100100000001001001101101010110000011111000110000101101000000101100100001011001111011111011111011110000101101000011100110100100101000001001001010001111010011010010101111110101111110000111110010010101111100111000011000100110110001011111111011100101111000011111010011010100100011011000000111100000100101110111001011101011110011110000100101000111001100100111111001011110000011100111001110100100011000010011010101100010000010010101111110111000000101101000011000101101011011111011110001101111010111011000111100101011110111100000110110011100001011000110110100000100000100110011001101011110000101111001011000010100001110001101010011000111100100100100100001001100011101100101100110100010011111100001111101110001010111010100101011111010001000010101111100010111000000011000100111110101010111111011011110000011000011101000010000100001011011110001010110100000110000010000010010110010010101001001001101011000101001100001011000111101011111010101111000111001101001001010000010010010100011110100110100101011111101000001010001111000000011000010011010110111011111100001000101001100111010000110010000000100101110111001011101011110011110000100101000111001100100111100001010010100011010100001001101010110001000001001010111111011100000010010011011000000100111001110000101101000011011110010001101111110011101100101000100001010111111000000001011010000000010110101111100100101011100010001011000100010101101010100110110010011110001010100100000100101110111001011101011110011110000010010011011001101010010100000101111110001010100101011111011100010100110101111011100111111000011110010100010001100000010100101011100010010010100010010110011100011010110001010011000010101001000001011000110110010001000001001101010110001000001001010111111010001100010001011000101111101101100000001101100111000000100011011010010000110100111011010110100001100100000001001100110110010000001001001101000011111001011110000011011001110000100001010000010001001100110100010110000000110000100111010010010100110000011100110001101101010011111000100001100000111000110101001100001011101001101101100101101000110011001000101001100000011011001110000110001011010110100110100001110111111000011111001001010111000100001100101111110101101110101101111000100100101011010001110010110000011101100010001001111110111100011011111001110001111100000001010111001110010101000100110100010000101001100001001011101110010111010111100111100000101101000001011101001101101100101101000010100001011110010110000101011111011000101110011110011101110000011100011010100110000010111111010101000111111110110000010011010101100010000010010101111110111000111111000001011001101101100011100010011110010111100001000111110010111000001000001101000011111001011110000011011001110000101001010010101111101000100100101011011011011000001010100100000011101000010000100001010011100110000100110101011010110001001011010000100100110101100010100110000010001110100010011001101110000011010100101000001011111100000100010101100110010000010111100111110000100011100001010010101010001001010100010010111011100101110101111001111001111110000000010110100000000101101100101110111001011101011110011110000001110011110010100001001000011010011000100111000010100110101111110000001001110110101100110001111101100010001001100110101100100101011011011000001001101010110001000001101011000001011110001001110000101010001001111110001011100100001100100111100010010111011100101110101111001111000010001100000001100000100101101000100101011111011100001110011000110001010011011001110001111010111010111100001011110010110000101000111011001001101001110111011100001000011000010010111011100101110101111001111000001001001101100000010111111000110111100110111000000110110011100001000011001011111101001000110111101011110111000000101110100010000010011010101100010000010010101111110111000001101001001011001101000011001000000001110100010010010010011111011110110010100010100010001111101001101011000011001000000011000100010110001011111011011000000011011001110000001001110010000110010000000011101000100100100100111110111101100101000001000110110100100001101001110110101101000011001000000010011001101100100000010010011010000111110001111110000111110010010101000110000010000110010111111011010001000111110100110010111100010011100001010100010000001010001111000000010111111111011000001001110100100101001100000111001100010000110000111011000110010111100111000011001000000001010010101110000100011100001110111110000100011011110011011100000011011101010001000010110001010011000010101001000000100110101011000100000100101011111101110000000110110011100000101001010111000100001100101111110101101110101011111100000000101101000000001011010111100110000001101110011101000010110001010011000001011010000100001011100000001000110100110100101011000010010011100110000011001000001000100111000010010011010110001010011011001011110001001011101110010111010111100111100000011100111100101000000101000111100000001100011001000010000011010001001001000010010011011001010000100100100110011010100100111010011100110001000101100010000011010010101011111100001111001010000010100011110000000111001110011101001000111110111000010010010101101000010100101011100001110100001000010000101100001111100001000110101100010100110000101010010000000101111010101001111101011001011110000110111010001111000010000011001000111110111001011110000011011001110000101001010010101111101000100100101011011011011000001010100100000001000110110000110010111101110000101101110111010000011101111000001000110111111001111101100101010000110010000000101000100111010000100100100110011010100100111010011100110000010101111101100001010011010000111000100100101011010001001000010010101001001110111100001000011101101111101001101011000101001101100011111100010010111011100101110101111001111001111101101110110010011010011101110111000001011110010110000101000010010011011000000101111110001110011110010001001101000100001100010110101101111011100000011011001110000001101001010011111100110011101110101111000101100011011111010100001100100000001000010100000100010101100110010000010111100111110000101101000011000010100011110111100010000110100100101011011011000110101100010100110000001101100111000010000111011100111100000001101001110110101100110000010010100011110001010000111010001100010100110110010011111100000001101100111000001110100010010010010011111011110110010100001001001001100110101001001110111111000111111000000001011010000000010110100101100110010101101111101010010101111101101111000100101110111001011101011110011110000001010001111000000010111111111011000000110000111011011000100101010000110010000000100100101011010100111100010000110000010100101011100001110111100101011110011101000100010110111011001010001010100100000010011010101100010000010010101111110100011000100110001110110010011110000011011001110000110010101001000101010001100111101111000111011001000010000110111111110110000110010000011111100001111001010000010100011110000000101111111110110000001001110011101110000001111000000011000001011100111100100011000001001001010110100001001110111100110101100001100100000001011010101101111000010011010101100010001001111100101100000001001001010001111010011010010101111110100011101100100110100111010000110010000000110001000101100010111110110110010111100011000010010100100101000000111100000011111000000110100001001101001110111101111000010001100101010000100100101000011000010010010000010001010110101010000011000010001110000111011111010001101001101100100001110001010100100000010010010100011110100110100101011111101110000000110110011100001001100110110010000001001001101000011111000111111000011111001001010111000001000110110000110010011000001000100111111100000001101100111000001001101001100100111100100001011100000001001011001110111110110011010001010000110010000000100101110111001011101011110011110000001010001110010101110000011100111100101000010000110100100100110010011110110010011110111000001010010101110000111011111001000010000011000010101001000001011000110110010001000010110101001110111110101100000010001101101001000011010011101101011010000110010000000100110011011001000000100100110100001111100000001101100111000001001101010110001000001001010111111010001100010001011000101111101101100011111100000000101101000000001011010111100110000010011010011001001100110010011100100010100110101111000100101110111001011101011110011110000010011010011101110001000110000000110001001000010010110011100011101101110110010100111001100011110011111101101110110010100010101001000001011010100111011111010110000010011010101100010000010010101100001100100000000011011001110000010011010101101011000100101101000010010011010110001010011001111110000111100101110000011001000101100010001001110000100110000101010001010111100000100101101000100101011111011100001110011100001010100010101111000001101100111000010000110010111111011010001000111110100110010111100010011100001010100010000001010001110010101110001011000110111011100010000101000001101100001000011001110100011000010001110110110000110010111100111110000101101010101001100010000110000010100101011100001110100001000010000101001110011000001101000011101111110000010110010101101101010110000111101100011100010011000010110100000100011111001011100000100000110100001111100000010110100000111010000100001000010100111001100010011100000101111101001111011000111110010111100010010111011100101110101111001111000000101000111100000111011000110010111100111110111000000100111000000101101000011000010100011110111100010000110100100101011011011000110101100010100110101111000010011010011001101110100101010000100011100001010010100001010010100101011111011100101111000001101100111000000100011011111100111110110010101110001010001001110100001001001001100110101001001110100111001100000101011111011000010100110100001110001000011000000110000100011001101000100000101001010001101010000100011001111000000110101001010000010001010110011001000001011110011111000001011111110010010101101111000010100101011100000110111010010101110110010100010101001000000100110101011000100000100101011111101110001111110001001011110000001000111110000100110100110010110000110001001111000010110100001011010101001010111000100001100101101000000110110000100001100100111011110011110010000110010000100110100010000011101000010000100001010011100110000100011101011011001101110101111000010100101011100010001011000100010101101010100110110010011110001010100100000100101110111001011101011110011110000001101100111000010000101110000011011010100111110001000011000011000010100011001000100001100100000001010001001110100001011110100100101001101010100111010010000100001001001110011000100110011011001110110001001001100100011100001000111000100111001100001001101001100101100001100010011110000101101000000101111110001001001111001101010111111

# 统计字符数和编码数，讨论编码效率。

原文出现字符数N： 3058

出现不同字符数M： 48

总编码比特数Nhuffman： 13591

源文件比特数N0（即N\*8）： 24464

编码效率：Nhuffman/N0 = 55.6 %

# 解码结果

ASCII Code: Its Importance and Applications

ASCII, the American Standard Code for Information Interchange, is a character encoding standard that has been instrumental in the development of modern electronic communication. It represents text in a way that computers can understand and process, paving the way for the digital revolution we experience today.

The ASCII code was first developed in the 1960s, and it remains an essential component of computer systems worldwide. Its simplicity and universality have made it a staple in the computing industry, despite the emergence of more complex character encoding systems like Unicode.

The ASCII code assigns a unique number, ranging from 0 to 127, to each character. These numbers, known as ASCII values, allow computers to store, retrieve, and manipulate text data efficiently. Whether it's a letter, a digit, or a punctuation mark, each character has a corresponding ASCII value that computers can use to represent it.

The importance of ASCII cannot be overstated. It is the foundation of many computer systems and programming languages, and it plays a crucial role in data transmission and storage. The internet, for example, relies heavily on ASCII to transmit text-based data between computers. Emails, web pages, and other forms of digital communication would not be possible without ASCII.

ASCII also facilitates cross-platform compatibility. Since ASCII is a universal standard, text encoded in ASCII can be read and interpreted by computers running different operating systems and using different programming languages. This interoperability has been crucial in enabling the widespread adoption of computers and the internet.

In addition to its practical applications, ASCII also has a significant cultural impact. It has allowed for the digitization of books, newspapers, and other forms of printed media, preserving our cultural heritage for future generations. ASCII-encoded text can be easily stored and accessed, making it possible to share information and ideas across vast distances and different cultures.

Moreover, ASCII has been a driving force in the development of computer science and software engineering. It has been used as a basis for creating more complex character encoding systems, such as Unicode, which support a wider range of characters and languages. The principles and concepts underlying ASCII have also influenced the design of many modern programming languages and computer systems.

In conclusion, ASCII code is a fundamental element of modern computing and communication. Its simplicity, universality, and interoperability have made it an indispensable tool in the digital age. From email to web pages to digital libraries, ASCII has enabled us to share information, connect with others, and preserve our cultural heritage in a way that was not possible before the advent of computers. As we continue to move into an increasingly digital world, the importance of ASCII and its role in shaping our technological landscape will continue to be felt.

# 程序代码：

#pragma GCC optimize(1)

#pragma GCC optimize(2)

#pragma GCC optimize(3,"Ofast","inline")

#include<bits/stdc++.h>

#include<regex>

#include <queue>

using namespace std;

const int maxN = 128;

int N = maxN;

typedef struct Data{

int index;

int value;

};

typedef struct BiTNode{

Data data;

struct BiTNode \*father, \*lchild, \*rchild;

}BiTNode, \*BiTree;

BiTree New\_HuffmanNode(BiTNode \*left, BiTNode \*right, int n){

BiTree newnode = (BiTree)malloc(sizeof(BiTNode));

newnode->father = nullptr;

newnode->lchild = left;

newnode->rchild = right;

newnode->data.index = n;

newnode->data.value = left->data.value + right->data.value;

return newnode;

}

// 定义两颗子数的比较原则

class Greater{

public:

bool operator()(const BiTree& a, const BiTree& b){

return a->data.value > b->data.value;

}

};

BiTree Create\_Huffman(int \*w){

std::priority\_queue<BiTree,std::vector<BiTree>,Greater> pq; // 优先队列，对value进行排序

int i;

BiTree temp,p1,p2;

for(i=0;i<maxN;i++){

if(w[i]==0) N--;

else{

temp = (BiTree)malloc(sizeof(BiTNode));

temp->lchild = nullptr;

temp->rchild = nullptr;

temp->data.index = i;

temp->data.value = w[i];

pq.push(temp);

}

}

while(pq.size()>1){

p1 = pq.top();

pq.pop();

p2 = pq.top();

pq.pop();

temp = New\_HuffmanNode(p1,p2,i+1);

p1->father = temp;

p2->father = temp;

pq.push(temp);

i++;

}

return pq.top();

}

void print\_Huffman(BiTree root,int level){

if(root->lchild == nullptr && root->rchild == nullptr){

for(int i=0;i<level;i++){

cout<<" ";

}

cout<<root->data.index<<" "<<root->data.value;

}

else {

print\_Huffman(root->lchild,level+1);

cout<<"\n";

for(int i=0;i<level;i++){

cout<<" ";

}

cout<<root->data.index<<" "<<root->data.value;

cout<<"\n";

print\_Huffman(root->rchild,level+1);

cout<<"\n";

}

}

void generate\_HuffmanCodes(BiTree root, std::map<char,std::string> \*pHuffmanCodes, std::string str){

if(root->lchild == nullptr && root->rchild == nullptr){

(\*pHuffmanCodes)[root->data.index] = str;

}

else{

// 递归生成左子树和右子树的编码

generate\_HuffmanCodes(root->lchild, pHuffmanCodes, str+"1");

generate\_HuffmanCodes(root->rchild, pHuffmanCodes, str+"0");

}

}

int main()

{

ios::sync\_with\_stdio(false);

int w[maxN]={0};

std::ifstream file("ascll.txt");

char ch;

while(file.get(ch)) {

if (ch >= 0 && ch < N){

w[ch]++;

}

}

file.close();

cout<<"character frequency\n";

for(int i=0;i<maxN;i++){

if(w[i]!=0){

if(i<=32||i==127)

cout<<" "<<"\\"<<i<<" "<<w[i]<<"\n";

else cout<<" "<<(char)i<<" "<<w[i]<<"\n";

}

}

// 生成Huffman树

BiTree huffmanTree;

huffmanTree = Create\_Huffman(w);

print\_Huffman(huffmanTree,0);

// 转化为编码

std::map<char,std::string> huffmanCodes; // 不要用string<bool>，用指针容易出问题，字符串可以轻易复制

std::string emptystr;

generate\_HuffmanCodes(huffmanTree, &huffmanCodes, emptystr);

for(std::map<char,std::string>::iterator it=huffmanCodes.begin();it!=huffmanCodes.end();it++){

if(it->first<=32||it->first==127)

cout<<"\\"<<(int)it->first<<" "<<it->second<<"\n";

else cout<<it->first<<" "<<it->second<<"\n";

}

cout<<"\n";

// 文本转编码

std::vector<bool> code;

file.open("ascll.txt");

while(file.get(ch)) {

for(std::string::iterator it=huffmanCodes[ch].begin();it!=huffmanCodes[ch].end();it++){

code.push\_back(\*it-48);

}

}

file.close();

for(std::vector<bool>::iterator it=code.begin();it!=code.end();it++){

cout<<\*it;

}

cout<<"\n"<<code.size();

cout<<"\n";

// 解码

BiTree p=huffmanTree;

for(std::vector<bool>::iterator it=code.begin();it!=code.end();it++){

if(\*it==1) p=p->lchild;

else p=p->rchild;

if(p->lchild==nullptr&&p->rchild==nullptr){

cout<<(char)p->data.index;

p=huffmanTree;

}

}

return 0;

}