

PROJECT 6

HUFFMAN ENCODING/ DECODING AND GREEDY ALGORITHM

Huffman Coding is an entropy encoding algorithm used for lossless data compression. It refers to the use of a variable-length code table for encoding a source symbol (such as a character in a file) where the variable-length code table has been derived in a particular way based on the estimated probability of occurrence for each possible value of the source symbol.

I. Variable length coding:

(a) Input: a text of 3 or more lines including 26 characters, comma, spaces and period.

Display the input texts.

Pre-determine frequencies of each character by looking up dictionary.

Using Huffman variable length encoding and decoding, implement a program in any language you desire (preferably C++ or java) to encode and decode a given text,

(b) Determine variable length binary code for each character and display the compressed encoded codes for the entire input text.

(c) Use the compressed binary code as input, decode it back to the original text and display the same.

II. Fixed length coding: (use 5 bits or 8 bits coding to represent each character; If you use 5 bits, you may create your own character representation. If you use 8 bits, you may use ASCII representation.)

Do the same as above.

Notes: Compare the numbers of bit representation of each text in both variable length and fixed length by inputting three different paragraphs of texts. For convenience, you may use the digit "1" to represent the bit "1", and the digit "0" to represent the bit "0".

Your program output must show proper information to be understood well by the reader/viewer.

Output

*****--> Original string:

qwertyu.iopas.dfghuuuuuuujkxz xcvb.nmdfghjkoiuytrewqsx.azxcfghjkertyu
ioqwe.rtyuiopzxcvbnm,sdfghjkzxcvbjsjsjsjsjswuwuwuwusisisisiakakakakwkwocicduuuu

uuuuunzxcvbn.kdhstqu.xgztalsdzia sdhsajz
jahqoxhhinw.ertyuios.dfgjhj.poiuytrlkj.hgfdsnbvzxcqazws.xedcrfvtgb.yhnujmikol,.vb

*****--> Huffman encoed (variable) binary string:

```
11101011011111011010011001001000011000110001010100000011010101100011100
0010100101111110011011011011011011011001011111110011001100111111001010
00011000111000100001000001110000101001011111100010111111010110000110100
01001001001111011110111110101011111000001110101100111100101000101001011
11110001011111111011010011001001000011100111100010101111010110111110110
00101001100100100001110001010100000011001111001010000110001110000100000
1100110010111100001010010111111000101111110011110010100001100011100101
01100101011001010110010101100101011110110111101101111011011110110111011
1000101110001011100010111000110101111111010111111101011111110101111110
11111111101110101101001000101000111100001101101101101101101101101100001
1100111100101000011000111000010001111111000111101011100101110100110001
11100010111100110010110101000101111000110011000110101001111001111011110
00111101011110100010110011001110010110101111011101010101111001111011110
10000000111011000111101101001100100100001110001010110110001110000101001
01111110001000010000001010110000110100010010010011001101111110010000111
11001011010101100010110000100111001101100111100101001110101101011001110
11101100011110011101111000101000100101010001101001001011001110001010001
1110000010110010000001100011111101011001101100110000010011000111
```

*****--> Huffman encoed (fixed) binary string:

```
00010011000110010000010100010100000101100001101100010111000011110000100
10001000100010010000000010001010100001111000001000000011000000111000010
00000101110001011100010111000101110001011100010111000101110000101000001
01100011010000111000000000000011010000000110001100000000010000011110001
00000000111000000100000001100000011100001000000010100000101100010001000
01001000101110001101100010110000101000000010100011001000100110001010100
01101000001111000000010001110000011010000000110000011000000111000010000
00010100000101100000101000101000001011000011011000101110000000000001001
00010001000100110001100100000101000011110001010000010110000110110001011
10000100100010001000100100001110000011010000000110001100000000010000100
00000011100000110000010101000001000000011000000111000010000000101000001
01100011100000110100000001100011000000000100000101000010101000010100001
01010000101000010101000010100001010100001010000101010001100100010111000
11001000101110001100100010111000110010001011100010101000010010001010100
00100100010101000010010001010100001001000000010000101100000001000010110
00000010000101100000001000010110001100100001011000110010001000100000011
00001001000000110001011100000100000101110001011100010111000101110001011
10001011100010111000101110001011100010000000111000001101000000011000110
```


11111001110111110100001011001110100110001001111000110001010101100010101
1001011000110100001011010100110011011101101111101110011100011110100101
11100101111100011110010000010011100101011011010111101001101010110001011
10100101111001011001100101001111100100000101111010011010001000010011111
01110110100010100011001111011101011010000011101001101001010101111011011
11101011000001101011110110110111010001101110110101111011011100111111010
11100100100111110000100001101110110011000001111100101111111101101111111
01011001001110101100010101100101101111110111001110001111010010111100101
11110001111001000001001110010101101101011110100110101011000101110100101
11100101101011011101000110100100100100000110011010100111000111111100100
11011100001000011011101100011111110011100100001101110011011110111001111
10111111101110001110110100101110100001001101110101000001100010101100101
10010100000100001100110111000100100110111000111101111010100101110100100
00111110000100100011101101010011001101010101111011100110011101110100000
10111010110011001010011111001011101011110110110001111101010001000000110
01011110010110111111011101111000110010110101001100110001010110010110100
01010001100111101110101101000001110100110100101010111101111100111111111
1

*****--> Huffman encoed (fixed) binary string:

00000001000110010000100100001001000100000000010000010010000000110000000
00001010000000111000011010001001000001010000000110000110100010111000000
11000001000001001000000011000010000001001000011000000101100001010000010
10100011100000000110000100000010010000001100001010000000111000011010001
00100000101000000011000001000000111100001010000101000001011000001101000
11000000010110001000000000011000110010001011100001000000001110000001100
00100100010100000101100000001100001111000101000001011100010111000011110
00010000001011100010111000000110000011100000100000110000000010000000011
00000110000101000001000000010101000101100000100000010111000101110000110
10001010000010010000100110000001100000010000110000000001100010110000010
00000010010000100000010110000101110000001100011000000101000000001100011
00000001011000010000000001100011001000101110000100000000011000101000000
10010000001100000100000000110001101000000100000101100000110100000100000
00101000011110000100000010001000011110000100000010010000010100001100000
00101100000011000001100001010000000111000010000000001100011000000001000
00001010000111100001000000000110000100100010100000101100000001100001000
00010010000001100001010000000111000011010001001000001010000000110000010
00000001100010111000101000001100100010110000001100000100000000011000101
11000111000001000000000101000101000000111100000011000011000001011100011
00100000110000010110000001100000100000101110000001100000100000000110000
01100000101100000100000101100000010000000110000110000000100000010110000
00011000011010001001000000011000001000000001100001001000011010000111100

```
00100000001110000000110001101100001011000010000001011000001000000000110
00110000000101100001000000000110001101000000100000101100000110100000100
00000101000011110000100000010001000011110000100000010010000010100001100
00000101100000011000001100001010000000111000010000000001100011000000001
00000001010000111100001000000000110000101100000100000101110000001100000
10100001000000010000001001000000011000001110000100000010110000011010001
10100000100000000111000000110000110100010010000000110000010000000011000
10101000001000001011000011000000011010000011000011001000011110000010000
01011000000011000110110000010000011100000000110000010100000100000101110
00010000000011100000011000101000001001000000011000110000000101100001000
00000011000010000001011100011000000011010001000000000100000110000000100
00000011100000011000101010001011000010100000001010000010000000101000011
01000011110000110100011000000111000000001100010100000010010000001100010
10000000110000001100001100100010110000101100000100000010010000001100000
10000000001100001001000101000001011000000011000010000000010000000110000
01011000000110001010100010100000101110001011100001101000001010000111100
001000000000110001101000000100000011110001100100001000000000011000101000
00010010000001100011000000010110000100000000011000101110001010000011001
00010110000001100000100000000011000101110001110000010000000001010001010
0000011110001001100011101
```

*****--> decoded string (variable) from binary string:

Huffman Coding is an entropy encoding algorithm used for lossless data compression. It refers to the use of a variable-length code table for encoding a source symbol (such as a character in a file) where the variable-length code table has been derived in a particular way based on the estimated probability of occurrence for each possible value of the source symbol.;

*****--> decoded string (fixed) from binary string:

Huffman Coding is an entropy encoding algorithm used for lossless data compression. It refers to the use of a variable-length code table for encoding a source symbol (such as a character in a file) where the variable-length code table has been derived in a particular way based on the estimated probability of occurrence for each possible value of the source symbol.;

*****--> Original string:

Mr. Trump appeared to be particularly angry at the W.H.O. for issuing a statement saying it did not support his decision on Jan. 31 to restrict some travel from China because of the virus. At the time, the group issued a statement saying that “restricting

the movement of people and goods during public health emergencies is ineffective in most situations and may divert resources from other interventions. Don't close your borders to China, don't do this," Mr. Trump said, paraphrasing the group and accusing the organization of "not seeing" the outbreak when it started in Wuhan, China. "They didn't see it, how do you not see it? They didn't see it. They didn't report it. If they did see it, they must have seen it, but they didn't report. In fact, the W.H.O. repeatedly issued statements about the emergence of the virus in China and its movement around the world.

*****--> Huffman encoded (variable) binary string:

```
11001111001001011001110111000010001111110101110010111110111100101100100
01110110100001101111110000110111100100100111111001011011010000010101001
01011111101000110110100110100010011011111011100010011101001001101111101
10001110001100000111101110010101100101101110101100101101111101100111011
10101100100111101001010101011111010100010011111111011111010100011011000
001110101001100000011101011101110011010101000100111111110100001111011110
10101111111000011000011101010111111001011001001100100000111110001010010
11111011100110010110100101101001101000111011010001111100111010110111000
10110011110110110001011010111111000011011101000010101000010010101001010
00111010101101101010011110000100110111101001001110100011101110101000110
11010111110110100110001010100011011111100100100110010111011011110101001
11101100111011110001100000111111010011010010001111010110110011110110101
10000111000110000011110001010110101001110011011100011000001111100111010
00110011111100101111010010101010111100110111111110111110101000110110000
01110101001100000011101011101110011010101000100111111000110001101100011
10111001101000010101000010010101001010001010100010011111100011000001111
11010101101101001001110101001100000011101100111011111100100010110110010
11010000011111101110001011111110011101100110101110101111101110111101001
01010001001111111100100111110010011101000101010010111111000001110111101
00000011000111001110101001010010011100110001001011010001010111110100101
11110101000001011101011101001100101000101011010010011111010100011111010
10110010100011101011010000011111101100010100110100001011111101110001011
11111101011101110011011110111101011010010010100000111010000101010110011
11010010010100101011110111010100011011010111101100001100000101001111010
10000000010100110100100110000001010011010000101101100111110011100101101
00010010000001111001011101000011001010011111001100110011110100111100100
10110010010111001010001011110000110111101101001100010101000110111100110
11110111011010001001000000111101110110111000110001010010111001101011011
01111110011110010010110011101110000100011111101011100101110101110111010
10111110011011111001011011010011011110010110000100110110101101010001001
11111000110000011111001110100011001111110010111110111000101111111101110
01011001010111101011010100010011111100011000001111011001001001111101110
```

00101010110110011101100010100110100011101100111011110111001110000110000
11101010010011010100010011110110110111100011000001111011001111000100100
1010000111011110011101111111001111110000011000111101000011101010001101
10100000001101111111010100011101110010011111100011011100011001101111011
01001100010101000110111011001110111001101110001100000110011011110111101
01011110001001000000111010100100111110100001100110111110000110110011111
11110111011011110011001100111111110000110000111010100100111110100001100
11100011101110001100000110011011110111101010111100010010000001110101001
00111110100001011001110111000110000011001101111011110101011110001001000
00011101000011100100110010000011110100001011001111011010100111011110001
10000011001101111011110101011111101010010011111010000110011011100011000
00110011011111010101111010100011111000110111101001001111010100100110001
11101000011001101111001001011110001110001100000110011011110111101010111
10001001000000111010000111001001100100000101100111101101010100011101110
11101110010100011001101110001100000111101110010101100101101110101100101
10111110110011101000011100100011101100000110111110100010011011110100101
01010111100110111111010100011011000001110101001100000001011111101110010
01011001111000111000110000011110011101010010100100111001100010010100111
10110011101111000110000011111101001101001000111101011111010100011110110
10011000101010001101111111011100010111111101000001011111101010110110100
100111010100110000001111101101000110011111000101111111000110000011111100
1111101100100110100010111101100

*****--> Huffman encoded (fixed) binary string:

00000110001000000001101100001101000010000010000000100100000110100001111
00000110100001110000111100001111000010010000011100010000000010010000100
01000011010010001100011101000011010000111100010010000011010001111000001
11000100000001000110001011000010000001001000001100000001110001000000001
10000010011100001101000011100001110000010100001000000010011100001101000
01110001000110000110100100011000101010001001000001101000010010001101100
00001100011011000001110001101100001101000100110001110100100000000011010
00101100010000100100001001001000001011000011100000101000000110100001110
000011010010000100100011100001110001000110001001000011010000100100001110
00010001100001101001000010000111000100111000101100001110000010100000011
01000101100010001100001101000100010001011000010001000011010001110000011
10100100011000011010010000100100100000111100001111000011101001000000010
00110000110100010101000101100010000100001101000100010001001000010000000
10110001000010001011000011101000111000000110100011101000111000000110100
00010100001110000111000001101100001101001000100001111100001101001000110
00111010000110100100000000100100010000100100011001000000001011000010000
00100011000011010010000100011101000110100001001000001101001000110010000
00000111000100101000100100001100000001101000100110010000000011101000110

10000011010000000100010101000101100001110000001110000011010000111100010
01000010000000011100010010000100001000100100000110100011101000100110000
11010010001100010101000100100000110100100101000101100010000000100100001
00001000110110000110100000000001000110000110100100011000101010001001000
00110100100011000101100001101000010010000110010000110100100011000101010
00100100000110100010100001000000001110100100100000111100000110100010110
00100001001000010010010000010010000100010000110100001110000011010010000
10010001100001110001000110001001000011010000100100001110000100011000011
01001000010000111000100111000101100001110000010100000011010010001100010
10100001110001000110000110100001011001000000001001000100001001000110010
00000001011000010000001000110001011000011100000101000000110100100011000
10101000100100000110100011010000111010010010100010010000110100001001000
01110000100011000011010001110100010011000011010001111000010010000111010
00111100001100000010010000011010000111000011100000100010000110100010100
00011101000111010001000100100001000011010001000100100100001000000001011
00001110000010100000011010001111000100100000011110001100000010110000100
00000011010001010100010010000011100001100000100011000101010000110100010
01000011010000100100010000000010100000100100001110000010000000101100001
00100010000100001101000101100010000100001101000101100001110000010010000
10011000100110001001000010000001000110001011000100101000100100000110100
01011000011100000011010001101000011101001000010010001100001101001000010
00101100010001100100100000011100010001100010110000111010001110000100001
00001101000011100001110000010001000011010001101000001110001001110000110
10001000100010110001001010001001000100000001000110000110100100000000100
1000100001000111010010010000100000000100000010010001000010000110100010
01100100000000111010001101000001101000111010010001100010101000100100010
00000000110100010110000111000010001100010010001000000010010100010010000
11100001000110001011000011101000111000010000100011011000011010000001000
01110100011100000010100010001100001101000100000001100000011101001000010
0010010000011010010011100011101001001000010000000011010000111100011101
00100000000100010001001000100000001000010000110100100011000111010000110
10000000100010101000101100001110000001110000110010000110100010001000111
01000111000000101000100011000011010001000100011101000011010010001100010
10100010110001000010001100100001100000011010000011000100000000110110000
11010000100000100000001001000001101000011110000011010010000100001110000
10110000100010001100100001101000111100000111000100000000011100001111000
01010100100000000011100010000100010110000111000001010000001101001000110
00101010001001000001101000101000010000000011101001001000001111000001101
00001110000111000001000100001101000011100001000000010000001001000010000
10001011000011100000101000000110100100011000101010001001000001101000111
01001000000001010000001110000111000001011000101000000011100010001100010
11000011101000111000000110100011101000100110000110100001011000001

11010010001100001101001000010001001000010010000101100001110000010100000
01100000011010010001100010101000100100000110100011101001001000010001100
00111100100000000100100000111000010111000011010010011000010101000100100
00111000000110100010110001000110000110100100001001000110000111000100000
00100011000100100001000100001101000101100001110000001101000010010010010
00001010100001110000111000001100100001101000000010001010100010110000111
00000011100001101100001101000010110000100000010101000100100010011100001
10100010001000101100001000100011100000010100010001100001101001000010001
00100001001000001101000101100010001100011001000011010001010100011101001
00110000011010001000100011101000011010010011100011101001001000000110100
01110000011101001000110000110100100001000100100001001000001101000101100
01000110010100100001101000010000001010100010010001001110000110100010001
00010110000100010001110000001010001000110000110100100001000100100001001
00000110100010110001000110001101100001101000010000001010100010010001001
11000011010001000100010110000100010001110000001010001000110000110100100
00000010010000111100001110100100000001000110000110100010110001000110001
10110000110100000100000100110000110100100011000101010001001000100111000
01101000100010001011000010001000011010010000100010010000100100000110100
01011000100011000110010000110100100011000101010001001000100111000011010
00110100010010000100001001000110000110100010101000011100010010100010010
00001101001000010001001000010010000111000000110100010110001000110001100
10000110100001111001001000010001100001101001000110001010100010010001001
11000011010001000100010110000100010001110000001010001000110000110100100
00000010010000111100001110100100000001000110001101100001101000001000001
11000000110100010011000011100001000000100011000110010000110100100011000
10101000100100000110100001001000110110000001100011011000001110001101100
00110100100000000100100001111000010010000011100010001100010010000100010
00110000010011100001101000101100010000100100001001001000001001000010001
00001101001000010010001100001110001000110001001000011010000100100001110
00010001100100001000011010000111000001111000111010010010000100011000011
01001000110001010100010010000011010001001000011010000100100010000000010
10000010010000111000001000000010010000011010001110100010011000011010010
00110001010100010010000011010010010100010110001000000010010000100001000
01101000101100001110000001101000000010001010100010110000111000000111000
00110100001110000111000001000100001101000101100010001100100001000011010
00110100001110100100101000100100001101000010010000111000010001100001101
00001110001000000001110100100100000111000001000100001101001000110001010
10001001000001101001001100001110100100000000110000001000100011011

*****--> decoded string (variable) from binariy string:

Mr. Trump appeared to be particularly angry at the W.H.O. for issuing a statement saying it did not support his decision on Jan. 31 to restrict some travel from China

because of the virus. At the time, the group issued a statement saying that “restricting the movement of people and goods during public health emergencies is ineffective in most situations and may divert resources from other interventions. Don’t close your borders to China, don’t do this,” Mr. Trump said, paraphrasing the group and accusing the organization of “not seeing” the outbreak when it started in Wuhan, China. “They didn’t see it, how do you not see it? They didn’t see it. They didn’t report it. If they did see it, they must have seen it, but they didn’t report. In fact, the W.H.O. repeatedly issued statements about the emergence of the virus in China and its movement around the world.

*****--> decoded string (fixed) from binary string:

Mr. Trump appeared to be particularly angry at the W.H.O. for issuing a statement saying it did not support his decision on Jan. 31 to restrict some travel from China because of the virus. At the time, the group issued a statement saying that “restricting the movement of people and goods during public health emergencies is ineffective in most situations and may divert resources from other interventions. Don’t close your borders to China, don’t do this,” Mr. Trump said, paraphrasing the group and accusing the organization of “not seeing” the outbreak when it started in Wuhan, China. “They didn’t see it, how do you not see it? They didn’t see it. They didn’t report it. If they did see it, they must have seen it, but they didn’t report. In fact, the W.H.O. repeatedly issued statements about the emergence of the virus in China and its movement around the world.

Process finished with exit code 0