

ENEE5304, INFORMATION AND CODING THEORY

Course Project on Source Coding

Due: January 10, 2024 (via ITC)

Course Project on Huffman Code

The Written Report

- 3-4 pages, double space, 12-point font.
- At least two recent references.
- Write the report in your own words. Do not just copy and paste. If you quote something, cite the reference
- Sections: Define the problem in the introduction, Method (or theoretical background), Results (or Simulations or implementation) and their analysis, the code (appendix), Conclusions, and References.
- Report is to be submitted via ITC by one student from the group
- Students can work in groups of no more than two.

Presentation

- Selected students will be required to present their work at designated dates, to be announced later.

Course Project on Huffman Code

- You will be given an English short story: **To Build A Fire by Jack London.**
- Write a computer program using matlab (or any language) to simulate the Huffman code, i.e., to generate the codewords given a certain set of symbols along with their probabilities.
 - a. Find the number of characters in the story along with their frequency of occurrence. Tabulate your results
 - b. Find the probabilities of the characters in the story. For simplicity do not distinguish between capital and small letters. You can also skip the “enter” character.
 - c. Find the entropy of the alphabet.

Course Project on Huffman Code

1. Use your program to find the codewords for the characters
2. If ASCII code is used, find the number of bits needed to encode the full story. Call this number “NASCII”
3. Find the average number of bits/character for the whole story using the Huffman code. Compare this to the entropy
4. Find the total number of bits needed to encode the entire story using Huffman code. Call this number “Nhuffman”
5. Find the percentage of compression accomplished by using the Huffman encoding as compared to ASCII code.
6. Fill in the table to showcase some of your results

Sample Calculations

- Find the probabilities, the lengths of the codewords, and the codewords for the following symbols

Symbol	Probability	codeword	Length of codeword
a			
b			
c			
d			
e			
f			
m			
z			
space			
. (dot)			