Course Project

Phase 1

Groups Deadline: November 1, 2021 (11:59 PM)

Phase 1 Deadline: November 25, 2021 (11:59 PM)

I. Project Description

1. Read the text file provided to you using MATLAB and compute the approximate probabilities of the different English characters (symbols) in this text file. The used characters are: lower case letters a-z, in addition to the following characters ( ) . , / -and the space character, i.e., 33 characters in total.
2. Calculate the entropy based on the symbol probabilities calculated in step one.
3. Calculate the number of bits/symbols required to construct a fixed length code and calculate the efficiency of that code.
4. Implement in MATLAB the Huffman encoder function and encode the file characters into a stream of zeros and ones. Show in a table each character in the source text file and its corresponding Huffman code.
5. Implement in MATLAB the Huffman decoder function and write the decoded stream of characters back to a separate text file, and make sure that it matches the original text file.
6. Calculate the efficiency of the Huffman code.

You should try to test your code with different text file to make sure it is functioning correctly and not dependent on the provided text file only.

II. Course Project Rules

- Make a team of 2 at most [larger numbers will need approval].
- Write your code so that it can be readable by others. Define your variables clearly (not abbreviated). Use comments as much as you want.
- The figures that you are going to show must be well presented. They must have clear labels, titles, and maybe legends.
- Your answers to the questions in the previous section must be appropriately enumerated.
- Any COPIED reports even one single part will take a **ZERO**.