**CIE 425 Information Theory and Coding Project (Part 1)**

**Software**: MATLAB, or any other suitable programming language

**Group**: 3-4 members

**Requirements**:

1. Read the text file provided to you using and compute an estimate of the 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. **(0.5 point)**
2. Calculate the entropy based on the symbol probabilities calculated in step one. **(0.5 point)**
3. Calculate the number of bits/symbol required to construct a fixed length code, and calculate the efficiency of that code. **(0.5 point)**
4. Implement the Huffman encoder function (don’t use the Huffman encoder function in Matlab, develop your own 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. **(1.5 points)**
5. Implement 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. **(1.5 points)**
6. Calculate the efficiency of the Huffman code. **(0.5 point)**
7. Implement the Shannon-Fano code (encoder/decoder) for the same data. **(2 points)**
8. Compare the performance (efficiency) of Shannon-Fano code with Huffman code. **(0.5 point)**

**Grading:**

Whole project: 20% from final grade

**Part 1: 7.5%**

Part 2: 7.5%

Presentation and discussion: 5%