MATHS 415 – Project 1

1. For a text file of at least 100,000 characters, compress it in each of the following three different ways. You only need to count the bits for each compression method, not produce the actual bits. However, you should include all overhead bits for code-table transmission, etc. in your count.
    a. Single Table Huffman Code (First-Order)
    b. Multiple Table Huffman Code (Second-Order)
    c. Symbol Pairing/Translation Code (Second-Order with a single table)
2. For each of the three methods, give a brief description of the method and the format of the compressed data.
3. For the two second-order methods, give a few additional details (the first ten or so symbol pairs, the first few symbols that trigger code table changes) that indicate how the method adapts to the statistics of your text.
4. Give an overall summary and interpretation of your results.