COSC 311 Programming Project #4 FALL 2015
Distributed 11/16/2015 **Due**: 11/30/2015 (2 weeks)

Precis: Implement and extend the code from Koffman & Wolfgang, Section 6.6, pp 345 – 352. Compare the amount of compression.

Encrypt (compress) the following English language files using
A Huffman tree based on frequency of letter in English Text (K&W p 346)
Huffman trees that are custom built on specific data files.

I. A Huffman code is represented as a binary tree. Create the Huffman code for generic English text as given on page 346.
II. Each data file will have a custom Huffman code built for it. The custom code starts on page 349. You are to use the following data files:
    a. *Odyssey*: http://classics.mit.edu/Homer/odyssey.mb.txt
    b. Watson & Crick, *A Structure for Deoxyribose Nucleic Acid* http://www.exploratorium.edu/origins/coldspring/printit.html
    c. Mystery data file – to be provided.
III. Only characters [a..z] are recognized (convert upper-case to lower-case). All other characters are ignored.
IV. The characters are read from the datafile. Do not read the entire file in to a string (in main memory). Clearly, you will have to take two passes through the data, once to accumulate the frequencies, build the Huffman tree and a second time to do the compression.
V. Compress the following data files using the four different Huffman code books:
    Generic English text
    Customized for Odyssey
    Customized for DNA paper.
    Customized for mystery data file
    Obtain only the original byte count and the final byte count for each compression (except when debugging, of course) – these values are available from the File API.
VI. Compare the amount of compression for each data file (three data files, each compressed using four different Huffman trees).
    Give a short paragraph that explains the results.

Turn in:
(1) Hardcopy of code
(2) Screen shot showing the I/O of creating the four Huffman code books.
(3) A listing of each of the Huffman code books.
(4) Screen shot showing compression of the Odyssey using the four Huffman code books. You should ***not*** display the compressed document. Just show the original character count and the final character count.
(5) Short paragraph describing the compression results.