CSCI B505 Fall 19: Programming assignment 4

Due online: Friday, November 08, 11:59pm EST

Submit your work online using Canvas (no assignments will be accepted on paper this time). You can use LaTeX, Word, or even pen and paper for the write-up. But please try to submit a PDF file. Use the lab sessions prior to the due date to get help from your AI/UI, if necessary.

What to do

You will be implementing a **Huffman code** (see slides or textbook for the description of the algorithm).

1. Go to `http://www.gutenberg.org` and find a book in English. Download the book as a plain text (.txt) file using a Plain Text UTF-8 link.
2. Calculate the frequency of each letter/symbol in the text.
3. Create the Huffman tree based on 128 ASCII characters (`https://ascii.cl/`). You can ignore the first 32 characters (with codes 0-31), so the total number of characters would be 96.
4. Print a table with your code, e.g:
   - a: 0010
   - b: 0011
   - c: 11111
   - ...
5. Compare the savings of your code with a fixed-length encoding using 7 bits per character. How many bits did you save?

What to submit

For this assignment you will be submitting two things:

1. Source code. Please, follow good coding practices: use indentation, write comments, etc.
2. Write-up. This should contain:
   - A link to the text file you used.
   - Description of how you constructed the code and the resulting code itself.
   - Justifications for any choices you've made, e.g. the tie-breaking rule you chose, how did you handle the characters which never appeared in the text, etc.
   - Comparison with the fixed-length 7 bit encoding, discussion of the savings.
   - Any other design choices you've made.