COMPSCI 145
EXTRA CREDIT HOMEWORK
Huffman Coding
Professor William T. Verts

Here is the first sentence of Abraham Lincoln's Gettysburg Address:

```
FOUR SCORE AND SEVEN YEARS AGO OUR FATHERS
BROUGHT FORTH ON THIS CONTINENT, A NEW NATION,
CONCEIVED IN LIBERTY, AND DEDICATED TO THE
PROPOSITION THAT ALL MEN ARE CREATED EQUAL.
```

Including a carriage return character and a line feed character at the end of each of the four lines, there are exactly 181 characters in this sentence. Using strict 8-bit ASCII, this file would be exactly 181×8=1448 bits in size.

Count all characters by type (how many **A**, how many **B**, etc., how many spaces, how many commas and periods, how many carriage returns and line feeds, and so on). Sort the list in descending order by frequency of occurrence. Generate a Huffman tree from the counts (see the Companion). When you replace each character with its path in the Huffman tree, how long in bits is the resulting compressed file?

Turn in your Huffman tree and the number of bits in the compressed file. Please draw neatly!

For example, the one line 25-character sentence **COMPUTER SCIENCE ROCKS!** would generate the following tree (#13 is the carriage return and #10 is the line feed):