UMass Boston CS 310 Homework 6
Due in class on Thursday, April 13, 2017

1. The following table has the frequencies of symbols in a file, where nl and sp stand for newline and space, respectively.

Symbols	:	sp	nl	,	0	1	2	3	4	5	6	7	8	9
Frequencies	100	605	101	705	431	242	176	59	185	250	174	199	205	217

   (a) Applying Huffman's algorithm to the data, draw the prefix code tree. The leaves should be annotated with the symbols and their weights, and internal nodes should be annotated with the total weights of their substrees. When merging two subtrees, put the lower-weight subtree on the right.

   (b) Make a table of four columns: symbols, codes, frequencies, total bits, like Slide 20 of lecture notes.

   (c) Compress this 7-symbol text: "04: 19,". Show the result in both binary and hexadecimal.

   (d) Decompress the binary string 11001110111100000.

   (e) In preparation for fast decoding, sort the codes by lexicographical order like Slide 27 of lecture notes.

   (f) Take the top three codes after sorting, and construct the decode array for these three codes, using ... to skip repetitive symbols in the decode array, like Slide 28 of lecture notes.

2. Apply the Burrows-Wheeler Transform on this string: "abracadabra", using the methods on Slides 47 and 48 of lecture notes.

   (a) Generate all rotations of the string.

   (b) Reorder the rotations by lexicographical order.

   (c) Output the last column and the index I of the original string. Note that I is zero-based. See Slide 47.

   (d) Write the contents of the three arrays F, L, and T. See Slide 48.