CS 240E: Structures and Data Management **Winter 2021**

Tutorial 11: Huffman Encoding, Run-Length Encoding, LZW

1. Apply Huffman encoding on the string $s =$ `CELESTEELA`. Show the resulting decoding trie and give the encoding for s.

2. Consider Run-Length Encoding compression.

a) Encode the string $s = 11111111000001111011111100000000000000000000$ using RLE.

b) Decode the string $c = 111001011010010011$ using RLE.

3. For the following LZW problems, consider the initial dictionary to be the ASCII table.

a) Encode the following string using LZW: `BANANA_BANDANA`

b) Decode the following encoded string using LZW:

$$71 - 73 - 86 - 69 - 95 - 77 - 131 - 82 - 69 - 128 - 137 - 65 - 83$$

4. Let $c_1, \ldots, c_k$ be the characters of a text, sorted by decreasing (non-increasing) frequencies. Let $s(c_1), \ldots, s(c_k)$ be the prefix-free encoding of these characters obtained with the Huffman encoding algorithm.

a) Prof. I.N. Correct thinks that $s(c_1)$ *must* have the shortest codeword, i.e., $|s(c_1)| \leq |s(c_i)|$ for all $i = 2, \ldots, k$. Show that the professor is incorrect.

b) Show that the professor is correct if the frequency of c_1 is *strictly* larger than all other frequencies.