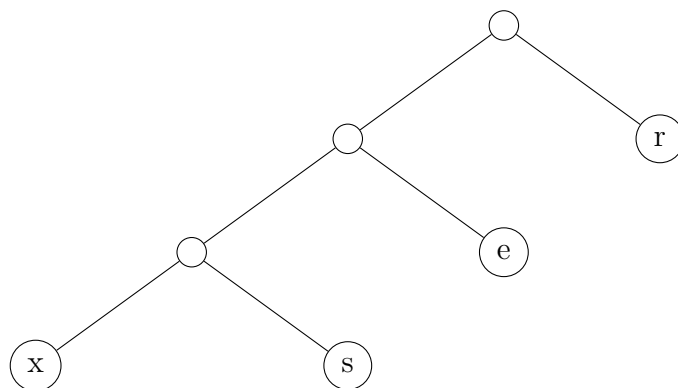


Tutorial 11 - Compression  
CS 240E Winter 2022  
University of Waterloo  
Monday, March 28th, 2022

1. **Huffman Encoding 1:**

- Build the Huffman tree for  $S = \text{“pusheen”}$
- Below is an encoding trie  $T$  for the string “xerxes”. Argue that this is *not* a Huffman tree.



2. **Huffman Encoding 2:**

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

- 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  $i = 2, \dots, k$ . Show that the professor is incorrect.
- Show that the professor is correct if the frequency of  $c_1$  is *strictly* larger than all other frequencies.

3. **LZW Encoding 1:**

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

- Encode the following string using LZW: BANANA\_BANDANA
- Decode the following encoded string using LZW:

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

4. **LZW Encoding 2:**

Let  $S$  be a string of length  $n$ . Argue that the LZW encoding of  $S$  must use  $\Omega(\sqrt{n})$  integers.