9 Information Retrieval (SHT)

In the inverted index of an information retrieval system, dictionary terms can be represented using different data structures.
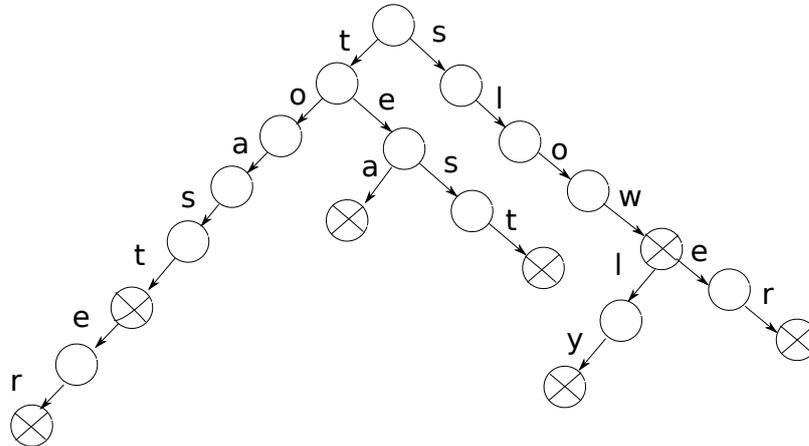
(a) Consider the *trie* in the figure above, which encodes several dictionary terms.

    (i) List the terms contained in this trie.   [2 marks]

    (ii) Explain how terms are looked up in a trie.   [2 marks]

(b) Alternatively, we could store the terms in a *binary search tree*.

    (i) Draw the binary search tree with minimal depth that stores the dictionary terms from the figure above.   [3 marks]

    (ii) Compare the worst-case time complexity of dictionary lookup for a binary tree and a trie. What are the conditions where the binary tree is preferable to a trie?   [3 marks]

(c) Next consider a *radix tree*, a space-optimised trie data structure where each internal node with only one child is merged with its child. (An internal node is one not associated with a term, and thus not pointing to any data.)

    (i) Draw the radix tree containing the dictionary terms from the figure above.   [2 marks]

    (ii) Give an algorithm for insertion of a new index term $t = t(0) \ldots t(k)$ into a radix tree. Use examples to illustrate your algorithm. You may use pseudocode as long as you clearly explain your thoughts.   [8 marks]