

CSc 445 — Questions about Tries and Suffix Trees

Not to be submitted.

1. Create a trie for the set of words $S = \{\text{ab, ba, ca, caa, caaa, baaa}\}$ over the alphabet $\Sigma = \{\text{a, b, c}\}$.
2. Consider a text B , and the suffix tree T for B . Show that a word w appears as a substring in B if and only if there is a path in T from the root to some nodes, and this path corresponds to w .
3. Create a suffix tree for the text $B = \{\text{abaabaaab}\}$ over the alphabet $\Sigma = \{\text{a, b}\}$.
4. How would you change the structure of the trie, so that you can perform the following operations on this trie:
 - (a) Given a set $S = \{w_1 \dots w_n\}$ of words, construct the trie for S in time $O(\sum_{i=1}^n |w_i|)$.
 - (b) Given a word w (not necessarily of S), find how many words in S have w as a prefix. You should be able to answer this query in time $O(|w|)$.
5. Given a text B of n characters, suggest a modification of the suffix tree data structure for B , such that the following query operation could be performed. Given a query word w , report how many times w appears (as a *contiguous* substring in B). For example

$$B = \text{"ccaaaabaaa"}$$

then the query word $w = \text{"c"}$ appears twice in B , the query word $w = \text{"cc"}$ appears once, and $w = \text{"aaa"}$ appears 3 times.

$$B = \text{"cc\underline{aaa}abaaa"} \quad B = \text{"cc\underline{aaa}ab\underline{aaa"}}, \text{"cc\underline{aaa}ab\underline{aaa"}}$$

The preprocessing time (the time for creating the structure) is $O(n^2)$, and the space required *after* for storing the data structure is $O(n)$.

6. Let k, n be given parameters, where $n = 2^i$ for some integer i . Suggest a set of words $S = \{w_1, \dots, w_n\}$ over an alphabet $\Sigma = \{a, b\}$, where $|w_i| \leq k$, such that the number of nodes in a trie storing S is as large as possible. What is this number ?