

Cs545 — Homework #5
Tries and Suffix Trees
Due 11/22/06

1. Create a trie for the set of words $S = \{ab, ba, ca, caa, caaa, baaa\}$ over the alphabet $\Sigma = \{a, b, c\}$
2. In the version of tries presented in class, every leaf of the tree is an array of size $1 + |\Sigma|$, (where Σ is the alphabet). Note that this is quite wasteful in space, since by definition, every pointer cell of a leaf contains a NULL.
 - (a) Suggest a modifications of the trie at which leafs are not stores at all, and the parent of a leaf store a special boolean field per every cell.
 - (b) Assume for simplicity that $|\Sigma| = 32$. Show how to implement the improvement while adding one extra computer word per each node.
 - (c) Show how to implement your improvement without any extra fields. Assume that there are areas in the computer memory at which cells of the tries cannot be stored.
 - (d) Show that for every value of an integer h , there is a trie of height h , for which your improvement saves at least $1/2$ of the memory needed, comparing to trie of the same set of words, that does not use this improvement. Prove your answer.
3. Let T be a tree with m leafs, and each internal (non-leaf) node has two or more children. Prove that T has $\leq m$ internal nodes.
4. Consider a text B , and the suffix trie T you are constructed 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 .
5. 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 $O(|w|)$.
6. **Bonus:** Given positive integers n and k , construct a set $S = \{w_1 \dots w_n\}$ of words over an alphabet Σ , such that $|w_i| \leq k$ for every $w_i \in S$, and the number of nodes in the trie T constructed for S is as large as possible. What is this number? Assume for simplicity that $n = |\Sigma|^\ell$ for some integer ℓ .