

3. Problem 11.4-1, page 277.

4. (Spell-checker) Assume that you are given a vocabulary — a list of $n$ words, and a sample text containing $m$ words, where $m \gg n$. Your goal is to find all words from the sample text that differ from a word from the vocabulary by replacing the first letter by another.

E.g. for the vocabulary “bat, cat, hat, is, sat” and the sample text “Pat sat on hat. That is sad.”, the output should be only the word “Pat”. (“Pat” differs from “bat” or “cat” just in the first letter, while words “sat” and “sad” in the last letter, not the first, and “that” and “hat” differ in the deletion of the letter “t”, but not in the replacement.)

Design an algorithm for the above problem, whose expected running time is $O(n + m)$. A verbal description of the idea of the algorithm is sufficient.

Some notes to simplify the argumentation: Assume that the number of possible letters is constant (like 26 for English), and also that the maximum word length is bounded by a constant. The expectation is measured across all possible sample texts. Assume also that each word may appear in the sample text with the same probability. The algorithm may use $O(n)$ extra memory.

For extra points: Extend the algorithm for all type of one-letter errors: replacement in any position, insertion, deletion. I.e., for the example above it should now output words ”Pat”, ”sad” (replacement) and ”That” (deletion).