CS 240: Data Structures and Data Management **Winter 2023**

Tutorial 10: March 27

This tutorial covers problems on string matching (Module 9). There are 4 problems in total – 3 easy [E] and 1 hard [H].

1. [E] Let $P =$ `abacabaca` and let $T =$ `abacabacdabaca`.

a) Compute the failure array

b) search for P in T using the KMP algorithm.

2. [E] Consider using the Boyer-Moore algorithm with only the Bad Character heuristic to search for a pattern P of length m in a text T of length n, with $n > m$, where P does **not** appear in T.

a) Give an example of a pattern P with length n and text T with length n that achieves the worst-case runtime for searching. Do not consider preprocessing time.

b) Same question, but for the best-case runtime.

3. [E] Let $P =$ `MOM` and let $T =$ `ALOMOMOLA`. Search for P in T using Suffix Arrays,

4. [H] Let s be a string of length n and let $\mathcal{T}_s$ denote the corresponding suffix tree. For an integer parameter $1 \leq \ell \leq n$, give an $O(n)$ time algorithm that finds the most commonly occurring substring of length ℓ in s.