SC2001/CX2101 Algorithm Design and Analysis

Tutorial 5: String Matching

College of Computing and Data Science *Nanyang Technological University*

This tutorial helps you develop skills in the learning outcome of the course: "Able to design algorithms using suitable strategies (preprocessing, etc) to solve a problem, able to analyse the efficiencies of different algorithms for problems like string matching, etc".

1. Rewrite the **simpleScan** algorithm in the lecture slides to eliminate the variable i.

2. How would you modify the **Rabin-Karp** algorithm to search for a given pattern with the additional condition that the middle character is a "wild card" (any text character at all can match it)?

3. Given pattern P = "AAA……AB" (m-1 A's followed by one B and text string T = "AAA…..A" (n A's)

   (1) Show the values of CharJump and matchJump arrays for P computed by the Boyer-Moore string matching algorithm. Assume that alpahbet is {A,B,…,Z}.

   (2) Find out exactly how many character comparisons are done by **simpleBMScan** and **BMScan** respectively to scan T for an occurrence of P.

4. Show the values of CharJump and matchJump arrays for the following patterns, which are computed by the Boyer-Moore string matching algorithm, assuming alpahbet is {A,B,…,Z}.

   (1) P = "BANANA"
   (2) P = "POTATO" // not covered if running out of time