

Tutorial 5: String Matching

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  $P = \text{“AAA.....AB”}$  ( $m-1$  A’s followed by one B and text string  $T = \text{“AAA.....A”}$  ( $n$  A’s)
  - (1) Show the values of CharJump and matchJump arrays for  $P$  computed by the Boyer-Moore string matching algorithm. Assume that alphabet 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 alphabet is  $\{A,B,...,Z\}$ .
  - (1)  $P = \text{“BANANA”}$
  - (2)  $P = \text{“POTATO”}$  // not covered if running out of time