COL106: Assignment 5

Due Date : April 15, 2019

Pattern Matching with wildcards

In this assignment, you will learn to implement suffix trees and use them for pattern matching. You are given a long text T containing English characters, numbers (and other special characters like blank, etc.). Given a string P, you have to find ALL occurences of P in T, i.e., you have to mention the locations in T where P occurs. Recall that we say P occurs in T if there is a sequence of consecutive characters in T which is same as P. To make your program more powerful, the pattern can contain the following wildcards :

- ? : the wildcard ? can match any character. For example suppose
  T = indian institute of technoligy, india
  P = i?d?

  In this case it should match this with indi in both the first word and the last word.
- * : it can match any sequence of characters. For example suppose
  T = indian institute of technology, india
  P = i*i

  In this case it should match this with "indi", "indian i", "indian insti", "indian institute of technology, i", "indian institute of technology, indi", "ian i", and so on.

You can assume for the sake of simplicity that P contains at most one * character. You can also assume that T does not contain the characters * and ?. Your implementation should be as efficient as possible. Details about input format and deliverables are here .