BIOINFORMATICS

Lab Assignment 3

Q 1: Print the number of times that the string Pattern = "TGATCA" occurs in the string Text corresponding to the DNA data present in DNA.txt file.

Use KMP algorithm of Exact matching and comment precisely on the time and space required by your algorithm.

Also Compare it with Boyre Moore Algorithm.

Q.2: Design a string-search algorithm to find if two substrings overlap in the larger string. What is the complexity of your algorithm?

Example (1):

String A: aaccggacgaaaccccgcggggaaa

Search StringB: gcgcgg

Search StringC: ggga

Output: Yes SearchString B and Search StringC overlap in StringA.

Example (2):

StringA: aaccggacgaaaccccgcggggaaa

SearchStringB: gcgcgg

SearchStringC: aa

Output: No Search String B and Search String C do not overlap in String A.

Note that there can be multiple copies of each substring in the larger string. Show this by replicating the StringA many times and showing the overlapping and non-overlapping occurances.

This assignment will be due on 8 Sept 2020.