

Exercise 6 – Karp-Rabin String Search

Due by the end of your Week-7 laboratory class.

(2 marks)

This exercise is to be done before or during your week 7 laboratory class. When you complete the exercise show your work to your lab tutor to have your work marked. The marking is based mainly on correct implementation and code readability. You should implement your code in one file (e.g. ex6.cpp, ex6.c, ex6.java). Make sure your program has a header comment block containing the name of the exercise, your name and your student login (e.g. jfk01). You may implement your solution in C, C++, Java or Python.

For this exercise, you are to implement the Karp-Rabin string search algorithm.

As usual, your program will prompt for the name of an input file and then read and process the data contained in this file.

The file contains two sequences of characters.

The first is the target sequence, *T*, the second is the search sequence *S*.

Read both strings and find all occurrences of sequence *S* in sequence *T* using the Karp-Rabin algorithm.

For each matched sequence, print the location of the first matched character in *T*.

For example, if the test data looked like this:

ACGTACGTACCAGTA

AC

The output should be:

0, 4, 8

Notes:

1. The sequence *T* is no longer than 5000 characters
2. The sequence *S* is no longer than 10 characters
3. The alphabet used in both sequences consists of the letters, A, C, G and T; the DNA base sequences.
4. Choose an appropriate modulus, *m*, for the hash function.
5. Try to make your hash computation as efficient as possible.

When you are finished, test your program using the provided text file named “Ex6.txt” and show your code and the output to your lab tutor to receive your mark.