Name: Nada Ali Ali Ahmed Keshk

Name of paper: Simple and Efficient Pattern Matching for Biological Sequence

The Authors:-

- Peyman Neamatollahi
- Montassir Hadi
- Mahmoud Naghibzadeh

Publisher: IEEE

Date of Publication: 23 January 2020

Abstract:

Pattern matching is important and practical operation in bioinformatics field.it enables users To find the locations of certain DNA sub sequences in databases or DNA sequences and because of the huge growth of biological data , We need to find high speed pattern matching algorthim to speed up the search process. This paper will introduce two pattern matching algorthims that is suitable for large data bases and DNA sequences.

Introduction:

There are many uses of pattern matching algorthims in Bioinformatics field such as local ailghment search ,Biomarker discovery and homologous series detection so this made it very important. There are already many pattern matching algorthims In Bioinformatics but we always need to develop this algorthims to make them more efficient as many of them may not be suitable for large Databaes and DNA sequences. The proposed algorthims in this paper tries to reduce the weekness of the exsiting algorithms. Like the exsiting algorthims , The proposed algorthims are divided into two phases . the first phase is preprocessing phase ,in this phase we try to find the parts of text to be matched with pattern and these parts are called windows. The second phase is matching phase , in this phase we see that if the windows that we choose in the preprocessing phase matches or mis matches with the pattern. It is logical that if we found fewer windows in the first phase (preprocessing phase) that wil reduce the time taken in the second phase(matching phase).

The proposed algorthims are:-

- First- last pattern matching algorthim (FLPM) and this algorthim look like the divide and conquer algorthim (DCPM) algorthim, but the difference between them that DCPM takes two passes to find the windows in preprocessing phase and FLPM takes one pass ,we will understand this in details in related work part and methods part.
- Process-aware pattern matching algorthim (PAPM) and this algorthim creates a new shape of pattern matching algorthim as it is a word based algorthim not a character based algorthim.This algorthim is specialized for utilizing the processing power of the processor.as we know nowadays all the processors are 32 or 64 bits so they can process 4 or 8 bytes in each excuetion cycle, so 4 or 8 characters (which is called a word) can be compared to 4 or 8 other charcters.