Data structures & Algorithms III

Assignment 01 - String Matching

Group Number 05

Deemantha H.K.T 17000254 (thdeema@gmail.com)

Dias M.D.H.R 17000272   (hashini.rangala.dias@gmail.com)

Dilushkumar V. 17000297   (dilushvasu1996@gmail.com)

Question 1

Naive string algorithm is used to implement the solution for the question1.

We used C++ language for the implantation.

Code explaination

- Receive text file name and pattern file name from user.

- Read text in text file to string text.Read pattern with wiildcard in pattern file to string wildCard

   Eg : Enter file name : text.txt

      Enter pattern file name : pattern.txt

- Used **fstream** to both read and write to/from files

- It the user mentioned text file and pattern file cannot be read it outputs the message "File not open"

- Take the length of the string to a variable named Tlength and the length of the pattern to a variable named Plength.

- If text file and pattern file are correctly read then by using ***naive algorithm*** find the starting index no of substring that matches the pattern with wildcard.

   if((text[i+j]==wildCard[j])||(wildCard[j]=='_'))

- In the inner for loop if 'j' matches the length of the pattern then it means there's a matching substring in the text

- By using **ofstream** the output will be written to the output file and it will display the indexes that the pattern can be found in the given string.

Question 2

We used Boyer Moore algorithm to solve this problem. The reason for use this algorithm is to minimize the searching time by moving the pattern through the text more than one step. So we used bad character heuristic find the steps.

This question there are only 'A', 'C', 'T', 'G'. But for the future improvement need for do it ease we bade the bad character table for all symbols. That's why we use 256 size table.

In the code we used functions.

- readQuery => this is used to read the query file and map into key, values.

- readDatab => this is used to read the database file and map into key,values

- badchar => this function is used to create the bad char table.

- Search => in this function we implement the boyer moore algorithm to find the matchings.