Course: IT 206

Project Title: ‘Text analyzer’

Group member :

Raval Kevin Kiritkumar(202001170)

Prajapati Harsh Mukundbhai(202001145)

OVERVIEW :

We made ‘text analyzer’ that analyzes our text and gives information about the character and word we use in the text also it analyzes two texts and detects if a text is copied or not from the source text.
For that, we use our data structure and object-oriented programming knowledge.

PROBLEM STATEMENT :

we all want to know and analyze our written text so that we can see how much time we use some words, how long we write, our text is copied or not from some other source or not, and many more other functionalities.

So the problem is we have to make a program that has below functionalities

1. **Character detail of the text**
2. **Word detail of the text**
3. **Find words in the text**
4. **Basic plagiarism checker**

So we have to write a program that performs the above tasks and gives the desired output.

Approach :

we decided to do this project in the small task which are.

- **Decide the logic, algorithms, and which data structure we are going to use for this project.**
- **Write pseudocode and code in VS code.**
- **Improve code using different test cases.**

How we implement different functionality

First of all, we made a character detail function which gives us information about a character we use an array to store characters and linked list for certain functions and display.

Then in word detail, we use a linked list to store words and we made a constructor that store relative information about each word, we made a function for finding the particular word in the text and its information so for that we use the stack data structure.

We make a basic plagiarism detector function(only for text) it is also our main function, For this, we take two input from the user

1. **Text on which you want to detect plagiarism**
2. **Text from where the text may be copied**

Both the inputs are in text format, after taking input we implement our own logic for plagiarism and show output as % that how much your text is copied from the source.

During this process we got errors and sometimes the code is not working. We learn from that and improve the code again and again.

In our plagiarism detection function, we use Rabin Karp algorithm for string pattern matching, and other than it we implement our own logic.

What we have learned during this project :

Doing a Project is always considered as one step ahead of learning new things, applies to us also. we learn many things from this project.

Now we have some considerable knowledge about LinkedList, Array, Stack, Queue also for our basic plagiarism checker we have learned many string pattern matching algorithms. one of them is Rabin-Karl Algorithm.

Rabin-Karp Algorithm is a pattern matching algorithm we use this algorithm for matching a pattern in a given text, it is really a good option than naive and other string matching algorithm .

Limitation of our project :

- **Our plagiarism detection option is not fully accurate.**
- **Also, the plagiarism detector does not tell that which sentences or lines from the text are copied.**
- **If we give input more than 1lakh characters then the time for giving output is very large.**

Our analysis on this is

For 1lakh character input = time taken is 3.30 min

For 2lakh character input = time taken is 13 min

But it runs perfectly for some average size text.

RESOURCES :

1. From this paper, we understood about plagiarism and the basic algorithm of how to check plagiarism.
2. Rabin Karp algorithm (geeks for geeks)
3. Rabin Karp (youtube video)
4. Positive words(we direct use from github)
5. Negative words (we direct use from github)

Thank You