Algorithm

We made our own algorithm for finding the plagiarized content between two texts (strings). At this stage, we didn’t apply the bonus task logic in our pseudocode. If we manage to make bonus task, we will include it in our implementation phase as well as provide its analysis and correctness.

In this algorithm, user will input two strings s1 and s2. We are finding the plagiarism between two strings by comparing them sentence by sentence. As we know that two piece of texts (strings) may have many words in common. For example, “the” is a common word that every piece of text has. So highlighting such little words will make so many substrings highlighted. Thus, instead of words, we are highlighting the whole sentence if found similar between them.

Example

Suppose we are giving

s1=Hi Ali. You are a boy.

s2=Hi Ahmad. You are a boy.

So in the above case, “Hi” will not be highlighted instead “You are a boy” will be highlighted.

Pseudocode

Let s1[1…n] and s2[1…n] are two input strings.

```
FIND_PLAGIARISM (s1, s2)
newf1[] <- SUBSTRING (s1)
newf2[] <- SUBSTRING (s2)
FOR i <-1 TO length[newf1]
  FOR j <- 1 TO length[newf2]
    IF newf1[i]==newf2[j]
      THEN highlight that portion
    ELSE don’t highlight anything
```

Definition of SUBSTRING () procedure:

Let s[1…n] be a string given to our procedure.

```
SUBSTRING (s)
temp1 <- length[s]
s1[] <- split[s]     //splitting our string character by character
fs1[] <- New_string []   //this string will be created at runtime for separating the whole
                         //sentences
FOR k <-1 TO temp1
    fs1[k] <- “ ”        //for ignoring NULL values
j <- 1
i <- 1
WHILE j != temp1 AND s1[j] != “.”     //because we want whole sentence to be highlighted
    fs1[i] += s1[j]
    j <- j+1
i <- i+1
return fs1
```