ISSUE

We have texts and we want to know similarity of the texts .We can add new word , delete words in the texts also we want know most 3 pairs in the text. The issue in this question ;How we can find most 3 pairs in the texts and similarity?

AİM

Find similarity and most 3 pairs words also when program work we can add new word and delete ,Of course when we did this things keep minimum in using memory.

SOLUTION

Firstly , I have to take input from user so ı make a while loop ,when user wrote –d command progress will finish. With "-r xxx.txt" command I take xxx.txt file and read and I put all word in nodes for each word a node so I can know order of words (So I can know which word is back or front

which word).When I build a link list I store head of this link list in a array so I can read how many I want .After that I send head of this link a new function , in this function I count a word how many time pass in text . After that , I add word of next node to down of which node in front of it so I have a link list but this link list have more than one same word (if this word have more than one in the text) .I have to delete node which one have same word and down of these nodes ,too ,except which one have max count .Which node is be deleting I have take to down it to which one have max count(if have same down add to count to word one otherwise I add a new node to down this node)

After all I sort the link list when my sort work like this; I did not move any node , I only move element in the nodes .After all this thing , I have a sorted link list and all of link list right.After that I look for command "-a xxx y filename" in this command ı firstly ı check in link list ı have this word or not .if ı have this word I add increasing count of this word node otherwise ı make a new node and resorted all nodes. The command "-n2 filename " is so simply in there ı look for which down have max count, of course when ı did it , if ı find the most used a pair ı marked this pairs so ı do not say again is most used a pair so ı get easily most three pairs .for "-d xxx"command I look for this word in link list after a remove it from link list when i did it ı do not forget remove down of this node, ı remove all of them ,too. Finally I can calculate similarity for this ı make a matrix and ı firstly ı put number of same word in after that ı put number of which word only in first link after that second link list .Logic is there is s(AUB)=s(A)+s(B)+s(A∩B) After that ı calculated **cosine** similarity .The last command is "-d " Remember that all this thing in a while loop and this file loop running with a int ,in there ı make value of this int 0.

What ı learned from this homework ? How ı can use delete ,add ,move a node in link list and more of it .