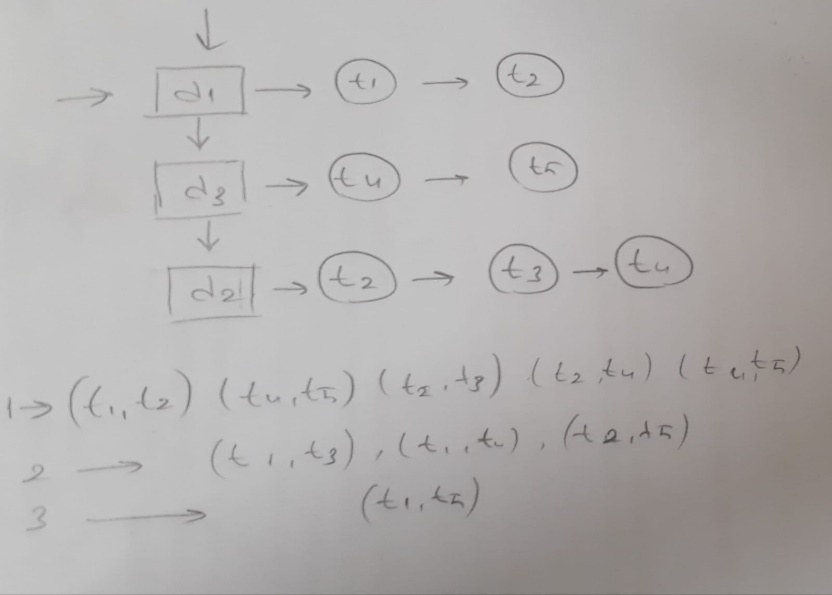
CSE225 Data Structures,2019(Fall)

Project #1 Report ***/*** Deadline : 6.11.2019

Ezgi Doğruer – 150117042

First off all , I tried to create a model in my head before I startedthe code and I installed this model.

I have 3 linked list. These holds terms, catagories and co-occurences.

There are a lot of functions in my code. Some of them are used for inserting, printing, and others for sorting.

Firstly, I tried to read directories and files but unfortunelly my code don’t run right. I can’t read all catagories in dataset directory.My code run only for econ category. I tried to hold term, catagories and co-occurances in linked lists. My code firstly open dataset and then open first category which is econ and read file names in it. Then it called the insertFile function. This function’s purpose is to add elements to nodeFile struct.

I called two function in this function.First called function is existNode . This function and the onlyExistNode function used find frequency words in option b of the Project .The other function I called is insert function. This function’s aim is to adds elements in node struct.

After these functions, I called insertFirstPath function in main function. This function is used to add term for first co-occurance. This function adds terms in nodePath struct. Then I called printPathList function in main and I printed first order term co-occurance. I could only find first order I couldn't find other second and third orders. And so I completed part a of the project, even if it wasn't completely done.

For part b of the Project , I writed a function which is frequencyList.   
This function first checks if the category contains the same word. Then call the insert function according to the situation and insert that word there.

Then for part c of the Project , function called the docCount function which finds the N-value given in the formula, ie the total number of documents. And then, calculates the idf value according to the given formula.

The function calls the sortNodeListB function after calculation. This function sorts the elements according to part b of the Project ie frequency and prints the first 5 most frequency words with how many.

After these things, the function calls the sortNodeListC function. This function also sorts elements according the value found in part of c of project whis is idf value and prints the first 5 words with the highest idf, together with their values.

Number of documents=m

Number of terms =n

The big o for part a of the project => O(n2/m)

The big o for part b of the project => O(n2)

The big o for part c of the project => O(n2)

The total big o for the Project => O(n2/m)

