Crawler

rob function : that returns a list of the disallowed urls inside the passed url.

Algorithm of this function : it searches about the (user-agent : *) in the robot.txt and append what is disallowed to url passed to this function and stores (for each url passed) all these links in ArrayList of string .

----

request function : that returns a doc (HTML) of

 the passed url.

Algorithm of this function : it takes a url as argument , and searches about this url in vistedLinks.txt (text file containing the urls visited before) and if it was visited before, it returns the doc of html to extract the links inside it ,

Else , it checks the connection through jsoup library , then

it creates a text file to put the contain of html doc in it and adds this url in visitedLinks.txt .

crawl function : recursive function that crawl every passed link for level=3

Algorithm of this function : it creates Arraylist of string to store the disallowed links from every passed url ,then it checks when reached 5000 page or not .

For each passed url , it extracts the hyperlinks inside it by

(href) , then checks if this extracted link is inside in vistedLinks or not , if not it creates a text file with this url inside it .

Indexer

Notes:

1-we take 5000 file(.txt) from crawler to indexer and make those algorithm.

PorterStemmer:

This is class which has function (stemWord)this make any word (travelling or travels….)=travel

Indexer:

It has for loop its loop 5000 time

In 1 time I take 1 document and remove all tags and remove all symbols and send to function ***countin***

Countin:

Now countin has string from doc but remove all I said first I split this string to words and I checked this array(split string) by for loop now in time I have 1 word I used hashmap<string,int> I add word in hashmab if I find it again I counter integer when I end I back to Indexer function by return this hashmap .

Here i make another hashmap<string,arraylist>inverse to save word and where it find it(computer ,find it in doc 0,5,8,9,…)

Indexer:

Now I have hashmap and number of document(counter loop) I created array list from hashmap<string ,int>myMap example

If position in array list [0] I have hashmap contain words and number of repeated in this document[0]

Now I in main I will call function fill

Fill:

To create DF and TF and store it I make this by use the two hashmap(myMap,inverse)

In main I call the last function (store) to store the four hashmap in 4 file .txt to use them in GUI and another file has links of documents

GUI:

I make class to read data from 5 file .txt those come from Indexer and store them in variables hashmap

Then I make same last lap I make home page html and css if user click search automatically run myMain class function doget this take input from user and send it to object from read class function(mai) and return string array has all links now I make new html has 10 first link from this array and this page has button next when user click it automatically run function doget from myMain class (it responsible next button) and now this page has next button if array has more links and back button to back last page of links

Notes:

I make Home button to back in Home page.