* Functions descriptions in your project (assume you have several functions in your project program)
* What is the purpose of this function?
* What is the input data of this function? ( one or two sentence description)
* What is the output of this function? ( one or two sentence description)
* What is the logic of this function? ( how does this function process, if you can write pseudocode )
* If you select word count or terasort, how do you use multiple threads or process to design your program?
* Milestone and time estimate (12/20 is the deadline for your final report and code submission )

Since (11/06), each week’s milestone (which part will you complete in that week?) What is your contribution to this milestone? (this part is separately done by yourself)

* Please sign up a github link for your project, make it private and add my name(chunyuyuan)
* Please put your group member names in your report
* Late penalty, 10% points reduce per day after deadline

Word Frequency Project

Alhamza Muswara

LLD

## Functions:

* getWordsDict(filename)

Purpose: The function getWordsDict(filename) is to return a dictionary containing meaningful words from a file and their frequency within a text.

Input: the input of the function is the name of the file you wish to read from.

Output: the function returns a dictionary with unique words as keys, and frequency as the value.

Logic:

1. def getWordsDict(filename):
2. regex = re.compile('[^a-z\'0-9]')
3. wordfreq={}
4. file= open(filename, 'r')
5. for line in file:
6. list=line.lower().split()
7. for item in list:
9. if(item in wordfreq):
10. wordfreq[regex.sub('',item)]+=1
11. else:
12. wordfreq[regex.sub('',item)]=1
13. return wordfreq

Line 2 initiates an object of re class that allows us to match literals in strings.

Line 5 loops for each line in the file; line wild hold string value.

Line 6 creates a list of strings containing each word separated by a space; and turns them into lower case before storing.

Line 7 goes through each word in the list and if there is no word already in the list it will create a new element in the dictionary with the specified word as key and set the value as 1, else it will increment the value by 1.

* def setWordsFile(wordfreq, filename)

Purpose: The function setWordsFile(wordfreq, filename) is to create a file containing the output of a dictionary list.

Input: the input of the function is the dictionary you wish to write into a file, and the second input is the string you wish to name said file.

Output: there is no return value, but there will be a new file in your current directory from wish the code runs.

1. def setWordsFile(wordfreq, filename):
2. f = open(filename, "w")
3. for word, freq in wordfreq.items():
4. f.write('%s:%s\n' % (word, freq))
5. f.close()

Line 2 creates a file with specified name.

Line 3 loops through each key and its respective value as word and freq.

Line 4 writes (word):(freq)(new-line)

Line 5 closes the file once we are finished writing.

## Threading:

As per the instruction we are to use threading to calculate the time spent on each function. We have two functions thus we use two threads, and time the beginning and the end of each function and subtract the two times respectively.