



Homework 1

Due by Mar 06 Friday 09:00

Subject: Hadoop

In this assignment you will solve the problem that was given in the exam, counting the words by their length. That is finding the number of words for each word length, #words for length 1, 2, 3, ...

Your input will be in piazza, a zip file containing 1000 Reuters news articles.

The output of the program will be a list in the following format and tab delimited:

1	4
2	12
3	45
4	78
5	123
6	...
...	

You will write your program in Java. Modify the famous Word Count example:

http://hadoop.apache.org/docs/stable1/mapred_tutorial.html

Deliverables:

1. Submit a zip file containing the MapReduce program, and named as **firstname-lastname-WordCountByLength.zip**. For example, erdogan-dogdu- WordCountByLength.zip. Zip file should include:
 - a. **WordCountByLength.java**: MapReduce program
 - b. **Output.txt**: Output list
 - c. **Report.pdf**: Report the performance of your program for 1,2,3,4,5,... nodes .

Document your programs with a header like the following:

BiL425 Software Development for Mobile Apps (Spring 2014)

Homework #

Date: Mar 06, 2014

Name: **Firstname Lastname**

2. Submit your homework to **edogdu@etu.edu.tr** by email w/ subject header "[BiL501] HW1" and email attachment containing the zip file.