Exercise 3

3D

Using the output from EX 3A in the form of `counter<tab>[words]` we did the following steps:

step 1 DATA

We loaded the merged output file from 3A and 3B namely
`frequencyTerms.txt frequencyNumberOfWords.txt` into a folder in hdfs

step 2 MAP-REDUCE and REDUCE - JOIN

We figured we didnt have to do much in the mapper as the two files has the same key:

```
"""mapper.py"""

import sys

# input comes from STDIN (standard input)
# in the form of <count><\t><words> or <count><\t><number of words>
for line in sys.stdin:
   print line.strip()
```

the join happends in the reduce phase:
the reducer can understand which type of record is parsing by checking the second column, if it is a number we just store the key in `f_key` otherwise we can print the right record having the `f_key` equals to the key of the currently parsed record:

```
#!/usr/bin/env python
"""reducer.py"""

from operator import itemgetter
import sys

n_words = 0
f_key   = 0

# input comes from STDIN
for line in sys.stdin:
    line = line.strip()
    # parse the input we got from mapper.py
    # <key> <words> or <key> <numofwords>
    values = line.split('\t')
    try:
       n_words = int(values[1])
       f_key = int(values[0])
    except ValueError:
        words = '\t'.join(values[1:])
        if f_key == int(values[0]):
            print "%d\t%s" % (n_words, words)
```

step 3 RUNNIN TASK

We ran the task the usual way using the data already loaded

```
hadoop jar /usr/local/hadoop/share/hadoop/tools/lib/hadoop-*streaming*.jar -mapper mapper.py -file ./mapper.py -r
```

step 4 CHECK RESULTS

using `hdfs dfs -text ex3d/output/* | tail -n 30` we checked our results

1	again									
1	by									
1	don									
1	eyes									
1	go									
1	harry									
1	he									
1	just									
1	know									
1	looking									
1	more									
1	mr									
1	of									
1	off									
1	re									
1	see									
1	ve									
10	burst	hogsmeade		flying	extremely		smiling	neck	lying	quiet	teacher	pain
10	case	brought	boys	doors	hold	hardly	glass	wide	today	summer		
10	walls	seat	skeeter	weren	soon	yer	real	low	continued		charlie	
5	lake	champions		funny	sent	returned						
5	task	hope	ah	dog	dropped							
6	madame	threw	stuff	seized	glasses	bill						
7	anyway	horrible		also	trees	turning	seamus	picked				
7	different		giant	beneath	covered	round	wands	thin				
7	enormous		lesson	curse	managed	pocket	sit	matter				
8	show	reckon	catch	chapter	completely		became	library	carrying			
8	snitch	stupid	twenty	granger	creatures		ago	friend	listening			

As we expected.