

Homework 06: Word Counter

CS 412

For this homework, your task is to implement a word frequency counter using Java. Your solution must read **two** inputs from the user (specifying the input and output files), count the number of unique words in the input file (ignoring case and all non-alphabetical characters), output the *total* number of words read, write an alphabetized list of all words and counts to the output file, and output the word and count of the most frequent word to the console.

REQUIREMENTS:

- Must read **two** inputs from the user:
 - inputFile: the input file to read from
 - outputFile: the output file to write the word frequencies to
- Must read words from the specified input file.
- Must convert all words to *lowercase*.
- Must ignore all non-alphabetical characters (e.g., numbers, periods, quotes, etc.,).
- Must maintain a data structure to count the number of unique words.
- Must output the number of total words read in.
- Must work for an input file of arbitrary size (e.g., Homer's Odyssey w/ 100,000+ words)
- Must write the alphabetized list of words (and counts) to the specified output file.
 - Output file must contain one word (and its count) per line
 - Format of each line- "word: count"
- Must output the *word* and *count* of the most frequent word to the console.
- Must error check file I/O.

HINTS:

- BufferedReader
- FileReader
- Scanner
- FileWriter
- String.format(..)
- try / catch
- Pre-sorted key:value data structure from Collections
- String.toLowerCase()
- String.replaceAll(*regular expression*, "")
- for-in loop
- Start small and work incrementally

EXAMPLE:

- Input file: **ItsyBitsySpider.txt**

The itsy bitsy spider crawled up the water spout.
Down came the rain, and washed the spider out.
Out came the sun, and dried up all the rain,
and the itsy bitsy spider went up the spout again

- Execution:

```
Enter input file: ItsyBitsySpider.txt  
Enter output file: stats.txt  
read          38 words  
"the" occurred 8 times
```

- Output file: **stats.txt**

again:	1
all:	1
and:	3
<u>bitsy</u> :	2
came:	2
crawled:	1
down:	1
dried:	1
<u>itsy</u> :	2
out:	2
rain:	2
spider:	3
spout:	2
sun:	1
the:	8
up:	3
washed:	1
water:	1
went:	1