**CMSC 113: Computer Science I**

**Homework: Files**

**due on Gradescope by the beginning of class on Monday, November 6, 2017**

For this assignment, you will read in text from a file that the user chooses and perform a few simple analyses on this text, reporting the results back to the user. You must perform the following 5 analyses:

1. Report the number of lines in the input file.
2. Report the number of characters in the input file. Note that files use a newline character to separate lines; these newline characters must be counted in your result.
3. Report the number of times the sequence of letters the occurs in the file. In this analysis, you are simply looking for that sequence of characters all in a row, so it will count the standalone word the as well as the word there.
4. Report the number of time the *word* the occurs in the file. Here, a *word* is a sequence of characters delimited either by the end/beginning of the string, or a non-letter. The Java method Character.isLetter detects characters, and the Java method charAt can extract a letter from a string. For example, if str is a String, then str.charAt(2) is the third letter in the string. You can test whether this character is a letter by saying Character.isLetter(str.charAt(2)).
5. Report the number of lines in the file that contain only whitespace. Whitespace is detected by the Java method Character.isWhitespace, which works similarly to Character.isLetter.

Here is a sample run of the program on the text of Lewis Carroll's *Alice's Adventures in Wonderland*, as downloaded from <http://www.gutenberg.org/ebooks/28885> and available on the course website.

What file should I read? alice.txt

1. Number of lines = 4046

2. Number of characters = 173365

3. Number of 'the' = 2334

4. Number of 'the' words = 1716

5. Number of blank lines = 1064