Lab Exercise (Python Practice)

Due Date: Tuesday 13th December in Lab

Exercise 1   Write a program that reads a file, breaks each line into words, strips whitespace and punctuation from the words, and converts them to lowercase.

Hint: The `string` module provides strings named `whitespace`, which contains space, tab, newline, etc., and `punctuation` which contains the punctuation characters.

```
>>> import string
>>> print string.punctuation
!"#$%&'()*+,-./:;<=>?@[\]^_`{|}~
```

Also, you might consider using the string methods `strip`, `replace` and `translate`.

Exercise 2 Use your program from the previous exercise to write a program that reads a plain text file and counts the number of times each unique word is used. (Hint: use the dictionary data-structure)

Exercise 3 Use the program from the previous exercise to print the 20 most frequently-used words in the book.

Exercise 4 Write a class which encapsulates the functionality you implemented earlier. We should be able to create an object of the class providing it the name of a specific text file and then call methods on the object to find information about the text file. You should implement the following methods for your class:

```
1. getFrequentWord(rank)
        # Returns the word with the passed 'rank'
        # E.g. if I call getFrequentWord(10) it
        # should return the 10th most frequent word
        # in the text file


2. getUniqueWords()
        # Returns a list of all unique words in the file

3. getWordRank(word)
        # For the given argument passed, it should return
        # the rank of the word or -1 if no such word
        # exists. For instance getWordRank('Pakistan')
        # should return 10 if 'Pakistan' is the 10th most
        # frequent word in the text file
```

Here is an example use of the class you created earlier (assuming the name of the class is WordCounter):

```
var1 = WordCounter("file1.txt")
print var1.getFrequentWord(12)
print var1.getUniqueWords()

var2 = WordCounter("file2.txt")
print var2.getFrequentWord(1)
print var2.getWordRank("the")
```