MATH5061: Week 3 Assignment

Instructions

Please submit answers to this assignment as a .tar.gz file of the directory containing the python solution source files via email to **math5061@temple.edu**. Make sure to use the subject line (without quotes) "**MATH5061:Assignment 03:ACCESSID**"
Where ACCESSID is your AccessNet ID, for example tue86537

Q1

Write a Python program that reads the included file rfc.txt and prints out a histogram of the number of occurrences of each **word**. Ignore case. And also ignore punctuation (comma, single and double quotes, period, and parenthesis, exclamation mark etc.). Sort the output so that the words with highest frequency are at the bottom of the histogram

- For the program
  - Write a function to generate a dictionary with words as keys and number of occurrences as values
  - Write a function that takes a dictionary returned by the previous function and prints a histogram that is sorted so that the highest frequencies appear last

Q1. Example Output of Your Program

Sample Input File:

```
The software tools philosophy was an important and integral concept
in the initial design and development of Unix (of which Linux and GNU
are essentially clones).
```

Output:

```
$ python3 countwords.py
integral : +
concept : +
are : +
software : +
was : +
design : +
an : +
unix : +
linux : +
development : +
clones : +
which : +
in : +
gnu : +
important : +
essentially : +
philosophy : +
initial : +
tools : +
the : ++
of : ++
and : +++
```

Hints for Q1.

- It is **very important** that you tackle this exercise in **stages.** Writing the entire program in one go will make your job **harder**
    - In class, I built the “histogram for characters” program in stages verifying that each stage works
    - For example, in the second stage I just write a program that splits the string into characters and prints each one. And then verify that it works correctly for different inputs.
    - Try a similar strategy
- Use `str.strip()` method to remove punctuation
- `The sorted()` built-in, when passed the items of a dictionary will return it sorted by the keys (we need it sorted by values)
- `sorted()` has a parameter called key which is the name of a function that takes one argument and returns the value that should be used for determining the sort order
- **Example of using `sorted()` with a key**
  The following program prints a list sorted in two different ways
    - With no key function specified. The items are sorted in their natural order
    - With key argument set to the name of a function that returns the absolute value of its argument. The resulting list is sorted by its absolute value because `sorted()` calls `myabs()` for each element in the list and uses the return value to determine sorting order.

- Example program and output:

```
def myabs(n):
    if n < 0:
        return -n
    else:
        return n

a = [-1, -3, 2, 4, -7,-5, 4]
print(sorted(a))
print(sorted(a, key=myabs))
```

```
$ python3 myabs.py
[-7, -5, -3, -1, 2, 4, 4]
[-1, 2, -3, 4, 4, -5, -7]
```

Q2

Write a program that creates a copy of a file

- Ask user for input pathname and output pathname
- Check if input file exists and inform user if it doesn’t exist
    - Ask user to try entering an input pathname again
- Print a message to the user if the output pathname already exists
    - Ask user to enter a pathname that does not already exist
- Read all contents of input file and write to the output file using only the methods mentioned in class
- Test your program with an image file ‘.JPG’ to see if the output file can be opened and is an exact copy of the input file