

Python for Informatics

Assignment 5

“Message Frequency Count”

Description:

1. Write a program that reads through the mail box data and when you find a line that starts with “From”, extract the address information from the line. Count the number of messages from each person by using a **dictionary**. Note that you might need to look at more than “From” because of duplicate instances of the address (hint: “From “ vs. “From:”). Otherwise, embedded email “thread histories” may cause your count to be incorrect.
2. After all of the data has been read, print (i.e., *print*) the address of the person with the highest number of messages, along with the number of messages of that person. To do this, create a list of tuples (count, email) from the dictionary, sort the list in reverse order and print out the person who has the highest number of messages.

Note: To succeed with this assignment, **know your data!** When your program counts the messages, how do you know if you are counting the messages correctly? Could you be counting the same message more than once? If your program were to operate correctly, how would you know it? Is there a smaller file that you can use to test your program? (Hint: There’s a file named mbox-short.txt.)

Deliverable:

Two files as attachments at our course shell assignment page. The first file should be a Python .py file with the specified functionality. The second file should be a screenshot image file (.png or .jpg) demonstrating the correct execution of your program with “mbox.txt”. Please ensure that your full name is specified as a Python comment at the top of the .py file.

Submission Deadline:

Please see the course schedule in our syllabus for all assignment submission deadlines.