Week 6 Lab Exercises

Blank notebook to be used for class exercises.

Name: Rudy Martinez

abc123: Lpe538

Exercise 1

Write code to loop over mbox.txt and counts, and prints, the number of lines thatc ontain the “From:” substring. Use the **re** package for this exercise.

file path: "./mbox.txt"

In [17]:
```
def text_counter(file_name):
    import re

    from_count = 0

    mbox_file = open(file_name)
    for word in mbox_file:
        if re.search('From:', word):
            from_count += 1
    mbox_file.close()

    return from_count
```

In [18]:
```
answer = text_counter('mbox.txt')
print(f"The number of times 'From:' is in the text is {answer} times")
```

```
The number of times 'From:' is in the text is 1797 times
```

Exercise 2

Write a program to look for lines of the form in the “mbox.txt” file:

New Revision: 39772

Extract the number from each of the lines using a regular expression andthe findall() method. Compute and print the average of the numbers.

file path: "./mbox.txt"

In [88]:

```
def num_extract(file_name):
    import re

    match_list = []
    extract = 0
    count = 0

    mbox_file = open(file_name)
    for word in mbox_file:
        if re.findall(r'New Revision:', word):
            match_list.append(word)

    for item in match_list:
        for num in item.split():
            if str(num).isdigit():
                extract += int(num)
                count += 1

    average = extract / count

    return f"The average of the numbers is: {round(average, 2)}"
```

```
    mbox_file.close()
```

In [89]:
```
answer = num_extract('mbox.txt')
answer
```

Out[89]: 'The average of the numbers is: 38549.79'

Exercise 3

For the following string:

text = "any machine with more than 6 GHz and 500 GB of disk space for less than $999.99"

Develop a regular expression that will extract **all** of the following information:

- 6 GHz
- 500 GB
- Mac
- $999.99

Hint: Use re.findall

NOTE: You should have code that is general (e.g., the regex would extrac "Mac" even though it does not appear in the string)

In [128…
```
text = "any machine with more than 6 GHz and 500 GB of disk space for less than $999.99"
```

In [129…
```
def str_extract(text):
    import re

    matches = re.findall(r"(\d+\s[Gg][Hh]z|\d+\s[Gg][Bb]|Mac|\$\d+(?:\.\d+)?)", text)

    return matches
```

In [130…
```
answer = str_extract(text)
answer
```

Out[130… ['6 GHz', '500 GB', '$999.99']

Exercise 4

Complete the lessons available at https://regexone.com/. Take a screenshot when you finish, and uplaod it along with this lab file.