# Week 6 Lab Exercises

Blank notebook to be used for class exercises.

Name: Dan Schumacher

abc123: hdd249

## 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 [94]:  import re

          mbox = open('mbox.txt')
          from_count = 0
          for line in mbox:
              if re.search('^From:',line):
                  from_count += 1


          print(from_count)
```

1797

## 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 [95]:  import re

          mbox = open('mbox.txt')
          new_rev_list = []

          for line in mbox:
              line = line.rstrip()
              new_rev = re.findall('^New Revision: (\d+)', line)
              if len(new_rev) > 0:
                  new_rev_list.append(int(new_rev[0]))
```

```
mean_rev = sum(new_rev_list)/len(new_rev_list)
print(f'Mean New Revision: {round(mean_rev,2)}')
```

Mean New Revision: 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 [96]:
```
text = "any machine with more than 6 GHz and 500 GB of disk space for less than $999.9
```

In [99]:
```
import re

for line in text:
    GHz = re.findall('(\d+)\s[Gg][Hh][Zz]\s', text)
    GB = re.findall('(\d+)\s[Gg][Bb]\s', text)
    Mac = re.findall('\s[mM]ac\s', text)
    price = re.findall('\$\d+\.\d+', text)

print('GHz:', GHz,
      '\nGB:', GB,
      '\nMac', Mac,
      '\nPrice:', price)
```

GHz: ['6']
GB: ['500']
Mac []
Price: ['$999.99']

In [98]:
```
# Teacher Solution

re.findall(r"[0-9] [Gg][Hh]z|[0-9]+ [Gg][Bb]|Mac|\$[0-9]+(?:\.[0-9]{2,2})?",text)
```

Out[98]:
['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.