***##### Roll No : 225229111***
***##### Name : Hariharan S***

## ### Lab.8 : Python Regular Expressions

Qunestion 1 : Using Email Collection file,mbox-short.txt,write a python program for the following queries

In [2]:

```
#Lab :8
#Pgm :1
#Qns :1

import re
mail=open("D:\PSPR\Python Coding\email.txt","r")
for line in mail:
    line=line.rstrip()
    if re.search('From:',line):
        print(line)
```

```
From: ds225229104@bhc.edu.in Thu Jan 3 19:51:21 2010
From: ds225229103@bhc.edu.in Sat Jan 5 09:14:16 2010
From: arulkumar1924@gmail.com Fri Jan 4 18:10:48 2011
From: ds225229101@bhc.edu.in Fri Jan 4 10:38:42 2011
From: ds225229126@bhc.edu.in Fri Jan 4 10:17:43 2015
From: swathi@caret.cam.ac.uk Fri Jan 4 10:04:14 2021
From: jumboarkk@gmail.com Fri Jan 4 09:05:31 2022
From: ds225229104@bhc.edu.in Thu Jan 3 19:51:21 2022
```

In [3]:

```
#Pgm :1
#Qns :2

import re
mail=open("D:\PSPR\Python Coding\email.txt","r")
for line in mail:
    line=line.rstrip()
    if re.search('F',line):
        print(line)
```

```
From ds225229103@bhc.edu.in Sat Jan 5 09:14:16 2008
From arulkumar1924@gmail.com Fri Jan 4 18:10:48 2008
From ds225229101@bhc.edu.in Fri Jan 4 10:38:42 2008
From ds225229126@bhc.edu.in Fri Jan 4 10:17:43 2008
From swathi@caret.cam.ac.uk Fri Jan 4 10:04:14 2008
From jumboarkk@gmail.com Fri Jan 4 09:05:31 2008
From: ds225229104@bhc.edu.in Thu Jan 3 19:51:21 2010
From: ds225229103@bhc.edu.in Sat Jan 5 09:14:16 2010
From: arulkumar1924@gmail.com Fri Jan 4 18:10:48 2011
From: ds225229101@bhc.edu.in Fri Jan 4 10:38:42 2011
From: ds225229126@bhc.edu.in Fri Jan 4 10:17:43 2015
From: swathi@caret.cam.ac.uk Fri Jan 4 10:04:14 2021
From: jumboarkk@gmail.com Fri Jan 4 09:05:31 2022
From: ds225229104@bhc.edu.in Thu Jan 3 19:51:21 2022
```

In [4]:

```
#Pgm :1
#Qns :3

import re
mail=open("D:\PSPR\Python Coding\email.txt","r")
for line in mail:
    line=line.rstrip()
    if re.search('F..m:',line):
        print(line)
```

```
From: ds225229104@bhc.edu.in Thu Jan 3 19:51:21 2010
From: ds225229103@bhc.edu.in Sat Jan 5 09:14:16 2010
From: arulkumar1924@gmail.com Fri Jan 4 18:10:48 2011
From: ds225229101@bhc.edu.in Fri Jan 4 10:38:42 2011
From: ds225229126@bhc.edu.in Fri Jan 4 10:17:43 2015
From: swathi@caret.cam.ac.uk Fri Jan 4 10:04:14 2021
From: jumboarkk@gmail.com Fri Jan 4 09:05:31 2022
From: ds225229104@bhc.edu.in Thu Jan 3 19:51:21 2022
```

In [5]:

```
#Pgm :1
#Qns :4

import re
mail=open("D:\PSPR\Python Coding\email.txt","r")
for line in mail:
    line=line.rstrip()
    if re.search('From.+@',line):
        print(line)
```

```
From ds225229103@bhc.edu.in Sat Jan 5 09:14:16 2008
From arulkumar1924@gmail.com Fri Jan 4 18:10:48 2008
From ds225229101@bhc.edu.in Fri Jan 4 10:38:42 2008
From ds225229126@bhc.edu.in Fri Jan 4 10:17:43 2008
From swathi@caret.cam.ac.uk Fri Jan 4 10:04:14 2008
From jumboarkk@gmail.com Fri Jan 4 09:05:31 2008
From: ds225229104@bhc.edu.in Thu Jan 3 19:51:21 2010
From: ds225229103@bhc.edu.in Sat Jan 5 09:14:16 2010
From: arulkumar1924@gmail.com Fri Jan 4 18:10:48 2011
From: ds225229101@bhc.edu.in Fri Jan 4 10:38:42 2011
From: ds225229126@bhc.edu.in Fri Jan 4 10:17:43 2015
From: swathi@caret.cam.ac.uk Fri Jan 4 10:04:14 2021
From: jumboarkk@gmail.com Fri Jan 4 09:05:31 2022
From: ds225229104@bhc.edu.in Thu Jan 3 19:51:21 2022
```

In [6]:

```
#Pgm :1
#Qns :5

s="Asha send a message from the address swathi@caret.cam.ac.uk to Arul address of arulkumar
fm=re.findall('\S+@\S+',s)
print("The Mail Adress :")
for i in fm:
    print(i)
```

```
The Mail Adress :
swathi@caret.cam.ac.uk
arulkumar1924@gmail.com
```

In [7]:

```
#Pgm :1
#Qns :6

import re
mail=open("D:\PSPR\Python Coding\email.txt","r")
for line in mail:
    line=line.rstrip()
    txt=re.findall('\S+@\S+',line)
    if len(txt)>0:
        print(txt)
```

```
['ds225229103@bhc.edu.in']
['arulkumar1924@gmail.com']
['ds225229101@bhc.edu.in']
['ds225229126@bhc.edu.in']
['swathi@caret.cam.ac.uk']
['jumboarkk@gmail.com']
['ds225229104@bhc.edu.in']
['ds225229103@bhc.edu.in']
['arulkumar1924@gmail.com']
['ds225229101@bhc.edu.in']
['ds225229126@bhc.edu.in']
['swathi@caret.cam.ac.uk']
['jumboarkk@gmail.com']
['ds225229104@bhc.edu.in']
```

In [8]:

```
#Pgm :1
#Qns :7

import re
mail=open("D:\PSPR\Python Coding\email.txt","r")
for line in mail:
    line=line.rstrip()
    txt=re.findall('[a-zA-Z0-9]\S*@\S*[a-zA-Z0-9]',line)
    if len(txt)>0:
        print(txt)
```

```
['ds225229103@bhc.edu.in']
['arulkumar1924@gmail.com']
['ds225229101@bhc.edu.in']
['ds225229126@bhc.edu.in']
['swathi@caret.cam.ac.uk']
['jumboarkk@gmail.com']
['ds225229104@bhc.edu.in']
['ds225229103@bhc.edu.in']
['arulkumar1924@gmail.com']
['ds225229101@bhc.edu.in']
['ds225229126@bhc.edu.in']
['swathi@caret.cam.ac.uk']
['jumboarkk@gmail.com']
['ds225229104@bhc.edu.in']
```

In [9]:

```
#Pgm :1
#Qns :8

import re
mail=open("D:\PSPR\Python Coding\mbox_short.txt")
for line in mail:
    line=line.rstrip()
    if re.search('X\S*: [0-9]+',line):
        print(line)
```

```
X-DSPAM-Confidence: 0.8475
X-DSPAM-Probability: 0.0000
X-DSPAM-Confidence: 0.6178
X-DSPAM-Probability: 0.0000
X-DSPAM-Confidence: 0.6961
X-DSPAM-Probability: 0.0000
X-DSPAM-Confidence: 0.7565
X-DSPAM-Probability: 0.0000
X-DSPAM-Confidence: 0.7626
X-DSPAM-Probability: 0.0000
X-DSPAM-Confidence: 0.7556
X-DSPAM-Probability: 0.0000
X-DSPAM-Confidence: 0.7002
X-DSPAM-Probability: 0.0000
X-DSPAM-Confidence: 0.7615
X-DSPAM-Probability: 0.0000
X-DSPAM-Confidence: 0.7601
X-DSPAM-Probability: 0.0000
X-DSPAM-Confidence: 0.7605
X-DSPAM-Probability: 0.0000
X-DSPAM-Confidence: 0.6959
X-DSPAM-Probability: 0.0000
X-DSPAM-Confidence: 0.7606
X-DSPAM-Probability: 0.0000
X-DSPAM-Confidence: 0.7559
X-DSPAM-Probability: 0.0000
X-DSPAM-Confidence: 0.7605
X-DSPAM-Probability: 0.0000
X-DSPAM-Confidence: 0.6932
X-DSPAM-Probability: 0.0000
X-DSPAM-Confidence: 0.7558
X-DSPAM-Probability: 0.0000
X-DSPAM-Confidence: 0.6526
X-DSPAM-Probability: 0.0000
X-DSPAM-Confidence: 0.6948
X-DSPAM-Probability: 0.0000
X-DSPAM-Confidence: 0.6528
X-DSPAM-Probability: 0.0000
X-DSPAM-Confidence: 0.7002
X-DSPAM-Probability: 0.0000
X-DSPAM-Confidence: 0.7554
X-DSPAM-Probability: 0.0000
X-DSPAM-Confidence: 0.6956
X-DSPAM-Probability: 0.0000
X-DSPAM-Confidence: 0.6959
X-DSPAM-Probability: 0.0000
X-DSPAM-Confidence: 0.7556
X-DSPAM-Probability: 0.0000
X-DSPAM-Confidence: 0.9846
```

```
X-DSPAM-Probability: 0.0000
X-DSPAM-Confidence: 0.8509
X-DSPAM-Probability: 0.0000
X-DSPAM-Confidence: 0.9907
X-DSPAM-Probability: 0.0000
```

In [10]:

```
#Pgm :1
#Qns :9

import re
mail=open("D:\PSPR\Python Coding\email.txt","r")
for line in mail:
    line=line.rstrip()
    txt=re.findall('Details:,*rev=([0-9]+)',line)
    if len(txt)>0:
        print(txt)
```

```
['225229104']
['225229106']
['225229114']
['225229150']
```

In [11]:

```
#Pgm :1
#Qns :10

import re
mail=open("D:\PSPR\Python Coding\email.txt","r")
for line in mail:
    line=line.rstrip()
    txt=re.findall('From.*([0-9][0-9])',line)
    if len(txt)>0:
        print(txt)
```

```
['08']
['08']
['08']
['08']
['08']
['08']
['10']
['10']
['11']
['11']
['15']
['21']
['22']
['22']
```

Qunestion 2 : Baby Names Popularrity Analysis

In [12]:

```
import re
import sys

def extract_names(filename):
    names=[]

    f=open(filename,'r')
    txt=f.read()
    year_match=re.search(r'Popularity\sin\s(\d\d\d\d)',txt)
    if not year_match:
        sys.stderr.write('Couldn\'\t find the year!\n\n')
        sys.exit(1)
    year=year_match.group(1)
    names.append(year)
    tuples=re.findall(r'<td>(\d+)</td><td>(\w+)</td>\<td>(\w+)</td>',txt)
    names_to_rank={}
    for rank_tuple in tuples:
        (rank,boyname,girlname)=rank_tuple
        if boyname not in names_to_rank:
            names_to_rank[boyname]=rank
        if girlname not in names_to_rank:
            names_to_rank[girlname]=rank
    sorted_name=sorted(names_to_rank.keys())
    for name in sorted_name:
        names.append(name+" "+names_to_rank[name])
    return (names)

#main:
extract_names("baby1990.html")
```

Out[12]:

```
['1990',
 'Aaron 34',
 'Abbey 482',
 'Abbie 685',
 'Abby 222',
 'Abdul 934',
 'Abel 384',
 'Abigail 90',
 'Abraham 246',
 'Abram 920',
 'Adam 32',
 'Adan 548',
 'Addison 645',
 'Adolfo 649',
 'Adrian 94',
 'Adriana 144',
 'Adrianna 325',
 'Adrianne 783',
```

In [ ]:

In [ ]: