

```
In [ ]: #DINESH
        #225229108
```

```
In [16]: #1

fhand = open('mbox-short.txt','r')
count = 0
for line in fhand:
    line = line.rstrip()
    if line == "": continue

    words = line.split()
    if "from" in words: continue
    print(words)
    count = count+1

print ("There were", count, "lines in the file which contains from in it.")

['From', 'stephen.marquard@uct.ac.za', 'Sat', 'Jan', '5', '09:14:16', '2008']
['From', 'louis@media.berkeley.edu', 'Fri', 'Jan', '4', '18:10:48', '2008']
['From', 'zqian@umich.edu', 'Fri', 'Jan', '4', '16:10:39', '2008']
['From', 'rjlowe@iupui.edu', 'Fri', 'Jan', '4', '15:46:24', '2008']
['From', 'zqian@umich.edu', 'Fri', 'Jan', '4', '15:03:18', '2008']
['From', 'rjlowe@iupui.edu', 'Fri', 'Jan', '4', '14:50:18', '2008']
['From', 'cwen@iupui.edu', 'Fri', 'Jan', '4', '11:37:30', '2008']
['From', 'cwen@iupui.edu', 'Fri', 'Jan', '4', '11:35:08', '2008']
['From', 'gsilver@umich.edu', 'Fri', 'Jan', '4', '11:12:37', '2008']
['From', 'gsilver@umich.edu', 'Fri', 'Jan', '4', '11:11:52', '2008']
['From', 'zqian@umich.edu', 'Fri', 'Jan', '4', '11:11:03', '2008']
['From', 'gsilver@umich.edu', 'Fri', 'Jan', '4', '11:10:22', '2008']
['From', 'wagnermr@iupui.edu', 'Fri', 'Jan', '4', '10:38:42', '2008']
['From', 'zqian@umich.edu', 'Fri', 'Jan', '4', '10:17:43', '2008']
['From', 'antranig@caret.cam.ac.uk', 'Fri', 'Jan', '4', '10:04:14', '2008']
['From', 'gopal.ramasammycook@gmail.com', 'Fri', 'Jan', '4', '09:05:31', '2008']
['From', 'david.horwitz@uct.ac.za', 'Fri', 'Jan', '4', '07:02:32', '2008']
['From', 'david.horwitz@uct.ac.za', 'Fri', 'Jan', '4', '06:08:27', '2008']
['From', 'david.horwitz@uct.ac.za', 'Fri', 'Jan', '4', '04:49:08', '2008']
['From', 'david.horwitz@uct.ac.za', 'Fri', 'Jan', '4', '04:33:44', '2008']
['From', 'stephen.marquard@uct.ac.za', 'Fri', 'Jan', '4', '04:07:34', '2008']
['From', 'louis@media.berkeley.edu', 'Thu', 'Jan', '3', '19:51:21', '2008']
['From', 'louis@media.berkeley.edu', 'Thu', 'Jan', '3', '17:18:23', '2008']
['From', 'ray@media.berkeley.edu', 'Thu', 'Jan', '3', '17:07:00', '2008']
['From', 'cwen@iupui.edu', 'Thu', 'Jan', '3', '16:34:40', '2008']
['From', 'cwen@iupui.edu', 'Thu', 'Jan', '3', '16:29:07', '2008']
['From', 'cwen@iupui.edu', 'Thu', 'Jan', '3', '16:23:48', '2008']
There were 27 lines in the file which contains from in it.
```

In [17]: #2

```
fhand = open(r'mbox-short.txt', 'r')
count = 0
for line in fhand:
    line = line.rstrip()
    if line == "": continue

    words = line.split()
    if words[0] != "From": continue
    print(words)
    count = count+1

print("There were", count, "lines in the file with From as the first word")
```

```
['From', 'stephen.marquard@uct.ac.za', 'Sat', 'Jan', '5', '09:14:16', '2008']
['From', 'louis@media.berkeley.edu', 'Fri', 'Jan', '4', '18:10:48', '2008']
['From', 'zqian@umich.edu', 'Fri', 'Jan', '4', '16:10:39', '2008']
['From', 'rjlowe@iupui.edu', 'Fri', 'Jan', '4', '15:46:24', '2008']
['From', 'zqian@umich.edu', 'Fri', 'Jan', '4', '15:03:18', '2008']
['From', 'rjlowe@iupui.edu', 'Fri', 'Jan', '4', '14:50:18', '2008']
['From', 'cwen@iupui.edu', 'Fri', 'Jan', '4', '11:37:30', '2008']
['From', 'cwen@iupui.edu', 'Fri', 'Jan', '4', '11:35:08', '2008']
['From', 'gsilver@umich.edu', 'Fri', 'Jan', '4', '11:12:37', '2008']
['From', 'gsilver@umich.edu', 'Fri', 'Jan', '4', '11:11:52', '2008']
['From', 'zqian@umich.edu', 'Fri', 'Jan', '4', '11:11:03', '2008']
['From', 'gsilver@umich.edu', 'Fri', 'Jan', '4', '11:10:22', '2008']
['From', 'wagnermr@iupui.edu', 'Fri', 'Jan', '4', '10:38:42', '2008']
['From', 'zqian@umich.edu', 'Fri', 'Jan', '4', '10:17:43', '2008']
['From', 'antranig@caret.cam.ac.uk', 'Fri', 'Jan', '4', '10:04:14', '2008']
['From', 'gopal.ramasammycook@gmail.com', 'Fri', 'Jan', '4', '09:05:31', '2008']
['From', 'david.horwitz@uct.ac.za', 'Fri', 'Jan', '4', '07:02:32', '2008']
['From', 'david.horwitz@uct.ac.za', 'Fri', 'Jan', '4', '06:08:27', '2008']
['From', 'david.horwitz@uct.ac.za', 'Fri', 'Jan', '4', '04:49:08', '2008']
['From', 'david.horwitz@uct.ac.za', 'Fri', 'Jan', '4', '04:33:44', '2008']
['From', 'stephen.marquard@uct.ac.za', 'Fri', 'Jan', '4', '04:07:34', '2008']
['From', 'louis@media.berkeley.edu', 'Thu', 'Jan', '3', '19:51:21', '2008']
['From', 'louis@media.berkeley.edu', 'Thu', 'Jan', '3', '17:18:23', '2008']
['From', 'ray@media.berkeley.edu', 'Thu', 'Jan', '3', '17:07:00', '2008']
['From', 'cwen@iupui.edu', 'Thu', 'Jan', '3', '16:34:40', '2008']
['From', 'cwen@iupui.edu', 'Thu', 'Jan', '3', '16:29:07', '2008']
['From', 'cwen@iupui.edu', 'Thu', 'Jan', '3', '16:23:48', '2008']
There were 27 lines in the file with From as the first word
```

In [18]: #3

```
fhand = open(r'mbox-short.txt', 'r')
count = 0
for line in fhand:
    words = line.split()
    for i in words:
        if i.startswith('F') and i.endswith('m'):
            print(words)
            count = count+1

print ("There were", count, "lines in the file with F as the first word and
```

```
['From', 'stephen.marquard@uct.ac.za', 'Sat', 'Jan', '5', '09:14:16', '2008']
['From', 'louis@media.berkeley.edu', 'Fri', 'Jan', '4', '18:10:48', '2008']
['From', 'zqian@umich.edu', 'Fri', 'Jan', '4', '16:10:39', '2008']
['From', 'rjlowe@iupui.edu', 'Fri', 'Jan', '4', '15:46:24', '2008']
['From', 'zqian@umich.edu', 'Fri', 'Jan', '4', '15:03:18', '2008']
['From', 'rjlowe@iupui.edu', 'Fri', 'Jan', '4', '14:50:18', '2008']
['From', 'cwen@iupui.edu', 'Fri', 'Jan', '4', '11:37:30', '2008']
['From', 'cwen@iupui.edu', 'Fri', 'Jan', '4', '11:35:08', '2008']
['From', 'gsilver@umich.edu', 'Fri', 'Jan', '4', '11:12:37', '2008']
['From', 'gsilver@umich.edu', 'Fri', 'Jan', '4', '11:11:52', '2008']
['From', 'zqian@umich.edu', 'Fri', 'Jan', '4', '11:11:03', '2008']
['From', 'gsilver@umich.edu', 'Fri', 'Jan', '4', '11:10:22', '2008']
['From', 'wagnermr@iupui.edu', 'Fri', 'Jan', '4', '10:38:42', '2008']
['From', 'zqian@umich.edu', 'Fri', 'Jan', '4', '10:17:43', '2008']
['From', 'antranig@caret.cam.ac.uk', 'Fri', 'Jan', '4', '10:04:14', '2008']
['From', 'gopal.ramasammycook@gmail.com', 'Fri', 'Jan', '4', '09:05:31', '2008']
['From', 'david.horwitz@uct.ac.za', 'Fri', 'Jan', '4', '07:02:32', '2008']
['From', 'david.horwitz@uct.ac.za', 'Fri', 'Jan', '4', '06:08:27', '2008']
['From', 'david.horwitz@uct.ac.za', 'Fri', 'Jan', '4', '04:49:08', '2008']
['From', 'david.horwitz@uct.ac.za', 'Fri', 'Jan', '4', '04:33:44', '2008']
['From', 'stephen.marquard@uct.ac.za', 'Fri', 'Jan', '4', '04:07:34', '2008']
['From', 'louis@media.berkeley.edu', 'Thu', 'Jan', '3', '19:51:21', '2008']
['From', 'louis@media.berkeley.edu', 'Thu', 'Jan', '3', '17:18:23', '2008']
['From', 'ray@media.berkeley.edu', 'Thu', 'Jan', '3', '17:07:00', '2008']
['From', 'cwen@iupui.edu', 'Thu', 'Jan', '3', '16:34:40', '2008']
['From', 'cwen@iupui.edu', 'Thu', 'Jan', '3', '16:29:07', '2008']
['From', 'cwen@iupui.edu', 'Thu', 'Jan', '3', '16:23:48', '2008']
There were 27 lines in the file with F as the first word and ends with m i
n second word
```

In [19]: #4

```

fhand = open(r'mbox-short.txt', 'r')
count = 0
for line in fhand:
    line = line.rstrip()
    if line == "": continue

    words = line.split()
    if words[0] != "From" and '@' in words: continue
    print(words)
    count = count+1

print ("There were", count, "lines in the file with From as the first word")

```

```

['From', 'stephen.marquard@uct.ac.za', 'Sat', 'Jan', '5', '09:14:16', '2008']
['From', 'louis@media.berkeley.edu', 'Fri', 'Jan', '4', '18:10:48', '2008']
['From', 'zqian@umich.edu', 'Fri', 'Jan', '4', '16:10:39', '2008']
['From', 'rjlowe@iupui.edu', 'Fri', 'Jan', '4', '15:46:24', '2008']
['From', 'zqian@umich.edu', 'Fri', 'Jan', '4', '15:03:18', '2008']
['From', 'rjlowe@iupui.edu', 'Fri', 'Jan', '4', '14:50:18', '2008']
['From', 'cwen@iupui.edu', 'Fri', 'Jan', '4', '11:37:30', '2008']
['From', 'cwen@iupui.edu', 'Fri', 'Jan', '4', '11:35:08', '2008']
['From', 'gsilver@umich.edu', 'Fri', 'Jan', '4', '11:12:37', '2008']
['From', 'gsilver@umich.edu', 'Fri', 'Jan', '4', '11:11:52', '2008']
['From', 'zqian@umich.edu', 'Fri', 'Jan', '4', '11:11:03', '2008']
['From', 'gsilver@umich.edu', 'Fri', 'Jan', '4', '11:10:22', '2008']
['From', 'wagnermr@iupui.edu', 'Fri', 'Jan', '4', '10:38:42', '2008']
['From', 'zqian@umich.edu', 'Fri', 'Jan', '4', '10:17:43', '2008']
['From', 'antranig@caret.cam.ac.uk', 'Fri', 'Jan', '4', '10:04:14', '2008']
['From', 'gopal.ramasammycook@gmail.com', 'Fri', 'Jan', '4', '09:05:31', '2008']
['From', 'david.horwitz@uct.ac.za', 'Fri', 'Jan', '4', '07:02:32', '2008']
['From', 'david.horwitz@uct.ac.za', 'Fri', 'Jan', '4', '06:08:27', '2008']
['From', 'david.horwitz@uct.ac.za', 'Fri', 'Jan', '4', '04:49:08', '2008']
['From', 'david.horwitz@uct.ac.za', 'Fri', 'Jan', '4', '04:33:44', '2008']
['From', 'stephen.marquard@uct.ac.za', 'Fri', 'Jan', '4', '04:07:34', '2008']
['From', 'louis@media.berkeley.edu', 'Thu', 'Jan', '3', '19:51:21', '2008']
['From', 'louis@media.berkeley.edu', 'Thu', 'Jan', '3', '17:18:23', '2008']
['From', 'ray@media.berkeley.edu', 'Thu', 'Jan', '3', '17:07:00', '2008']
['From', 'cwen@iupui.edu', 'Thu', 'Jan', '3', '16:34:40', '2008']
['From', 'cwen@iupui.edu', 'Thu', 'Jan', '3', '16:29:07', '2008']
['From', 'cwen@iupui.edu', 'Thu', 'Jan', '3', '16:23:48', '2008']
There were 27 lines in the file with From as the first word and have @ character

```

In [20]: #5

```
import re
fhand = open(r'mbox-short.txt', 'r')
count = 0
for line in fhand:
    x=re.findall("@",line)
    count = count+1
    print(line)
if x:
    print ("\nThere were", count, "lines in the file have @ character")
else:
    print("not a match")
```

From stephen.marquard@uct.ac.za Sat Jan 5 09:14:16 2008

From louis@media.berkeley.edu Fri Jan 4 18:10:48 2008

From zqian@umich.edu Fri Jan 4 16:10:39 2008

From rjlowe@iupui.edu Fri Jan 4 15:46:24 2008

From zqian@umich.edu Fri Jan 4 15:03:18 2008

From rjlowe@iupui.edu Fri Jan 4 14:50:18 2008

From cwen@iupui.edu Fri Jan 4 11:37:30 2008

From cwen@iupui.edu Fri Jan 4 11:35:08 2008

From gsilver@umich.edu Fri Jan 4 11:12:37 2008

From gsilver@umich.edu Fri Jan 4 11:11:52 2008

From zqian@umich.edu Fri Jan 4 11:11:03 2008

From gsilver@umich.edu Fri Jan 4 11:10:22 2008

From wagnermr@iupui.edu Fri Jan 4 10:38:42 2008

From zqian@umich.edu Fri Jan 4 10:17:43 2008

From antranig@caret.cam.ac.uk Fri Jan 4 10:04:14 2008

From gopal.ramasammycook@gmail.com Fri Jan 4 09:05:31 2008

From david.horwitz@uct.ac.za Fri Jan 4 07:02:32 2008

From david.horwitz@uct.ac.za Fri Jan 4 06:08:27 2008

From david.horwitz@uct.ac.za Fri Jan 4 04:49:08 2008

From david.horwitz@uct.ac.za Fri Jan 4 04:33:44 2008

From stephen.marquard@uct.ac.za Fri Jan 4 04:07:34 2008

From louis@media.berkeley.edu Thu Jan 3 19:51:21 2008

From louis@media.berkeley.edu Thu Jan 3 17:18:23 2008

From ray@media.berkeley.edu Thu Jan 3 17:07:00 2008

From cwen@iupui.edu Thu Jan 3 16:34:40 2008

From cwen@iupui.edu Thu Jan 3 16:29:07 2008

From cwen@iupui.edu Thu Jan 3 16:23:48 2008

There were 27 lines in the file have @ character

In [5]: #6

```
import re
f=open('mbox-short.txt')
for line in f:
    line=line.strip()
    x=re.findall('[a-zA-Z0-9]\S+@\S+[a-zA-Z]',line)
    if len(x)>0:
```

```
['stephen.marquard@uct.ac.za']
['louis@media.berkeley.edu']
['zqian@umich.edu']
['rjlowe@iupui.edu']
['zqian@umich.edu']
['rjlowe@iupui.edu']
['cwen@iupui.edu']
['cwen@iupui.edu']
['gsilver@umich.edu']
['gsilver@umich.edu']
['zqian@umich.edu']
['gsilver@umich.edu']
['wagnermr@iupui.edu']
['zqian@umich.edu']
['antranig@caret.cam.ac.uk']
['gopal.ramasammycook@gmail.com']
['david.horwitz@uct.ac.za']
['david.horwitz@uct.ac.za']
['david.horwitz@uct.ac.za']
['david.horwitz@uct.ac.za']
['stephen.marquard@uct.ac.za']
['louis@media.berkeley.edu']
['louis@media.berkeley.edu']
['ray@media.berkeley.edu']
['cwen@iupui.edu']
['cwen@iupui.edu']
['cwen@iupui.edu']
```

In [11]: #7

```
import re
f = open('mbox-short1.txt')
for line in f:
    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
```


In [15]: #8

```
import re
hand = open('mbox-short1.txt')
for line in hand:
    line = line.rstrip()
    x = re.findall('^X\S*: ([0-9.]+)', line)
    if len(x) > 0:
        print(x)
```

```
['0.8475']
['0.0000']
['0.6178']
['0.0000']
['0.6961']
['0.0000']
['0.7565']
['0.0000']
['0.7626']
['0.0000']
['0.7556']
['0.0000']
['0.7002']
['0.0000']
['0.7615']
['0.0000']
['0.7601']
['0.0000']
['0.7605']
['0.0000']
['0.6959']
['0.0000']
['0.7606']
['0.0000']
['0.7559']
['0.0000']
['0.7605']
['0.0000']
['0.6932']
['0.0000']
['0.7558']
['0.0000']
['0.6526']
['0.0000']
['0.6948']
['0.0000']
['0.6528']
['0.0000']
['0.7002']
['0.0000']
['0.7554']
['0.0000']
['0.6956']
['0.0000']
['0.6959']
['0.0000']
['0.7556']
['0.0000']
```

```
['0.9846']  
['0.0000']  
['0.8509']  
['0.0000']
```

In [13]: #9

```
import re  
hand = open('mbox-short1.txt')  
for line in hand:  
    line = line.rstrip()  
    x = re.findall('^Details:.*rev=([0-9.]+)', line)  
    if len(x) > 0:
```

```
['39772']  
['39771']  
['39770']  
['39769']  
['39766']  
['39765']  
['39764']  
['39763']  
['39762']  
['39761']  
['39760']  
['39759']  
['39758']  
['39757']  
['39756']  
['39755']  
['39754']  
['39753']  
['39752']  
['39751']  
['39750']  
['39749']  
['39746']  
['39745']  
['39744']
```

In [9]: #10

```
import re
hand = open('mbox-short.txt')
for line in hand:
    line = line.rstrip()
    x = re.findall('^From .* ([0-9][0-9]):', line)
    if len(x) > 0:
```

```
['09']
['18']
['16']
['15']
['15']
['14']
['11']
['11']
['11']
['11']
['11']
['10']
['10']
['10']
['09']
['07']
['06']
['04']
['04']
['04']
['19']
['17']
['17']
['16']
['16']
['16']
```

In []:

In []:

In []: