

In []:

In []: In [3]:

```
# In order to read any file  
# Where is file Located: Folder name  
# what is the name of the file  
# what is the file extension
```

File name **type** of the file

Text file .txt

word document .doc **or** .docx

Excel sheet .csv (Comma seperated value)

Excel sheet .xlsx

Pdf .pdf

json .json

```
##### Encoded ##### xml .xml
```

Delta .delta (DE)

Parquet .delta (DE)

tf record '\x\b\29'

image .jpg .png

Text file

Excel sheet .csv (Comma seperated value)

Excel sheet .xlsx

json .json

Pdf .pdf

xml .xml

tf record '\x\b\29'

images

mbox-short

#folder_name="C:\Users\omkar\OneDrive\Documents\Data science\Naresh IT\Dataf

#file_name="mbox-short"

#extension=".txt"

#<folder_name>\<file_name><.txt>

"C:\\Users\\omkar\\OneDrive\\Documents\\Data science\\Naresh IT\\Datafiles\\

"C:\\Users\\omkar\\OneDrive\\Documents\\Data science\\Naresh IT\\Datafiles\\

"C:\\Users\\omkar\\OneDrive\\Documents\\Data science\\Naresh IT\\Datafiles\\

"C:\\Users\\omkar\\OneDrive\\Documents\\Data science\\Naresh IT\\Datafiles\\

Out[3]: 'C:\\Users\\omkar\\OneDrive\\Documents\\Data science\\Naresh IT\\Datafiles
\\IMDB Dataset.csv'

In [4]: ta science\Naresh IT\Datafiles"
r"C:\Users\omkar\OneDrive\Documents\Da

Out[4]: 'C:\\Users\\omkar\\OneDrive\\Documents\\Data science\\Naresh IT\\Datafile
s'

In [5]: In [6]:

Cell In[5], line 1

"C:\Users\omkar\OneDrive\Documents\Data
science\Naresh IT\Datafiles" ^

SyntaxError: (unicode error)

'unicodeescape' codec can't decode bytes
in p osition 2-3: truncated \UXXXXXXX
escape

Unicode error: single slash
either make into double slash
or keep an 'r' befor the paths

Textfiles

"C:\Users\omkar\OneDrive\Documents\Data
science\Naresh IT\Datafiles"

path="C:\\Users\\omkar\\OneDrive\\Document
s\\Data science\\Naresh IT\\Datafi path

```
Out[6]: 'C:\\Users\\omkar\\OneDrive\\Documents\\Data science\\Naresh IT\\Datafiles
\\mbox-short.txt'
file=open(path,
th) file
```

In [15]:

```
Out[15]: <_io.TextIOWrapper name='C:\\Users\\omkar\\OneDrive\\Documents\\Data scien
ce\\Naresh IT\\Datafiles\\mbox-short.txt' mode='r' encoding='cp1252'>
```

In [13]:

```
print(file.read())
```

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

Return-Path: <postmaster@collab.sakaiproject.org>

Received: from murder (mail.umich.edu [141.211.14.90])

by frankenstein.mail.umich.edu (Cyrus v2.3.8) with LMTPA;

Sat, 05 Jan 2008 09:14:16 -0500

X-Sieve: CMU Sieve 2.3

Received: from murder ([unix socket])

by mail.umich.edu (Cyrus v2.2.12) with LMTPA;

Sat, 05 Jan 2008 09:14:16 -0500

Received: from holes.mr.itd.umich.edu (holes.mr.itd.umich.edu [141.211.14.79])

by flawless.mail.umich.edu () with ESMTP id m05EEFR1013674;

Sat, 5 Jan 2008 09:14:15 -0500

Received: FROM paploo.uhi.ac.uk (app1.prod.collab.uhi.ac.uk [194.35.219.184])

BY holes.mr.itd.umich.edu ID 477F90B0.2DB2F.12494 ;

5 Jan 2008 09:14:10 -0500

Received: from paploo.uhi.ac.uk (localhost [127.0.0.1])

by paploo.uhi.ac.uk (Postfix) with ESMTP id 5F919BC2F2;

Sat 5 Jan 2008 14:10:05 +0000 (GMT)

unicode error

```
mode='r' : read the file
mode='w' : write the file
```

In []: In [16]:

```
mode='rb' : read bytes the file
mode='wb' : write bytes the file
```

- encoding

- some time data has unusual characters

- encoding='utf-8-sig'

```
path="C:\\Users\\omkar\\OneDrive\\Documents\\Data science\\Naresh IT\\Datafi
le=open(path,
mode='r',
encoding='utf-8-sig')
file.read()
```

file not found error

open()

```
Out[16]: 'From stephen.marquard@uct.ac.za Sat Jan 5 09:14:16 2008\nReturn-Path:
```

```
<postmaster@collab.sakaiproject.org>\nReceived: from murder (mail.umich.
edu [141.211.14.90])\n\t by frankenstein.mail.umich.edu (Cyrus v2.3.8) w
ith LMTPA;\n\t Sat, 05 Jan 2008 09:14:16 -0500\nX-Sieve: CMU Sieve 2.3\n
Received: from murder ([unix socket])\n\t by mail.umich.edu (Cyrus v2.2.
```

```
12) with LMTPA;\n\t Sat, 05 Jan 2008 09:14:16 -0500\nReceived: from hole
s.mr.itd.umich.edu (holes.mr.itd.umich.edu [141.211.14.79])\n\tby flawle
ss.mail.umich.edu () with ESMTP id m05EEFR1013674;\n\tSat, 5 Jan 2008 0
9:14:15 -0500\nReceived: FROM paploo.uhi.ac.uk (app1.prod.collab.uhi.ac.
uk [194.35.219.184])\n\tBY holes.mr.itd.umich.edu ID 477F90B0.2DB2F.1249
4 ; \n\t 5 Jan 2008 09:14:10 -0500\nReceived: from paploo.uhi.ac.uk (loc
alhost [127.0.0.1])\n\tby paploo.uhi.ac.uk (Postfix) with ESMTP id 5F919
BC2F2;\n\tSat, 5 Jan 2008 14:10:05 +0000 (GMT)\nMessage-ID: <2008010514
12.m05ECIaH010327@nakamura.uits.iupui.edu>\nMime-Version: 1.0\nContent-T
ransfer-Encoding: 7bit\nReceived: from prod.collab.uhi.ac.uk ([194.35.21
9.182])\n by paploo.uhi.ac.uk (JAMES SMTP Server 2.1.3) with SM TP ID
899\n for <source@collab.sakaiproject.org>;\n Sa t, 5 Jan 2008 14:09:50
+0000 (GMT)\nReceived: from nakamura.uits.iupui.e du
(nakamura.uits.iupui.edu [134.68.220.122])\n\tby shmi.uhi.ac.uk (Post
```

```
fi ) ith ESMTP id A215243002\ \tf < @ ll b k i j t > S
In [17]: file=open(path)
path="C:\\Users\\omkar\\OneDrive\\Document file.read()
s\\Data science\\Naresh IT\\Datafi
```

```
Out[17]: 'From stephen.marquard@uct.ac.za Sat Jan 5 09:14:16 2008\nReturn-Path:
```

```
<postmaster@collab.sakaiproject.org>\nReceived: from murder (mail.umich.
edu [141.211.14.90])\n\t by frankenstein.mail.umich.edu (Cyrus v2.3.8) w
ith LMTPA;\n\t Sat, 05 Jan 2008 09:14:16 -0500\nX-Sieve: CMU Sieve 2.3\n
Received: from murder ([unix socket])\n\t by mail.umich.edu (Cyrus v2.2.
12) with LMTPA;\n\t Sat, 05 Jan 2008 09:14:16 -0500\nReceived: from hole
s.mr.itd.umich.edu (holes.mr.itd.umich.edu [141.211.14.79])\n\tby flawle
ss.mail.umich.edu () with ESMTP id m05EEFR1013674;\n\tSat, 5 Jan 2008 0
9:14:15 -0500\nReceived: FROM paploo.uhi.ac.uk (app1.prod.collab.uhi.ac.
uk [194.35.219.184])\n\tBY holes.mr.itd.umich.edu ID 477F90B0.2DB2F.1249
4 ; \n\t 5 Jan 2008 09:14:10 -0500\nReceived: from paploo.uhi.ac.uk (loc
alhost [127.0.0.1])\n\tby paploo.uhi.ac.uk (Postfix) with ESMTP id 5F919
BC2F2;\n\tSat, 5 Jan 2008 14:10:05 +0000 (GMT)\nMessage-ID: <2008010514
12.m05ECIaH010327@nakamura.uits.iupui.edu>\nMime-Version: 1.0\nContent-T
ransfer-Encoding: 7bit\nReceived: from prod.collab.uhi.ac.uk ([194.35.21
9.182])\n by paploo.uhi.ac.uk (JAMES SMTP Server 2.1.3) with SM TP ID
899\n for <source@collab.sakaiproject.org>;\n Sa t, 5 Jan 2008 14:09:50
+0000 (GMT)\nReceived: from nakamura.uits.iupui.e du
(nakamura.uits.iupui.edu [134.68.220.122])\n\tby shmi.uhi.ac.uk (Post
```

```
fi ) ith ESMTP id A215243002\ \tf < @ ll b k i j t > S
In [19]: file=open(path,
path="C:\\Users\\omkar\\OneDrive\\Document mode='w')
s\\Data science\\Naresh IT\\Datafi file.write("hello")
```

```
Out[19]: 5
```

with keywordwith

```
In [ ]:
# Same by using
```

```
In [21]:
path="C:\\Users\\omkar\\OneDrive\\Document with open(path) as file:
s\\Data science\\Naresh IT\\Datafi print(file.read())
```

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

```
Return-Path: <postmaster@collab.sakaiproject.org>
Received: from murder (mail.umich.edu [141.211.14.90])
by frankenstein.mail.umich.edu (Cyrus v2.3.8) with LMTPA;
Sat, 05 Jan 2008 09:14:16 -0500
X-Sieve: CMU Sieve 2.3
Received: from murder ([unix socket])
```

```

        by mail.umich.edu (Cyrus v2.2.12) with LMTPA;
        Sat, 05 Jan 2008 09:14:16 -0500
Received: from holes.mr.itd.umich.edu (holes.mr.itd.umich.edu [141.211.1
4.79])
        by flawless.mail.umich.edu () with ESMTP id m05EEFR1013674;
        Sat, 5 Jan 2008 09:14:15 -0500
Received: FROM paploo.uhi.ac.uk (app1.prod.collab.uhi.ac.uk [194.35.219.
184])
        BY holes.mr.itd.umich.edu ID 477F90B0.2DB2F.12494 ;
        5 Jan 2008 09:14:10 -0500
Received: from paploo.uhi.ac.uk (localhost [127.0.0.1])
        by paploo.uhi.ac.uk (Postfix) with ESMTP id 5F919BC2F2;

S t 5 J 2008 14 10 05 +0000 (GMT)

```

```

In [22]:
path="C:\\Users\\omkar\\OneDrive\\Document
s\\Data science\\Naresh IT\\Datafi
file=open(path)
file.read() # session will close

```

```

path="C:\\Users\\omkar\\OneDrive\\Document

```

```

Out[22]: 'From stephen.marquard@uct.ac.za Sat Jan 5 09:14:16 2008\\nReturn-Path:

```

```

<postmaster@collab.sakaiproject.org>\\nReceived: from murder (mail.umich.
edu [141.211.14.90])\\n\\t by frankenstein.mail.umich.edu (Cyrus v2.3.8) w
ith LMTPA;\\n\\t Sat, 05 Jan 2008 09:14:16 -0500\\nX-Sieve: CMU Sieve 2.3\\n
Received: from murder ([unix socket])\\n\\t by mail.umich.edu (Cyrus v2.2.
12) with LMTPA;\\n\\t Sat, 05 Jan 2008 09:14:16 -0500\\nReceived: from hole
s.mr.itd.umich.edu (holes.mr.itd.umich.edu [141.211.14.79])\\n\\tby flawle
ss.mail.umich.edu () with ESMTP id m05EEFR1013674;\\n\\tSat, 5 Jan 2008 0
9:14:15 -0500\\nReceived: FROM paploo.uhi.ac.uk (app1.prod.collab.uhi.ac.
uk [194.35.219.184])\\n\\tBY holes.mr.itd.umich.edu ID 477F90B0.2DB2F.1249
4 ; \\n\\t 5 Jan 2008 09:14:10 -0500\\nReceived: from paploo.uhi.ac.uk (loc
alhost [127.0.0.1])\\n\\tby paploo.uhi.ac.uk (Postfix) with ESMTP id 5F919
BC2F2;\\n\\tSat, 5 Jan 2008 14:10:05 +0000 (GMT)\\nMessage-ID: <2008010514
12.m05ECIaH010327@nakamura.uits.iupui.edu>\\nMime-Version: 1.0\\nContent-T
ransfer-Encoding: 7bit\\nReceived: from prod.collab.uhi.ac.uk ([194.35.21
9.182])\\n by paploo.uhi.ac.uk (JAMES SMTP Server 2.1.3) with SM TP ID
899\\n for <source@collab.sakaiproject.org>;\\n Sa t, 5 Jan 2008 14:09:50
+0000 (GMT)\\nReceived: from nakamura.uits.iupui.e du
(nakamura.uits.iupui.edu [134.68.220.122])\\n\\tby shmi.uhi.ac.uk (Post

```

```

fi ) ith ESMTP id A215243002\\ \\tf < @ ll b k i j t > S

```

```

In [23]:
file # session

```

```

Out[23]: <_io.TextIOWrapper name='C:\\Users\\omkar\\OneDrive\\Documents\\Data scien
ce\\Naresh IT\\Datafiles\\mbox-short.txt' mode='r' encoding='cp1252'>
print(file.write("python"))

```

```

In [25]: In [ ]:

```

```

6

```

```

path="C:\\Users\\omkar\\OneDrive\\Document
s\\Data science\\Naresh IT\\Datafi
file=open(path)
file.read() # session will close

path="C:\\Users\\omkar\\OneDrive\\Document
s\\Data science\\Naresh IT\\Datafi
with open(path,mode='w') as file:

```

```
with open(path) as file:
    print(file.read())
```

Case - 3

```
In [27]:
# you written 500 lines of code
# that code use one text file as a input

path="C:\\Users\\omkar\\OneDrive\\Documents\\Data science\\Naresh IT\\Datafi
file=open(path,mode='w')
file.write('hello') # session will close

path="C:\\Users\\omkar\\OneDrive\\Documents\\Data science\\Naresh IT\\Datafi
file=open(path)
file.read()

with open(path,mode='w') as file:
    print(file.write('python'))

# and you shared this code to client
# and you mentioned just click run all
```

Out[27]: 'From stephen.marquard@uct.ac.za Sat Jan 5 09:14:16 2008\nReturn-Path:

<postmaster@collab.sakaiproject.org>\nReceived: from murder (mail.umich.edu [141.211.14.90])\n\t by frankenstein.mail.umich.edu (Cyrus v2.3.8) with LMTPA;\n\t Sat, 05 Jan 2008 09:14:16 -0500\nX-Sieve: CMU Sieve 2.3\nReceived: from murder ([unix socket])\n\t by mail.umich.edu (Cyrus v2.2.12) with LMTPA;\n\t Sat, 05 Jan 2008 09:14:16 -0500\nReceived: from holes.mr.itd.umich.edu (holes.mr.itd.umich.edu [141.211.14.79])\n\t by flawless.mail.umich.edu () with ESMTP id m05EEFR1013674;\n\t Sat, 5 Jan 2008 09:14:15 -0500\nReceived: FROM paploo.uhi.ac.uk (app1.prod.collab.uhi.ac.uk [194.35.219.184])\n\t BY holes.mr.itd.umich.edu ID 477F90B0.2DB2F.12494 ; \n\t 5 Jan 2008 09:14:10 -0500\nReceived: from paploo.uhi.ac.uk (localhost [127.0.0.1])\n\t by paploo.uhi.ac.uk (Postfix) with ESMTP id 5F919BC2F2;\n\t Sat, 5 Jan 2008 14:10:05 +0000 (GMT)\nMessage-ID: <200801051412.m05ECIaH010327@nakamura.uits.iupui.edu>\nMime-Version: 1.0\nContent-Transfer-Encoding: 7bit\nReceived: from prod.collab.uhi.ac.uk ([194.35.219.182])\n by paploo.uhi.ac.uk (JAMES SMTP Server 2.1.3) with SMTP ID 899\n for <source@collab.sakaiproject.org>;\n Sat, 5 Jan 2008 14:09:50 +0000 (GMT)\nReceived: from nakamura.uits.iupui.edu (nakamura.uits.iupui.edu [134.68.220.122])\n\t by shmi.uhi.ac.uk (Post

fi) ith ESMTP id A215243002\ \tf < @ ll b k i j t > S

```
In [26]:
# if your python file and data
file is in same location # no
need to provide entire path
# only filename.extension is
enough

# check where your python file
```

Out[26]: 'C:\\Users\\omkar\\OneDrive\\Documents\\Data science\\Naresh IT\\Data science\\Batch-6_Jan\\Python sessions'
file.read()

In []: In []:
Note: keep python file and

data file in sam location #

read the data

```
path="mbox-short.txt"
file=open(path)
In [1]: rt.txt"
path="mbox-sho file=open(path
```

```
) file.read()
```

```
Out[1]: 'From stephen.marquard@uct.ac.za Sat Jan 5 09:14:16 2008\nReturn-Path:
```

```
<postmaster@collab.sakaiproject.org>\nReceived: from murder (mail.umich.edu [141.211.14.90])\n\tby frankenstein.mail.umich.edu (Cyrus v2.3.8) with LMTPA;\n\tSat, 05 Jan 2008 09:14:16 -0500\nX-Sieve: CMU Sieve 2.3\nReceived: from murder ([unix socket])\n\tby mail.umich.edu (Cyrus v2.2.12) with LMTPA;\n\tSat, 05 Jan 2008 09:14:16 -0500\nReceived: from holes.mr.itd.umich.edu (holes.mr.itd.umich.edu [141.211.14.79])\n\tby flawless.mail.umich.edu () with ESMTP id m05EEFR1013674;\n\tSat, 5 Jan 2008 09:14:15 -0500\nReceived: FROM paploo.uhi.ac.uk (app1.prod.collab.uhi.ac.uk [194.35.219.184])\n\tBY holes.mr.itd.umich.edu ID 477F90B0.2DB2F.12494 ;\n\t5 Jan 2008 09:14:10 -0500\nReceived: from paploo.uhi.ac.uk (localhost [127.0.0.1])\n\tby paploo.uhi.ac.uk (Postfix) with ESMTP id 5F919BC2F2;\n\tSat, 5 Jan 2008 14:10:05 +0000 (GMT)\nMessage-ID: <200801051412.m05ECIaH010327@nakamura.uits.iupui.edu>\nMime-Version: 1.0\nContent-Transfer-Encoding: 7bit\nReceived: from prod.collab.uhi.ac.uk ([194.35.219.182])\n\tby paploo.uhi.ac.uk (JAMES SMTP Server 2.1.3) with SMTP ID 899\n\tfor <source@collab.sakaiproject.org>;\n\tSat, 5 Jan 2008 14:09:50 +0000 (GMT)\nReceived: from nakamura.uits.iupui.edu (nakamura.uits.iupui.edu [134.68.220.122])\n\tby shmi.uhi.ac.uk (Post
```

```
fix) with ESMTP id A215243002\n\tfor <@llbkijt> S
```

```
In [2]: In [ ]:
```

```
# Both python file and data file in same folder
path="mbox-short.txt"
file=open(path)
#file.read()
for i in file:
    print(i)
```

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

```
Return-Path: <postmaster@collab.sakaiproject.org>
```

```
Received: from murder (mail.umich.edu [141.211.14.90])
```

by frankenstein.mail.umich.edu (Cyrus v2.3.8) with LMTPA; Sat, 05 Jan 2008

09:14:16 -0500

X-Sieve: CMU Sieve 2.3

Received: from murder ([unix socket])

by mail.umich.edu (Cyrus v2.2.12) with LMTPA;

Sat, 05 Jan 2008 09:14:16 -0500

Received: from holes.mr.itd.umich.edu (holes.mr.itd.umich.edu [141.211.14.79])

#Q1) Find the number of lines in a given text
Idea : Counter program on above code

In [3]: In [4]:

```
path="mbox-short.txt"
file=open(path)
count=0
for i in file:
    count=count+1
print(count)
```

by frankenstein.mail.umich.edu (Cyrus v2.3.8) with LMTPA; Sat, 05 Jan 2008 09:14:16 -0500

X-Sieve: CMU Sieve 2.3

Received: from murder ([unix socket])

by mail.umich.edu (Cyrus v2.2.12) with LMTPA;

Sat, 05 Jan 2008 09:14:16 -0500

Received: from holes.mr.itd.umich.edu (holes.mr.itd.umich.edu [141.211.14.79])

Those who are getting 2125: Empty Lines
It is not harm

1910

#Q2) Print first 10 Lines

```
path="mbox-short.txt"
file=open(path)
count=0
for i in file:
    print(i)
    count=count+1
    if count==10:
        break
```

enumerate

enumerate will provide both index and line
In [10]:

From stephen.marquard@uct.ac.za Sat Jan 5

09:14:16 2008 Return-Path:

<postmaster@collab.sakaiproject.org>

Received: from murder (mail.umich.edu [141.211.14.90])

v2.3.8) with LMTPA; 4 Sat, 05 Jan 2008

09:14:16 -0500

5 X-Sieve: CMU Sieve 2.3

6 Received: from murder ([unix socket])

7 by mail.umich.edu (Cyrus v2.2.12) with LMTPA;

8 Sat, 05 Jan 2008 09:14:16 -0500

9 Received: from holes.mr.itd.umich.edu (holes.mr.itd.umich.edu [141.21

In []: In [9]:

```
path="mbox-short.txt"
file=open(path)
for index,line in enumerate(file):
    print(index,line)
```

0 From stephen.marquard@uct.ac.za Sat Jan

5 09:14:16 2008

1 Return-Path:

<postmaster@collab.sakaiproject.org>

1 14 79])

```
for i,j in zip(11,12)
```

2 Received: from murder (mail.umich.edu [141.211.14.90])

```
num=10,20
num
```

3 by frankenstein.mail.umich.edu (Cyrus

Out[9]: (10, 20)

In [18]:

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18

In [34]: In [38]:

```
# Q3)
# we got the number of lines by using
counter
# can you get the same answer using index

# _ means we are not using that output

path="mbox-short.txt"
file=open(path)
for index,_ in enumerate(file):
    print(index)
```

0

```

file=open(path)
for index,_ in enumerate(file):
    pass
    for i in range(25):
        pass

print("the number of lines are:",index+1) i
the number of lines are: 1910

```

Out[38]: 24

In [25]:
a

Out[25]: 4

```

In [39]:
path = "mbox-short.txt"
file = open(path)
for index,line in enumerate(file):
    print(index,line)
    if index==10:
        break

0 From stephen.marquard@uct.ac.za Sat Jan
5 09:14:16 2008 1 Return-Path:
<postmaster@collab.sakaiproject.org>
2 Received: from murder (mail.umich.edu
[141.211.14.90]) 3 by
frankenstein.mail.umich.edu (Cyrus
v2.3.8) with LMTPA; 4 Sat, 05 Jan 2008
09:14:16 -0500
5 X-Sieve: CMU Sieve 2.3
6 Received: from murder ([unix socket])
7 by mail.umich.edu (Cyrus v2.2.12) with
LMTPA;
8 Sat, 05 Jan 2008 09:14:16 -0500
9 Received: from holes.mr.itd.umich.edu
(holes.mr.itd.umich.edu [141.211.1 4.79])
10 by flawless.mail.umich.edu () with
ESMTP id m05EEFR1013674;

```

In [41]:

```

# Q5) Align all the lines
# From stephen.marquard@uct.ac.za Sat Jan
5 09:14:16 2008 # Return-Path:
<postmaster@collab.sakaiproject.org> #
Received: from murder (mail.umich.edu
[141.211.14.90]) # by
frankenstein.mail.umich.edu (Cyrus
v2.3.8) with LMTPA; # Sat, 05 Jan 2008
09:14:16 -0500
# Idea: strip the lines
path = "mbox-short.txt"
file = open(path)
for line in file:
    print(line.strip())

```

```

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

```

```

Return-Path: <postmaster@collab.sakaiproject.org>
Received: from murder (mail.umich.edu [141.211.14.90])
by frankenstein.mail.umich.edu (Cyrus v2.3.8) with LMTPA;
Sat, 05 Jan 2008 09:14:16 -0500
X-Sieve: CMU Sieve 2.3

```

```

Received: from murder ([unix socket])
by mail.umich.edu (Cyrus v2.2.12) with LMTPA;
Sat, 05 Jan 2008 09:14:16 -0500
Received: from holes.mr.itd.umich.edu (holes.mr.itd.umich.edu [141.211.1
4.79])
by flawless.mail.umich.edu () with ESMTP id m05EEFR1013674;
Sat, 5 Jan 2008 09:14:15 -0500
Received: FROM paploo.uhi.ac.uk (app1.prod.collab.uhi.ac.uk [194.35.219.
184])
BY holes.mr.itd.umich.edu ID 477F90B0.2DB2F.12494 ;
5 Jan 2008 09:14:10 -0500
Received: from paploo.uhi.ac.uk (localhost [127.0.0.1])
by paploo.uhi.ac.uk (Postfix) with ESMTP id 5F919BC2F2;

S t 5 J 2008 14 10 05 +0000 (GMT)

```

In [44]:

with From

```

# Idea= startswith keyword
path = "mbox-short.txt"
file = open(path)
for line in file:
    if line.startswith("From"):
        print(line)

```

```

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

```

Q6) I want the lines which are start

```

From: stephen.marquard@uct.ac.za

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

From: louis@media.berkeley.edu

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

From: zqian@umich.edu

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

From: rjlowe@iupui.edu

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

From: zqian@umich.edu

```

In [46]:

```

#[<o/p> <for loop> <if
condition>]
path = "mbox-short.txt"

```

```

file = open(path)
list1=[line for line in file if
line.startswith("From")] list1

```

```

Out[46]: ['From stephen.marquard@uct.ac.za Sat Jan 5 09:14:16 2008\n',
          'From: stephen.marquard@uct.ac.za\n',
          'From louis@media.berkeley.edu Fri Jan 4 18:10:48 2008\n',
          'From: louis@media.berkeley.edu\n',
          'From zqian@umich.edu Fri Jan 4 16:10:39 2008\n',
          'From: zqian@umich.edu\n',
          'From rjlowe@iupui.edu Fri Jan 4 15:46:24 2008\n',
          'From: rjlowe@iupui.edu\n',
          'From zqian@umich.edu Fri Jan 4 15:03:18 2008\n',
          'From: zqian@umich.edu\n',
          'From rjlowe@iupui.edu Fri Jan 4 14:50:18 2008\n',
          'From: rjlowe@iupui.edu\n',
          'From cwen@iupui.edu Fri Jan 4 11:37:30 2008\n',

```

```

'From: cwen@iupui.edu\n',
'From cwen@iupui.edu Fri Jan 4 11:35:08 2008\n',
'From: cwen@iupui.edu\n',
'From: gsilver@umich.edu Fri Jan 4 11:12:37 2008\n',
'From: gsilver@umich.edu\n',
'From: gsilver@umich.edu Fri Jan 4 11:11:52 2008\n',
'From: gsilver@umich.edu\n',
'From: zqian@umich.edu Fri Jan 4 11:11:03 2008\n',
'From: zqian@umich.edu\n',
'From: gsilver@umich.edu Fri Jan 4 11:10:22 2008\n',
'From: gsilver@umich.edu\n',
'From: wagnermr@iupui.edu Fri Jan 4 10:38:42 2008\n',
'From: wagnermr@iupui.edu\n',
'From: zqian@umich.edu Fri Jan 4 10:17:43 2008\n',
'From: zqian@umich.edu\n',
'From: antranig@caret.cam.ac.uk Fri Jan 4 10:04:14 2008\n',
'From: antranig@caret.cam.ac.uk\n',
'From: gopal.ramasammycook@gmail.com Fri Jan 4 09:05:31 2008\n',
'From: gopal.ramasammycook@gmail.com\n',
'From: david.horwitz@uct.ac.za Fri Jan 4 07:02:32 2008\n',
'From: david.horwitz@uct.ac.za\n',
'From: david.horwitz@uct.ac.za Fri Jan 4 06:08:27 2008\n',
'From: david.horwitz@uct.ac.za\n',
'From: david.horwitz@uct.ac.za Fri Jan 4 04:49:08 2008\n',
'From: david.horwitz@uct.ac.za\n',
'From: david.horwitz@uct.ac.za Fri Jan 4 04:33:44 2008\n',
'From: david.horwitz@uct.ac.za\n',
'From: stephen.marquard@uct.ac.za Fri Jan 4 04:07:34 2008\n',
'From: stephen.marquard@uct.ac.za\n',
'From: louis@media.berkeley.edu Thu Jan 3 19:51:21 2008\n',
'From: louis@media.berkeley.edu\n',
'From: louis@media.berkeley.edu Thu Jan 3 17:18:23 2008\n',
'From: louis@media.berkeley.edu\n',
'From: ray@media.berkeley.edu Thu Jan 3 17:07:00 2008\n',
'From: ray@media.berkeley.edu\n',
'From: cwen@iupui.edu Thu Jan 3 16:34:40 2008\n',
'From: cwen@iupui.edu\n',
'From: cwen@iupui.edu Thu Jan 3 16:29:07 2008\n',
'From: cwen@iupui.edu\n',
'From: cwen@iupui.edu Thu Jan 3 16:23:48 2008\n',
'From: cwen@iupui.edu\n']

```

In [51]: *#Q7) remove \n from above output list1*

```

list2=[]
for line in list1:
    list2.append(line.rstrip('\n'))

```

```

list2=[line.rstrip('\n') for line in list1]
In [61]: line.startswith("From")]
path = "mbox-short.txt" list2=[line.rstrip('\n') for
file = open(path) line in list1] list2
list1=[line for line in file if

```

```

Out[61]: ['From: stephen.marquard@uct.ac.za Sat Jan 5 09:14:16 2008',
'From: stephen.marquard@uct.ac.za',
'From: louis@media.berkeley.edu Fri Jan 4 18:10:48 2008',
'From: louis@media.berkeley.edu',
'From: zqian@umich.edu Fri Jan 4 16:10:39 2008',
'From: zqian@umich.edu',
'From: rjlowe@iupui.edu Fri Jan 4 15:46:24 2008',
'From: rjlowe@iupui.edu',
'From: zqian@umich.edu Fri Jan 4 15:03:18 2008',
'From: zqian@umich.edu',

```

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

```
In [71]: rjlowe@iupui.edu
#Q8) Get the email ids from rjlowe@iupui.edu
list2 cwen@iupui.edu
for line in list2: cwen@iupui.edu
    print(line.split()[1]) cwen@iupui.edu
                                cwen@iupui.edu
emails_list=[line.split()[1]gsilver@umich.edu
for line in list2] gsilver@umich.edu
                                gsilver@umich.edu
stephen.marquard@uct.ac.za gsilver@umich.edu
stephen.marquard@uct.ac.za zqian@umich.edu
louis@media.berkeley.edu zqian@umich.edu
louis@media.berkeley.edu gsilver@umich.edu
zqian@umich.edu gsilver@umich.edu
zqian@umich.edu wagnermr@iupui.edu
rjlowe@iupui.edu wagnermr@iupui.edu
rjlowe@iupui.edu zqian@umich.edu
zqian@umich.edu zqian@umich.edu
zqian@umich.edu antranig@caret.cam.ac.uk
```

```

antranig@caret.cam.ac.uk      stephen.marquard@uct.ac.za
gopal.ramasammycook@gmail.colouis@media.berkeley.edu
m                               louis@media.berkeley.edu
gopal.ramasammycook@gmail.colouis@media.berkeley.edu
m                               louis@media.berkeley.edu
david.horwitz@uct.ac.za       ray@media.berkeley.edu
david.horwitz@uct.ac.za       ray@media.berkeley.edu
david.horwitz@uct.ac.za       cwen@iupui.edu
david.horwitz@uct.ac.za       cwen@iupui.edu
david.horwitz@uct.ac.za       cwen@iupui.edu
david.horwitz@uct.ac.za       cwen@iupui.edu
david.horwitz@uct.ac.za       cwen@iupui.edu
david.horwitz@uct.ac.za       cwen@iupui.edu
david.horwitz@uct.ac.za       cwen@iupui.edu
stephen.marquard@uct.ac.za

```

```

In [ ]: In [74]: emails_list=[line.split()[1] for line
in list2]

```

```

#Q9) Unique email id
unique_emails=[]
for i in emails_list: # we are
iterating each line if i not in
unique_emails: # we are checking the
line in l1 unique_emails.append(i) #
if it is not there push into l1
list2=[line.rstrip('\n') for line in len(unique_emails)
list1]

```

```

Out[74]: 11

```

```

In [ ]:

```

```

In [ ]: In [ ]: In [ ]:

```

```

In [ ]: In [ ]: In [ ]:

```

```

#Q1) count the number of
lines
#Q2) print first 10 lines
#Q3) Count the number of
lines using index #Q4)
print first 10 lines
using index #Q5)
Alignment of all the
lines #Q6) Start with
From
#Q7) Remove /n
#Q8) Extract email id
#Q9) Unique email id

```