

## Connecting the Colab notebook to Google Drive

```
from google.colab import drive
drive.mount('/content/gdrive')

Mounted at /content/gdrive
```

```
!ls
```

```
drive  gdrive  mail_20110402.tgz  sample_data
```

As we can see through the output of "LS" command, our Mail dataset is present in the directory.

### Unzipping the dataset:

```
!tar -xvzf mail_20110402.tgz
```



```
enron_mail_20110402/maildir/neal-s/discussion_threads/172.
enron_mail_20110402/maildir/neal-s/discussion_threads/76.
enron_mail_20110402/maildir/neal-s/discussion_threads/401.
enron_mail_20110402/maildir/neal-s/discussion_threads/256.
enron_mail_20110402/maildir/neal-s/discussion_threads/188.
enron_mail_20110402/maildir/neal-s/discussion_threads/208.
enron_mail_20110402/maildir/neal-s/discussion_threads/83.
enron_mail_20110402/maildir/neal-s/discussion_threads/348.
enron_mail_20110402/maildir/neal-s/discussion_threads/310.
enron_mail_20110402/maildir/neal-s/discussion_threads/337.
enron_mail_20110402/maildir/neal-s/discussion_threads/194.
enron_mail_20110402/maildir/neal-s/discussion_threads/91.
enron_mail_20110402/maildir/neal-s/discussion_threads/207.
enron_mail_20110402/maildir/neal-s/discussion_threads/201.
enron_mail_20110402/maildir/neal-s/discussion_threads/54.
enron_mail_20110402/maildir/neal-s/discussion_threads/335.
enron_mail_20110402/maildir/neal-s/discussion_threads/69.
enron_mail_20110402/maildir/neal-s/discussion_threads/354.
enron_mail_20110402/maildir/neal-s/discussion_threads/302.
enron_mail_20110402/maildir/neal-s/discussion_threads/393.
enron_mail_20110402/maildir/neal-s/discussion_threads/259.
enron_mail_20110402/maildir/neal-s/discussion_threads/392.
enron_mail_20110402/maildir/neal-s/discussion_threads/195.
enron_mail_20110402/maildir/neal-s/discussion_threads/149.
enron_mail_20110402/maildir/neal-s/discussion_threads/285.
enron_mail_20110402/maildir/neal-s/discussion_threads/329.
enron_mail_20110402/maildir/neal-s/discussion_threads/413.
enron_mail_20110402/maildir/neal-s/discussion_threads/254.
enron_mail_20110402/maildir/neal-s/discussion_threads/443.
enron_mail_20110402/maildir/neal-s/discussion_threads/107.
enron_mail_20110402/maildir/neal-s/discussion_threads/231.
enron_mail_20110402/maildir/neal-s/discussion_threads/89.
enron_mail_20110402/maildir/neal-s/discussion_threads/190.
enron mail 20110402/maildir/neal-s/discussion threads/427.
```

```

.....
enron_mail_20110402/maildir/neal-s/discussion_threads/189.
enron_mail_20110402/maildir/neal-s/discussion_threads/10.
enron_mail_20110402/maildir/neal-s/discussion_threads/287.
enron_mail_20110402/maildir/neal-s/discussion_threads/51.
enron_mail_20110402/maildir/neal-s/discussion_threads/341.
enron_mail_20110402/maildir/neal-s/discussion_threads/1.
enron_mail_20110402/maildir/neal-s/discussion_threads/19.
enron_mail_20110402/maildir/neal-s/discussion_threads/150.
enron_mail_20110402/maildir/neal-s/discussion_threads/384.
enron_mail_20110402/maildir/neal-s/discussion_threads/18.
enron_mail_20110402/maildir/neal-s/discussion_threads/169.

enron_mail_20110402/maildir/neal-s/discussion_threads/45.
enron_mail_20110402/maildir/neal-s/discussion_threads/291.
enron_mail_20110402/maildir/neal-s/discussion_threads/349.
enron_mail_20110402/maildir/neal-s/discussion_threads/328.
enron_mail_20110402/maildir/neal-s/discussion_threads/363.
enron_mail_20110402/maildir/neal-s/discussion_threads/357.
enron_mail_20110402/maildir/neal-s/discussion_threads/55.
enron_mail_20110402/maildir/neal-s/discussion_threads/358.
enron_mail_20110402/maildir/neal-s/discussion_threads/151.
enron_mail_20110402/maildir/neal-s/discussion_threads/319.
enron_mail_20110402/maildir/neal-s/discussion_threads/368.
enron_mail_20110402/maildir/neal-s/discussion_threads/433.
enron_mail_20110402/maildir/neal-s/discussion_threads/388.

```

Example of a user's Email (some users have more folders):

```
!ls enron_mail_20110402/maildir/lay-k
```

```

all_documents  compaq          elizabeth  inbox        sent
business       deleted_items   enron      notes_inbox  _sent
calendar       discussion_threads  family     sec_panel    sent_items

```

Using the GLOB package to read all files from the folders recursively. Using the EMAIL package to parse the emails.

```
import email
import glob
```

```
files = glob.glob('enron_mail_20110402/maildir/**/*/*.*')
```

```
out = open("emails.txt", "a")
```

Writing the output in emails.txt

```
for file in files:
    with open(file) as f:
```

```
    try:
        lines = f.readlines()
        lines = ''.join(lines)
        e = email.message_from_string(lines)
        out.write(e.get_payload())
    except:
        pass
```

```
out.close()
```

```
inb = open("emails_inbox.txt", "a")
```

Only parsing through "inbox" folder for all users and writing the output in "emails\_inbox.txt":

```
files = glob.glob('enron_mail_20110402/maildir/**/inbox/*.')
```

```
for file in files:
    with open(file) as f:
        try:
            lines = f.readlines()
            lines = ''.join(lines)
            e = email.message_from_string(lines)
            inb.write(e.get_payload())
        except:
            pass
```

```
inb.close()
```

[Colab paid products](#) - [Cancel contracts here](#)

