

5)extracting all emails from txt file

```
In [28]: path="mbox-short.txt"
emails=[]
dup=[]
unique=[]
with open(path,'r') as file:
    for line in file:
        words=line.split()
        for word in words:
            if '@' in word:
                i1=word.index('@')
                if '.' in word[i1:]:
                    email=word.strip('<>;')
                    emails.append(email)
print(emails)
print()
print(f'Number of total emails:{len(emails)})')
```

['stephen.marquard@uct.ac.za', 'postmaster@collab.sakaiproject.org', '200801051412.m05ECIAH010327@nakamura.uits.iupui.edu', 'source@collab.sakaiproject.org', 'source@collab.sakaiproject.org', 'stephen.marquard@uct.ac.za', 'source@collab.sakaiproject.org', 'stephen.marquard@uct.ac.za', 'source@collab.sakaiproject.org', 'louis@media.berkeley.edu', 'postmaster@collab.sakaiproject.org', '200801042308.m04N8v60008125@nakamura.uits.iupui.edu', 'source@collab.sakaiproject.org', 'source@collab.sakaiproject.org', 'source@collab.sakaiproject.org', 'louis@media.berkeley.edu', 'louis@media.berkeley.edu', 'louis@media.berkeley.edu', 'zqian@umich.edu', 'postmaster@collab.sakaiproject.org', '200801042109.m04L92hb007923@nakamura.uits.iupui.edu', 'source@collab.sakaiproject.org', 'source@collab.sakaiproject.org', 'zqian@umich.edu', 'source@collab.sakaiproject.org', 'zqian@umich.edu', 'zqian@umich.edu', 'rjlowe@iupui.edu', 'postmaster@collab.sakaiproject.org', '200801042044.m04Kiem3007881@nakamura.uits.iupui.edu', 'source@collab.sakaiproject.org', 'source@collab.sakaiproject.org', 'rjlowe@iupui.edu', 'source@collab.sakaiproject.org', 'rjlowe@iupui.edu', 'rjlowe@iupui.edu', 'zqian@umich.edu', 'postmaster@collab.sakaiproject.org', '200801042001.m04K1c00007738@nakamura.uits.iupui.edu', 'source@collab.sakaiproject.org', 'source@collab.sakaiproject.org', 'zqian@umich.edu', 'source@collab.sakaiproject.org', 'zqian@umich.edu', 'zqian@umich.edu', 'rjlowe@iupui.edu', 'postmaster@collab.sakaiproject.org', '200801041948.m04Jmdw0007705@nakamura.uits.iupui.edu', 'source@collab.sakaiproject.org', 'source@collab.sakaiproject.org', 'source@collab.sakaiproject.org', 'rjlowe@iupui.edu', 'source@collab.sakaiproject.org', 'rjlowe@iupui.edu', 'cwen@iupui.edu', 'postmaster@collab.sakaiproject.org', '200801041635.m04GZQGZ007313@nakamura.uits.iupui.edu', 'source@collab.sakaiproject.org', 'source@collab.sakaiproject.org', 'cwen@iupui.edu', 'source@collab.sakaiproject.org', 'cwen@iupui.edu', 'cwen@iupui.edu', 'hu2@iupui.edu', 'cwen@iupui.edu', 'postmaster@collab.sakaiproject.org', '200801041633.m04GX6eG007292@nakamura.uits.iupui.edu', 'source@collab.sakaiproject.org', 'source@collab.sakaiproject.org', 'source@collab.sakaiproject.org', 'cwen@iupui.edu', 'cwen@iupui.edu', 'hu2@iupui.edu', 'gsilver@umich.edu', 'postmaster@collab.sakaiproject.org', '200801041611.m04GB1Lb007221@nakamura.uits.iupui.edu', 'source@collab.sakaiproject.org', 'source@collab.sakaiproject.org', 'source@collab.sakaiproject.org', 'gsilver@umich.edu', 'source@collab.sakaiproject.org', 'gsilver@umich.edu', 'gsilver@umich.edu', 'gsilver@umich.edu', 'postmaster@collab.sakaiproject.org', '200801041610.m04GA5KP007209@nakamura.uits.iupui.edu', 'source@collab.sakaiproject.org', 'source@collab.sakaiproject.org', 'source@collab.sakaiproject.org', 'gsilver@umich.edu', 'source@collab.sakaiproject.org', 'gsilver@umich.edu', 'gsilver@umich.edu', 'zqian@umich.edu', 'postmaster@collab.sakaiproject.org', '200801041609.m04G9EuX007197@nakamura.uits.iupui.edu', 'source@collab.sakaiproject.org', 'source@collab.sakaiproject.org', 'source@collab.sakaiproject.org', 'zqian@umich.edu', 'source@collab.sakaiproject.org', 'zqian@umich.edu', 'zqian@umich.edu', 'gsilver@umich.edu', 'postmaster@collab.sakaiproject.org', '200801041608.m04G8d7w007184@nakamura.uits.iupui.edu', 'source@collab.sakaiproject.org', 'source@collab.sakaiproject.org', 'source@collab.sakaiproject.org', 'gsilver@umich.edu', 'source@collab.sakaiproject.org', 'source@collab.sakaiproject.org', 'wagnermr@iupui.edu', 'postmaster@collab.sakaiproject.org', '200801041537.m04Fb6Ci007092@nakamura.uits.iupui.edu', 'source@collab.sakaiproject.org', 'source@collab.sakaiproject.org', 'source@collab.sakaiproject.org', 'source@collab.sakaiproject.org', 'source@collab.sakaiproject.org', 'source@collab.sakaiproject.org', 'wagnermr@iupui.edu', 'source@collab.sakaiproject.org', 'wagnermr@iupui.edu', 'zqian@umich.edu', 'postmaster@collab.sakaiproject.org', '200801041515.m04FFv42007050@nakamura.uits.iupui.edu', 'source@collab.sakaiproject.org', 'source@collab.sakaiproject.org', 'source@collab.sakaiproject.org']


```
en@iupui.edu', 'wagnermr@iupui.edu', 'cwen@iupui.edu', 'postmaster@collab.sakaiproject.org', '200801032122.m03LMFo4005148@nakamura.uits.iupui.edu', 'source@collab.sakai project.org', 'source@collab.sakaiproject.org', 'source@collab.sakaiproject.org', 'source@collab.sakaiproject.org', 'cwen@iupui.edu', 'source@collab.sakaiproject.org', 'cwen@iupui.edu', 'cwen@iupui.edu', 'wagnermr@iupui.edu']
```

Number of total emails:305

6)unique emails

```
In [39]: path="mbox-short.txt"
emails=[]
dup=[]
unique=[]
with open(path,'r',encoding='utf-8') as file:
    for line in file:
        words=line.split()
        for word in words:
            if '@' in word:
                i1=word.index('@')
                if '.' in word[i1:]:
                    email=word.strip('<>;')
                    emails.append(email)
                    for i in emails:
                        if i not in unique:
                            unique.append(i)
                        elif i not in dup:
                            dup.append(i)
print(unique)
print()
print(f'Number of unique emails:{len(unique)}')
```

```
['stephen.marquard@uct.ac.za', 'postmaster@collab.sakaiproject.org', '200801051412.m05ECIAH010327@nakamura.uits.iupui.edu', 'source@collab.sakaiproject.org', 'louis@media.berkeley.edu', '200801042308.m04N8v60008125@nakamura.uits.iupui.edu', 'zqian@umich.edu', '200801042109.m04L92hb007923@nakamura.uits.iupui.edu', 'rjlowe@iupui.edu', '200801042044.m04Kiem3007881@nakamura.uits.iupui.edu', '200801042001.m04K1c00007738@nakamura.uits.iupui.edu', '200801041948.m04Jmdw0007705@nakamura.uits.iupui.edu', 'cwen@iupui.edu', '200801041635.m04GZQGZ007313@nakamura.uits.iupui.edu', 'hu2@iupui.edu', '200801041633.m04GX6eG007292@nakamura.uits.iupui.edu', 'gsilver@umich.edu', '200801041611.m04GB1Lb007221@nakamura.uits.iupui.edu', '200801041610.m04GA5KP007209@nakamura.uits.iupui.edu', '200801041609.m04G9Eux007197@nakamura.uits.iupui.edu', '200801041608.m04G8d7w007184@nakamura.uits.iupui.edu', 'wagnermr@iupui.edu', '200801041537.m04Fb6Ci007092@nakamura.uits.iupui.edu', '200801041515.m04FFv42007050@nakamura.uits.iupui.edu', 'antranig@caret.cam.ac.uk', '200801041502.m04F21J0007031@nakamura.uits.iupui.edu', 'gopal.ramasammycook@gmail.com', '200801041403.m04E3psW006926@nakamura.uits.iupui.edu', 'david.horwitz@uct.ac.za', '200801041200.m04C0gfK006793@nakamura.uits.iupui.edu', '200801041106.m04B61K3006677@nakamura.uits.iupui.edu', '200801040947.m049lUxo006517@nakamura.uits.iupui.edu', 'josrodrri@iupui.edu', '200801040932.m049W2i5006493@nakamura.uits.iupui.edu', '200801040905.m0495rWB006420@nakamura.uits.iupui.edu', '200801040023.m040NpCc005473@nakamura.uits.iupui.edu', '200801032216.m03MGhDa005292@nakamura.uits.iupui.edu', 'ray@media.berkeley.edu', '200801032205.m03M5Ea7005273@nakamura.uits.iupui.edu', '200801032133.m03LX3gG005191@nakamura.uits.iupui.edu', '200801032127.m03LRUqH005177@nakamura.uits.iupui.edu', '200801032122.m03LMFo4005148@nakamura.uits.iupui.edu']
```

Number of unique emails:42

Reading emails.txt file

```
In [31]: path="emails.txt"
with open(path,mode='r') as file:
    data=file.read()
print(data)
```

Santhanam, Jayendhran <jayendhran.ms@accenture.com>; Bhattacharjee, Sankalan <s.t.bhattacharjee@accenture.com>; Bansal, Vaishnavi <vaishnavi.a.bansal@accenture.com>; Chandra Shekhar, Pavuluri <p.a.chandra.shekhar@accenture.com>; J Doshi, Hiren <hiren.j.doshi@accenture.com>; Shimoga Prakash, Jeevan <j.shimoga.prakash@accenture.com>; Akhouri, Shreyash <shreyash.akhouri@accenture.com>; Nath, Nupanga <nupanga.nath@accenture.com>; Ponna, Ramesh <r.ponna@accenture.com>; Singhal, Aditya <aditya.d.singhal@accenture.com>; Gorai, Sumit <sumit.gorai@accenture.com>; Valdez, Mariele Dyan Francely <mariele.dyan.valdez@accenture.com>; Cancel, Abigail Kyle <abigail.kyle.cancel@accenture.com>; BONDALAKUNTA, POOJITHA <p.bondalakunta@accenture.com>; priyanka uppala, sai <sai.priyanka.uppala@accenture.com>; Nitin SHINDE, NIKHIL <nikhil.nitin.shinde@accenture.com>; Chidre, Swapna <swapna.chidre@accenture.com>; Sahu, Swastik <s.satyapragyan.sahu@accenture.com>; Sikka, Ankit <ankit.sikka@accenture.com>; Dara, Sai M adhukar <sai.madhukar.dara@accenture.com>; Pattanshetti, Sampreet <s.pattanshetti@accenture.com>; Sharma, Ayush <ayush.ad.sharma@accenture.com>; Latorena, Zairous Jay <zairous.jay.latorena@accenture.com>; Abhimanyu Pradhan, Roshankumar <r.abhimanyu.prahan@accenture.com>; Jaiswal, Lehar <lehar.jaiswal@accenture.com>; Gupta, Retik <retik.gupta@accenture.com>; Satti, Phanindra Reddy <phanindra.r.satti@accenture.com>; Kumar, Manish <manish.dw.kumar@accenture.com>; Balakrishnan, Devi Priya <devi.p.balakrishnan@accenture.com>; M S, Roshni Priya <roshni.priya.m.s@accenture.com>; Cheerala, Vanaja <vanaja.cheerala@accenture.com>; G D, Chandrashekhar <chandrashekhar.g.d@accenture.com>; Nikhil Reddy, Madithati <m.nikhil.reddy@accenture.com>; Santos, Jeroshelle V. <jeroshelle.v.santos@accenture.com>; Milan, Gerald John <gerald.john.milan@accenture.com>; Gupta, Vivek <vivek.u.gupta@accenture.com>; Mandal, Sumit <s.sunilkumar.mandal@accenture.com>; Prakash Shukla, Ashok <ashok.prakash.shukla@accenture.com>; Naganathan, Siva <sivakumar.naganathan@accenture.com>; GAHLOT, NIKHIL <nikhil.gahlot@accenture.com>; Kanjilal, Sourav <s.a.kanti.kanjilal@accenture.com>; Nallagoni, Omkar <omkar.nallagoni@accenture.com>; Sahajan, Sandra <sandra.sahajan@accenture.com>; Nelluri, Chaitanya <chaitanya.nelluri@accenture.com>

extracting emails

In [165...]

```
path="emails.txt"
emailss=[]
with open(path,'r') as file:
    for line in file:
        words=line.split()
        for word in words:
            if '@' in word:
                i1=word.index('@')
                if '.' in word[i1+1:]:
                    email=word.strip('<>;')
                    emailss.append(email)
print(emailss)
```

```
['jayendhran.ms@accenture.com', 's.t.bhattacharjee@accenture.com', 'vaishnavi.a.bansal@accenture.com', 'p.a.chandra.shekhar@accenture.com', 'hiren.j.doshi@accenture.com', 'j.shimoga.prakash@accenture.com', 'shreyash.akhouri@accenture.com', 'nupanga.nathan@accenture.com', 'r.ponna@accenture.com', 'aditya.d.singhal@accenture.com', 'sumit.gorai@accenture.com', 'mariele.dyan.valdez@accenture.com', 'abigail.kyle.cancel@accenture.com', 'p.bondalakunta@accenture.com', 'sai.priyanka.uppal@accenture.com', 'nikhil.nitin.shinde@accenture.com', 'swapna.chidre@accenture.com', 's.satyapragyan.sahu@accenture.com', 'ankit.sikka@accenture.com', 'sai.madhukar.dara@accenture.com', 's.pattanshetti@accenture.com', 'ayush.ad.sharma@accenture.com', 'zairous.jay.latorena@accenture.com', 'r.abhimanyu.pradhan@accenture.com', 'lehar.jaiswal@accenture.com', 'retik.gupta@accenture.com', 'phanindra.r.satti@accenture.com', 'manish.dw.kumar@accenture.com', 'devi.p.balakrishnan@accenture.com', 'roshni.priya.m.s@accenture.com', 'vanaja.cheeralu@accenture.com', 'chandrashekhar.g.d@accenture.com', 'm.nikhil.reddy@accenture.com', 'jeroshelle.v.santos@accenture.com', 'gerald.john.milan@accenture.com', 'vivek.u.gupta@accenture.com', 's.sunilkumar.mandal@accenture.com', 'ashok.prakash.shukla@accenture.com', 'sivakumar.naganathan@accenture.com', 'nikhil.gahlot@accenture.com', 's.a.kanti.kanjilal@accenture.com', 'omkar.nallagoni@accenture.com', 'sandra.sahajan@accenture.com', 'chaitanya.nelluri@accenture.com']
```

extracting firstname,secondname and company name

In [170...]

```
fnamee = []
snamee = []
cnamee = []
for email in emailss:
    # Find the positions of '.' and '@'
    i1 = email.find('.')
    i2 = email.find('@')
    fname = email[:i1]
    fnamee.append(fname)
    sname = email[i1+1:i2]
    snamee.append(sname)
    cname = email[i2+1:email.index('.', i2)]
    cnamee.append(cname)
print("First Names :", fnamee)
print()
print("Second Names :", snamee)
print()
print("Company Names :", cnamee)
```

First Names : ['jayendhran', 's', 'vaishnavi', 'p', 'hiren', 'j', 'shreyash', 'nupanga', 'r', 'aditya', 'sumit', 'mariele', 'abigail', 'p', 'sai', 'nikhil', 'swapna', 's', 'ankit', 'sai', 's', 'ayush', 'zairous', 'r', 'lehar', 'retik', 'phanindra', 'm anish', 'devi', 'roshni', 'vanaja', 'chandrashekhar', 'm', 'jeroshelle', 'gerald', 'vivek', 's', 'ashok', 'sivakumar', 'nikhil', 's', 'omkar', 'sandra', 'chaitanya']

Second Names : ['ms', 't.bhattacharjee', 'a.bansal', 'a.chandra.shekhar', 'j.doshi', 'shimoga.prakash', 'akhouri', 'nath', 'ponna', 'd.singhal', 'gorai', 'dyan.valdez', 'kyle.cancel', 'bondalakunta', 'priyanka.uppala', 'nitin.shinde', 'chidre', 'satyapr agyan.sahu', 'sikka', 'madhukar.dara', 'pattanshetti', 'ad.sharma', 'jay.latorena', 'abhimanyu.pradhan', 'jaiswal', 'gupta', 'r.satti', 'dw.kumar', 'p.balakrishnan', 'priya.m.s', 'cheerala', 'g.d', 'nikhil.reddy', 'v.santos', 'john.milan', 'u.gupta', 'sunilkumar.mandal', 'prakash.shukla', 'naganathan', 'gahlot', 'a.kanti.kanjilal', 'nallagoni', 'sahajan', 'nelluri']