Last login: Thu Jun 22 10:53:46 on ttys000

dhcp-892b9f9f:~ LiangXS$ ssh student@137.43.49.13

student@137.43.49.13's password:

Welcome to Ubuntu 16.04.2 LTS (GNU/Linux 4.4.0-81-generic x86\_64)

\* Documentation: https://help.ubuntu.com

\* Management: https://landscape.canonical.com

\* Support: https://ubuntu.com/advantage

60 packages can be updated.

0 updates are security updates.

Last login: Thu Jun 22 09:49:07 2017 from 83.136.43.209

student@csi6220-1-vm2:~$ ls

correct setuptools-33.1.1 setuptools-33.1.1.zip

student@csi6220-1-vm2:~$ cd correct/

student@csi6220-1-vm2:~/correct$ ls

sql virtualenv

student@csi6220-1-vm2:~/correct$ cd virtualenv/

student@csi6220-1-vm2:~/correct/virtualenv$ ls

correctbackend correctpro

student@csi6220-1-vm2:~/correct/virtualenv$ python3 -m venv backendvirtual

student@csi6220-1-vm2:~/correct/virtualenv$ source backendvirtual/bin/activate

(backendvirtual)

student@csi6220-1-vm2:~/correct/virtualenv$ pip install --upgrade pipCollecting pip

Using cached pip-9.0.1-py2.py3-none-any.whl

Installing collected packages: pip

Found existing installation: pip 8.1.1

Uninstalling pip-8.1.1:

Successfully uninstalled pip-8.1.1

Successfully installed pip-9.0.1

(backendvirtual) student@csi6220-1-vm2:~/correct/virtualenv$ pip install django psycopg2 beautifulsoup4 celery django-celery

Collecting django

Using cached Django-1.11.2-py2.py3-none-any.whl

Collecting psycopg2

Using cached psycopg2-2.7.1-cp35-cp35m-manylinux1\_x86\_64.whl

Collecting beautifulsoup4

Downloading beautifulsoup4-4.6.0-py3-none-any.whl (86kB)

100% |████████████████████████████████| 92kB 1.7MB/s

Collecting celery

Downloading celery-4.0.2-py2.py3-none-any.whl (396kB)

100% |████████████████████████████████| 399kB 1.5MB/s

Collecting django-celery

Downloading django-celery-3.2.1.tar.gz (91kB)

100% |████████████████████████████████| 92kB 3.9MB/s

Collecting pytz (from django)

Using cached pytz-2017.2-py2.py3-none-any.whl

Collecting billiard<3.6.0,>=3.5.0.2 (from celery)

Downloading billiard-3.5.0.2-py3-none-any.whl (102kB)

100% |████████████████████████████████| 102kB 4.4MB/s

Collecting kombu<5.0,>=4.0.2 (from celery)

Downloading kombu-4.0.2-py2.py3-none-any.whl (178kB)

100% |████████████████████████████████| 184kB 2.8MB/s

Collecting amqp<3.0,>=2.1.4 (from kombu<5.0,>=4.0.2->celery)

Downloading amqp-2.1.4-py2.py3-none-any.whl (49kB)

100% |████████████████████████████████| 51kB 4.1MB/s

Collecting vine>=1.1.3 (from amqp<3.0,>=2.1.4->kombu<5.0,>=4.0.2->celery)

Downloading vine-1.1.3-py2.py3-none-any.whl

Installing collected packages: pytz, django, psycopg2, beautifulsoup4, billiard, vine, amqp, kombu, celery, django-celery

Running setup.py install for django-celery ... done

Successfully installed amqp-2.1.4 beautifulsoup4-4.6.0 billiard-3.5.0.2 celery-4.0.2 django-1.11.2 django-celery-3.2.1 kombu-4.0.2 psycopg2-2.7.1 pytz-2017.2 vine-1.1.3

(backendvirtual) student@csi6220-1-vm2:~/correct/virtualenv$ ls

backendvirtual correctbackend correctpro

(backendvirtual) student@csi6220-1-vm2:~/correct/virtualenv$ cd backendvirtual/

(backendvirtual) student@csi6220-1-vm2:~/correct/virtualenv/backendvirtual$ ls

bin include lib lib64 pip-selfcheck.json pyvenv.cfg share

(backendvirtual) student@csi6220-1-vm2:~/correct/virtualenv/backendvirtual$ mkdir Scraping

(backendvirtual) student@csi6220-1-vm2:~/correct/virtualenv/backendvirtual$ ls

bin include lib lib64 pip-selfcheck.json pyvenv.cfg Scraping share

(backendvirtual) student@csi6220-1-vm2:~/correct/virtualenv/backendvirtual$ cd Scraping/

(backendvirtual) student@csi6220-1-vm2:~/correct/virtualenv/backendvirtual/Scraping$ vi cd ..

2 files to edit

(backendvirtual) student@csi6220-1-vm2:~/correct/virtualenv/backendvirtual/Scraping$ cd ..

(backendvirtual) student@csi6220-1-vm2:~/correct/virtualenv/backendvirtual$ ls

bin include lib lib64 pip-selfcheck.json pythondb pyvenv.cfg Scraping share

(backendvirtual) student@csi6220-1-vm2:~/correct/virtualenv/backendvirtual$ cd ..

(backendvirtual) student@csi6220-1-vm2:~/correct/virtualenv$ ls

backendvirtual correctbackend correctpro

(backendvirtual) student@csi6220-1-vm2:~/correct/virtualenv$ pip install requests

Collecting requests

Downloading requests-2.18.1-py2.py3-none-any.whl (88kB)

100% |████████████████████████████████| 92kB 1.7MB/s

Collecting certifi>=2017.4.17 (from requests)

Downloading certifi-2017.4.17-py2.py3-none-any.whl (375kB)

100% |████████████████████████████████| 378kB 1.2MB/s

Collecting urllib3<1.22,>=1.21.1 (from requests)

Downloading urllib3-1.21.1-py2.py3-none-any.whl (131kB)

100% |████████████████████████████████| 133kB 2.1MB/s

Collecting idna<2.6,>=2.5 (from requests)

Downloading idna-2.5-py2.py3-none-any.whl (55kB)

100% |████████████████████████████████| 61kB 3.6MB/s

Collecting chardet<3.1.0,>=3.0.2 (from requests)

Downloading chardet-3.0.4-py2.py3-none-any.whl (133kB)

100% |████████████████████████████████| 143kB 2.3MB/s

Installing collected packages: certifi, urllib3, idna, chardet, requests

Successfully installed certifi-2017.4.17 chardet-3.0.4 idna-2.5 requests-2.18.1 urllib3-1.21.1

(backendvirtual) student@csi6220-1-vm2:~/correct/virtualenv$ pwd

/home/student/correct/virtualenv

(backendvirtual) student@csi6220-1-vm2:~/correct/virtualenv$ ls

backendvirtual correctbackend correctpro

(backendvirtual) student@csi6220-1-vm2:~/correct/virtualenv$ cd backendvirtual/

(backendvirtual) student@csi6220-1-vm2:~/correct/virtualenv/backendvirtual$ pwd

/home/student/correct/virtualenv/backendvirtual

(backendvirtual) student@csi6220-1-vm2:~/correct/virtualenv/backendvirtual$ ls

bin include lib lib64 pip-selfcheck.json pythondb pyvenv.cfg Scraping share

(backendvirtual) student@csi6220-1-vm2:~/correct/virtualenv/backendvirtual$ cd Scraping/

(backendvirtual) student@csi6220-1-vm2:~/correct/virtualenv/backendvirtual/Scraping$ pwd

/home/student/correct/virtualenv/backendvirtual/Scraping

(backendvirtual) student@csi6220-1-vm2:~/correct/virtualenv/backendvirtual/Scraping$ scp /Users/LiangXS/Desktop/FakeNewsScraping.py student@csi6220-1-vm2:/home/student/correct/virtualenv/backendvirtual/Scraping

The authenticity of host 'csi6220-1-vm2 (137.43.49.13)' can't be established.

ECDSA key fingerprint is SHA256:6g1PteOXzae1sj1x+XAvNPxhMZgBJligV1/9P9T+ZGE.

Are you sure you want to continue connecting (yes/no)? yes

Warning: Permanently added 'csi6220-1-vm2,137.43.49.13' (ECDSA) to the list of known hosts.

student@csi6220-1-vm2's password:

/Users/LiangXS/Desktop/FakeNewsScraping.py: No such file or directory

(backendvirtual) student@csi6220-1-vm2:~/correct/virtualenv/backendvirtual/Scraping$ scp /Users/LiangXS/Desktop/FakeNewsScraping.py student@csi6220-1-vm2:.

student@csi6220-1-vm2's password:

/Users/LiangXS/Desktop/FakeNewsScraping.py: No such file or directory

(backendvirtual) student@csi6220-1-vm2:~/correct/virtualenv/backendvirtual/Scraping$ pwd

/home/student/correct/virtualenv/backendvirtual/Scraping

(backendvirtual) student@csi6220-1-vm2:~/correct/virtualenv/backendvirtual/Scraping$ ls

FakeNewsScraping.py

(backendvirtual) student@csi6220-1-vm2:~/correct/virtualenv/backendvirtual/Scraping$ python FakeNewsScraping.py

http://activistpost.com

Traceback (most recent call last):

File "FakeNewsScraping.py", line 103, in <module>

get\_news\_firstpage(link)

File "FakeNewsScraping.py", line 26, in get\_news\_firstpage

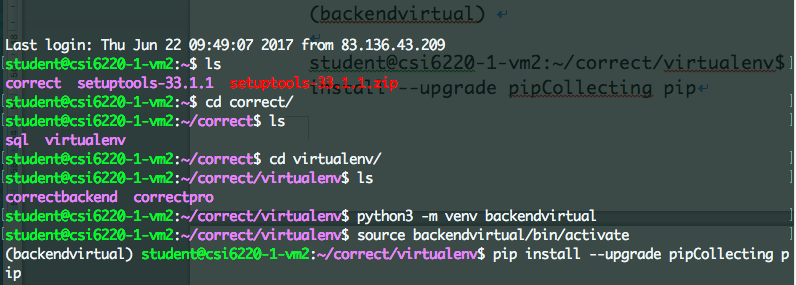
soup = BeautifulSoup(source.content, "lxml")

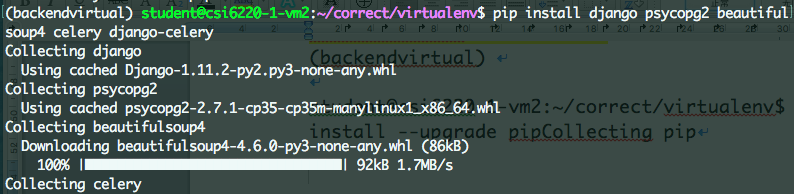
File "/home/student/correct/virtualenv/backendvirtual/lib/python3.5/site-packages/bs4/\_\_init\_\_.py", line 165, in \_\_init\_\_

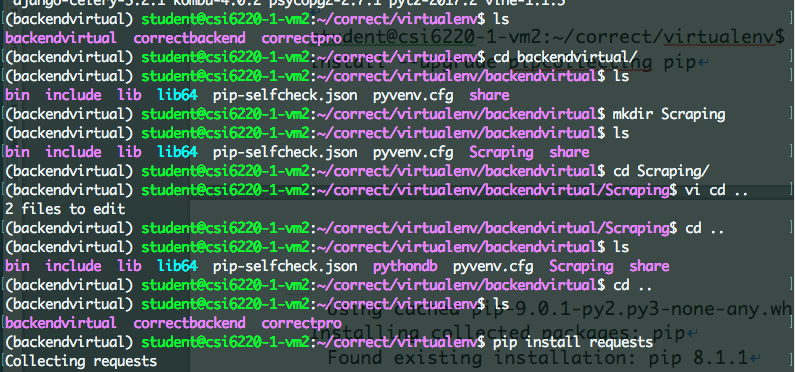
% ",".join(features))

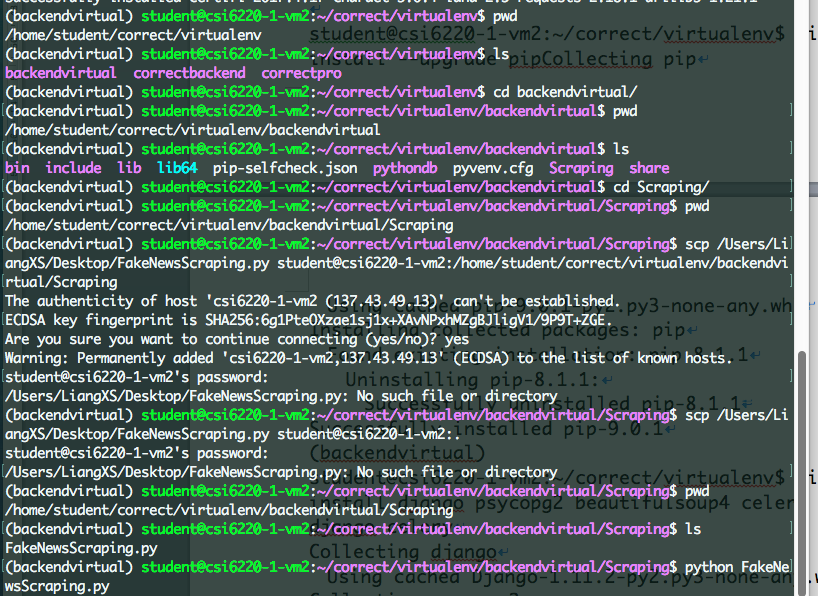
bs4.FeatureNotFound: Couldn't find a tree builder with the features you requested: lxml. Do you need to install a parser library?

(backendvirtual) student@csi6220-1-vm2:~/correct/virtualenv/backendvirtual/Scraping$









# for i in range(len(link)):

# link[i] = 'http://' + link[i]

# headers = {'user-agent' : 'Mozilla/5.0'}

# source = requests.get(link[i], headers = headers)

# if source.status\_code == 200:

# soup = BeautifulSoup(source.content, "lxml")

# if (soup.findAll('article')):

# for article in soup.findAll('article'):

# if (article.find('a')):

# href = article.find('a').get('href')

# if href.startswith('http'):

# href = href

# else:

# href = link + href

# print (href)

# get\_news\_content(href)

# else:

# print ("first page can not be scaped")

# else:

# print ('The first page can not be found!')

# else:

# print ("Oops, the website can not be found!!!")

# 'www.clickhole.com',

# 'www.theonion.com']

# link = 'americannews.com'

# link = 'activistpost.com'

# link = 'thedailysheeple.com'

# link = 'waterfordwhispersnews.com'

# link = 'clickhole.com'

# link = 'theonion.com'

# link = 'thebeaverton.com' # no use

# link = 'dailynewsbin.com' # not exist

# link = 'abcnews.com.co' # not aceptable

# link = 'fakingnews.com' #not able to

# link = 'usapoliticstoday.com' #not able to

# link = 'usherald.com' #not able to

# link = 'borowitzreport.com' #not able to

# link = 'satirewire.com' #not able to

# link = 'theshovel.com.au' #not able to