

1. Purpose:

This module implements DataSet Creation Using Python.

2. Module Location:

1_DataSet_Creation/Codes

3. Approaches:

3.1-Get Feeds from NEWS Websites.

Tool Used:

- a)FeedParser
- b)BeautifulSoup

Code Location:

1_DataSet_Creation/Codes

Program file name:

parse_feeds.py

input:

No Input required.

Output:

1_DataSet_Creation/Data/Parse_Feeds_Files

Process to Execute:

- a)use command `python parse_feeds.py`

3.2-Copy Data from OSINT lab

Tool Used:

- a)MySQLdb

Code Location:

1_DataSet_Creation/Codes

Program file name:

get_data.py
extract_all_files.py

input:

No Input required.

Output:

- a)a table named NEWS in Database containing Articles and its features.
- b) 1_DataSet_Creation/Data/All_Files with .txt files

Process to Execute:

- a)Install SQL Database(remember the credentials you used while setting up database)
- b)create a database named TESTDB.
- c)open get_data.py and enter credential to access Database.
- d)now, in terminal execute command `get_data.py`
- e)To get the Database files in .txt format use '`python extract_all_files.py`'.
- f)this will create .txt files in 1_DataSet_Creation/Data/All_Files .

4. Software and Tools Requirement:

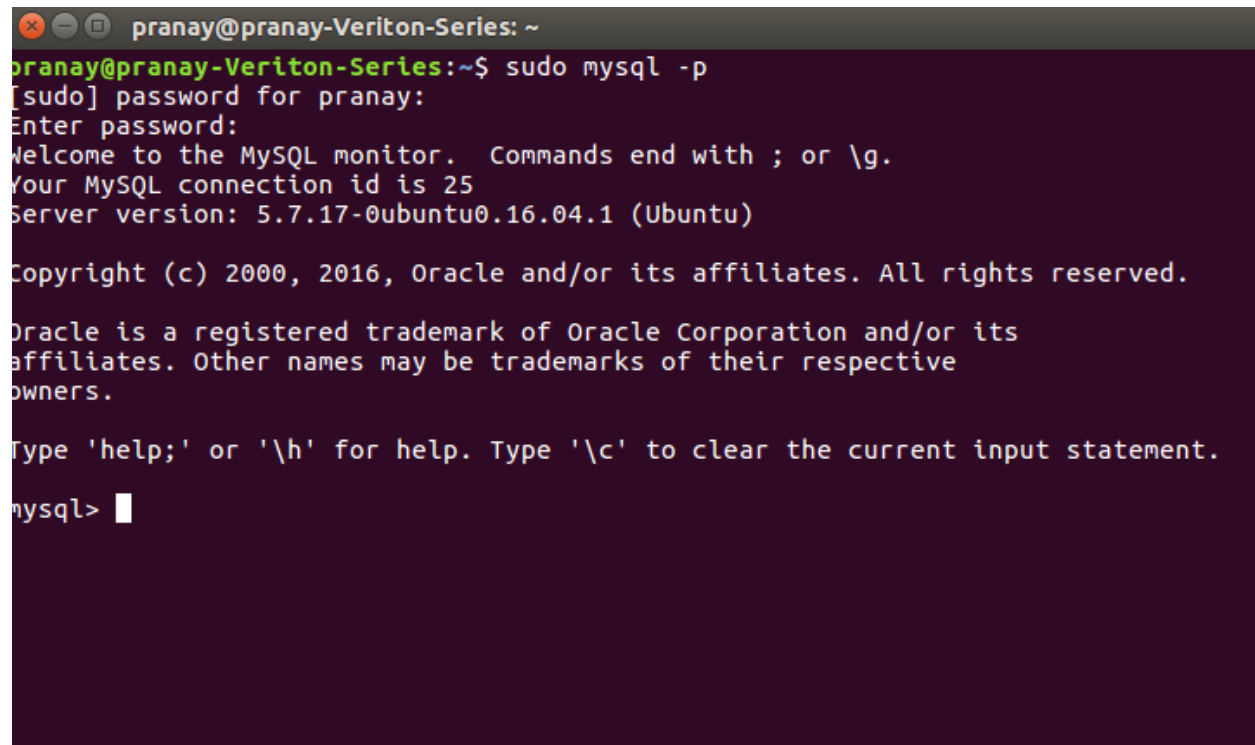
python 2.7,MySQLdb

(to install MySQLdb use `pip install pymysql`)

5. Process:

Before the mid term evaluation we worked on Feed Parser approach but after evaluation as instructed by instructor we used second approach.

a) Install MySQL:

A terminal window with a dark background and light-colored text. The window title is 'pranay@pranay-Veriton-Series: ~'. The user enters 'sudo mysql -p' at the prompt. The terminal shows the password prompt, the MySQL welcome message, the connection ID (25), the server version (5.7.17-0ubuntu0.16.04.1), and copyright information. The prompt then changes to 'mysql>' with a cursor.

```
pranay@pranay-Veriton-Series: ~  
pranay@pranay-Veriton-Series:~$ sudo mysql -p  
[sudo] password for pranay:  
Enter password:  
Welcome to the MySQL monitor.  Commands end with ; or \g.  
Your MySQL connection id is 25  
Server version: 5.7.17-0ubuntu0.16.04.1 (Ubuntu)  
  
Copyright (c) 2000, 2016, Oracle and/or its affiliates. All rights reserved.  
  
Oracle is a registered trademark of Oracle Corporation and/or its  
affiliates. Other names may be trademarks of their respective  
owners.  
  
Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.  
mysql> █
```

After installing you will get a screen like this

b) Create Database using *Create Database TESTDB;* command

```

pranay@pranay-Veriton-Series: ~
pranay@pranay-Veriton-Series:~$ sudo mysql -p
Enter password:
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 26
Server version: 5.7.17-0ubuntu0.16.04.1 (Ubuntu)

Copyright (c) 2000, 2016, Oracle and/or its affiliates. All rights reserved.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> create database TESTDB;

```

c)open get_data.py and enter credential to access Database.

```

db_local = MySQLdb.connect("127.0.0.1","YOUR USERNAME","YOUR
PASSWORD","TESTDB");

```

```

db_local = MySQLdb.connect("127.0.0.1","root","password","TESTDB")
cursor_local = db_local.cursor()
#Query to create table
sql_create_table = "CREATE TABLE News (id int NOT NULL,Link CHAR(200),published_on int,Title
BLOB,news_paper char(200),article_text BLOB )"
cursor_local.execute(sql_create_table)

```

d)in terminal execute command get_data.py

```

pranay@pranay-Veriton-Series:~/Desktop/Final Code /DataSet Creation/Codes$ python get_data.py

```

wait till you get

```
line no. %d copying 101549
Finished
pranay@pranay-Veriton-Series:~/Desktop/Final Code /DataSet Creation/Codes$
```

Now the database is ready. It has a table named News

```
mysql> use TESTDB;
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A

Database changed
mysql> describe News
-> ;
+-----+-----+-----+-----+-----+
| Field | Type   | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+
| id     | int(11) | NO   |     | NULL    |       |
| Link   | char(200) | YES  |     | NULL    |       |
| Date   | int(11) | YES  |     | NULL    |       |
| Title  | char(100) | YES  |     | NULL    |       |
| source | char(200) | YES  |     | NULL    |       |
| content | blob    | YES  |     | NULL    |       |
+-----+-----+-----+-----+-----+
6 rows in set (0.32 sec)
```

Next part is to create .txt files using this database.

e) To get the Database files in .txt format use 'python extract_txt.py'.

Simply Run this command *python extract_txt.py*

```
pranay@pranay-Veriton-Series:~/Desktop/Final Code /DataSet Creation/Codes$ python extract_txt.py
```