Alexander Wilson

2/11/2016

1)

I ran my twitter api on multiple occasions to see a wide variety of data. After all the twitter scraping was done I was left with around 10 g of data. I then picked a 1g sample to use to extract links. I was able to extract many links but I consolidated a list of 1000.

http://adilmoujahid.com/posts/2014/07/twitter-analytics/ This is the website that guided me through this assignment.

```
#Import the necessary methods from tweepy library
from tweepy.streaming import StreamListener
from tweepy import OAuthHandler
from tweepy import Stream

#Variables that contains the user credentials to access Twitter API
access_token = "XXXX"
access_token_secret = "XXXX"
consumer_key = "kXXXXX"
consumer_secret = "XXXX"


#This is a basic listener that just prints received tweets to stdout.
class StdOutListener(StreamListener):

    def on_data(self, data):
        print (data)
        return True

    def on_error(self, status):
        print (status)


if __name__ == '__main__':

    #This handles Twitter authetification and the connection to Twitter Streaming API
    l = StdOutListener()
    auth = OAuthHandler(consumer_key, consumer_secret)
    auth.set_access_token(access_token, access_token_secret)
    stream = Stream(auth, l)

    #This line filter Twitter Streams to capture data by the keywords: 'python', 'javascript', 'ruby'
    stream.filter(track=['.org', '.com', '.net', 'tinyurl','t.co', 'bit.ly', 'goo.gl', 'http' ] )
```
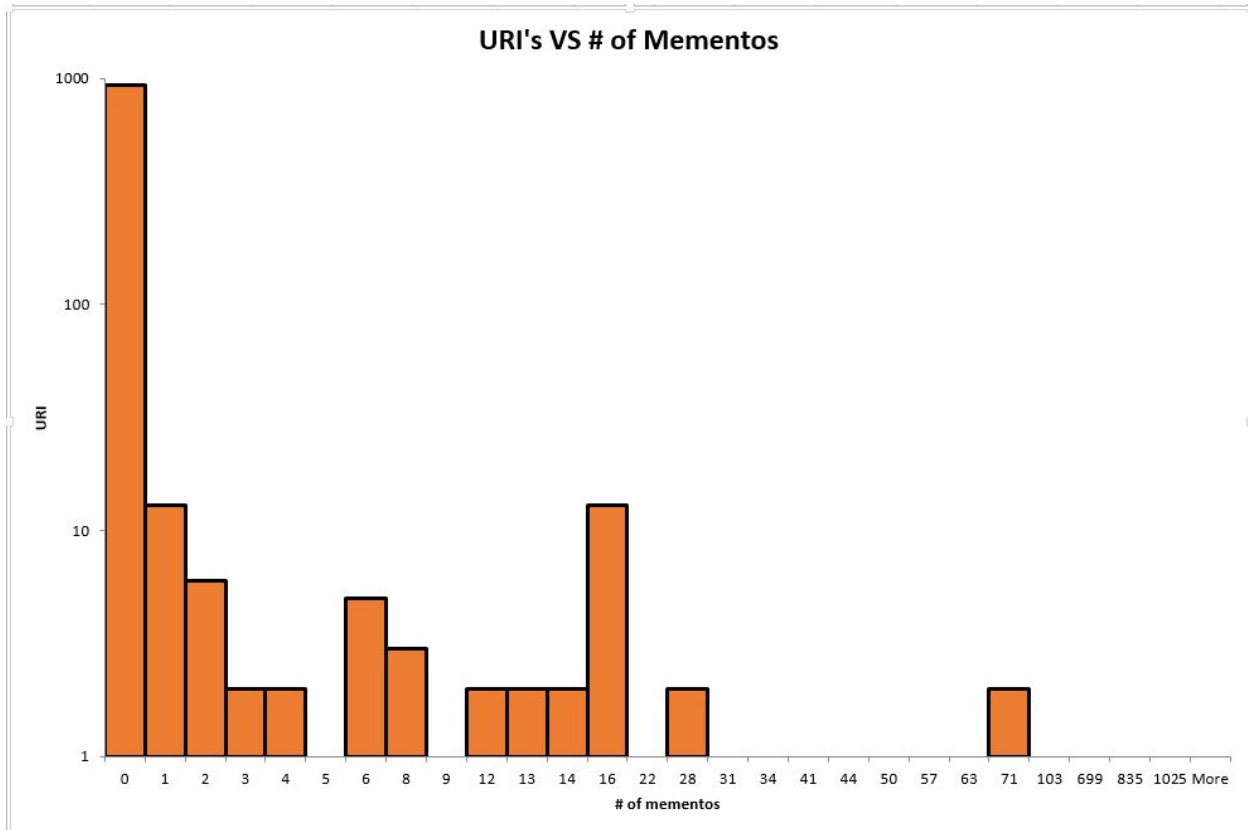
```python
1   #Auired from Adil Moujahid
2   #minimual fixes were required
3   #Import the necessary methods from tweepy library
4   #It allwed me to extract the data I needed.
5   #I was able to parse after collecting it.
6   from tweepy.streaming import StreamListener
7   from tweepy import OAuthHandler
8   from tweepy import Stream
9
10  #Variables that contains the user credentials to access Twitter API
11  access_token = "3379644472-1s8Se7xJTsrsVyOAN4N6PXRZtO9C92EuIdnAS3i"
12  access_token_secret = "76X6vml96MdpgYzfciUgVH2L32wMfaKIbe8Q834s8ld6A"
13  consumer_key = "3t6s2AykC8jBOZ0H7zxImwgpk"
14  consumer_secret = "K7MqLm8qfCWhQCDf9SLh7lYsZyXEldJDu045CYisdRAFhNz65e"
15
16
17  #This is a basic listener that just prints received tweets to stdout.
18  class StdOutListener(StreamListener):
19
20      def on_data(self, data):
21          print (data)
22          return True
23
24      def on_error(self, status):
25          print (status)
26
27
28  if __name__ == '__main__':
29
30      #This handles Twitter authetification and the connection to 4Twitter Streaming API
31      l = StdOutListener()
32      auth = OAuthHandler(consumer_key, consumer_secret)
33      auth.set_access_token(access_token, access_token_secret)
34      stream = Stream(auth, l)
35
36      #This line filter Twitter Streams to capture data by the keywords: 'python', 'javascript', 'ruby'
37      stream.filter(track=['.org', '.com', '.net', 'tinyurl','t.co', 'bit.ly', 'goo.gl', 'http' ] )
```

2)

0 930

1 10

2 7

3 3

4 2

5 4

6 5

8 3

9 1

12 2

13 2

14 2

16 13

22 1

28 2

31 1

34 1

41 1

44 1

50 1

57 1

63 1

71 2

103 1
699 1
835 1
1025 1

I was able to make the graph in R however it was as not pretty.

3)
I was able to use a parser to separate links with no archive history and with archive history. I then got links with an archive history to run through the carbon dater. I ran into trouble trying to collect the dates from the archived uris.

Usage: carbontool.py url backlinksOnOffFlag (e.g: carbontool.py http://www.cs.odu.edu [--compute-backlinks])
Done Bitly
Done Last Modified
Done Google
Done Archives 0
runtime in seconds: 17
{

```
  "URI":
"http://mementoproxy.cs.odu.edu/aggr/timemap/link/1/https://twitter.com/kayefofficial/status/627
497266081886208\n",
  "Estimated Creation Date": "",
  "Last Modified": "",
  "Bitly.com": "",
  "Topsy.com": "Topsy is out of service",
  "Backlinks": "",
  "Google.com": "",
  "Archives": [
    [
      "Earliest",
      ""
    ],
    [
      "By_Archive",
      {}
    ]
  ]
}
```

Done Bitly
Done Last Modified
Done Google
Done Archives 1
runtime in seconds: 0

```
{
  "URI": "http://\n",
  "Estimated Creation Date": "2013-01-09T16:32:13",
  "Last Modified": "",
  "Bitly.com": "",
  "Topsy.com": "Topsy is out of service",
  "Backlinks": "",
  "Google.com": "",
  "Archives": [
    [
      "Earliest",
      "2013-01-09T16:32:13"
    ],
    [
      "By_Archive",
      {
        "http://web.archive.org/web/20130109213213/http:////": "2013-01-09T16:32:13"
```

```
        }
      ]
    ]
}

Done Bitly
Done Last Modified
Done Google
Done Archives 0
runtime in seconds:  30
{
  "URI": "http://mementoproxy.cs.odu.edu/aggr/timemap/link/1/http://ticketchatter.net\n",
  "Estimated Creation Date": "",
  "Last Modified": "",
  "Bitly.com": "",
  "Topsy.com": "Topsy is out of service",
  "Backlinks": "",
  "Google.com": "",
  "Archives": [
    [
      "Earliest",
      ""
    ],
    [
      "By_Archive",
      {}
    ]
  ]
}

Done Last Modified
Done Bitly
Done Google
Done Archives 1
runtime in seconds:  0
{
  "URI": "http://\n",
  "Estimated Creation Date": "2013-01-09T16:32:13",
  "Last Modified": "",
  "Bitly.com": "",
  "Topsy.com": "Topsy is out of service",
  "Backlinks": "",
  "Google.com": "",
```

    "Archives": [
     [
       "Earliest",
       "2013-01-09T16:32:13"
     ],
     [
       "By_Archive",
       {
         "http://web.archive.org/web/20130109213213/http:////": "2013-01-09T16:32:13"
       }
     ]
   ]
}

Done Last Modified
Done Bitly
Done Google
Done Archives 0
runtime in seconds: 13
{
  "URI": "http://mementoproxy.cs.odu.edu/aggr/timemap/link/1/http://www.nba.com\n",
  "Estimated Creation Date": "",
  "Last Modified": "",
  "Bitly.com": "",
  "Topsy.com": "Topsy is out of service",
  "Backlinks": "",
  "Google.com": "",
  "Archives": [
   [
     "Earliest",
     ""
   ],
   [
     "By_Archive",
     {}
   ]
  ]
}

Done Bitly
Done Last Modified
Done Google
Done Archives 1

runtime in seconds: 0
{
  "URI": "http://\n",
  "Estimated Creation Date": "2013-01-09T16:32:13",
  "Last Modified": "",
  "Bitly.com": "",
  "Topsy.com": "Topsy is out of service",
  "Backlinks": "",
  "Google.com": "",
  "Archives": [
    [
      "Earliest",
      "2013-01-09T16:32:13"
    ],
    [
      "By_Archive",
      {
        "http://web.archive.org/web/20130109213213/http:////": "2013-01-09T16:32:13"
      }
    ]
  ]
}

Done Last Modified
Done Bitly
Done Google
Done Archives 0
runtime in seconds: 17
{
  "URI":
"http://mementoproxy.cs.odu.edu/aggr/timemap/link/1/http://neuvoo.ca/jobs/?k=legal&l=Montreal
%2C+QC%2C+Canada&f=&p=1&r=15\n",
  "Estimated Creation Date": "",
  "Last Modified": "",
  "Bitly.com": "",
  "Topsy.com": "Topsy is out of service",
  "Backlinks": "",
  "Google.com": "",
  "Archives": [
    [
      "Earliest",
      ""
    ],

```
        [
          "By_Archive",
          {}
        ]
      ]
    }

Done Bitly
Done Last Modified
Done Google
Done Archives 1
runtime in seconds:  0
{
  "URI": "http://\n",
  "Estimated Creation Date": "2013-01-09T16:32:13",
  "Last Modified": "",
  "Bitly.com": "",
  "Topsy.com": "Topsy is out of service",
  "Backlinks": "",
  "Google.com": "",
  "Archives": [
    [
      "Earliest",
      "2013-01-09T16:32:13"
    ],
    [
      "By_Archive",
      {
        "http://web.archive.org/web/20130109213213/http:////": "2013-01-09T16:32:13"
      }
    ]
  ]
}

Done Last Modified
Done Bitly
Done Google
Done Archives 0
runtime in seconds:  30
{
  "URI": "http://mementoproxy.cs.odu.edu/aggr/timemap/link/1/http://www.radioemotion.be\n",
  "Estimated Creation Date": "",
  "Last Modified": "",
```

    "Bitly.com": "",
    "Topsy.com": "Topsy is out of service",
    "Backlinks": "",
    "Google.com": "",
    "Archives": [
      [
        "Earliest",
        ""
      ],
      [
        "By_Archive",
        {}
      ]
    ]
}

Done Last ModifiedDone Bitly

 Done Google
Done Archives 1
runtime in seconds: 0
{
  "URI": "http://\n",
  "Estimated Creation Date": "2013-01-09T16:32:13",
  "Last Modified": "",
  "Bitly.com": "",
  "Topsy.com": "Topsy is out of service",
  "Backlinks": "",
  "Google.com": "",
  "Archives": [
    [
      "Earliest",
      "2013-01-09T16:32:13"
    ],
    [
      "By_Archive",
      {
        "http://web.archive.org/web/20130109213213/http:////": "2013-01-09T16:32:13"
      }
    ]
  ]
}

```
Done Last Modified
Done Bitly
Done Google
Done Archives 0
runtime in seconds:  30
{
  "URI": "http://mementoproxy.cs.odu.edu/aggr/timemap/link/1/http://www.radioemotion.be\n",
  "Estimated Creation Date": "",
  "Last Modified": "",
  "Bitly.com": "",
  "Topsy.com": "Topsy is out of service",
  "Backlinks": "",
  "Google.com": "",
  "Archives": [
   [
     "Earliest",
     ""
   ],
   [
     "By_Archive",
     {}
   ]
  ]
}

Done Bitly
Done Last Modified
Done Google
Done Archives 1
runtime in seconds:  0
{
  "URI": "http://\n",
  "Estimated Creation Date": "2013-01-09T16:32:13",
  "Last Modified": "",
  "Bitly.com": "",
  "Topsy.com": "Topsy is out of service",
  "Backlinks": "",
  "Google.com": "",
  "Archives": [
   [
     "Earliest",
     "2013-01-09T16:32:13"
   ],
```

```
      [
        "By_Archive",
        {
          "http://web.archive.org/web/20130109213213/http:////": "2013-01-09T16:32:13"
        }
      ]
    ]
  }
```

Done Last Modified
Done Bitly
Done Archives 0
Done Google
runtime in seconds: 0

```
  {
    "URI": "http://mementoproxy.cs.odu.edu/aggr/timemap/link/1/http://www.radioemotion.be\n",
    "Estimated Creation Date": "",
    "Last Modified": "",
    "Bitly.com": "",
    "Topsy.com": "Topsy is out of service",
    "Backlinks": "",
    "Google.com": "",
    "Archives": [
      [
        "Earliest",
        ""
      ],
      [
        "By_Archive",
        {}
      ]
    ]
  }
```

Done Last ModifiedDone Bitly

Done Google
Done Archives 1
runtime in seconds: 0

```
  {
    "URI": "http://\n",
    "Estimated Creation Date": "2013-01-09T16:32:13",
    "Last Modified": "",
```

    "Bitly.com": "",
    "Topsy.com": "Topsy is out of service",
    "Backlinks": "",
    "Google.com": "",
    "Archives": [
      [
        "Earliest",
        "2013-01-09T16:32:13"
      ],
      [
        "By_Archive",
        {
          "http://web.archive.org/web/20130109213213/http:////": "2013-01-09T16:32:13"
        }
      ]
    ]
}

Done Last Modified
Done Bitly
Done Google
Done Archives 0
runtime in seconds: 11
{
  "URI": "http://mementoproxy.cs.odu.edu/aggr/timemap/link/1/http://www.templeton.org/\n",
  "Estimated Creation Date": "",
  "Last Modified": "",
  "Bitly.com": "",
  "Topsy.com": "Topsy is out of service",
  "Backlinks": "",
  "Google.com": "",
  "Archives": [
    [
      "Earliest",
      ""
    ],
    [
      "By_Archive",
      {}
    ]
  ]
}

Done Bitly
Done Last ModifiedDone Google

Done Archives 1
runtime in seconds: 0
{
  "URI": "http://\n",
  "Estimated Creation Date": "2013-01-09T16:32:13",
  "Last Modified": "",
  "Bitly.com": "",
  "Topsy.com": "Topsy is out of service",
  "Backlinks": "",
  "Google.com": "",
  "Archives": [
    [
      "Earliest",
      "2013-01-09T16:32:13"
    ],
    [
      "By_Archive",
      {
        "http://web.archive.org/web/20130109213213/http:////": "2013-01-09T16:32:13"
      }
    ]
  ]
}

Done Last Modified
Done Bitly
Done Google
Done Archives 0
runtime in seconds: 17
{
  "URI": "http://mementoproxy.cs.odu.edu/aggr/timemap/link/1/http://triblive.com/\n",
  "Estimated Creation Date": "",
  "Last Modified": "",
  "Bitly.com": "",
  "Topsy.com": "Topsy is out of service",
  "Backlinks": "",
  "Google.com": "",
  "Archives": [
    [
      "Earliest",

```
          ""
      ],
      [
        "By_Archive",
        {}
      ]
   ]
}

Done Last Modified
 Done Bitly
Done Google
Done Archives 1
runtime in seconds:  0
{
  "URI": "http://\n",
  "Estimated Creation Date": "2013-01-09T16:32:13",
  "Last Modified": "",
  "Bitly.com": "",
  "Topsy.com": "Topsy is out of service",
  "Backlinks": "",
  "Google.com": "",
  "Archives": [
    [
      "Earliest",
      "2013-01-09T16:32:13"
    ],
    [
      "By_Archive",
      {
        "http://web.archive.org/web/20130109213213/http:////": "2013-01-09T16:32:13"
      }
    ]
  ]
}

Done Last Modified
Done Bitly
Done Google
Done Archives 0
runtime in seconds:  11
{
  "URI": "http://mementoproxy.cs.odu.edu/aggr/timemap/link/1/http://citytweet.nl\n",
```

    "Estimated Creation Date": "",
    "Last Modified": "",
    "Bitly.com": "",
    "Topsy.com": "Topsy is out of service",
    "Backlinks": "",
    "Google.com": "",
    "Archives": [
      [
        "Earliest",
        ""
      ],
      [
        "By_Archive",
        {}
      ]
    ]
}

Done BitlyDone Last Modified

Done Google
Done Archives 1
runtime in seconds: 0
{
  "URI": "http://\n",
  "Estimated Creation Date": "2013-01-09T16:32:13",
  "Last Modified": "",
  "Bitly.com": "",
  "Topsy.com": "Topsy is out of service",
  "Backlinks": "",
  "Google.com": "",
  "Archives": [
    [
      "Earliest",
      "2013-01-09T16:32:13"
    ],
    [
      "By_Archive",
      {
        "http://web.archive.org/web/20130109213213/http:////": "2013-01-09T16:32:13"
      }
    ]
  ]

```
}

Done Last Modified
Done Bitly
Done Google
Done Archives 0
runtime in seconds:  10
{
  "URI": "http://mementoproxy.cs.odu.edu/aggr/timemap/link/1/http://www.goal.com/fr/\n",
  "Estimated Creation Date": "",
  "Last Modified": "",
  "Bitly.com": "",
  "Topsy.com": "Topsy is out of service",
  "Backlinks": "",
  "Google.com": "",
  "Archives": [
    [
      "Earliest",
      ""
    ],
    [
      "By_Archive",
      {}
    ]
  ]
}

Done Bitly
 Done Last ModifiedDone Google

Done Archives 1
runtime in seconds:  0
{
  "URI": "http://\n",
  "Estimated Creation Date": "2013-01-09T16:32:13",
  "Last Modified": "",
  "Bitly.com": "",
  "Topsy.com": "Topsy is out of service",
  "Backlinks": "",
  "Google.com": "",
  "Archives": [
    [
      "Earliest",
```

      "2013-01-09T16:32:13"
    ],
    [
     "By_Archive",
     {
       "http://web.archive.org/web/20130109213213/http:////": "2013-01-09T16:32:13"
     }
    ]
  ]
}