

Problem 1

To search for links within tweets in my twitter account maturban1, I have used exactly the same way suggested in problem 1 in our assignment: <http://thomassileo.com/blog/2013/01/25/using-twitter-rest-api-v1-dot-1-with-python/> List of key words, around 70, is used for the search. The idea is that the program will start searching the first word, collect possible links, and then continue searching the second word and so on. It stops either if collecting 1000 URIs or there is no more key words. In order to approve a link and add it to a final list, its effective url must be checked for the following:

- HTML code 200
- It is not already in the final list
- Content type 'text/html'
- It does not belong to "instagram.com" because, at the beginning, I got around 95 percent of links belong to this website!
- If the URI is split around '/'. the last part of the URI must contains one of the special characters or a white space.

The program will store the results in a file called "twitterLinks.txt", the result could be something like:

```
1 http://tavernkeepers.com/video-2011-john-mccain-visit-to-benghazi-praises-libyan-rebels-with-ambassador-steve
2 http://tavernkeepers.com/part-xxii-benghazi-today-is-a-source-of-hope-john-mccain-april-2011/
3 http://www.youtube.com/watch?v=F69PBQ42yNw&feature=youtu.be
4 http://www.youtube.com/watch?v=hBCKy_Evbfc&feature=youtu.be
5 http://avimeo.com/watch.php?vid=12dd9a60d#.UkGBQQq-ONk.twitter
6 http://www.redstate.com/2013/09/23/sitting-out-the-battle-fighting-a-fake-fight/
7 http://abchits.playtheradio.com
8 http://jainei.tumblr.com/post/62178500497/epicworkaholic-teenagedream-katyperry-song#_=_
9 http://www.hiphoptima.com/2013/09/18/director-sheldon-candis-discusses-j-coles-crooked-smile-video/
10 http://www.newyorkpost.com/2013/09/22/miley-cyrus-basically-bares-all/
11 http://www.yenewyorkpost.com/2013/09/22/miley-cyrus-basically-bares-all/
12 http://www.hiphoptima.com/2013/09/17/pusha-t-talks-kanye-rant/
13 https://www.youtube.com/watch?v=jqoHPzqUkZo&feature=youtube_gdata_player
14 https://itunes.apple.com/us/music-video/roar/id700911241?ign-mpt=uo%3D4
15 .
16 .
17 .
18
```

Figure 1: Output URIs

"getTwitterLinks.py" contains Python code for solving problem 1. "links.txt" is the output file.

Problem 2

Curl is used within Python to get number of mementos for each URI. The Curl command looks like:

```
$curl http://mementoproxy.cs.odu.edu/aggr/timemap/link/[link is here]
```

Because some of the returned lines should not be counted as a memento, only lines that have the key word "datetime=" will do. Unfortunately. The following histogram illustrates that most of the URIs come with Zero memento:

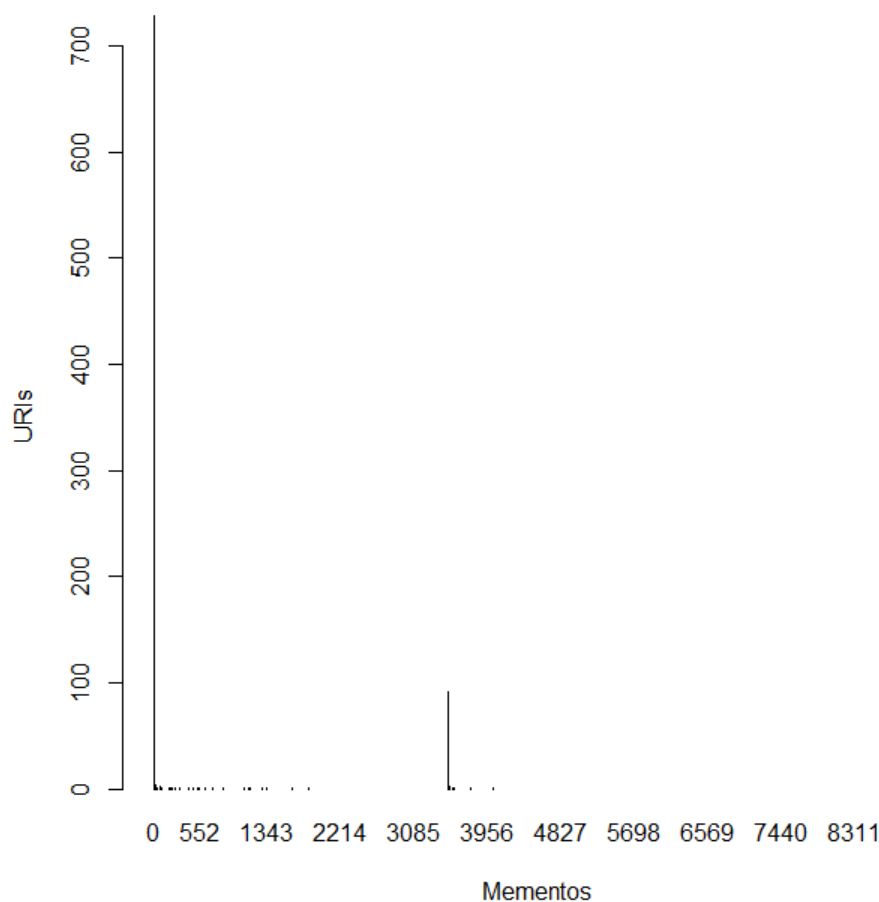


Figure 2: Mementos vs URIs

"getMementos.py" contains Python code for solving problem 2. "mementos.txt" is the output file.

Problem 3

Carbon Date tool is installed and used to estimate the age of the 1000 URIs. Also, Curl within Python code is used to connect the Carbon Date server, and because it takes time, several minutes, for the server to respond and send back results, five copies of this tool are run in parallel. This will help to finish estimating the 1000 URIs age more fast. The following plot shows how the number of mementos is effected by the age of the URIs. Only URIs that have mementos are included in this chart. Two URIs with age (0 days) are included in the plot also. I did not remove them because (0 days) indicates that the age of these URIs is less than 24 hours.

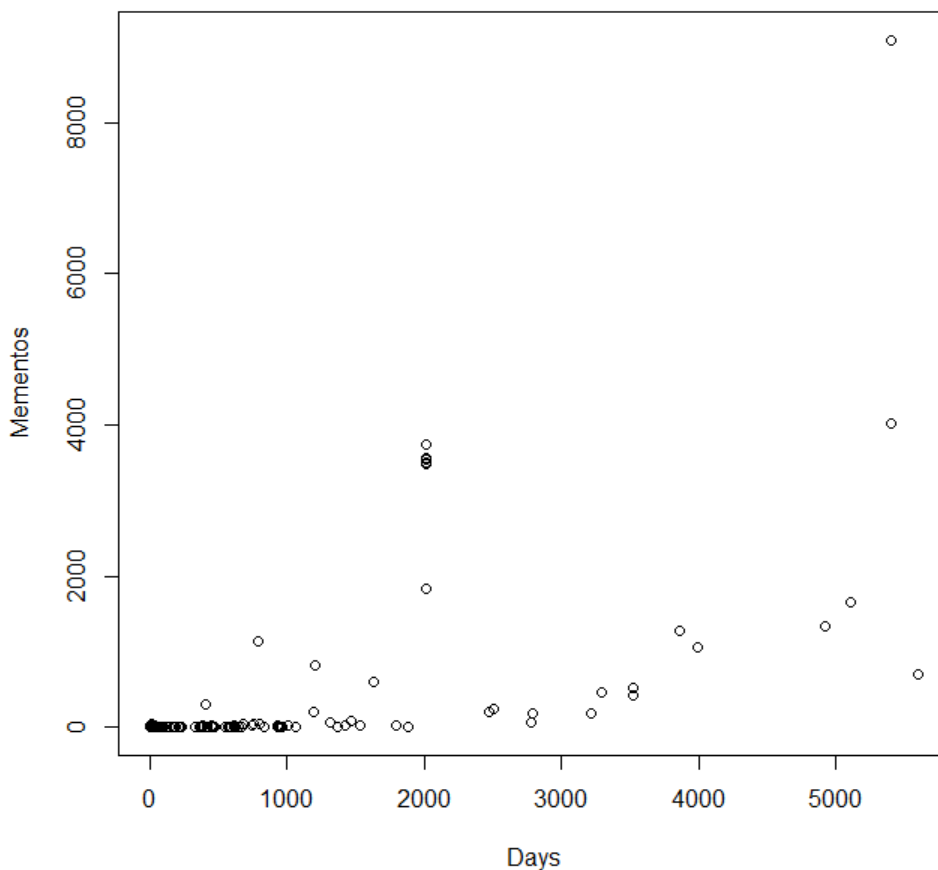


Figure 3: URIs age (in Days) vs Mementos

The plot shows that the young URIs get less chance of having more mementos! Finally, I would like to indicate that R language is used for plotting. I am still new in this environment. Next, I will try to understand how to use good scale to represent data more clear in plots.

"getURIage.py" contains Python code for solving problem 3. "age.txt" is the output file.