

Job Task Analysis.

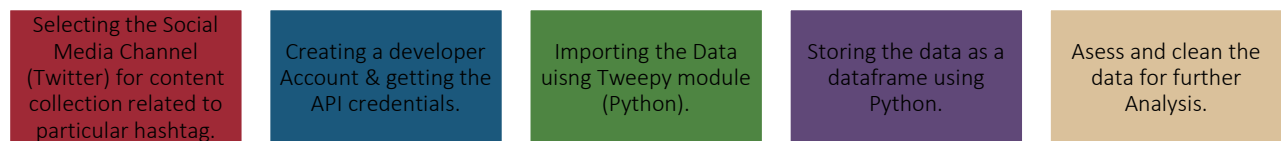
CONTENT COLLECTION FROM SOCIAL MEDIA CHANNELS

Problem Statement: Define a workflow to collect content from social media channels corresponding to hashtags.

OVERVIEW

This report briefly describes a methodological approach to collect content from social media channel(s) corresponding to hashtags.

SUMMARY OF WORKFLOW



We are planning to extract tweets containing a particular hashtag (#ODSCEurope) from twitter.

Following are the steps involved:

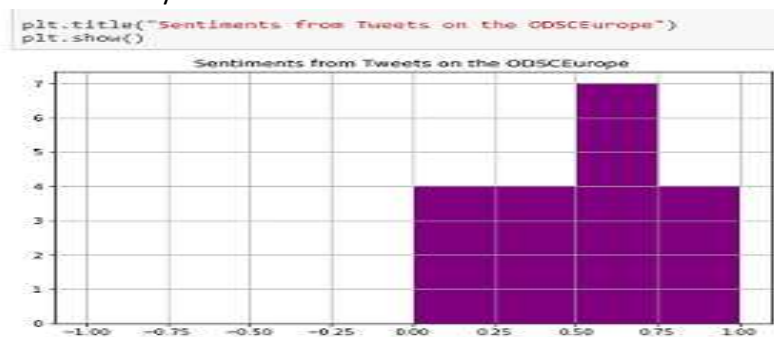
1. Sign up for a developer account via your Twitter account in order to generate the Consumer API keys, and the Access Token keys.
2. Import the relevant libraries in python.
3. Use the Tweepy module in python to interact with the twitter API.
4. Authentication and Extraction - Authenticate your account to access twitter's API. Once the authentication is completed, extract tweets for the hashtag 'ODSCEurope'. Here we have extracted 2000 tweets.
5. The type of content collected includes unique id of the tweet (id), Tweet date and time (created_at), Number of retweets (retweet_count), user name, user followers count, user location, Tweet text(text).
6. Process Results: Store the relevant tweets data in a data frame using Python and remove tweets that have duplicate text.
7. Sentiment Analysis: In order to add another level to the analysis, we can perform sentiment analysis of the tweets. Python has an inbuilt library (textblob) to do this.
8. After gathering the above data, assess them visually and programmatically for quality and tidiness issues.

SAMPLE SCREENSHOTS

a) Sample Data Set

	id	created_at	retweet_count	user_screen_name	user_followers_count	user_location	Hashtags	text
0	1159279337046233088	2019-08-08 01:45:06	11	DamiSammaro	1929	Buenos Aires - Argentina	{{'text': 'AI', 'indices': [81, 84]}, {'text': ...	RT @KirkDBorne: The trend is clear. With some ...
1	1159207366170152960	2019-08-07 20:59:07	0	stanford_ai	441	Cambridge, MA	{{'text': 'ODSCEurope', 'indices': [7, 18]}, { ...	Attend #ODSCEurope, the largest applied data s...
2	1159193190865522690	2019-08-07 20:02:48	1	JustBeMentalist	1216	Shropshire	{{'text': 'ODSCEurope', 'indices': [17, 28]}, ...	RT @odsc: Attend #ODSCEurope, the largest appl...

b) Sentiment Analysis Plot



REFERENCES.

1. <https://github.com/ptwobrussell/Mining-the-Social-Web-2nd-Edition/blob/master/ipynb/Chapter%20-%20Mining%20Twitter.ipynb>
2. <https://www.toptal.com/python/twitter-data-mining-using-python>
3. <https://gist.github.com/vickyqian/f70e9ab3910c7c290d9d715491cde44c>
4. <https://www.kdnuggets.com/2017/03/beginners-guide-tweet-analytics-pandas.html>
5. <https://pypi.org/project/imageio/>

COMPLICATIONS

1. Unable to collect the content related to the hashtag #ODSCEast2019 due to the Twitter Official API limitation of time constraints. Typically tweets older than 2 weeks will not be returned.
2. Tried different analytics tool like Netlytic, but not able to download data corresponding to hashtag#ODSCEast2019. Since the conference were held from April 30 to May 3rd 2019, its difficult to download the data.

ASSUMPTIONS

1. Assuming that the content for hashtag to be collected is associated with the hashtag #ODSCEurope.
2. Only 2000 tweets are included in this dataset.