Goal: Getting people's impact factor for each subject keyword Using Twitter API

(ex. science, startup, biotech)

- **Approach using python code (using Tweepy module)**

  1. Search Tweet by using search_recent_tweets, keyword filtering *(use twitter API v2 endpoints)*
  - Code example:
  query= 'science -is:retweet lang:en'   *# query can be modified to filter results by the rule of twitter API docs*
  response = client.search_recent_tweets(query=query, max_results=10, tweet_fields=['created_at', 'lang', 'author_id']
                                                          # 'author_id' field provide the user id information to access User Object for further process

  2. Get User object by using 'author_id'.
  - Code example:
  user = api.get_user(user_id =userid)   *# return User object which contains the unique information of the user*

  3. Create data frame of the user information by using the User object
  - using Pandas module
  - *Code example:*   *# user information such as user id, description, followers can be stored by this process*
    list_of_tuples = list(zip(user_id, name, description, fav_cnt)
    df_user = pd.DataFrame(list_of_tuples, columns=['User Id', 'Screen name', 'Description', 'favorite count'])

  4. Get the recent tweets using user_timeline of each user *(use twitter API v1.1 endpoints)*
  - Code example:
  public_tweets = api.user_timeline(user_id=userid)   *# return recent 20 tweets of the user (*TBD: need to filter contents and get 10 tweets)*

  5. Get statistics information(sum, average, median, stdev) of each user
  - using numpy module to get statistics   *# get the retweet count and the favorite count of each tweet and store into a list.*
  - Code example: *# and get the statistics using the values in the list.*
    avg_ret.append(np.average(retweet_cnt)); avg_fav.append(np.average(favorite_cnt))
    median_ret.append(np.median(retweet_cnt)); std_ret.append(np.std(retweet_cnt))

  6. Merge the information into a dataframe and store it as csv file (TBD: need to store AWS SQL database server)**
   ex) df_user_statistic = df_user[['Screen name', 'Description', 'Total retweets', 'Total favorite', 'Average retweets',
                    'Average Favorites', 'Median retweets', 'Median favorites',
                    'stdev retweets', 'stdev favorites']]
      df_user_statistic.to_csv(f"log/{file_name}")

Anna Ha

Example of the data frame after processing

	Screen name	Description	Total retweets	Total favorite	Average retweets	Average Favorites	Median retweets	Median favorites	stdev retweets	stdev favorites
0	ethicypatriot	Retired system engineer and inventor. Adding t...	7902	0	395.10	0.00	0.0	0.0	1220.509521	0.000000
1	MentalWelcome	Curing your mental illnesses one ignorant obse...	0	0	0.00	0.00	0.0	0.0	0.000000	0.000000
2	fkanori		90533	2	4526.65	0.10	854.0	0.0	7670.751464	0.300000
3	flowsdoc	Ex-pro @Dodgers. Doctor of Physical Therapy. B...	21	150	1.05	7.50	0.5	2.0	1.430909	12.431814
4	es_iode_ai	AI powered scientific document search engine f...	7	8	0.35	0.40	0.0	0.0	0.726292	0.734847
5	GHalv	Truth isn't relative. https://t.co/2Qghb62MKl	469	27	23.45	1.35	0.0	0.0	67.797843	4.992745
6	Newsgrasp	NewsGrasp feeds you with the latest news from ...	1	0	0.05	0.00	0.0	0.0	0.217945	0.000000
7	keepinthescore	OU Alu '17 https://t.co/xoBGRtQUto... #Pittsburg...	292	1	14.60	0.05	0.0	0.0	61.163224	0.217945
8	NicoleZiege	Government & politics reporter for @PCMyrtleBe...	14386	5	719.30	0.25	0.0	0.0	3082.717001	0.536190
9	YKainis	It speaks for itself	0	0	0.00	0.00	0.0	0.0	0.000000	0.000000
10	CarminaFajardo4	This too shall pass	0	0	0.00	0.00	0.0	0.0	0.000000	0.000000
11	truhoopa4		0	0	0.00	0.00	0.0	0.0	0.000000	0.000000
12	rrustonyourdoor	#DRUKKARI golden lovebot | just close your eye...	21974	5	1098.70	0.25	0.0	0.0	3559.910913	0.433013
13	nanosplainer	techno biblical prophecy world view	0	2	0.00	0.10	0.0	0.0	0.000000	0.300000
14	ItachiU53932583	astrophille	193	79	9.65	3.95	0.0	0.0	42.063375	14.800253
15	chasingangels4	singer but not famous. goldigga	396	0	19.80	0.00	0.0	0.0	21.889724	0.000000
16	CryptoTiges	Navigating through life and sending good vibes...	19	27	0.95	1.35	0.0	1.0	2.418160	1.194780
17	SydneyG28807524	رعبة""\n، الثقة عشره عُمر ، لَيسَت مُجرد أمر ...إلى	0	0	0.00	0.00	0.0	0.0	0.000000	0.000000
18	PerryPe87700113	the same items frm here	295	0	14.75	0.00	0.0	0.0	18.066198	0.000000
19	DanielBundred	English teacher; KS5 Coordinator; opinions are...	4	12	0.20	0.60	0.0	0.0	0.509902	0.916515

Reference
https://developer.twitter.com/en/docs/twitter-api
https://docs.tweepy.org/en/stable/
https://numpy.org/doc/stable/reference/routines.statistics.html

Anna Ha