**1. Collect all COVID-related tweets by keywords (using Twitter search API)**

**→Have done the test, if the keywords have no problem, could start at any time.**

According to the related references or dataset, I choose these as the primary keywords in investigated platforms.

|  |  |
| --- | --- |
| Keywords | RFs |
| '#corona', '#coronavirus', '#covid', '#covid19', '#sarscov2', '#covid-19' | (1) |
| '#corona virus', ‘#Koronawirus’, ‘#Coronavirus’ | (2) |
| '#COVID19Italia', '#COVID19Pandemic' | (3) |
| ‘#2019nCoV’, ‘#codvid\_19’, ‘#codvid19’, ‘#conronaviruspandemic’, ‘#coronaflu’, ‘#coronaoutbreak’, ‘#coronapandemic’, ‘#Coronapanik’, ‘#coronavid19’, ‘#コロナ’, ‘#コロナウィルス’, ‘#新型コロナウイルス’, ‘#新型肺炎’ | (4) |

References

(1) IEEE Dataport <https://ieee-dataport.org/open-access/corona-virus-covid-19-tweets-dataset>

(2) Andrzej Jarynowski, Wojta-Kempa, Belik (2020) Trends in Perception of COVID-19 in Polish Internet. doi: <https://doi.org/10.1101/2020.05.04.20090993>

(3) http://twita.di.unito.it/dataset/40wita

(4) Twitter COVID-19 Stream. <https://developer.twitter.com/en/docs/labs/covid19-stream/filtering-rules>

**2. Collect & Track tweets by geographic location (using Twitter streaming API)**

**→I'm writing this code, still need some modifies for now.**

By one purely anecdotal estimate, somewhere between 1% of Twitter users have set up precise location tracking and 10% have set up tracking to a wider area, such as a city.

The Tweets could be tracked by location, language, and text by passing the three arrays shown below into the stream.filter() method of Tweepy, which is a Python API for Twitter. The location is a rectangle whose first two coordinates (longitude and latitude) are the bottom left corner and the last two are the top right corner.

Therefore, I collected the extreme points of Italy, Poland and Japan.

In this way, I think we could collect tweets that meet the language, location and keywords conditions at the same time.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Italy | Poland | Japan |
| Northernmost | [47.083333, 12.183333] | [54.833333, 18.066667] | [148.4508, 45.3326] |
| Southernmost | [35.483333, 12.6] | [49.00238,  22.8471] | [136.0411, 20.2531] |
| Easternmost | [45.1, 6.616667] | [50.86852, 24.14585] | [122.5557, 24.2705] |
| Westernmost | [40.1, 18.516667] | [52.83827, 14.12298] | [153.5912, 24.1659] |

**In order to get suitable data for the project, I’ll start the collection using 1 in the next week and then select tweets by their location. At the same time, continue to fix my code in 2.**