

MANUAL

Rumor Tracking Tools

COVID-19 GLOBAL RESPONSE

Corina Markodimitraki, Melissa El Hamouch and Georgios Roullis

2021.05.20 – Version 1

With support of



For more information regarding the use of this product, please contact the 510 Data Team by sending an email to support@510.global

Contents

- Introduction 1
 - Product in a nutshell..... 1
 - What are the rumor tracking tools and what can you do with them?..... 1
 - Who are rumor tracking tools for? 2
 - Why use rumor tracking? 2
- Case study 3
- Data responsibility..... 4
 - Datasets 4
 - Data processing 4
 - Non-discrimination 4
 - Human oversight 4
 - Risks 5
- Requirements 6
 - User knowledge 6
 - To collect and analyze Twitter data the user will need the following skills: 6
 - To collect data through KoBo, the user will need the following skills: 6
 - Hardware 7
 - For rumor tracking through Twitter:..... 7
 - For rumor tracking with KoBo:..... 8
 - Software..... 8
 - For rumor tracking through Twitter..... 8
 - For rumor tracking through KoBo 9
 - Time 10
 - Rumor tracking through Twitter 10
 - Rumor tracking through KoBo 10
 - Cost 10
- The product..... 12
 - What do rumor tracking tools consist of? 12
 - Rumor tracking through Twitter 12

Rumor tracking through KoBo	12
Locations to download the product	12
Rumor tracking through Twitter	12
Rumor tracking through KoBo	13
Advantages & Limitations	13
Rumor tracking through Twitter	13
Rumor tracking through KoBo	13
Setting up rumor tracking: a step-by-step guide.....	15
Rumor tracking through Twitter	15
Rumor tracking through KoBo.....	21
Abbreviations.....	25
Resources.....	25

Introduction

Since March 2020, the 510 Data Team of the Netherlands Red Cross (NLRC) has supported different National Societies (NS) by providing a framework that supports tracking of rumors spread through Twitter or word of mouth communication. This manual provides the information that one needs to know when planning on using this framework to track rumors.

Product in a nutshell

The framework presented in this manual consists of data collection platforms and programming scripts that support the collection and analysis of rumors spread by the general population of the country of interest, either through Twitter or word of mouth communication.

What are the rumor tracking tools and what can you do with them?

The 510 Data Team uses rumor tracking tools to support rumor data collection and analysis for NS across the globe. This is done in either of the following ways:

1. By collecting rumors from Twitter
2. By collecting word-of-mouth rumors with Kobo forms

To collect rumors on Twitter, the 510 Data Team uses Twitter queries, Python scripts and Google API's (Application Programming Interface) to download the rumor data from Twitter for the country of interest. These are then analyzed for sentiment (positive or negative connotation of the Tweets) and split on topic. The data can be visualized with a dashboard built on a platform such as PowerBI. For more information on data visualization, please refer to the manual "Dashboards".

To collect word-of-mouth rumors, the 510 Data Team uses KoBo forms that contain structured questions, allowing categorization of data on the spot. Until now, data has been collected on different instances, e.g., during a sensibilization campaign by volunteers of the NS. Word-of-mouth data can also be visualized with a dashboard built on a platform such as PowerBI or Google Data Studio.

Some examples of the collected rumor data until now are:

- Rumors about vaccines against COVID-19
- Questions towards the NS

- Suggestions towards the NS

Who are rumor tracking tools for?

Rumor tracking tools can be used by any NS or other non-governmental organisation looking to understand the opinions circulating amongst the country's general population. This can help with decision making e.g., targeting campaigns to inform the population about COVID-19 vaccination programs. Examples of National Societies that have used rumor tracking until now are:

- The Philippine Red Cross
- The Mali Red Cross
- The Uganda Red Cross Society

Why use rumor tracking?

The Twitter rumor tracking tool presented in this manual offers an open source, fast and automated way of tracking rumors in a country or area of interest. Due to the nature of the Twitter platform, the user can collect vast amounts of data, and repeat the process of collection and analysis as often as they desire once the pipeline has been set up.

The KoBo rumor collection on the other hand, offers an open source, automated and accessible way of tracking rumors, even in areas with low to no internet connection or phone coverage. Due to the manual aspect of the data collection, this manner of tracking rumors often offers more precise and granular geographical information, which allows very targeted decision-making down the line. These are considerable advantages for NS that often work in the field, and save the NS from time-consuming manual work, thereby increasing efficiency. The flipside to this is that the amount of the collected data can be limited compared to the Twitter data and that the collection process is more time-consuming.

All in all, the rumor tracking tools offer the user a clear overview of the rumors and this can accelerate decision-making processes. For example, a considerable amount of rumors showing mistrust in vaccines against SARS-CoV-2, the virus responsible for COVID-19, allows the NS to design a targeted campaign informing the general population of the benefits of vaccination, with scientifically-based information.

Case study

Rumor Tracking Tools are a great way for NS to understand the opinions and thoughts of the general population. This tool is particularly integrated in the Community Engagement and Accountability (CEA) work in an integral program in NS. CEA works directly with the community and is important in understanding the population's perception on specific topics.

In the case of the Philippines, the general population were hesitant about undergoing vaccination, which in turn led to outbreaks in the country. With the rollout of several vaccination programs as well as the COVID-19 vaccine coming to market, the CEA unit at the Philippine Red Cross needed to know what the population thinks. For this particular case, Twitter Rumor tracking was used, as it was a more efficient way for the NS to gather information, especially given the social distancing regulations. Based on the data gathered, the NS' CEA department was able to address the misinformation regarding vaccines at a national level, but also at a geo-specific and community level. Moreover, the tool is tangible and can be adapted to fit any particular program and context, making it a sustainable community engagement method.

To read more about how rumor tracking was implemented in the Uganda and in Jordan please read our blog post:

<https://www.510.global/rumor-tracking-a-digital-transformation-approach/>

Data responsibility

Datasets

The data used for the rumor tracking tools, could include personal information, such as gender, location, and identifier/name of data subject. The data, however, can be minimized and anonymized to protect the rights of data subjects. In the case of the Twitter rumor tracking, Twitter handles of users appear in the data collection script but are not used later on.

Data processing

In the case of the Twitter rumor tracking: the programming script collects data by searching for key words, translates the tweets, adds geolocation whenever possible, removes emojis and other markings and classifies them based on topic and sentiment. After that, the data can be visualized, for example by showing positive or negative tweets, date and location of the tweet, and frequency of tweet.

In the case of the KoBo rumor tracking: The data is collected manually, but can then automatically be linked to a visualization dashboard. Based on what information is collected and the indicators present in the dashboard, different categories could also be formed.

Non-discrimination

The tools are non-discriminatory. However, the data cannot always be generalized for the entire population if for example Twitter is not a nation-wide used platform.

Human oversight

Human involvement occurs at the data collection and analysis phase. For data collection, when using a KoBo form, data collectors need to abide by data responsibility guidelines and protect the rights of the data subjects, ensuring that no harm is done. When using a programming script, the developer will need to assure that personal information or identifiers are not included in the output. For data analysis, the end users will need to prepare specific campaigns and programs that will address the rumors obtained.

Risks

Depending on the tool and data used, it is important to notice if the information can or cannot be generalized as the overall population's opinion. When addressing the rumors, at a national, sub-national or even community level, the end user needs to be careful on how the misinformation is addressed, without censure.

Requirements

User knowledge

In this section we describe the skills that someone would need to use the tools presented in this manual. In addition to the skills mentioned below, it is crucial the user has the following skills:

- Analytical thought
- Problem-solving attitude
- Data responsibility (see also section “Data responsibility” of this manual)
- Visualization skills (optional)

To collect and analyze Twitter data the user will need the following skills:

- **Computer skills:** the user will need to have advanced computer skills to use and browse the internet, access databases and use spreadsheets.
- **Twitter:** the user will need to be familiar with Twitter and know how to use its functions such as searching for keywords.
- **APIs:** the user will need to know how to use Twitter and Google APIs (Application Programming Interface).
- **Programming language:** that the user will need to have knowledge and experience with a programming language (such as Python) to use the API.
- **JSON:** the user will need to have experience with using json files.
- **PowerBI (optional):** in case the user wants to visualize the collected data, they will need to have knowledge of PowerBI (or any other dashboard-supporting program) to visualize the data.

To collect data through KoBo, the user will need the following skills:

- **Computer skills:** the user will need to know how to use a computer, a mobile phone or tablet, an internet browser and optionally, any spreadsheet-based program such as Google sheets or Microsoft Excel.

- **PowerBI (optional):** in case the user wants to visualize the collected data, they will need to have knowledge of PowerBI dashboards (or any other dashboard-supporting program) to visualize the data.
- **Mailchimp (optional):** in case the user wants to automate the process of sending the form to the target audience, they can link Kobo to Mailchimp. Therefore, experience with the Mailchimp platform is desirable.

Hardware

Below you can find the hardware requirements for each of the tools mentioned in this manual.

For rumor tracking through Twitter:

- **Python** (versions: 2.7.X, 3.6.X):
 - Processor: Intel Atom® processor or Intel® Core™ i3 processor
 - Disk space: 1 GB
 - Operating systems: Windows* 7 or later, macOS, or Linux
- **Excel** Microsoft Office (2019):
 - Processor: minimum Windows OS: 1.6 GHz, 2-core / macOS: Intel processor
 - RAM: Windows OS: 4 GB RAM; 2 GB RAM (32-bit) / macOS: 4 GB RAM
 - Disk Space: Windows OS: 4 GB available disk space / macOS: 10 GB of available disk space
 - Screen resolution: Windows OS: 1280 x 768 (32-bit requires hardware acceleration for 4K and higher) / macOS: 1280 x 800
 - Graphics: Windows OS: DirectX 9 or later, with WDDM 2.0 or higher for Windows 10 (or WDDM 1.3 or higher for Windows 10 Fall Creators Update) / macOS: No graphics requirements.
- **PowerBI Desktop:**
 - Processor: 1 GHz 64-bit (x64) recommended.
 - RAM: At least 1 GB available, 1.5 GB or more recommended.
 - Operating systems: Windows 7 / Windows Server 2008 R2, or later
 - Display: At least 1440x900 or 1600x900 (16:9) required. Lower resolutions such as 1024x768 or 1280x800 aren't supported, as certain controls (such as closing the startup screen) display beyond those resolutions.
 - Windows display settings: If you set your display settings to change the size of text, apps, and other items to more than 100%, you may not be able to see certain dialogs that you must interact with to continue using Power BI Desktop. If you encounter this

issue, check your display settings in Windows by going to Settings > System > Display, and use the slider to return display settings to 100%.

For rumor tracking with KoBo:

- **To create the form:** a computer with internet connection and an internet browser. Alternatively, the user could build the form offline with a spreadsheet program such as Excel (Microsoft Office, hardware requirements below) but they would need to have internet connection once the form is designed to upload it to the Kobo server. If the user wishes to automatize the process of sending the form to the target audience, they need to have internet connection.
- **To fill in the form:** a mobile phone, tablet, or laptop/personal computer. No internet connection is necessary.
- **To gather the data from the form:** a computer with internet connection to upload the collected data to the server.
- **To analyze the data from the form:** Python (see above).
- **To visualize the data:** PowerBI Desktop (see above)

Software

Below we list the software required for each of the tools in this manual.

For rumor tracking through Twitter

To collect, analyze and visualize the data

- Internet browser (for computers) such as
 - Mozilla Firefox
www.mozilla.org/en-US/firefox/new/
 - Google Chrome
www.google.com/chrome/
- Access to Twitter
<https://twitter.com/?lang=en>
- Spreadsheet-based program such as
 - Excel (Microsoft Office)
www.microsoft.com/en-us/microsoft-365/excel
- Installed Python
<https://www.python.org/>

- (Optional) PowerBI or any other dashboard-supporting program
<https://www.microsoft.com/en-us/download/details.aspx?id=58494>

For rumor tracking through KoBo

To create the form

- Internet browser as
 - Mozilla Firefox
www.mozilla.org/en-US/firefox/new/
 - Google Chrome
www.google.com/chrome/
- (Optional) Spreadsheet-based program such as
 - Excel (Microsoft Office)
www.microsoft.com/en-us/microsoft-365/excel
 - Calc (Apache OpenOffice)
www.openoffice.org/download/index.html
 - Google Sheets (used in an internet browser)
www.google.com/sheets/about/
- (Optional) Mailchimp platform
<https://mailchimp.com/>

To fill in the form:

- The KoBo app
<https://play.google.com/store/apps/details?id=org.koboc.collect.android>

To gather the data from the form, analyze and visualize them (optional):

- Internet browser
- (Optional) Spreadsheet-based program
- Installed Python
<https://www.python.org/>
- (Visualisation - optional) PowerBI
<https://www.microsoft.com/en-us/download/details.aspx?id=58494>
- (Visualisation - optional) Google Data Studio
<https://datastudio.google.com/>

- (Visualisation - optional) ArcGIS
<https://pro.arcgis.com/en/pro-app/latest/get-started/download-arcgis-pro.htm>

Time

Here we indicate which parameters influence the duration of usage of the tools and state real-life examples of usage by NS.

Rumor tracking through Twitter

The duration of usage is the total duration of downloading the Twitter data, preprocessing the data, adding the geolocation, analysis of the data for sentiment and topics and optionally, visualizing the data. All this depends on user experience with an internet browser, APIs and programming skills in Python.

For reference, the 510 Data Team of the NLRC needed about 24 hrs to collect and analyze the Twitter data for the Philippine Red Cross. To estimate the amount of time that is needed for the visualization of the data please refer to the 510 manual “Dashboards”.

Rumor tracking through KoBo

The duration of usage is the total duration of creating, testing and filling in the form by the target audience. All this depends on user experience with an internet browser, the amount of data that needs to be collected (and therefore the number of questions that need to be added to the form), the number of persons that need to fill out the form and the time they take to do so.

For reference, the 510 data team of the NLRC department needed about 24 hrs to build the form for Uganda Red Cross. To estimate the amount of time that is needed for the visualization of the data please refer to the 510 manual “Dashboards”.

Cost

Here we indicate which parameters influence the costs of usage of the rumor tracking tools.

The usage of Google Translation API is tied to a price per month the amount of which depends on the number of characters needing to be translated. The first 500,000 characters are free of charge. Other prices start at 20 USD/million characters. For more information, please refer to <https://cloud.google.com/translate/pricing>.

The usage of Google Natural Language API for sentiment analysis is tied to a price per month the amount of which depends on the number of characters needing to be analysed. The first 5,000 units (1 unit = 1,000 Unicode characters or 1 request to the API) are free of charge. Other prices start at 1 USD/month for 5,000-1,000,000 units. For more information, please refer to <https://cloud.google.com/natural-language/pricing>.

We would like to note that Google Cloud grants new users 300 USD free credit for the first 90 days.

The product

What do rumor tracking tools consist of?

Rumor tracking through Twitter

Rumor tracking through Twitter can be achieved by using all of the following tools:

- Python scripts & libraries
- Twitter and Google APIs

Rumor tracking through KoBo

Rumor tracking through a Kobo form can be achieved with:

- An online form made through the website of Kobo
- An Excel spreadsheet. To deploy the form, the Excel sheet has to be uploaded to the Kobo website first.
- The KoBoCollect app to use on smartphone/tablet to fill in the form.
- Python scripts & libraries for analysis

Locations to download the product

Here we provide the links and locations to find the tools needed to set up and use rumor tracking.

Rumor tracking through Twitter

- Github page containing the Python scripts and information needed to run these
<https://github.com/rodekruis/rumor-tracker-philippines-covid>
- Python package to clean up the tweets
<https://pypi.org/project/tweet-preprocessor/>
- Model that categorizes tweets into topics (GSDMM: short text clustering)
<https://github.com/rwalk/gsdmm>

Rumor tracking through KoBo

- KoBo Toolbox
www.kobotoolbox.org
- The KoBoCollect app
<https://play.google.com/store/apps/details?id=org.koboc.collect.android>
- (Optional) Github page containing the Python scripts and information needed to run these
<https://github.com/rodekruijs/rumor-tracker-uganda-covid>
- (Optional) Model that categorizes tweets into topics (short text clustering)
<https://github.com/rwalk/gsdmm>

Advantages & Limitations

There are several advantages and limitations linked to each of the strategies for rumor tracking. Here we briefly discuss these.

Rumor tracking through Twitter

- Twitter data can be biased: only people with internet connection and a smartphone and/or computer have access.
- The demographics of Twitter users are not always available.
- Exact location of the Twitter user is not always available.
- Twitter provides large amounts of data at a short time
- The pipeline described in this manual requires expertise in Python to setup, although less to run once it is set up.

Rumor tracking through KoBo

- Rumor tracking through KoBo can be less biased than Twitter, as almost anyone can be approached to fill in the questionnaire.
- Demographics information is easier to collect, as this can be incorporated as questions in the form (age, gender, exact geolocation, religion etc).
- It provides exact geolocation if questions about the location are incorporated into the form.

- It is more time-consuming and provides less data compared to rumor tracking through Twitter.
- Rumor tracking through KoBo requires less expertise than the rumor tracking through Twitter if topics are not analyzed with a Python script.

Setting up rumor tracking: a step-by-step guide

Rumor tracking through Twitter

1. Search for Tweets using a specified query

For this step, the Python script `get_past_tweets.py` is needed, found on the Github page mentioned in the section “The product”. By using the `get_past_tweets.py` script:

- a. Initialize a Twitter API and build a query. You can specify the location, language, as well as other parameters. The page below offers more information about performing a standard search API <https://developer.twitter.com/en/docs/twitter-api/v1/tweets/search/api-reference/get-search-tweets>
- b. Prompt Twitter to return the results in json format. This file will contain information such as the tweet text, the date of creation etc. Part of an example dataset concerning tweets about COVID-19 vaccines in the Philippines is shown in **Figure 1**. This dataset will be shown in all the subsequent steps of the “Rumor tracking through Twitter” section of the step-by-step guide. The json file contains the data (Tweets) that will be used from this step onward in the analysis.

```
{
  "created_at": "Wed Feb 03 07:31:25 +0000 2021",
  "id": "1356867876183891971",
  "id_str": "1356867876183891971",
  "full_text": "Metro mayors want to buy other vaccine brands once authorized
https://t.co/XfoMvmRgSO",
  "truncated": false,
  "display_text_range": [
    0,
    85
  ],
  "entities": {
    "hashtags": [
    ],
    "symbols": [
    ],
    "user_mentions": [
    ],
    "urls": [
      {
        "url": "https://t.co/XfoMvmRgSO",
        "expanded_url": "https://mb.com.ph/2021/02/03/metro-mayors-want-to-buy-
other-vaccine-brands-once-authorized/",
        "display_url": "mb.com.ph/2021/02/03/met\u2026",
        "indices": [

```

Figure 1 • Part of the collected Twitter data within the json file, opened in <http://json.parser.online.fr/>

2. Preprocess the data and add a geolocation

For this step, the Python script `parse_geolocate_tweets.py` is needed, found on the Github page mentioned in the section “The product”. First, the data contains a lot of information such as emojis, hashtags etc, which are not necessary for the analyses and need to be removed. What’s more, some tweets contain a geolocation by default and others don’t, depending on the Twitter user settings. About 20% of tweets contain a geolocation by default or mention one in the tweet text. To understand and/or visualize tweets in a geographical context, only tweets with an attached geolocation can be used (example visualization in **figure 2, left panel**). To understand the general trend of the tweets in the region specified in step 1, all tweets regardless of the presence of a geolocation can be included (example visualization in **figure 2, right panel**). By using the `parse_geolocate_tweets.py` script:

- a. Clean the data from emojis, hashtags, URLs etc., by using the “tweet-preprocessor 0.6.0” Python library (<https://pypi.org/project/tweet-preprocessor/>). An example result of the cleaned data is shown in **figure 3**.
- b. Assign a geolocation to tweets that do not contain one by default. For this, the tweet text is searched for mentions of a location. If a location is found to be mentioned in the tweet text, it will be assigned as the geolocation of the tweet. The script will add a column to the csv file containing the geolocation. An example of the added column containing the geolocation of the tweets is shown in **figure 4**.

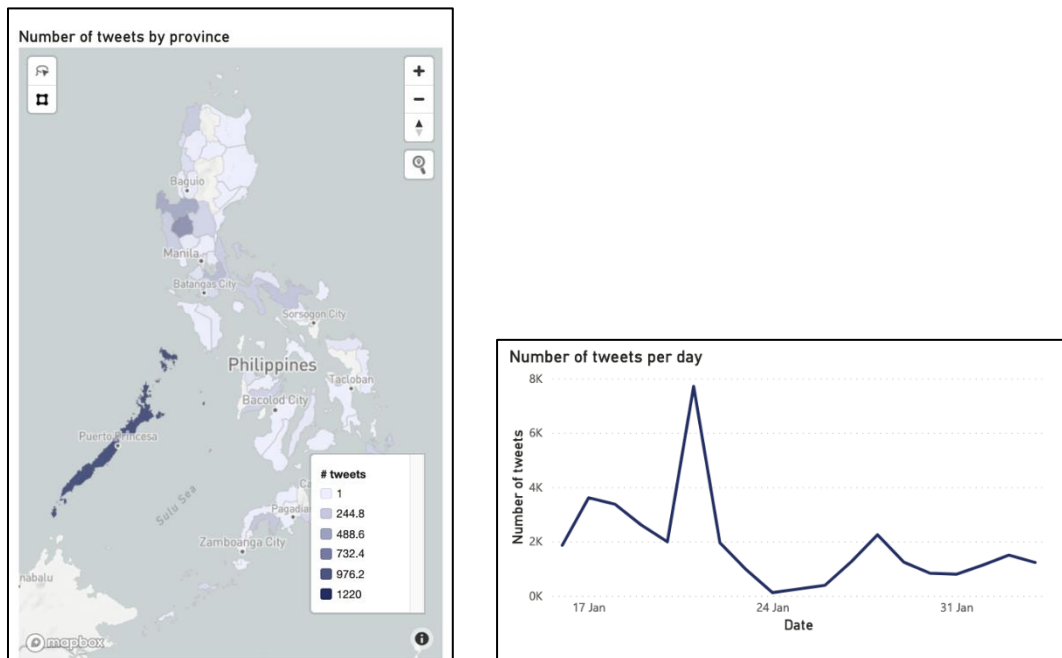


Figure 2 • Examples of visualization of Twitter data for the Philippines. The left panel shows the number of tweets by province on a map, which can only be done if there is a geolocation attached to the tweet. The right panel shows the number of tweets without specifying the exact geolocation. Data is visualized with PowerBI.

	X	created_at	id	id_str	full_text	truncated	full_text_clean
77	76	2021-02-02 12:51:58+00:00	1.356586e+18	1.356586e+18	RT @smnnews: Nangako ang China na magdo-donat...	False	: nangako ang china na magdo-donate ito
78	77	2021-02-02 12:39:34+00:00	1.356583e+18	1.356583e+18	PANORIN Ano nga ba ang vaccine o bakuna, at paan...	False	panoorinano nga ba ang vaccine o bakuna,
79	78	2021-02-02 12:25:09+00:00	1.356579e+18	1.356579e+18	Nangako ang China na magdo-donate ito ng 500K do...	False	nangako ang china na magdo-donate ito n
80	79	2021-02-02 08:48:49+00:00	1.356525e+18	1.356525e+18	Hindi handang LGU 'di bibigyan ng bakuna - vaccine ...	False	hindi handang lgu 'di bibigyan ng bakuna -
81	80	2021-02-02 08:21:33+00:00	1.356518e+18	1.356518e+18	RT @juanelyas: Ang daming kuda ni Thinking Pinoy at...	False	: ang daming kuda ni thinking pinoy at sas
82	81	2021-02-02 08:01:33+00:00	1.356513e+18	1.356513e+18	Naghihintay na lamang ng emergency use authorizati...	False	naghihintay na lamang ng emergency use a
83	82	2021-02-02 06:43:12+00:00	1.356493e+18	1.356493e+18	Naniniwala si Pangulong Duterte na iniipit umano ng ...	False	naniniwala si pangulong duterte na iniipit u
84	83	2021-02-02 06:42:29+00:00	1.356493e+18	1.356493e+18	Sinabi ni Health Undersecretary at spokesperson Mari...	False	sinabi ni health undersecretary at spokespe
85	84	2021-02-02 06:29:01+00:00	1.356490e+18	1.356490e+18	RT @dzb: #BantayBakuna: Pagbabakuna ng COVID-1...	False	: pagbabakuna ng covid-19 vaccine kay p
86	85	2021-02-02 04:57:06+00:00	1.356467e+18	1.356467e+18	Once may bakuna na daw, okey na. Tapos ngayong m...	False	once may bakuna na daw, okey na. tapos n
87	86	2021-02-02 04:41:22+00:00	1.356463e+18	1.356463e+18	Vaccine czar Galvez assures that the vaccine distribut...	False	vaccine czar galvez assures that the vaccin
88	87	2021-02-02 03:43:44+00:00	1.356448e+18	1.356448e+18	Inaasahang darating sa bansa ngayong Pebrero ang h...	False	inaasahang darating sa bansa ngayong peb
89	88	2021-02-02 03:32:02+00:00	1.356445e+18	1.356445e+18	Kinopo ng European Union ang COVID-19 vaccine na ...	False	kinopo ng european union ang covid-19 va
90	89	2021-02-02 03:28:00+00:00	1.356444e+18	1.356444e+18	Vaccine czar Galvez backs Finance Chief Carlos Domi...	False	vaccine czar galvez backs finance chief carl
91	90	2021-02-02 03:19:21+00:00	1.356442e+18	1.356442e+18	Binigyang-linaw ni Pangulong Rodrigo Duterte ang m...	False	binigyang-linaw ni pangulong rodrigo dute
92	91	2021-02-02 02:41:05+00:00	1.356432e+18	1.356432e+18	Olarte sa COVID-19 vaccine: But merong mga sinasab...	False	olarte sa covid-19 vaccine: but merong mg
93	92	2021-02-02 02:40:28+00:00	1.356432e+18	1.356432e+18	Olarte sa COVID-19 vaccine: Generally, ang mga baku...	False	olarte sa covid-19 vaccine: generally, ang r
94	93	2021-02-02 01:42:53+00:00	1.356418e+18	1.356418e+18	RT @News5PH: #TedFailonandDJChaCha Inakusahan...	False	: inakusahan ni pres. duterte ang europe
95	94	2021-02-02 01:29:00+00:00	1.356414e+18	1.356414e+18	#TedFailonandDJChaCha Inakusahan ni Pres. Dutert...	False	inakusahan ni pres. duterte ang european
96	95	2021-02-01 20:34:42+00:00	1.356340e+18	1.356340e+18	RT @SyLicoNgaAko: Matindi talaga ang pag-aahente ...	False	: matindi talaga ang pag-aahente ng rehim
97	96	2021-02-01 17:52:06+00:00	1.356299e+18	1.356299e+18	RT @SyLicoNgaAko: Matindi talaga ang pag-aahente ...	False	: matindi talaga ang pag-aahente ng rehim
98	97	2021-02-01 14:07:11+00:00	1.356243e+18	1.356243e+18	RT @aileentaliping: Pres.Duterte to critics on covid va...	False	: pres.duterte to critics on covid vaccine : p
99	98	2021-02-01 14:05:09+00:00	1.356242e+18	1.356242e+18	Pres.Duterte to critics on covid vaccine : patapusin ni...	False	pres.duterte to critics on covid vaccine : pa
100	99	2021-02-01 13:18:15+00:00	1.356230e+18	1.356230e+18	RT @SyLicoNgaAko: Matindi talaga ang pag-aahente ...	False	: matindi talaga ang pag-aahente ng rehim
101	100	2021-02-01 12:23:34+00:00	1.356217e+18	1.356217e+18	Sinabi ni Health Undersecretary Maria Rosario Vergeir...	False	sinabi ni health undersecretary maria rosar
102	101	2021-02-01 11:12:25+00:00	1.356199e+18	1.356199e+18	RT @SyLicoNgaAko: Matindi talaga ang pag-aahente ...	False	: matindi talaga ang pag-aahente ng rehim

Figure 3 • Part of the csv file generated with the parse_geolocate_tweets.py script. The right panel shows a column containing the cleaned text. *n* indicates the number of columns between the columns shown in the left panel, and the one shown in the right panel. The example data is opened in R Studio.

coord
POINT (123.2948675780001 13.58852384600004)
POINT (119.9769508930001 16.29121255600006)
POINT (120.0365199700001 15.36772434900007)
POINT (123.2948675780001 13.58852384600004)
POINT (120.0365199700001 15.36772434900007)
POINT (123.634337393 9.189841767000075)
POINT (122.70966018 7.841985388000069)
POINT (119.9769508930001 16.29121255600006)
POINT (119.9769508930001 16.29121255600006)

Figure 4 • Examples of added geographical information to the tweets. The screenshot only shows the column that was added to dataset after assigning the geolocation of the tweets. The example data is opened in R Studio.

3. Analyze the data for sentiment

For this step, the Python script `get_sentiment_tweets.py` is needed, found on the Github page mentioned in the section “The product”. Analysis of sentiment of the collected tweets can identify the prevailing emotional opinion within the text tweet. It can determine the tweet’s writers’ attitude as positive, neutral or negative. This gives an understanding of the general trend of the tweets. By using the `get_sentiment_tweets.py` script:

- a. Translate the data to English (if applicable) by using the Google Translation API (<https://cloud.google.com/translate>). The API will add a column to the csv file containing the English translation.
- b. Analyze the sentiment of the tweets by using the Google Natural Language API (<https://cloud.google.com/natural-language/docs/analyzing-sentiment#language-sentiment-file-python>). The API will add two columns to the csv file containing the sentiment score, which indicates the overall emotion of the tweet, as well as the magnitude of the sentiment which indicates the amount of emotional content present in the tweet. Using the example data shown in the previous figures, we demonstrate the output of this step in **figure 5**. Further explanation on how to interpret the results of the sentiment analysis can be found in the section “Interpreting sentiment analysis” of the link mentioned above.

Note that usage of the translation API and the sentiment analysis API by Google are tied to user costs. For more information, please refer to the section “Costs” of this manual.

sentiment_score	sentiment_magnitude
0.5	0.5
-0.1	0.4
-0.5	1.0
-0.5	1.0
-0.5	1.0
-0.5	1.0
-0.5	1.0
-0.5	1.0
0.0	0.0
-0.1	1.0
-0.5	1.0
0.0	0.1
0.0	0.2
-0.5	1.0

Figure 5 • Examples of sentiment score and sentiment magnitude added to the tweets. The screenshot only shows the column that was added to the csv file after assigning the sentiment. The file is opened in R Studio.

4. Cluster the data into topics and identify the prevailing themes

For this step, the Python script `topic_modelling.py` is needed, found on the Github page mentioned in the section “The product”. Grouping the tweets based on their content can identify the prevailing themes present in the dataset. This gives an understanding of the general trend of the tweets, and thus, this specific population. By using the `topic_modelling.py` script:

- a. Cluster the tweets based on their content by using the “GSDMM: Short text clustering” Python library (<https://github.com/rwalk/gsdmm>) which is based on a short text clustering model¹. The script will group tweets with a similar theme together, creating theme clusters and will output two extra columns to the csv file. One will contain the theme cluster number to which each tweet was assigned, and the other will contain the confidence score, indicating the certainty of the model about the classification of the tweet.
- b. Manually identify the themes present in the dataset. For this, we suggest investigating each cluster individually. In particular, the 10 tweets with the highest confidence score for each cluster can reveal an overarching theme, which can be assigned to the cluster. With the `topic_modelling.py` script, one can assign a theme to each tweet. The output csv file should have substituted the numbers of the cluster with the topic name of each tweet. This file is also the final product of the rumor tracking analysis (**figure 6**).



topic
Opinions of politicians over COVID-19 vaccine rollout
Government and/or others are profiting from the situ...
Vaccine is ineffective or even dangerous
Vaccine is ineffective or even dangerous
Vaccine is ineffective or even dangerous
Vaccine is ineffective or even dangerous
Vaccine is ineffective or even dangerous
Vaccine is ineffective or even dangerous
Vaccine is ineffective or even dangerous
Vaccine is ineffective or even dangerous
Vaccine is ineffective or even dangerous
Opinions of politicians over COVID-19 vaccine rollout
Government and/or others are profiting from the situ...
Vaccine is ineffective or even dangerous
Unclear / unknown
Vaccine is ineffective or even dangerous

Figure 6 • Examples of assigned topic names to the tweets. The screenshot only shows the column that was added to the csv file after assigning the topic.

¹ Yin, J., & Wang, J. (2014). A dirichlet multinomial mixture model-based approach for short text clustering. Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. DOI:10.1145/2623330.2623715

5. Visualize the results

The results of the rumor tracking analysis can be visualized in a number of ways. A visualization by means of a dashboard can provide an interactive and accessible interface for others to interpret the data. In **figure 7**, we present an example visualization of the sample data, which has been produced with the program PowerBI. For more information on data visualization through dashboards, please refer to the manual “Dashboards”.

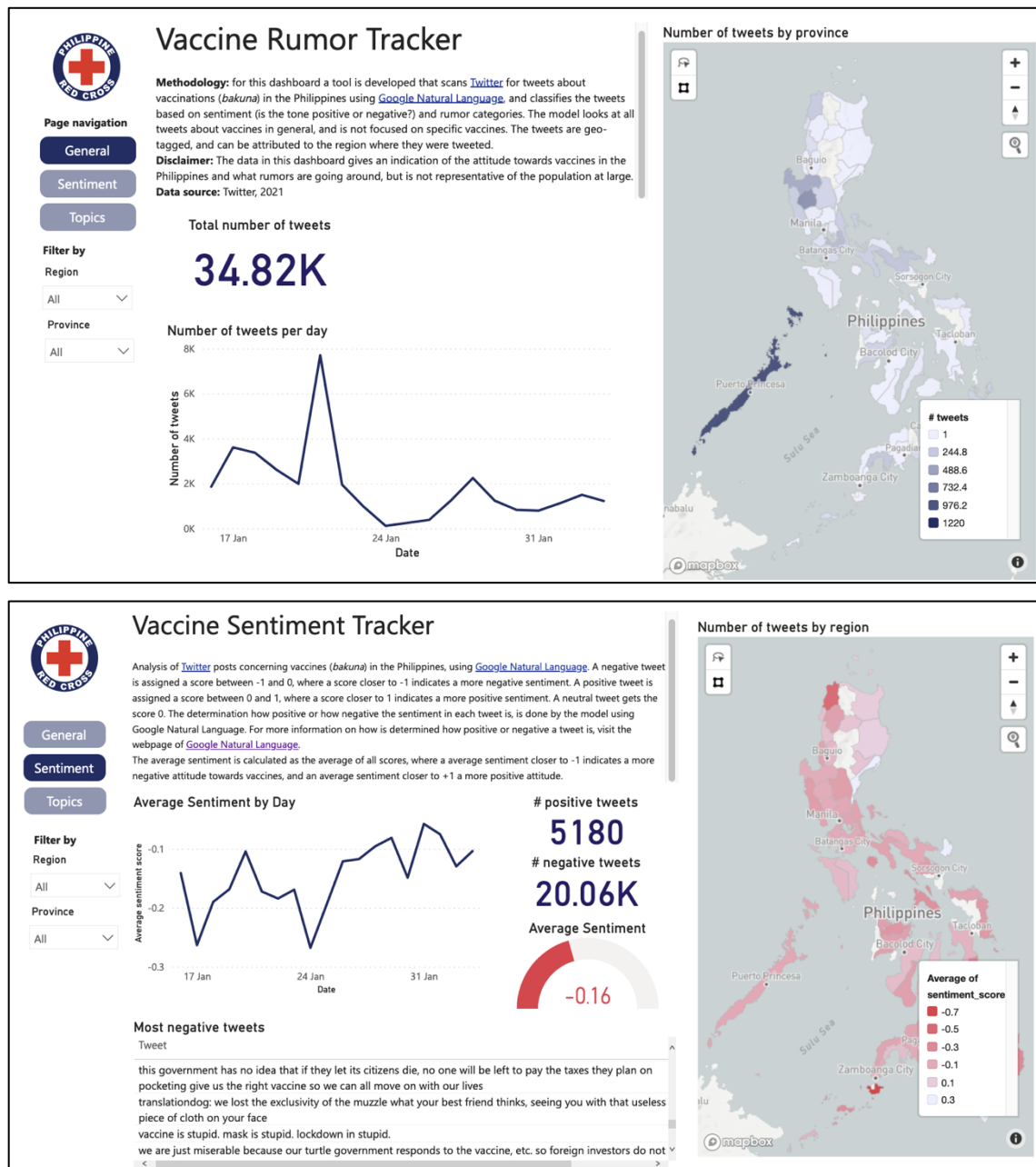


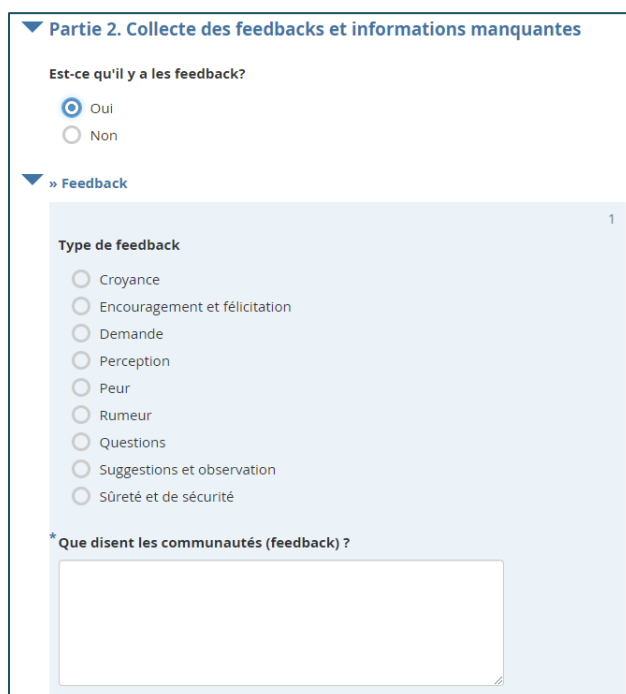
Figure 7 • Example of rumor tracking data visualization in PowerBI. The shown data are tweets collected from Twitter users in the Philippines.

Rumor tracking through KoBo

Additionally, several websites offer instructions on how to use Kobo, see section “Resources” of this manual. To avoid replicating already available information, the present step-by-step guide for rumor tracking through KoBo will be brief.

1. Create a KoBo form

Create a form that will be used to collect the information from the general population. It is crucial that the questions in the form are well thought of and formulated well. A co-design session between the NS and staff with KoBo expertise is recommended. It is of importance the questions will gather maximum information for evidence-based decisions down the road. The form can be created on the KoBoToolbox website <http://www.kobotoolbox.org/>. KoBo also offers the option of creating a form template in a spreadsheet-based program such as Excel (Microsoft Office). **Figure 8** shows an example of a KoBo form as made through the website (in French).



▼ **Partie 2. Collecte des feedbacks et informations manquantes**

Est-ce qu'il y a les feedback?

Oui
 Non

▼ **» Feedback**

Type de feedback

Croyance
 Encouragement et félicitation
 Demande
 Perception
 Peur
 Rumeur
 Questions
 Suggestions et observation
 Sécurité et de sécurité

* **Que disent les communautés (feedback) ?**

Figure 8 • Example of a data collection form made with KoBo (in French). The form is shown on the KoBo website.

2. Test the form

Make sure the questions that are stated in the form are clear by asking a test group to fill in the questionnaire. If results show ambiguous or off-topic answers, the questions will need to be adjusted. If necessary, repeat a test round after revising the form.

3. Install the KoBoCollect app

If planning on collecting data in the field with low to no internet connection, we recommend installing the KoBoCollect app on the smartphone or tablet that will be used (<https://play.google.com/store/apps/details?id=org.koboc.collect.android&hl=en>). The gathered data from the filled-in forms will be uploaded to the KoBo server once the smartphone or tablet are reconnected to the internet.

4. Collect information with the form

Address the questions in the form to the persons of interest to collect the data. This can be done through the KoBo website or the KoBoCollect app. **Figure 9** showcases collected data of a KoBo form, viewed in Excel. The strategy of collection depends on the target population, meaning data can be collected by visiting villages specific villages, specific parts of a city, by visiting markets etc.

	CY	CZ	DA	DB	DC
1	What top three rumours or misconceptions are circulating in the community?	What top three rumours or misconceptions are circulating in the community?/Beliefs (Black people can not get Covid-19, Covid-19 is for rich people)	What top three rumours or misconceptions are circulating in the community?/Preventive Measure (Covid-19 can be prevented by taking beer, ginger, lemon and other)	What top three rumours or misconceptions are circulating in the community?/Perceptions (COVID-19 is not real)	What top three rumours or misconceptions are circulating in the community?/Other
2					
3				1	
4	Preventive Measure (Covid-19 can be prevented by taking beer, ginger, lemon and other)	0	1	0	0
5				1	
6	Beliefs (Black people can not get Covid-19, Covid-19 is for white people)	0	0	1	1
7	Preventive Measure (Covid-19 can be prevented by taking beer, ginger, lemon and other)	0	0	1	1
8		1		1	
9	Perceptions (COVID-19 is not real) Preventive Measure (Covid-19 can be prevented by taking beer, ginger, lemon and other)		1	1	1
10	Preventive Measure (Covid-19 can be prevented by taking beer, ginger, lemon and other) Perceptions (COVID-19 is not real) Other (Specify)		1	1	0
11	Perceptions (COVID-19 is not real) Other (Specify) Beliefs (Black people can not get Covid-19, Covid-19 is for white people)		0	1	1
12	Beliefs (Black people can not get Covid-19, Covid-19 is for white people)	0	0	0	1
13	Preventive Measure (Covid-19 can be prevented by taking beer, ginger, lemon and other) Perceptions (COVID-19 is not real)	1	1	1	1
14	Beliefs (Black people can not get Covid-19, Covid-19 is for white people)	1	0	1	0
15	Preventive Measure (Covid-19 can be prevented by taking beer, ginger, lemon and other)	0	1	0	0
16		1	1	1	1
17	Perceptions (COVID-19 is not real) Preventive Measure (Covid-19 can be prevented by taking beer, ginger, lemon and other)		1	1	0
18			1	1	1
19				1	
20	Beliefs (Black people can not get Covid-19, Covid-19 is for white people) Perceptions (COVID-19 is not real)	0	1	1	0
21	Preventive Measure (Covid-19 can be prevented by taking beer, ginger, lemon and other) Perceptions (COVID-19 is not real)		1	1	0
22	Perceptions (COVID-19 is not real) Preventive Measure (Covid-19 can be prevented by taking beer, ginger, lemon and other)		1	1	0
23	Preventive Measure (Covid-19 can be prevented by taking beer, ginger, lemon and other) Perceptions (COVID-19 is not real)		1	1	0
24					
25					
26	Other (Specify)	0	0	0	1

Figure 9 • Example of data collected with KoBo, opened as a spreadsheet in Excel.

5. (Optional) Analyze the data

Depending on how the questions are formulated in the questionnaire, data can be analyzed for topic, as explained in the “Rumor tracking through Twitter” section above. For this step, the Python script `run_model.py` is needed, found on the Github page mentioned in the section “The product”. Grouping the tweets based on their content can identify the prevailing themes present in the dataset. This gives an understanding of the general trend of the tweets, and thus, the population. By using the `run_model.py` script:

- a. Cluster the tweets based on their content by using the “GSDMM: Short text clustering” Python library (<https://github.com/rwalk/gsdmm>) which is based on a short text clustering model². The script will group tweets with a similar theme together, creating theme clusters and will output two extra columns to the csv file. One will contain the theme cluster number to which each tweet was assigned, and the other will contain the confidence score, indicating the certainty of the model about the classification of the tweet.
- b. Manually identify the themes present in the dataset. For this, we suggest investigating each cluster individually. In particular, the 10 tweets with the highest confidence score for each cluster can reveal an overarching theme, which can be assigned to the cluster. With the `run_model.py` script, one can assign a topic description to each tweet. The output csv file should have substituted the numbers of the cluster with the topic name of each tweet. This file is also the final product of the rumor tracking analysis (**figure 6**).

6. Visualize the results

The results of the rumor tracking analysis can be visualized in a number of ways. A visualization by means of a dashboard can provide an interactive and accessible interface for others to interpret the data. In **figure 10**, we present an example visualization of the sample rumor data, which is produced with Google Data Studio. For more information on data visualization through dashboards, please refer to the manual “Dashboards”.

² Yin, J., & Wang, J. (2014). A dirichlet multinomial mixture model-based approach for short text clustering. Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. DOI:10.1145/2623330.2623715

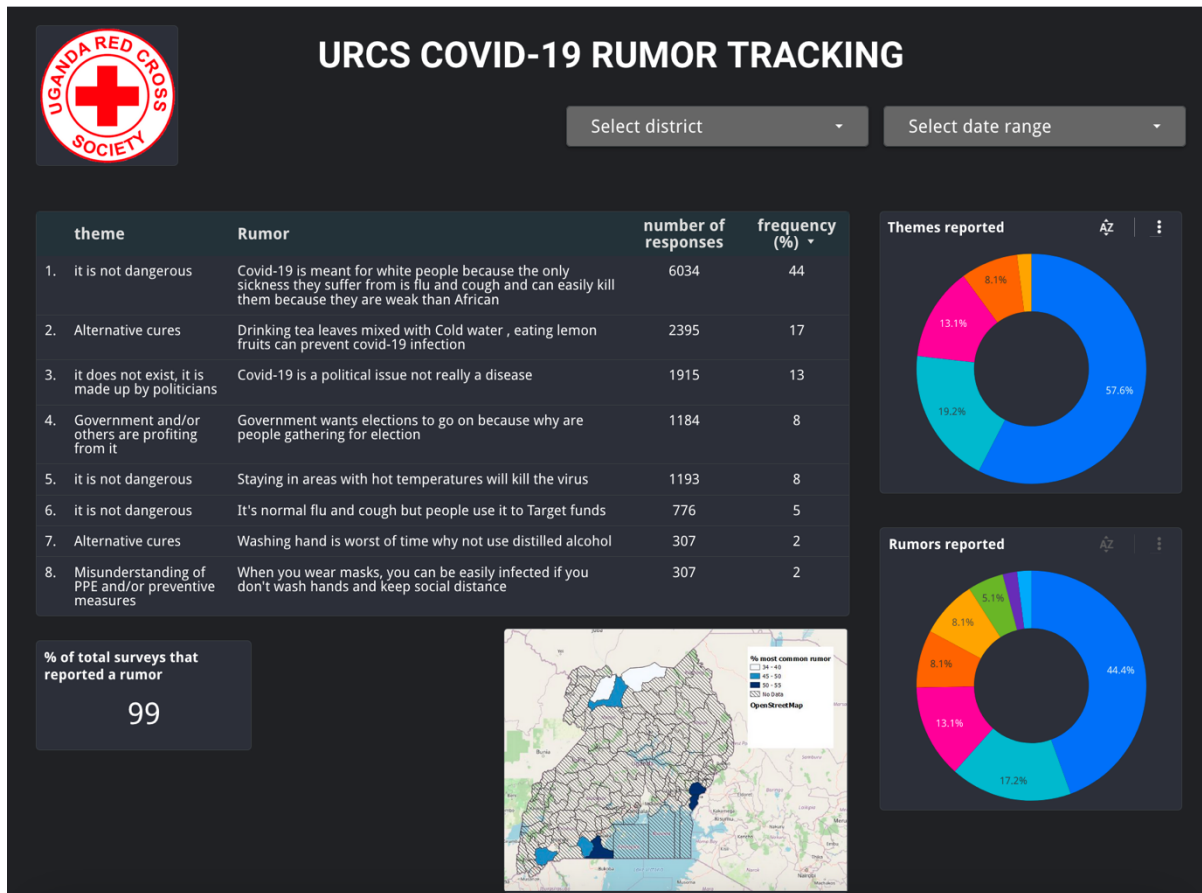


Figure 10 • Example of rumor data collected with KoBoCollect and visualized in Google Data Studio.

Abbreviations

NLRC	Netherlands Red Cross
NS	National Society/Societies
SARS-CoV-2	Severe Acute Respiratory Syndrome coronavirus 2
COVID-19	The disease caused by the SARS-CoV-2 virus
CEA	Community Engagement and Accountability
API	Application Programming Interface
GSDMM	Gibbs Sampling algorithm for the Dirichlet Multinomial Mixture model

Resources

- Twitter API
<https://developer.twitter.com/en/docs/twitter-api/v1/tweets/search/api-reference/get-search-tweets>
- Tweet-preprocessor 0.6.0 Python library
<https://pypi.org/project/tweet-preprocessor/>
- Google Translation API
<https://cloud.google.com/translate>
- Cloud Natural Language API (Google)
<https://cloud.google.com/natural-language>
- Short text clustering model
<https://github.com/rwalk/gsdmm>
- KoBo Toolbox
<https://www.kobotoolbox.org/>
- More information on KoBo Toolbox
<https://www.humanitarianresponse.info/en/applications/kobotoolbox/frequently-asked-questions>