Introduction to Twitter APIs

Taehee Kim

University of Oldenburg

Twitter APIs

Twitter's APIs

What is Twitter's API?

- Application programming interface (API) : a set of functions and procedures that allow the creation of applications which access the features or data of an operating system etc..
- About Twitter's APIs: Link
- We request Twitter data by calling to endpoint

Developer Agreement and Policy

- Before using APIs, visit the website and read their policy: Link
- ...If you provide Twitter Content to third parties,..., you will only distribute or allow download of Tweet IDs, Direct Message IDs, and/or User IDs.

Twitter's APIs

As of Dec 2020, Twitter is rebuilding API, *Twitter API v2*. Since they are not fully ready, this course assumes v1.1, which is still standard.

Three APIs

- Standard APIs (free): Search tweets published in 7 days, 3200 recent tweets of a user etc.
- Premium APIs (not free) : Search historical data from 2006, charge based on number of requests, monthly paid, from $149
- Enterprise APIs (expensive)

Twitter's APIs

As of Dec 2020, Twitter is rebuilding API, *Twitter API v2*. Since they are not fully ready, this course assumes v1.1, which is still standard.

Three APIs

- Standard APIs (free): Search tweets published in 7 days, 3200 recent tweets of a user etc.
- Premium APIs (not free) : Search historical data from 2006, charge based on number of requests, monthly paid, from $149
- Enterprise APIs (expensive)

Standard APIs

Standard APIs

- Streaming APIs
- Rest APIs

Rate limits

- a maximum number of requests per time (15 min) that user can request
- rate limits are differ depending on requests $\rightarrow$ check Twitter Doc.

Rest APIs

Rest APIs is used to retrieve, publish, and delete tweets.

Search Tweets

- search tweets published in 7 days using queries (keywords)
- https://developer.twitter.com/en/docs/tweets/search/api-reference/get-search-tweets.html
- rate limits: 180 requests in 15 min (per one requests, we can get 100 tweets, so 18,000 per 15min)
- Note: search API is focused on relevance and not completeness

User Timelines

- retrieve the most recent Tweets posted by the user indicated by the screen_name or user_id
- up to 3,200 of a user's most recent tweets.
- https://developer.twitter.com/en/docs/tweets/timelines/api-reference/get-statuses-user_timeline.html

Streaming APIs

Streaming APIs

- real-time data
- we are allowed to retrieve 1% of all the tweets at any given time
- about 500 million Tweets are posted per day $\rightarrow$ 350,000 per minutet $\rightarrow$ 1% : 3500 per minute?
- If you are using the filter and number of tweets matching your filter don't exceed 1%, you can get all tweets.
- https://developer.twitter.com/en/docs/tweets/filter-realtime/api-reference/post-statuses-filter.html

Twitter Object

Twitter Object

- https://developer.twitter.com/en/docs/tweets/data-dictionary/overview/tweet-object
- we receive Twitter in JSON

JSON

- JavaScript Object Notation
- common data format that can store structured data
- The JSON will be a mix of root-level attributes, and child objects (which are represented here with the {} notation) (from Twitter documentation)

Twitter JSON

Twitter JSON

```
1  {"created_at": "Wed Nov 21 08:39:49 +0000 2018", "id": 1065162858479185921, "id_str": "1065162858479185921",
   "full_text": "Angela #Merkel: Wir haben in den Verhandlungen um den #UNMigrationspakt, Wert darauf gelegt,
   dass die Souver\u00e4nit\u00e4t unseres Landes und unsere eigene Rechtssprechung nicht ber\u00fchrt wird.
   #Generaldebatte #Bundestag #Haushalt", "truncated": false, "display_text_range": [0, 220], "entities":
   {"hashtags": [{"text": "Merkel", "indices": [7, 14]}, {"text": "UNMigrationspakt", "indices": [54, 71]},
   {"text": "Generaldebatte", "indices": [184, 199]}, {"text": "Bundestag", "indices": [200, 210]}, {"text":
   "Haushalt", "indices": [211, 220]}], "symbols": [], "user_mentions": [], "urls": []}, "source": "<a
   href=\"http://twitter.com\" rel=\"nofollow\">Twitter Web Client</a>", "in_reply_to_status_id": null,
   "in_reply_to_status_id_str": null, "in_reply_to_user_id": null, "in_reply_to_user_id_str": null,
   "in_reply_to_screen_name": null, "user": {"id": 20429858, "id_str": "20429858", "name": "CDU Deutschlands",
   "screen_name": "CDU", "location": "Berlin", "description": "Die #CDU ist die Volkspartei der Mitte. Seit
   1945. – Redaktion: https://t.co/5uW0YgzHvo Impressum: https://t.co/AoefcbGyX1", "url": "https://t.co/
   GGABUUSnFy", "entities": {"url": {"urls": [{"url": "https://t.co/GGABUUSnFy", "expanded_url": "http://
   www.cdu.de", "display_url": "cdu.de", "indices": [0, 23]}]}, "description": {"urls": [{"url": "https://t.co/
   5uW0YgzHvo", "expanded_url": "https://www.cdu.de/redaktion", "display_url": "cdu.de/redaktion", "indices":
   [64, 87]}, {"url": "https://t.co/AoefcbGyX1", "expanded_url": "https://www.cdu.de/impressum", "display_url":
   "cdu.de/impressum", "indices": [99, 122]}]}}, "protected": false, "followers_count": 269114, "friends_count":
   1249, "listed_count": 1744, "created_at": "Mon Feb 09 11:43:27 +0000 2009", "favourites_count": 4405,
   "utc_offset": null, "time_zone": null, "geo_enabled": true, "verified": true, "statuses_count": 15496,
   "lang": "de", "contributors_enabled": false, "is_translator": false, "is_translation_enabled": false,
   "profile_background_color": "CDCCBB", "profile_background_image_url": "http://abs.twimg.com/images/themes/
   theme1/bg.png", "profile_background_image_url_https": "https://abs.twimg.com/images/themes/theme1/bg.png",
   "profile_background_tile": true, "profile_image_url": "http://pbs.twimg.com/profile_images/
   903206605537333250/13iosvQp_normal.jpg", "profile_image_url_https": "https://pbs.twimg.com/profile_images/
   903206605537333250/13iosvQp_normal.jpg", "profile_banner_url": "https://pbs.twimg.com/profile_banners/
   20429858/1542897116", "profile_link_color": "E95D0F", "profile_sidebar_border_color": "FFFFFF",
   "profile_sidebar_fill_color": "FFFFFF", "profile_text_color": "333333", "profile_use_background_image": true,
   "has_extended_profile": false, "default_profile": false, "default_profile_image": false, "following": false,
   "follow_request_sent": false, "notifications": false, "translator_type": "none"}, "geo": null, "coordinates":
   null, "place": null, "contributors": null, "is_quote_status": false, "retweet_count": 19, "favorite_count":
   35, "favorited": false, "retweeted": false, "lang": "de", "text": "Angela #Merkel: Wir haben in den
   Verhandlungen um den #UNMigrationspakt, Wert darauf gelegt, dass die Souver\u00e4nit\u00e4t unseres Landes
   und unsere eigene Rechtssprechung nicht ber\u00fchrt wird. #Generaldebatte #Bundestag #Haushalt"}
```

Twitter JSON

```
{
 "created_at":"Thu Apr 06 15:24:15 +0000 2017",
 "id": 850006245121695744,
 "id_str": "850006245121695744",
 "text": "1/ Today we're sharing our vision for the future of the Twitter API pla
 "user": {},
 "entities": {}
}
```

Figure 1: See: https://developer.twitter.com/en/docs/tweets/
data-dictionary/overview/tweet-object

Twitter JSON

1 {"created_at": "Wed Nov 21 08:39:49 +0000 2018", "id": 1065162858479185921,
   "id_str": "1065162858479185921", "full_text": "Angela #Merkel: Wir haben in den
   Verhandlungen um den #UNMigrationspakt, Wert darauf gelegt, dass die
   Souver\u00e4nit\u00e4t unseres Landes und unsere eigene Rechtssprechung nicht
   ber\u00fchrt wird. #Generaldebatte #Bundestag #Haushalt", "truncated": false,
   "display_text_range": [0, 220], "entities": {"hashtags": [{"text": "Merkel",
   "indices": [7, 14]}, {"text": "UNMigrationspakt", "indices": [54, 71]}, {"text":
   "Generaldebatte", "indices": [184, 199]}, {"text": "Bundestag", "indices": [200,
   210]}, {"text": "Haushalt", "indices": [211, 220]}], "symbols": [],
   "user_mentions": [], "urls": []}, "source": "<a href=\"http://twitter.com\"
   rel=\"nofollow\">Twitter Web Client", "in_reply_to_status_id": null,
   "in_reply_to_status_id_str": null, "in_reply_to_user_id": null,
   "in_reply_to_user_id_str": null, "in_reply_to_screen_name": null, "user": {"id":
   20429858, "id_str": "20429858", "name": "CDU Deutschlands", "screen_name":
   "CDU", "location": "Berlin", "description": "Die #CDU ist die Volkspartei der
   Mitte. Seit 1945. – Redaktion: https://t.co/5uW0YgzHvo Impressum: https://t.co/

Twitter JSON

```
1  {"created_at": "Wed Nov 21 08:39:49 +0000 2018", "id": 1065162858479185921,
      "id_str": "1065162858479185921", "full_text": "Angela #Merkel: Wir haben in den
      Verhandlungen um den #UNMigrationspakt, Wert darauf gelegt, dass die
      Souver\u00e4nit\u00e4t unseres Landes und unsere eigene Rechtssprechung nicht
      ber\u00fchrt wird. #Generaldebatte #Bundestag #Haushalt", "truncated": false,
      "display_text_range": [0, 220], "entities": {"hashtags": [{"text": "Merkel",
      "indices": [7, 14]}, {"text": "UNMigrationspakt", "indices": [54, 71]}, {"text":
      "Generaldebatte", "indices": [184, 199]}, {"text": "Bundestag", "indices": [200,
      210]}, {"text": "Haushalt", "indices": [211, 220]}], "symbols": [],
      "user_mentions": [], "urls": []}, "source": "<a href=\"http://twitter.com\"
      rel=\"nofollow\">Twitter Web Client</a>", "in_reply_to_status_id": null,
      "in_reply_to_status_id_str": null, "in_reply_to_user_id": null,
      "in_reply_to_user_id_str": null, "in_reply_to_screen_name": null, "user": {"id":
      20429858, "id_str": "20429858", "name": "CDU Deutschlands", "screen_name":
      "CDU", "location": "Berlin", "description": "Die #CDU ist die Volkspartei der
      Mitte. Seit 1945. - Redaktion: https://t.co/5uW0YgzHvo Impressum: https://t.co/
```

Twitter JSON

Important Attributes

- "created_at" : (string) UTC time
- "id" : (int) The integer representation of the unique identifier for this Tweet ("status_id" in rtweet)
- "full_text" : The actual UTF-8 text ("text" in rtweet)
- "user" : User object which contains "id", "screen_name", "followers_count" etc. ("user_id" in rtweet)

!!! Note: "id" is unique, but not "screen_name"

Twitter JSON

Important Attributes

- "entities" : which have been parsed out of the text of the Tweet (e.g., hashtag)
- "lang" : (String) the machine-detected language of the Tweet text
- "place" : place that the tweet is associated
- "retweeted_status" : (Tweet) the original Tweet that was retweeted.
- "quoted_status" : (Tweet) the original Tweet that was quoted.

Exercise:

Find information from Twitter documentation and understand Twitter data we will receive.

1. Find the documentation page explaining "entities" of Twitter object.
2. From the page Q1, find the information on retweets and quotes. What are the differences between them?
3. Find the documentation page explaining "user object".
4. There are two attributes included "user object", "name" and "screen_name". What is the difference?
5. To confirm if an account is verified or not, which attributes of "user object" should we check?

UTC time

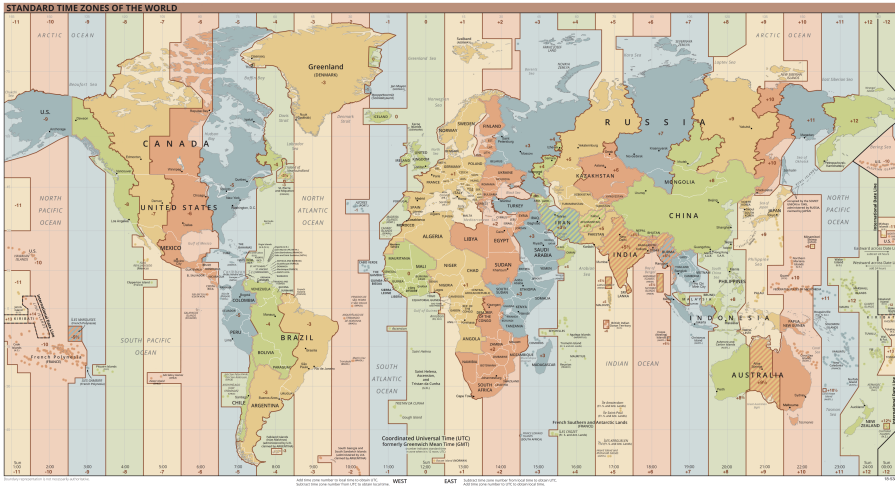- Coordinated Universal Time
- Germany : $+1$

UTC

Figure 2: https://en.wikipedia.org/wiki/Coordinated_Universal_Time

15

How to request?

OAuth

- Twitter uses OAuth to provide authorized access to its API.
- We can get OAuth access tokens from Twitter's app, which will create together later.

Endpoint (Resource URL)

- You can find URL's in Twitter's document.
- e.g., search endpoint:
  https://api.twitter.com/1.1/search/tweets.json

How to request?

Parameters

- We can change search condition by setting up parameters.

Important parameters for search (REST API).

- 'q' : query !!! NOTE: required parameter
- 'since' and 'until' : Returns tweets created after/before the given date
- 'result_type' : mixed or recent or popular
- 'lang' : Restricts tweets to the given language (e.g., en, de)

Understanding rate limit

Q. Suppose that we want to search and retrieve tweets using a specific query by connecting REST APIs. So, how many tweets we can get by one requests?

Understanding rate limit

Q. Suppose that we want to search and retrieve tweets using a specific query by connecting REST APIs. So, how many tweets we can get by one requests?

A. 100

Understanding rate limit

Q. Then how many times we can request in 15 min?

Understanding rate limit

Q. Then how many times we can request in 15 min?

A. 180

Understanding rate limit

Q. So what is the max number of tweets we can get in 15min?

Understanding rate limit

Q. So what is the max number of tweets we can get in 15min?

A. 18,000

Understanding rate limit

Q. So what is the max number of tweets we can get in 15min?

A. 18,000

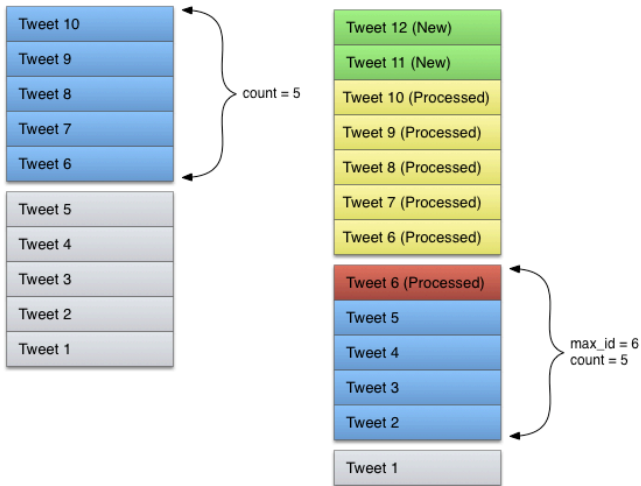??? By the way.... how to iterate through all list of tweets?

max_id

Figure 3: https://developer.twitter.com/en/docs/tweets/timelines/guides/working-with-timelines

21

max_id parameter for iteration

- APIs will only return Tweets with IDs lower(older) than or equal to the value of the max_id parameter.
- set max_id = minimum number of Tweet ID among recently obtained

loop for obtaining tweets more than 18,000

- wait until reset time (max 15 min) $\rightarrow$ continue requests

Twitter in R: 'rtweet' package

'rtweet' package

- An implementation of calls designed to collect and organize Twitter data via Twitter's REST and stream

- documentation: https://cran.r-project.org/web/packages/rtweet/rtweet.pdf

Set up Twitter Applications

- Go to Twitter developer page
  https://developer.twitter.com/en.html
- Create an app
- Get Consumer API keys and Access token & access token secret

If you have any questions..

First, read Twitter Documentation

https://developer.twitter.com/en/docs

Second, 'rtweet' documentation

Third, go to Forum

https://twittercommunity.com

Further material

- Jürgens, Pascal and Andreas Jungherr. 2016 A Tutorial for Using Twitter Data in the Social Sciences: Data Collection, Preparation, and Analysis
- Zachary C. Steinert-Threlkeld. 2018. Twitter as Data. Cambridge University Press (check his recent workshop materials: `https://github.com/ZacharyST/APSA2020_EventDataFromSocialMedia`
- Conover, Michael, Jacob Ratkiewicz, Matthew R Francisco, Bruno Gon calves, Filippo Menczer and Alessandro Flammini. 2011. Political Polarization on Twitter. Icwsm 133:8996.