

Introduction to Twitter data and APIs

Taehee Kim

December 9, 2021

University of Oldenburg

What is Twitter data?

User information



← **Joe Biden**

7,312 Tweets



Follow

Joe Biden

@JoeBiden

United States government official

Husband to @DrBiden, proud father and grandfather. Ready to build back better for all Americans. Official account is @POTUS.

📍 Washington, DC 🌐 joebiden.com 📅 Joined March 2007

47 Following **30M Followers**

Not followed by anyone you're following

User information

- Name
 - @username
 - Description
 - Location
 - URL
 - Joined date
 - Following, Followers
 - Verified or not
 - Total number of Tweets published
- For more detail, check Twitter developer doc:: [User object](#)

Tweet



Joe Biden

@JoeBiden

United States government official

...

With the first 100 days of our administration under our belt, America is getting back on track—all thanks to the support of grassroots volunteers. Help us keep the momentum going by joining our Week of Action:
democrats.org/100-days

11:19 PM · April 27, 2021 · Twitter Web App

1,371 Retweets **118** Quote Tweets **13.5K** Likes



Kaitlyn Quintana @quinn_hungry · 14h

...

Replying to @JoeBiden

America is stronger than ever with Biden in charge!



39



2



49



Unbelievable Ron @UnbelievableRon · 14h

...

Mention



Joe Biden @JoeBiden · Apr 23

United States government official

...

Heading back to Georgia with [@DrBiden](#) next Thursday to mark 100 days of our administration and thank folks for helping us build back better as a nation. See you soon!



7.6K



4.3K



51.7K



Media info



Joe Biden @JoeBiden · Apr 27

...

► United States government official

The American Jobs Plan will strengthen unions, help protect workers, and rebuild our economy.

“

**UNIONS BUILT
the
MIDDLE CLASS.
IT'S ABOUT TIME
THEY START TO
GET A PIECE
of
THE ACTION.**

AMERICAN JOBS PLAN



2.2K



2.6K



20.8K



Retweet and Quote

Retweeted by Joe Biden



President Biden  @POTUS · Apr 25

...

United States government official

Just as India sent assistance to the United States as our hospitals were strained early in the pandemic, we are determined to help India in its time of need.



Jake Sullivan  @JakeSullivan46 · Apr 25

Spoke today with National Security Advisor Ajit Doval about the spike in COVID cases in India and we agreed to stay in close touch in the coming days. The United States stands in solidarity with the people of India and we are deploying more supplies and resources:

available for India. To help treat COVID-19 patients and protect front-line health workers in India, the United States has identified supplies of therapeutics, rapid diagnostic test kits, ventilators, and Personal Protective Equipment (PPE) that will immediately be made available for India. The United States also is pursuing options to provide oxygen generation and related supplies on an urgent basis. The U.S. Development Finance Corporation (DFC) is funding a substantial expansion of manufacturing capability for BioE, the vaccine manufacturer in India, enabling BioE to ramp up to produce at least 1 billion doses of

- Text (late-2017, the limit changed from 140 to 280)
 - In a text: url, mention (@username), hashtab (#)
 - Media: image, GIF, or video
 - Location (only if users decide to assign a location to their tweets)
 - [Retweet or Quote](#)
- For more detail, check Twitter developer doc: [Tweet Object](#)

Twitter APIs

What is Twitter's API?

Application programming interface (API)

- About Twitter's APIs: [Here](#)

"APIs are the way computer programs talk to each other so that they can request and deliver information. This is done by allowing a software application to call what's known as an endpoint: an address that corresponds with a specific type of information we provide (endpoints are generally unique like phone numbers).

- We request Twitter data by calling to endpoint

...But why do we use APIs? Isn't there other ways?

Developer Agreement and Policy

- **Developer Agreement:**
"This Twitter Developer Agreement (Agreement) is made between you (either an individual or an entity, referred to herein as you) and Twitter (as defined below) and governs your access to and use of the Licensed Material (as defined below)."
- More about restricted uses of the Twitter APIs: [Here](#)

Developer Agreement and Policy

Redistribution of Twitter content

"If you need to share Twitter content you obtained via the Twitter APIs with another party, the best way to do so is by sharing Tweet IDs, Direct Message IDs, and/or User IDs, which the end user of the content can then rehydrate (i.e. request the full Tweet, user, or Direct Message content) using the Twitter APIs. This helps ensure that end users of Twitter content always get the most current information directly from us.

We permit limited redistribution of hydrated Twitter content via non-automated means. If you choose to share hydrated Twitter content with another party in this way, you may only share up to 50,000 hydrated public Tweet Objects and/or User Objects per recipient, per day, and should not make this data publicly available (for example, as an attachment to a blog post or in a public Github repository). (copied 8th Dec 2021)"

Twitter's APIs

Twitter API v1.1

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

Twitter API v2 (Released Nov. 15th, 2021)

- compared to v1
- Essential: v2 access only
- Elevated: v1.1 and v2
- Academic Research: v1.1, v2, Full-archive search

Twitter's APIs

Twitter API v1.1

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

Twitter API v2 (Released Nov. 15th, 2021)

- compared to v1
- Essential: v2 access only
- Elevated: v1.1 and v2
- Academic Research: v1.1, v2, Full-archive search

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 → 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 → 350,000 per minute
→ 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

CDU Deutschlands  @CDU

Follow

Angela **#Merkel**: Wir haben in den Verhandlungen um den **#UNMigrationspakt**, Wert darauf gelegt, dass die Souveränität unseres Landes und unsere eigene Rechtssprechung nicht berührt wird.

#Generaldebatte #Bundestag #Haushalt

Translate Tweet

9:39 AM - 21 Nov 2018

19 Retweets 35 Likes



71 19 35

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",  

```

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\u00fcft 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

```
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\u00fcchert 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

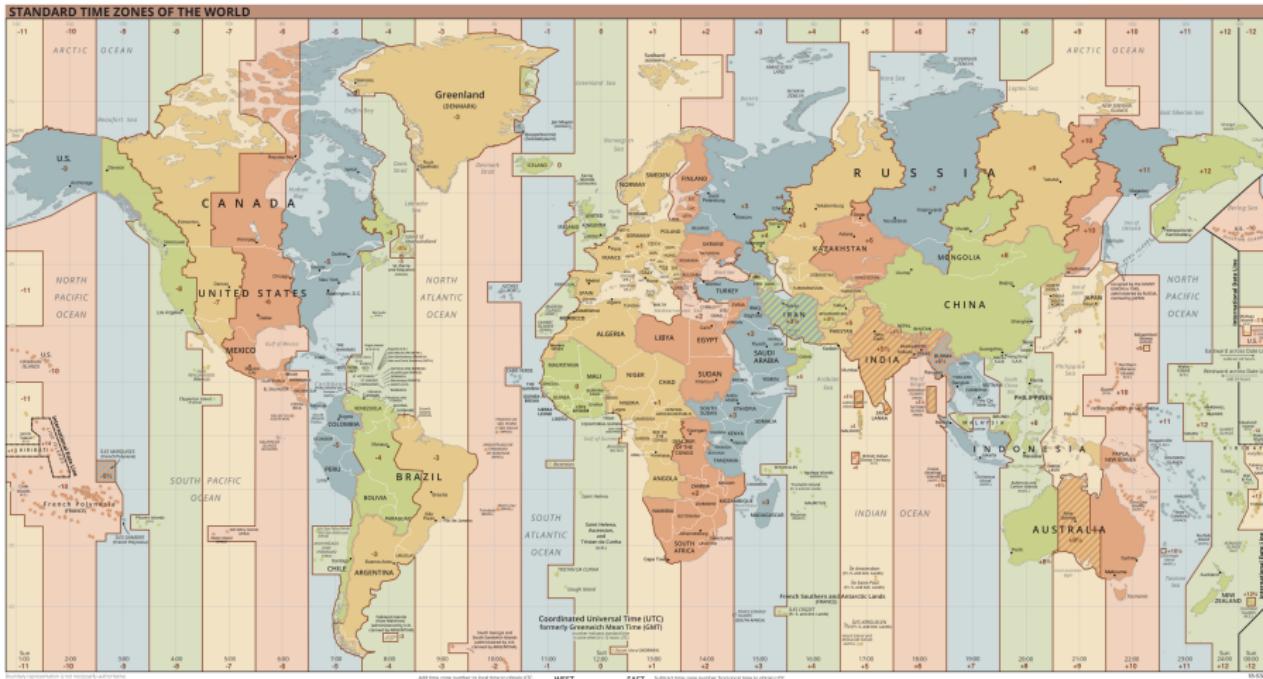


Figure 2: https://en.wikipedia.org/wiki/Coordinated_Universal_Time

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).

- Search API [documentation](#)
- '`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

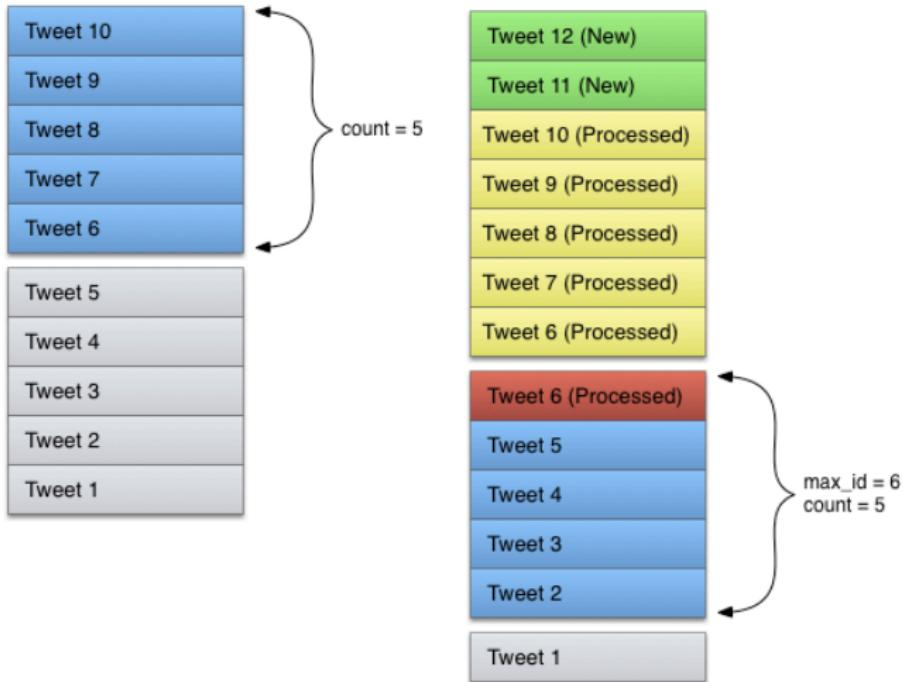


Figure 3: <https://developer.twitter.com/en/docs/tweets/timelines/guides/working-with-timelines>

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) → 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>

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>

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