

Deep into the tweet

Introduction to the twitter meta-data
Mining Twitter API

Why Twitter?



Image by Gerd Altmann from [Pixabay](#)

By Dina Bavli



Fertile ground for
many types of
research

Why Mine Twitter?

*One size
doesn't fit all*



Image by [lisa runnels](#) from [Pixabay](#)

By Dina Bawli

Where to Start?

By Dina Bavli

Twitter developer account

Authenticate and integrate with most Twitter developer products. Create an app to get your API key.

Please apply for a Twitter developer account



You may continue managing your existing apps but if you would like to create new apps or use Twitter premium APIs, please apply for a developer account.

As a developer platform, our first responsibility is to our users: to provide a place that supports the health of conversation on Twitter. To continue to prevent misuse of our platform, we have introduced a few new requirements for developers.

Apply

Creating an app for credentials

The screenshot shows a user interface for creating a new application. At the top, there are two buttons: 'Apply' and 'Apps', followed by a user profile icon. Below these is a blue button labeled 'Create an app'. The main area has a header 'Apps >' and three tabs: 'App details' (which is active and highlighted in blue), 'Keys and tokens', and 'Permissions'. A yellow callout box contains an important notice about access tokens.

Important notice about your access token and access token secret

To make your API integration more secure, we will no longer show your access token and access token secret beyond the first time that you generate it starting **January 20, 2020**. You will be able to regenerate it at anytime here, which will invalidate your current access token and secret. Please save this information if you need to access it. This does not affect your consumer API keys, which will still be shown here as they are below. To learn more, [visit the Forums](#).

By Dina Bavli

Getting account credentials

Apps >

App details Keys and tokens Permissions

Important notice about your access token and access token secret

To make your API integration more secure, we will no longer show your access token and access token secret beyond the first time that you generate it starting **January 20, 2020**. You will be able to regenerate it at anytime here, which will invalidate your current access token and secret. Please save this information if you need to access it. This does not affect your consumer API keys, which will still be shown here as they are below. To learn more, [visit the Forums](#).

Keys and tokens
Keys, secret keys and access tokens management.

Consumer API keys

(API key)
(API secret key)

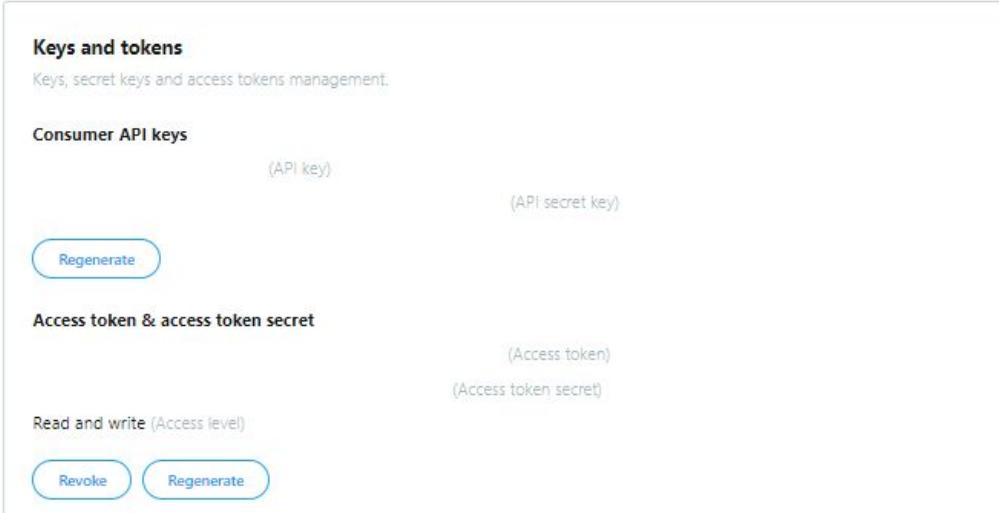
[Regenerate](#)

Access token & access token secret

(Access token)
(Access token secret)

Read and write (Access level)

[Revoke](#) [Regenerate](#)



By Dina Bavli

What for?

What are the credentials for?

```
# Twitter API credentials
consumer_key = "your_consumer_key"
consumer_secret = "your_consumer_secret"
access_key = "your_access_key"
access_secret = "your_access_secret"

OAUTH_KEYS = {'consumer_key':consumer_key, 'consumer_secret':consumer_secret,
              'access_token_key':access_key, 'access_token_secret':access_secret}
auth = tweepy.OAuthHandler(OAUTH_KEYS['consumer_key'], OAUTH_KEYS['consumer_secret'])
api = tweepy.API(auth, wait_on_rate_limit=True, wait_on_rate_limit_notify=True)
```



Now what?

Installs & Imports

```
! pip install tweepy
```

```
import tweepy #https://github.com/tweepy/tweepy
import pandas as pd
import time
import json
```

Getting to know API search

- Queries
- Filters
- Tweet_mode
- Include_rts
- Lang
- Items
- Search time limit

```
filters = '-filter:media' # any filter
query_words = ['love']#, 'hate', 'makeup', 'maddona'] # any query words
for i in range(len(query_words)):
    search = tweepy.Cursor(api.search, q = query_words[i]+filters,
                           tweet_mode = 'extended', # length of tweet text
                           include_rts = True, # to include retweets
                           lang = "en").items(1) # any number of tweets- up to about 4000

    all_data = []
    for tweet in search:
        status = tweet
        all_data.append(tweet._json)

    print(query_words[i])
    time.sleep(1) # deals with the API's search limits

pprint.pprint(all_data)
```

Closer look at extended tweet

```
"text": "Can't fit your Tweet into 140 characters? 🤯\n\nWe're trying something new with a small group, and increasing the char... https://t.co/y1rJlHsVB5",
```

```
"extended_tweet": {  
  "full_text": "Can't fit your Tweet into 140 characters? 🤯\n\nWe're trying something new with a small group, and increasing the character limit to 280! Excited about the possibilities? Read our blog to find out how it all adds up.  
👉\nhttps://t.co/C6hjsB9nbL",
```

What do you see?

What would you like to retrieve from this Example tweet tread?

- Text
- Emojis
- Dates
- Retweet
- Mentions
- users
- Hashtags
- Replying to
- Likes
- URLs
- Images

← Tweet

Dina Bavli @BavliDina · Jan 1 towardsdatascience.com/mining-twitter... I published. A great opening to a new decade 😊 Happy new year! 🎉 Thanks @Baot_IL @lkd10 @AnatShk @OphirHarpaz @Dafna_Mordechai @DalyaGar @TDataScience @IBM #DataScience #datamining #python #IBM



Mining Twitter on IBM Cloud Platform
Tweet Object renders as JSON so it's only sensible to save them as such.
Saving as JSON on IBM cloud storage is not like saving locally.
© towardsdatascience.com

1 7 12

Replies

Ophir Harpaz @OphirHarpaz · Jan 1 Replying to @BavliDina @Baot_IL and 6 others



GIF

1 1 3 1

How about user data?

What fields would you like to retrieve from a profile?

- Name
- Screen name
- Followers
- Following
- Joined at
- Likes
- Tweets

Dina Bavli
8 Tweets

 Edit profile

Dina Bavli
@BavliDina

Passionate about data, sharing knowledge, and contributing to society and open source. An NLP & SNA researcher at the University of Haifa.

Joined December 2019

46 Following 17 Followers

Tweets Tweets & replies Media Likes

Dina Bavli @BavliDina · Feb 25
Sometimes we forget the importance of "Begin at the beginning" (Alice's Adventures in Wonderland, L.C.). See you next session where we'll continue our journey through the data science pipeline. #datascience #machinelearning #outliers #meetup @DataHackIL



Image by [intographics](#) from Pixabay

By Dina Bavli

By Dina Bauli

```
{  
    "created_at": "Tue Sep 26 21:00:22 +0000 2017",  
    "id": 912783930431905797,  
    "id_str": "912783930431905797",  
    "text": "Can't fit your Tweet into 140 characters? 🤯\n\nWe're trying something new with a small group, and increasing the character limit to 280! Excited about  
the possibilities? Read our blog to find out how it all adds up. 🎉\nhttps://t.co/C6hjsB9nbL",  
    "source": "Twitter Web Client",  
    "truncated": true,  
    "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": 783214,  
        "id_str": "783214",  
        "name": "Twitter",  
        "screen_name": "Twitter",  
        "location": "Everywhere",  
        "url": "https://about.twitter.com/",  
        "description": "What's happening?",  
        "translator_type": null,  
        "protected": false,  
        "verified": true,  
        "followers_count": 56119543,  
        "friends_count": 140,  
        "listed_count": 91049,  
        "favourites_count": 5911,  
        "statuses_count": 10132,  
        "created_at": "Tue Feb 20 14:35:54 +0000 2007",  
        "utc_offset": null,  
        "time_zone": null,  
        "geo_enabled": true,  
        "lang": "en",  
        "contributors_enabled": false,  
        "is_translator": false,  
        "profile_background_color": "null",  
        "profile_background_image_url": "null",  
        "profile_background_image_url_https": "null",  
        "profile_background_tile": null,  
        "profile_link_color": "null",  
        "profile_sidebar_border_color": "null",  
        "profile_sidebar_fill_color": "null",  
        "profile_text_color": "null",  
        "profile_use_background_image": null,  
        "profile_image_url": "null",  
        "profile_image_url_https": "https://pbs.twimg.com/profile_images/1111729635610382336/_65QFj7B_normal.png",  
        "profile_banner_url": "https://pbs.twimg.com/profile_banners/783214/1556918042",  
        "default_profile": false,  
        "default_profile_image": false,  
        "following": null,  
        "follow_request_sent": null,  
        "notifications": null  
    },  
    "geo": null,  
    "coordinates": null,  
    "place": null,  
    "contributors": null,  
    "is_quote_status": false,  
    "extended_tweet": {  
        "full_text": "Can't fit your Tweet into 140 characters? 🤯\n\nWe're trying something new with a small group, and increasing the character limit to 280! Excited about  
the possibilities? Read our blog to find out how it all adds up. 🎉\nhttps://t.co/C6hjsB9nbL",  
        "display_text_range": [0, 239]  
    },  
    "entities": {  
        "hashtags": [  
            {"url": "https://t.co/C6hjsB9nbL",  
             "expanded_url": "https://cards.twitter.com/cards/gsby/4ubsj",  
             "display_url": "cards.twitter.com/cards/gsby/4ub..."},  
            {"url": "https://cards.twitter.com/cards/gsby/4ubsj",  
             "status": 200,  
             "title": "Giving you more characters to express yourself",  
             "description": null},  
            {"indices": [216, 239]}],  
        "user_mentions": [],  
        "symbols": []},  
        "quote_count": 46955,  
        "reply_count": 17263,  
        "retweet_count": 50214,  
        "favorite_count": 89601,  
        "entities": {  
            "hashtags": []},  
        "urls": [  
            {"url": "https://t.co/y1rlHsVB5",  
             "expanded_url": "https://twitter.com/i/web/status/912783930431905797",  
             "display_url": "twitter.com/i/web/status/9...",  
             "indices": [117, 140]}],  
        "user_mentions": [],  
        "symbols": []},  
        "favorited": false,  
        "retweeted": false,  
        "possibly_sensitive": false,  
        "filter_level": "low",  
        "lang": "en",  
        "matching_rules": [  
            {"tag": null}]}]
```

By Dina Bawli

Closer look at the text

```
"text": "Can't fit your Tweet into 140 characters? 🤯\n\nWe're trying something new with a small group, and increasing the char... https://t.co/y1rJlHsVB5",
```

```
"extended_tweet": {  
  "full_text": "Can't fit your Tweet into 140 characters? 🤯\n\nWe're trying something new with a small group, and increasing the character limit to 280! Excited about the possibilities? Read our blog to find out how it all adds up.  
👉\nhttps://t.co/C6hjsB9nbL",
```

Retrieving tweet text

```
tweet_mode='extended'  
  
text.append(tweet.full_text)  
  
use text as a fall-back.
```

Is it enough?
Apparently NOT

```
for tweet in search:  
    status = tweet  
    if 'extended_tweet' in status._json: status_json = status._json['extended_tweet']['full_text']  
    elif 'retweeted_status' in status._json and 'extended_tweet' in status._json['retweeted_status']:  
        status_json = status._json['retweeted_status']['extended_tweet']['full_text']  
    elif 'retweeted_status' in status._json: status_json = status._json['retweeted_status']['full_text']  
    else: status_json = status._json['full_text']  
    status = status._json  
    tweet_data['original_text'].append(status_json)
```

Retrieving Tweet Data

```
tweet_data['tweet_id'].append(status['id'])
tweet_data['tweet_created_at'].append(status['created_at'])
tweet_data['source'].append(status['source'])
tweet_data['tweet_favorite_count'].append(status['favorite_count'])
tweet_data['retweet_count'].append(status['retweet_count'])
```

Retrieving mentions and hashtags data

```
mentions = status["entities"]["user_mentions"]
mentioned_users = ", ".join([mention['screen_name'] for mention in mentions])
tweet_data['user_mentions'].append(mentioned_users)
mentions_id =[mention['id'] for mention in mentions]
tweet_data['user_mentions_id'].append(mentions_id)

hashtags = ", ".join([hashtag_item['text'] for hashtag_item in status['entities']['hashtags']])
tweet_data['hashtags'].append(hashtags)
```

In reply to

```
tweet_data['in_reply_to_status_id'].append(tweet.in_reply_to_status_id)
tweet_data['in_reply_to_user_id'].append(tweet.in_reply_to_user_id)
tweet_data['in_reply_to_screen_name'].append(tweet.in_reply_to_screen_name)
```

Retrieving user data

```
tweet_data['user_screen_name'].append(status['user']['screen_name'])#
tweet_data['user_id'].append(status['user']['id'])
tweet_data['user_followers_count'].append(status['user']['followers_count'])
tweet_data['user_listed_count'].append(status['user']['listed_count'])
tweet_data['user_favourites_count'].append(status['user']['favourites_count'])
tweet_data['user_statuses_count'].append(status['user']['statuses_count'])
tweet_data['user_created_at'].append(status['user']['created_at'])
```

Location data

```
#get location of the tweet if possible
try:
    location = status['user']['location']
    user_loc=geolocator.geocode(location)

    if user_loc:

        lat=user_loc.latitude
        lon=user_loc.longitude
    else:
        lat=''
        lon=''

except TypeError:
    location = ''


tweet_data['latitude_corr'].append(lat)
tweet_data['longitude_corr'].append(lon)
tweet_data['place'].append(location)

try:
    coordinates = [coord for loc in status['place']['bounding_box']['coordinates'] for coord in loc]
except TypeError:
    coordinates = None
tweet_data['place_coord_boundaries'].append(coordinates)
```



Done with retrieving?

Encoding may create garbage

```
df.to_csv("file.csv", encoding="utf-8")
```

Is there any workaround?

df.head() is this:

```
demography  Adults ≥49 yrs  Adults 18-49 yrs at high risk||  \
state
Alabama          32.7           38.6
Alaska           31.2           33.2
Arizona          22.9           38.8
Arkansas         31.2           34.0
California       29.8           38.8
```

csv output is this

```
state,  Adults ≥49 yrs,  Adults 18-49 yrs at high risk||
0,  Alabama, 32.7, 38.6
1,  Alaska, 31.2, 33.2
2,  Arizona, 22.9, 38.8
3,  Arkansas,31.2, 34
4,  California,29.8, 38.8
```

Dealing with emojis

```
pip install emoji
```

```
import emoji
```

```
print(emoji.emojize('Python is :thumbs_up:'))
```

Python is 

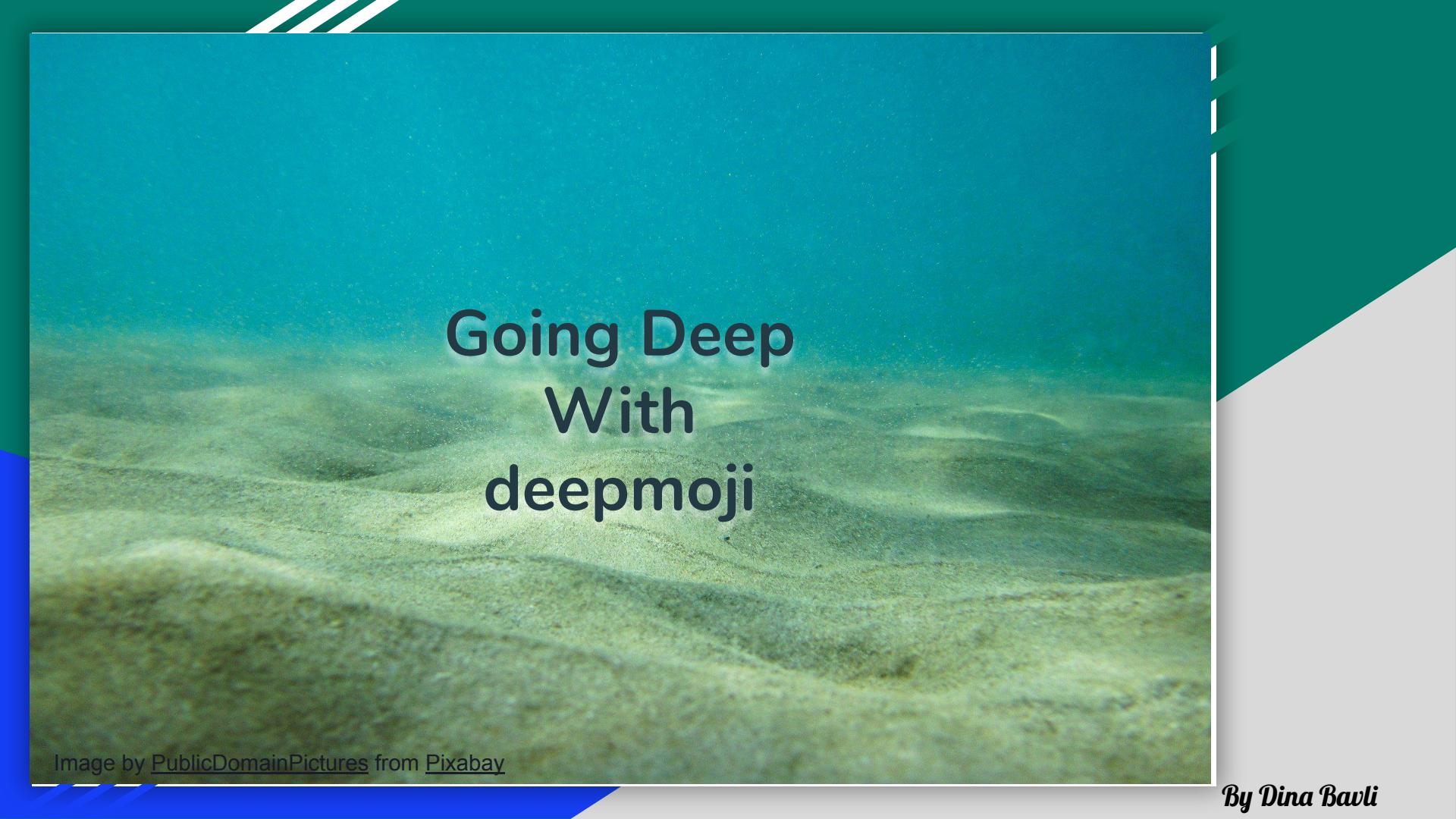
```
print(emoji.emojize('Python is :thumbsup:', use_aliases=True))
```

Python is 

```
print(emoji.demojize('Python is '))
```

Python is :thumbs_up:

```
tweet = """ #startspreadingthenews yankees win great start by 🎉 going 5strong innings with 5k's🔥⚾️  
solo homerun 💣💣 with 2 solo homeruns and🔥 3run homerun... 🎉DAC with rbi's ... 🔥🔥  
MX and NI to close the game🔥🔥!!!...  
WHAT A GAME!!!.. """  
  
print(emoji.emojize(tweet))  
  
#startspreadingthenews yankees win great start by 🎉 going 5strong innings with 5k's🔥⚾️  
solo homerun 💣💣 with 2 solo homeruns and🔥 3run homerun... 🎉DAC with rbi's ... 🔥🔥  
MX and NI to close the game🔥🔥!!!...  
WHAT A GAME!!!..  
  
print(emoji.emojize(tweet, use_aliases=True))  
  
#startspreadingthenews yankees win great start by 🎉 going 5strong innings with 5k's:fire: :ox:  
solo homerun :volcano::volcano: with 2 solo homeruns and:orange: 3run homerun... :clown_face: :person_rowing_boat_medium-light_skin_tone: :man_judge_medium_skin_tone: with rbi's ... :fire::fire:  
:Mexico: and :Nicaragua: to close the game:fire::fire:!!!...  
WHAT A GAME!!!..
```

A photograph of an underwater environment. The bottom is a sandy ocean floor with some small, scattered debris. The water is a clear, pale blue-green color, with light rays filtering down from the surface. The overall atmosphere is calm and mysterious.

Going Deep With deepmoji

Image by [PublicDomainPictures](#) from [Pixabay](#)

By Dina Bavli

Setting up deepmoji

```
[2] !pip3 install torch==1.0.1 -f https://download.pytorch.org/wheel/cpu/stable
!git clone https://github.com/huggingface/torchMoji
import os
os.chdir('torchMoji')
!pip3 install -e .
```

type "yes" when prompted

```
[ ] !python3 scripts/download_weights.py
```

```
[ ] import numpy as np
import emoji, json
from torchmoji.global_variables import PRETRAINED_PATH, VOCAB_PATH
from torchmoji.sentence_tokenizer import SentenceTokenizer
from torchmoji.model_def import torchmoji_emojis

EMOJIS = '''':joy: :unamused: :weary: :sob: :heart_eyes: :pensive: :ok_hand: :blush: :heart: :smirk: :grin: :notes: :flushed: :100: :sleeping:
:relieved: :relaxed: :raised_hands: :two_hearts: :expressionless: :sweat_smile: :pray: :confused: :kissing_heart: :heartbeat: :neutral_face:
:information_desk_person: :disappointed: :see_no_evil: :tired_face: :v: :sunglasses: :rage: :thumbsup: :cry: :sleepy: :yum: :triumph: :hand:
:mask: :clap: :eyes: :gun: :persevere: :smiling_imp: :sweat: :broken_heart: :yellow_heart: :musical_note: :speak_no_evil: :wink: :skull:
:confounded: :smile: :stuck_out_tongue_winking_eye: :angry: :no_good: :muscle: :facepunch: :purple_heart: :sparkling_heart: :blue_heart:
:grimacing: :sparkles:''.split(' ')
model = torchmoji_emojis(PRETRAINED_PATH)
with open(VOCAB_PATH, 'r') as f:
    vocabulary = json.load(f)
st = SentenceTokenizer(vocabulary, 30)

def deepmojify(sentence,top_n =5):
    def top_elements(array, k):
        ind = np.argpartition(array, -k)[-k:]
        return ind[np.argsort(array[ind])][::-1]

    tokenized, _ , _ = st.tokenize_sentences([sentence])
    prob = model(tokenized)[0]
    emoji_ids = top_elements(prob, top_n)
    emojis = map(lambda x: EMOJIS[x], emoji_ids)
    return emoji.emojize(f'{sentence} {"".join(emojis)}", use_aliases=True)
```

Deepmoji example

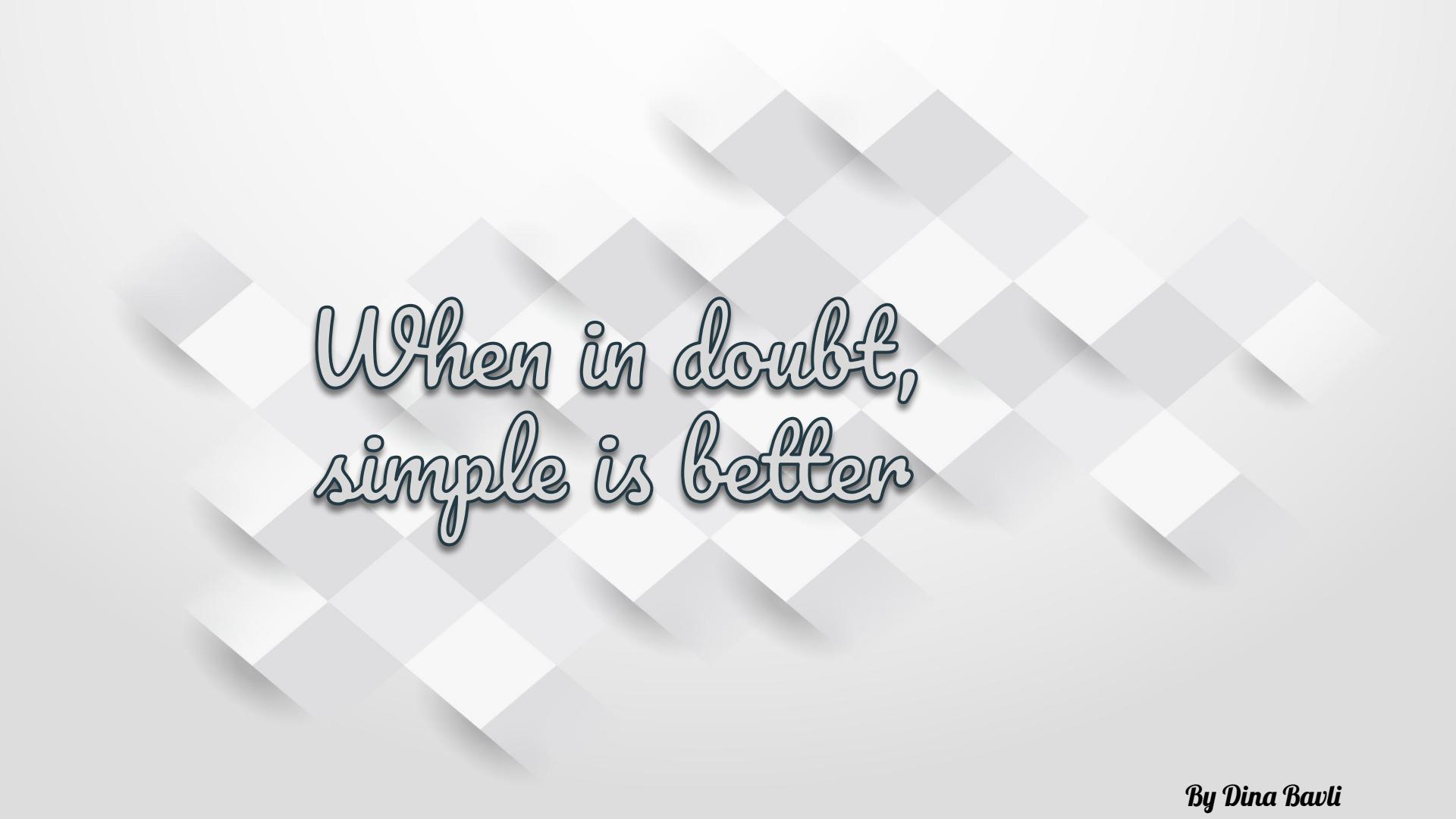
```
sentence = "i really like you but you can be so dumb some times"
```

```
deepmojify(sentence, top_n = 5)
```

```
/usr/local/lib/python3.6/dist-packages/torch nn/functional.py:1320: UserWarning
  warnings.warn("nn.functional.tanh is deprecated. Use torch.tanh instead")
/usr/local/lib/python3.6/dist-packages/torch nn/modules/container.py:92:
    input = module(input)
'i really like you but you can be so dumb some times 😢 😢 😔 😢 😢'
```

```
deepmojify('😢', top_n = 3)
```

```
/usr/local/lib/python3.6/dist-packages/torch nn/functional.py:1320: UserWarning
  warnings.warn("nn.functional.tanh is deprecated. Use torch.tanh instead")
/usr/local/lib/python3.6/dist-packages/torch nn/modules/container.py:92:
    input = module(input)
'😢 ❤️ 🌻 💙'
```



*When in doubt,
simple is better*

Introducing Emoji

```
pip install demoji
```

```
import demoji
```

```
demoji.download_codes()
```

```
tweet = """ #startspreachingthenews yankees win great start by 🎉 going 5strong innings with 5k's🔥 🐻 solo homerun 🚨🚨 with 2 solo homeruns and 🦸 3run homerun... 🎉 🚤 🧑 with rbi's ... 🔥🔥 MX and NI to close the game🔥🔥!!!....  
WHAT A GAME!!!.. """
```

```
demoji.findall(tweet)
```

```
{'MX': 'flag: Mexico',
'NI': 'flag: Nicaragua',
'🌋': 'volcano',
'🎅': 'Santa Claus: medium-dark skin tone',
'🐂': 'ox',
'👨\u200d🧑': 'man judge: medium skin tone',
'👹': 'ogre',
'🔥': 'fire',
'🚣': 'person rowing boat: medium-light skin tone',
'🤡': 'clown face'}
```

Appending emojis using demoji

```
tweet_data[ 'emojis' ].append(demoji.findall(status_json))
```

Questions?





Image by [Alexas Fotos](#) from [Pixabay](#)



<https://www.linkedin.com/in/dina-bavli-502430158/>



<https://github.com/dinbav>

By Dina Bavli