Twitter Mining with Python

Dajun Tian and Connie Zabarovskaya

Center for Biomedical Informatics

Washington University in St. Louis
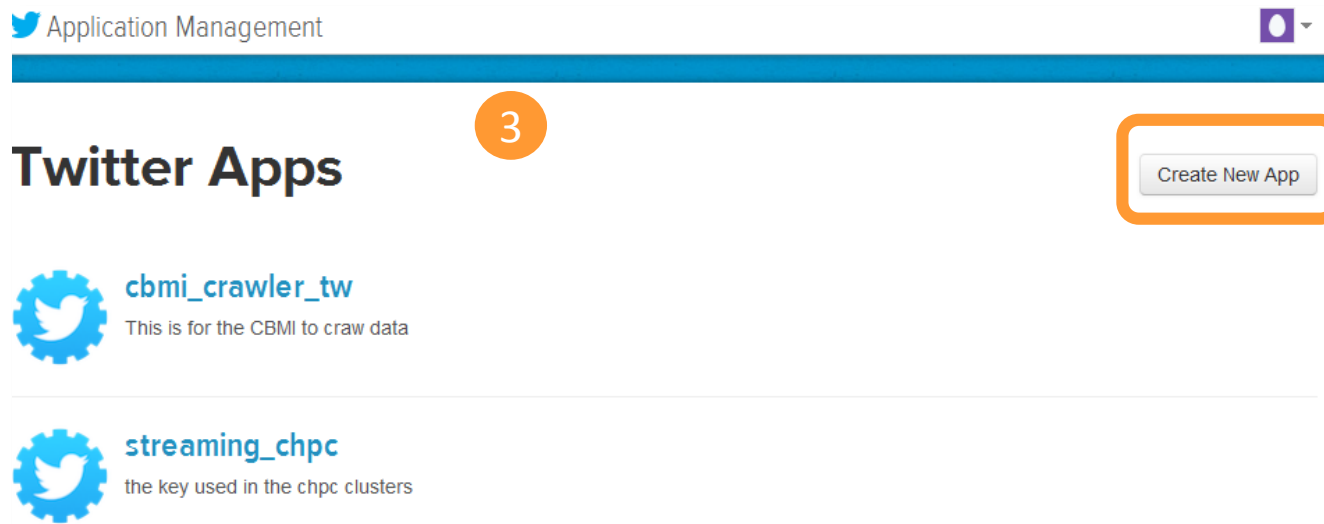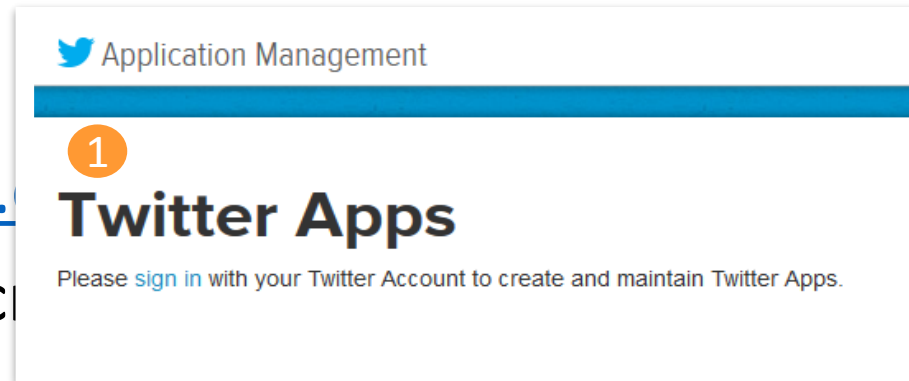
What you will learn

- Examples of research using Twitter
- How to create a developer account with Twitter
- How to import Twitter data into Python
- Understanding Twitter data and what JSON format is
- How to search Twitter data
- How to analyze Twitter data
- How to graph results
- How to export results into CSV format

Research Using Twitter

- "What can we learn about the Ebola outbreak from tweets?" (pubmed/26042846)

- Disease detection or public opinion reflection? Content analysis of tweets, other social media, and online newspapers during the measles outbreak in The Netherlands in 2013.: (pubmed/26013683)

- Tweet content related to sexually transmitted diseases: no joking matter. (pubmed/25289463)

- Discovering health topics in social media using topic models.: (pubmed/25084530)

How to create a developer account with Twitter

1. Go to: https://apps.twitter.c

2. Sign in using your Twitter cr

3. Click button to create new App (Application)

4. Fill out the new App form (see next slide)

Application Management

①

Twitter Apps

Please sign in with your Twitter Account to create and maintain Twitter Apps.

Application Management

Twitter Apps

③

Create New App

cbmi_crawler_tw
This is for the CBMI to craw data

streaming_chpc
the key used in the chpc clusters

How to create a developer account with Twitter (cont.)

Create an application

Application Details

Name *

Your application name. This is used to attribute the source of a tweet and in user-facing authorization screens. 32 characters max.

Description *

Your application description, which will be shown in user-facing authorization screens.

Website *

Your application's publicly accessible home page, where users can go to download, make use of, or find out more information about your application. This fully-qualified URL is used in the source attribution for tweets created by your application and will be shown in user-facing authorization screens.
(If you don't have a URL yet, just put a placeholder here but remember to change it later.)

Any URL relevant to you (e.g. biography page or social media page)

Callback URL

Where should we return after successfully authenticating? OAuth 1.0a applications should explicitly specify their oauth_callback URL on the request token step, regardless of the value given here. To restrict your application from using callbacks, leave this field blank.

How to create a developer account with Twitter (cont.)

- Note your Consumer Key and Consumer Secret

foo_bar

| Details | Settings | Keys and Access Tokens | Permissions |

Application Settings

Keep the "Consumer Secret" a secret. This key should never be human-readable in your application.

Consumer Key (API Key)	███████████████████	APP_KEY
Consumer Secret (API Secret)	██████████████████████	APP_SECRET
Access Level	Read and write (modify app permissions)	
Owner	DavidTianNull	
Owner ID	2738389812	

How to create a developer account with Twitter (cont.)

• Note your Access Token and Access Token Secret

Your Access Token

This access token can be used to make API requests on your own account's behalf. Do not share your access token secret with anyone.

Access Token	██████████████████	OAUTH_TOKEN
Access Token Secret	██████████████████	OAUTH_TOKEN_SECRET
Access Level	Read and write	
Owner	DavidTianNull	
Owner ID	2738389812	

Token Actions

Regenerate My Access Token and Token Secret   Revoke Token Access

Twitter API Types

- API: application program interface, or in simple terms accessing a program using functions

1.REST APIs

- programmatic access to read and write Twitter data

2.Streaming API

- monitor or process Tweets in real-time

- Parameters: **track**, **locations,** follow, delimited, stall_warnings

How to import Twitter data into Python

1. Import TwythonStreamer and json packages

2. Save twitter credentials into variables

3. Create a new class based on TwythonStreamer class with default or adjusted parameters

4. Initiate a connection to Twitter using the new class, filtering tweets based on either track, location, or source

5. Save tweets, while they come in through the connection

6. Optionally, save tweets into a json or text file, and then read in the json file or txt file into Python to analyze it

A Single Tweet in JSON Format

{"contributors": null, "truncated": false, "text": "Rue de la Fraternity 1/1\nTownscape acrylic painting\nSOLD available print on canvas\nhttp://t.co/dxIVBvJvMG http://t.co/u2SXS2enjm", "in_reply_to_status_id": null, "id": 544588839302340608, "favorite_count": 0, "source": "Twitter Web Client", "retweeted": false, "coordinates": null, "timestamp_ms": "1418675071918", "entities": {"symbols": [], "media": [{"expanded_url": "http://twitter.com/bobheatly/status/544588839302340608/photo/1", "display_url": "pic.twitter.com/u2SXS2enjm", "url": "http://t.co/u2SXS2enjm", "media_url_https": "https://pbs.twimg.com/media/B47EqdlCEAAgtZG.jpg", "id_str": "544588838731911168", "sizes": {"small": {"h": 340, "resize": "fit", "w": 340}, "large": {"h": 625, "resize": "fit", "w": 624}, "medium": {"h": 600, "resize": "fit", "w": 600}, "thumb": {"h": 150, "resize": "crop", "w": 150}}, "indices": [106, 128], "type": "photo", "id": 544588838731911168, "media_url": "http://pbs.twimg.com/media/B47EqdlCEAAgtZG.jpg"}], "hashtags": [], "user_mentions": [], "trends": [], "urls": [{"url": "http://t.co/dxIVBvJvMG", "indices": [83, 105], "expanded_url": "http://www.bobheatlyartist.com/#!town-/c18r3", "display_url": "bobheatlyartist.com/#!town-/c18r3"}]}, "in_reply_to_screen_name": null, "id_str": "544588839302340608", "retweet_count": 0, "in_reply_to_user_id": null, "favorited": false, "user": {"follow_request_sent": null, "profile_use_background_image": true, "default_profile_image": false, "id": 2382261656, "verified": false, "profile_image_url_https": "https://pbs.twimg.com/profile_images/443045908637155329/_n-9wh94_normal.jpeg", "profile_sidebar_fill_color": "DDEEF6", "profile_text_color": "333333", "followers_count": 1336, "profile_sidebar_border_color": "000000", "id_str": "2382261656", "profile_background_color": "C0DEED", "listed_count": 43, "profile_background_image_url_https": "https://pbs.twimg.com/profile_background_images/444501687168806912/R53r3-il.jpeg", "utc_offset": null, "statuses_count": 27551, "description": "ARTIST and award winning design educator", "friends_count": 2001, "location": "Joplin MO", "profile_link_color": "080808", "profile_image_url": "http://pbs.twimg.com/profile_images/443045908637155329/_n-9wh94_normal.jpeg", "following": null, "geo_enabled": true, "profile_banner_url": "https://pbs.twimg.com/profile_banners/2382261656/1404759282", "profile_background_image_url": "http://pbs.twimg.com/profile_background_images/444501687168806912/R53r3-il.jpeg", "name": "bob heatly", "lang": "en", "profile_background_tile": true, "favourites_count": 801, "screen_name": "bobheatly", "notifications": null, "url": "http://www.bobheatlyartist.com/wwwhttp://www.bobheatlyartist..bobheatlyartist.com/#!slide-show/c1hlx", "created_at": "Mon Mar 10 14:56:57 +0000 2014", "contributors_enabled": false, "time_zone": null, "protected": false, "default_profile": false, "is_translator": false}, "geo": null, "in_reply_to_user_id_str": null, "possibly_sensitive": false, "lang": "en", "created_at": "Mon Dec 15 20:24:31 +0000 2014", "filter_level": "none", "in_reply_to_status_id_str": null, "place": {"country_code": "US", "url": "https://api.twitter.com/1.1/geo/id/504ec2713c70e7a3.json", "country": "United States", "place_type": "city", "bounding_box": {"type": "Polygon", "coordinates": [[[-94.579066, 37.019415], [-94.579066, 37.173623], [-94.4173905, 37.173623], [-94.4173905, 37.019415]]]}, "full_name": "Joplin, MO", "attributes": {}, "id": "504ec2713c70e7a3", "name": "Joplin"}, "extended_entities": {"media": [{"expanded_url": "http://twitter.com/bobheatly/status/544588839302340608/photo/1", "display_url": "pic.twitter.com/u2SXS2enjm", "url": "http://t.co/u2SXS2enjm", "media_url_https": "https://pbs.twimg.com/media/B47EqdlCEAAgtZG.jpg", "id_str": "544588838731911168", "sizes": {"small": {"h": 340, "resize": "fit", "w": 340}, "large": {"h": 625, "resize": "fit", "w": 624}, "medium": {"h": 600, "resize": "fit", "w": 600}, "thumb": {"h": 150, "resize": "crop", "w": 150}}, "indices": [106, 128], "type": "photo", "id": 544588838731911168, "media_url": "http://pbs.twimg.com/media/B47EqdlCEAAgtZG.jpg"}]}}

JSON Format Explained

- JSON: **J**ava**S**cript **O**bject **N**otation.
- JSON is a syntax for storing and exchanging data.
- JSON is an easier-to-use alternative to XML.

Dictionary

{ "KEY": "VALUE" } ⟶ { "NAME": "JOHN" }

Dictionary containing Array

```
{ "Employees": [
        { "FirstName" : "John", "LastName" : "Doe"},
        { "FirstName" : "Jane", "LastName" : "Doe"},
]}
```

JSON in Pretty Format

- Sample only:

```
{
  "contributors": null,
  "truncated": false,
  "text": "Rue de la Fraternity 1\/1\nTownscape acrylic painting\nSOLD available print on canvas\nhttp:\/\/t.co\/dxIVBvJvMG http:\/\/t.co\/u2SXS2enjm",
  "in_reply_to_status_id": null,
  "id": 5.4458883930234e+17,
  "favorite_count": 0,
  "source": "<a href=\"http:\/\/twitter.com\" rel=\"nofollow\">Twitter Web Client<\/a>",
  "retweeted": false,
  "coordinates": null,
  "timestamp_ms": "1418675071918",
… }
```

http://jsonprettyprint.com

Common Analyses of Twitter Data

- **Sentiment analysis:** extracting opinions, emotions, and sentiments as text, and evaluate them as negative or positive
  - Packages: **NLTK**, **textblob, vaderSentiment**
- **Network analysis**: identify important users and important topics
  - Packages: **NetworkX, Graph-tool**
- **Topic modeling:** understand the topics the users talk about – e.g. through automatic topic discovery with latent Dirichlet allocation (LDA)
  - Packages: **lda, gensim**

Hands-on

Resources

- http://jsonprettyprint.com/

- https://docs.python.org/2/library/json.html

- https://instagram.com/developer

- https://github.com/Instagram/python-instagram

- https://dev.twitter.com/

- https://twython.readthedocs.org/en/latest/index.html

- Analyses:
  http://tweettracker.fulton.asu.edu/tda/TwitterDataAnalytics.pdf

Thank you!

Questions?