

Gathering Data from Twitter

Diogo Pacheco – dpacheco2013@my.fit.edu

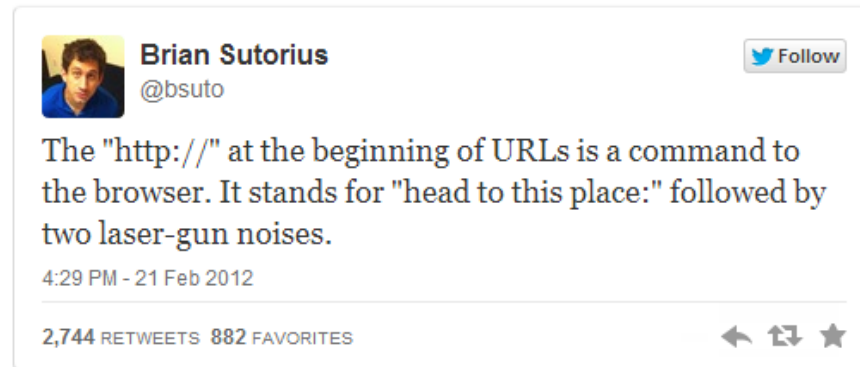
Marcos Oliveira - moliveirajun2013@my.fit.edu

Adapted from 2013 Hugo Serrano's Lecture



Introduction

- What is Twitter
 - Twitter is an online social network that allows users to send and read messages of up to 140 characters, known as **tweets**.



<http://twitter.com>



Introduction

- Twitter
 - online social network
 - Users can follow and be followed by other users.
 - a real time information network
 - Tweets can be replied, retweeted



Twitter API

- Search API
 - Finding a set of tweets
 - Tweets from a particular user
 - Tweets with specific keywords
 - Has a rate limit
- REST API
 - Access core primitives of Twitter
 - Timelines
 - Retweet other tweets
- Streaming API
 - Access a sample of Twitter Firehose for Data intensive Applications
 - Data Mining
 - Analytics research
 - Large quantity of keywords
 - Requires a long-lived HTTP connection



<https://dev.twitter.com/start>

Twitter API

- Twitter API limits the number of requests in a given time window
 - <https://dev.twitter.com/docs/rate-limiting/1.1>
 - <https://dev.twitter.com/docs/rate-limiting/1.1/limits>



Twitter platform objects



Twitter API

- The most frequently observed objects in Twitter ecosystem:
 - Tweets
 - Users
 - Entities
 - Places



Twitter API

- Tweets
 - Basic building blocks
 - Also known as Status Updates
 - Fields such as
 - **created_at**: UTC time when this Tweet was created.
 - **coordinates**: the geographic location of this Tweet
 - **in_reply_to_screen_name**: the name of the original Tweet's author.
 - **in_reply_to_status_id**: the id of the original Tweet's author.
 - **retweeted_status**: a representation of the *original* Tweet that was retweeted.

<https://dev.twitter.com/docs/platform-objects/tweets>



Twitter API

- Users
 - They tweet, follow, retweet, are mentioned etc.
 - Fields such as:
 - id: the unique identification of a user
 - screen_name: the screen name, handle, or alias that this user identifies themselves with
 - followers_count: the number of followers this account currently has

<https://dev.twitter.com/docs/platform-objects/users>



Twitter API

- Entities
 - provide metadata and additional information about content posted on Twitter.
 - media
 - urls
 - hashtags
 - mentions

<https://dev.twitter.com/docs/platform-objects/entities>



Twitter API

- Places
 - specific, named locations with corresponding geo coordinates
 - country
 - bounding_box
 - coordinates
 - full_name
 - name
 - place_type

<https://dev.twitter.com/docs/platform-objects/places>



Twitter API

- Authentication
 - Create an application
 - Create Access Tokens
 - Use the OAuth tool to get your keys, tokens and secrets



Useful Hints

- For some libraries (python), the Twitter API is the documentation for developers - <https://dev.twitter.com/docs/api/1.1;>

Search

Find relevant Tweets based on queries performed by your users.

Resource	Description
GET search/tweets	Returns a collection of relevant Tweets matching a specified query. Please note that Twitter's search service and, by extension, the Search API is not meant to be an exhaustive source of Tweets. Not all Tweets will be indexed or made available via the search interface. In API v1.1, the response...
<pre>-> api.search.tweets(q="#HowToComplexNetworks")</pre>	

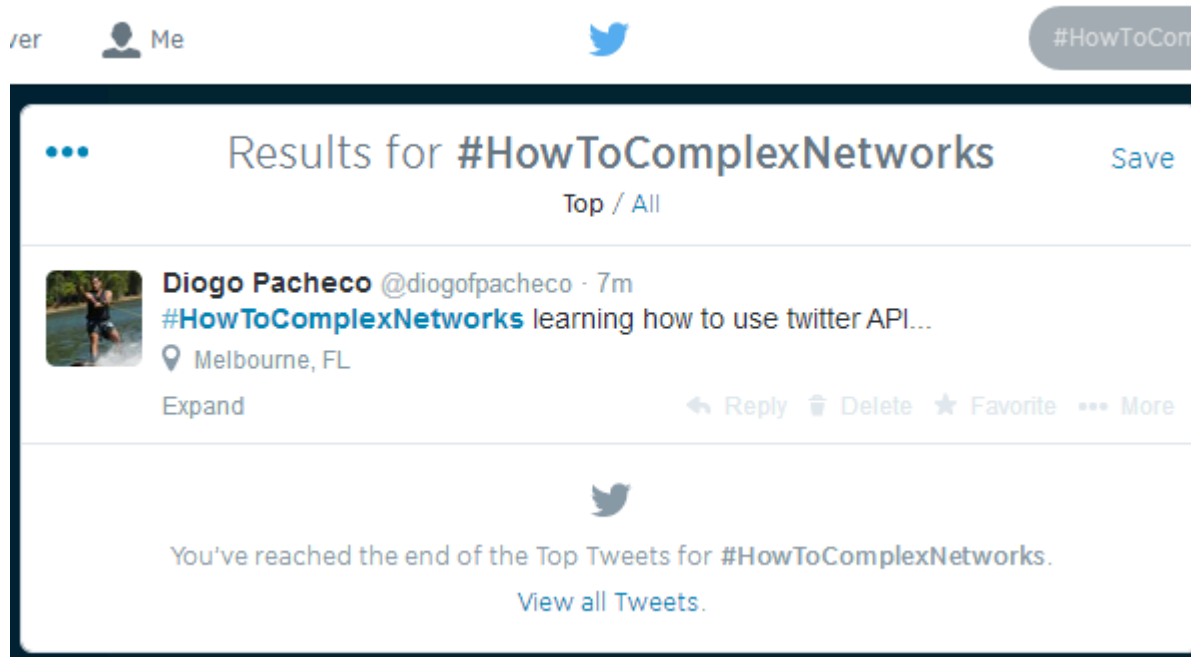
Streaming

Resource	Description
POST statuses/filter	Returns public statuses that match one or more filter predicates. Multiple parameters may be specified which allows most clients to use a single connection to the Streaming API. Both GET and POST requests are supported, but GET requests with too many parameters may cause the request to be...
<pre>-> api.statuses.filter(track="#HowToComplexNetworks")</pre>	



Useful Hints

- Try first Twitter App vs. Console before coding - <https://dev.twitter.com/console>;



Useful Hints

Exploring the Twitter API

Getting Started

View

What links here

Updated on Thu, 2013-01-17 08:58

Explore the API on your own machine using [Twitter for Mac](#)'s developer console or our command-line [Twurl](#) tool.

Explore the API here by using this console provided by [@apigee](#).

Service

https://api.twitter.com/1.1

Authentication

twitter-diogofpacheco

powered by **apigee**

Select an API method

Search methods...

Timelines

GET

/statuses/mentions_timeline.json

GET

/statuses/user_timeline.json

GET

/statuses/home_timeline.json

Tweets

GET

/statuses/retweets/{id}.json

GET

/statuses/show/{id}.json

POST

/statuses/destroy/{id}.json

POST

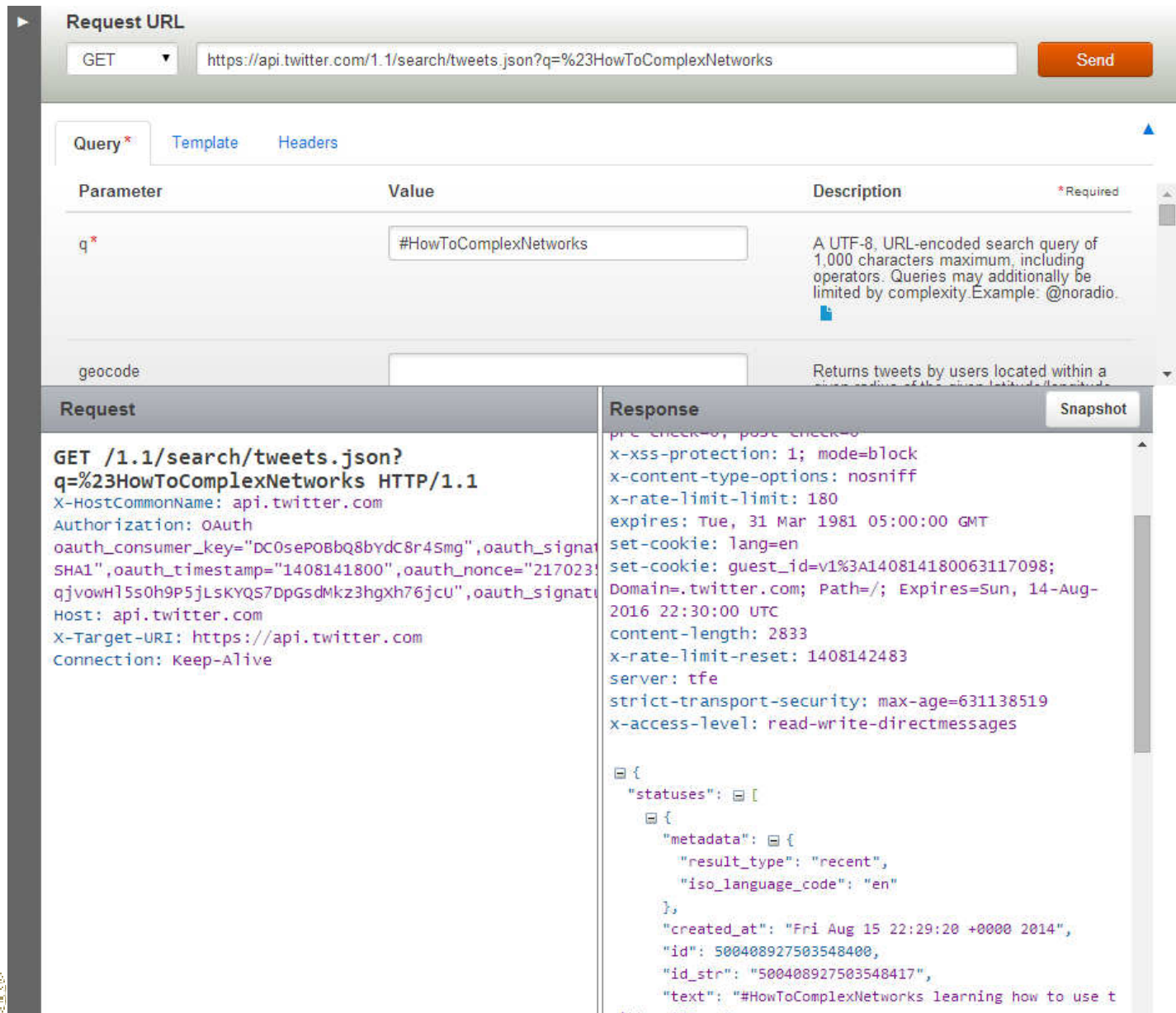
/statuses/update.json

Response

Snapshot



Useful Hints



Request URL

GET

Query **Template** **Headers**

Parameter	Value	Description	* Required
q	<input type="text" value="#HowToComplexNetworks"/>	A UTF-8, URL-encoded search query of 1,000 characters maximum, including operators. Queries may additionally be limited by complexity. Example: @noradio.	
geocode	<input type="text"/>	Returns tweets by users located within a given radius of the given latitude/longitude.	

Request

```
GET /1.1/search/tweets.json?
q=%23HowToComplexNetworks HTTP/1.1
X-HostCommonName: api.twitter.com
Authorization: OAuth
oauth_consumer_key="dc0sePO8bq8bydc8r4smg",oauth_signature="...",
oauth_timestamp="1408141800",oauth_nonce="217023...",
oauth_version="1.0",oauth_consumer_key="dc0sePO8bq8bydc8r4smg",
oauth_timestamp="1408141800",oauth_nonce="217023...",
Host: api.twitter.com
X-Target-URI: https://api.twitter.com
Connection: Keep-Alive
```

Response

```
HTTP/1.1 200 OK
x-xss-protection: 1; mode=block
x-content-type-options: nosniff
x-rate-limit-limit: 180
expires: Tue, 31 Mar 1981 05:00:00 GMT
set-cookie: lang=en
set-cookie: guest_id=v1%3A140814180063117098; Domain=.twitter.com; Path=/; Expires=Sun, 14-Aug-2016 22:30:00 UTC
content-length: 2833
x-rate-limit-reset: 1408142483
server: tfe
strict-transport-security: max-age=631138519
x-access-level: read-write-directmessages

{
  "statuses": [
    {
      "metadata": {
        "result_type": "recent",
        "iso_language_code": "en"
      },
      "created_at": "Fri Aug 15 22:29:20 +0000 2014",
      "id": 500408927503548400,
      "id_str": "500408927503548417",
      "text": "#HowToComplexNetworks learning how to use t"
    }
  ]
}
```



Useful Hints

- Check rate limits in your code – try/catch loops
- Do you need database? Relational one?
- Remember 2-step process:
 - Collect data
 - Construct the network
- Intensive traffic/volume, try MongoDB



Useful Hints?

- 2014 FIFA World Cup Dataset
 - +50 millions tweets
 - +1 million geo-tagged tweets
 - 1 month window period
- Database Optimization
- Memory Consuming
- Programs Crashes



Twitter api resources



Twitter API - Resources

- Timelines
- Search
- Tweets
- Streaming
- Direct Messages
- Friends & Followers
- Users
- Suggested Users
- Favorites
- Lists
- Saved Searches
- Places & Geo
- Trends
- Spam Reporting
- OAuth
- Help



<https://dev.twitter.com/docs/api/1.1>

Twitter API - Resources

- Search
 - Let's search for 5 most recent tweets with “Florida Tech” on it;
 - Let's search for 5 most recent tweets within 1mi;

<https://dev.twitter.com/console>



Twitter API – Resources

- Friends & Followers
 - Let's query for the friends of a particular user
 - Let's query for the first 100 followers of a particular user



Twitter API - Resources

- Users
 - Let's query for the usernames of the first 100 followers of a particular user



Twitter API - Resources

- Streaming
 - Let's track tweets with the hashtag #cse5656 and #netsci
 - Let's track the tweets in Brevard County



Questions?

Diogo Pacheco – dpacheco2013@my.fit.edu

Marcos Oliveira - moliveirajun2013@my.fit.edu

<http://my.fit.edu/~dpacheco2013/#classes>

