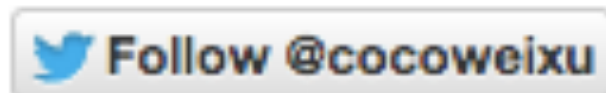


# Social Media & Text Analysis

## lecture 2 - Twitter API



**CSE 5539-0010 Ohio State University**

**Instructor: Wei Xu**

**Website: [socialmedia-class.org](http://socialmedia-class.org)**

# Course Website

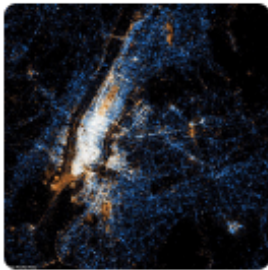
[socialmedia-class.org](http://socialmedia-class.org)

Social Media & Text Analytics

Syllabus

Twitter API Tutorial

Homework ▾



*A visualization showing the location of Twitter messages (blue) and Flickr photos (orange) in New York City by Eric Fischer*

Social media provides a massive amount of valuable information and shows us how language is actually used by lots of people. This course will give an overview of prominent research findings on language use in social media. The course will also cover several machine learning algorithms and the core natural language processing techniques for obtaining and processing Twitter data.

## Instructor

[Wei Xu](#) is an assistant professor in the Department of Computer Science and Engineering at the Ohio State University. Her research interests lie at the intersection of machine learning, natural language processing, and social media. She holds a PhD in Computer Science from New York University. Prior to joining OSU, she was a postdoc at the University of Pennsylvania. She is organizing the ACL/COLING [Workshop on Noisy User-generated Text](#), serving as a workshop co-chair for ACL 2017, an area chair for EMNLP 2016 and the publicity chair for NAACL 2016.

## Time/Place new

**Fall 2016, CSE 5539-0010 The Ohio State University**

**Cockins Hall Room 218 | Wednesday 2:20PM – 4:10PM**

dual-listed undergraduate and graduate course

## Prerequisites

In order to succeed in this course, you should know basic probability and statistics, such as the chain rule of probability and Bayes' rule. On the programming side, all projects will be in Python. You should understand basic computer science concepts (like recursion), basic data structures (trees, graphs), and basic algorithms (search, sorting, etc).

## Course Readings

[Various academic papers](#)

## Previous Offerings

Summer 2016, [The North American Summer School on Logic, Language, and Information \(NASSLLI\)](#)

Teaching evaluation was 5.72 out of 6 at NASSLLI; average across all instructors was 5.23.

Summer 2015, University of Pennsylvania (where this course was first designed and taught)

# This is a **Special Topic Class**

- It is about NLP **research**, not programming.  
(pre-requirements: familiar with Python programming)
- As students request, I am planning to teach deep learning and have a homework on implementing word2vec — see **Quiz #2**
- It is **difficult** (meant for **PhD** students, though open to advanced master and undergraduate students).
- **No TA and no grader**

# Alternatives

- **audit** the course (won't hurt your GPA)
- take this course **next year** (most likely will re-occur)
- **other related courses:**
  - CSE 5525 Foundations of Speech and Language Processing
  - CSE 5523 Machine Learning
  - CSE 5522 Survey of Artificial Intelligence II: Advanced Techniques
  - CSE 5526 Introduction to Neural Networks

# Have a Question?

- **Ask in class!**
- **Office Hour:** Wed 4:15 pm — 5:15 pm, Dreese 495
- **Piazza Q&A Board** (a Module within OSU Canvas)

The screenshot shows the Piazza interface for the course AU 16 5539. The top navigation bar includes the Piazza logo, the course ID, and links to Q & A, Resources, Statistics, and Manage Class. Below this, there are tabs for polls, homework (hw1-hw4), and a filter bar with Unread, Updated, Unresolved, and Following. A 'New Post' button and a search bar are also visible. The main content area is divided into sections: 'PINNED' with a post titled 'Search for Teammates!' dated 8/30/16, and 'TODAY' with a post titled 'Welcome to CSE 5539-0010 (...)' dated 12:59AM. The 'YESTERDAY' section is partially visible at the bottom. The right sidebar shows a 'note' section with the same welcome message as the pinned post.

piAZZA AU 16 5539 Q & A Resources Statistics Manage Class

polls hw1 hw2 hw3 hw4

Unread Updated Unresolved Following

New Post Search or add a post...

PINNED

Private Search for Teammates! 8/30/16 1

TODAY

Instr Welcome to CSE 5539-0010 (... 12:59AM

Hi All, Welcome to CSE 5539-0010: Social Media and Text Analytics. Please remember to bring your laptop (if you have o

YESTERDAY

note

Welcome to CSE 5539-0010 (and bring yo

Hi All,

Welcome to CSE 5539-0010: Social Media and Text Analytics.

Please remember to bring your laptop (if you have one) to the cla and try it out in the class!

The course homepage is: <http://socialmedia-class.org/>

# Updated Grading

- two programming assignments (50 points)
- a 3rd assignment or a research project (30 points)
- in-class presentation (20 points)
- several Quiz/Survey (bonus 10 points)
- **in-class and Piazza participation (bonus 10 points)**

# Quiz #1

- For events  $A$  and  $B$ , prove

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

# Quiz #1

- What does this regular expression mean?

147	Hashtag = "#[a-zA-Z0-9_]+"
-----	----------------------------



# Quiz #1

- Softmax function is defined as  $\text{softmax}(\mathbf{x})_i = \frac{e^{x_i}}{\sum_j e^{x_j}}$
- prove

$$\text{softmax}(\mathbf{x}) = \text{softmax}(\mathbf{x} + c)$$

# Quiz #2

- implement Softmax function in Python (need to be computationally efficient)

```
1 import numpy as np
2
3 def softmax(x):
4
5     # Handle the special case: when vector x is only 1-dimensional
6     if x.ndim <= 1:
7         x = x - np.max(x)
8         ex = np.exp(x)
9         return ex / np.sum(ex)
10
11
12     ### YOUR CODE HERE -- for vectors that are not 1-dimensional
13
14
15
16
17
18
19     ### END YOUR CODE
20
21     return dist
22
23
24 # Check your Softmax implementation
25
26 print softmax(np.array([[101,102],[-1,-2]]))
```

# Homework #1 is out

## Due in a week



- **Canvas:** submit homework
- **Piazza:**
  - Q&A Discussion
  - Announcements

# Twitter API Tutorial: [socialmedia-class.org](https://socialmedia-class.org)

Social Media & Text Analytics

Syllabus

Twitter API Tutorial

Homework Assignments ▾



*Twitter's 404 error page --  
the Fail Whale*

## Twitter API tutorial

by [Wei Xu](#) (July 1, 2015)

[Follow @cocoweixu](#)

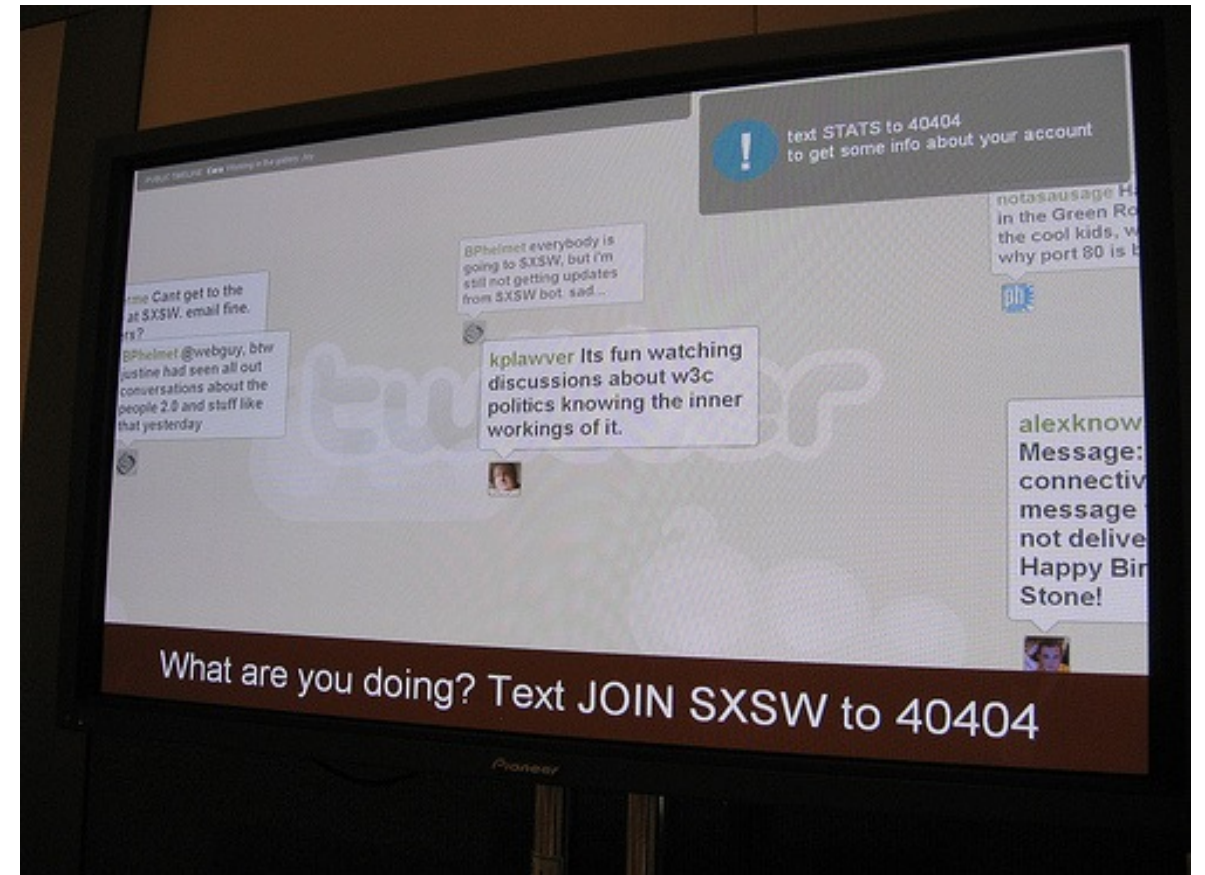
### 1. Getting Twitter API keys

To start with, you will need to have a Twitter account and obtain credentials from the Twitter developer site to access the Twitter API, following these steps:

- Create a Twitter user account if you do not already have one.
- Go to <https://apps.twitter.com/> and log in with your Twitter user account.
- Click "Create New App"

# Twitter History

- Jack Dorsey's idea (a NYU undergraduate then)
- 1st tweet on March 21, 2006
- exploded at SXSW 2007 (20k → 60k tweets/day)
- 100m tweets/quarter in 2008, 50m tweets/day in 2010, 400m tweets/day in 2013
- Huge API usage was unexpected as was the rise of the @ sign for replies



*Twitter staff received the festival's Web Award prize with the remark "we'd like to thank you in 140 characters or less. And we just did!"*



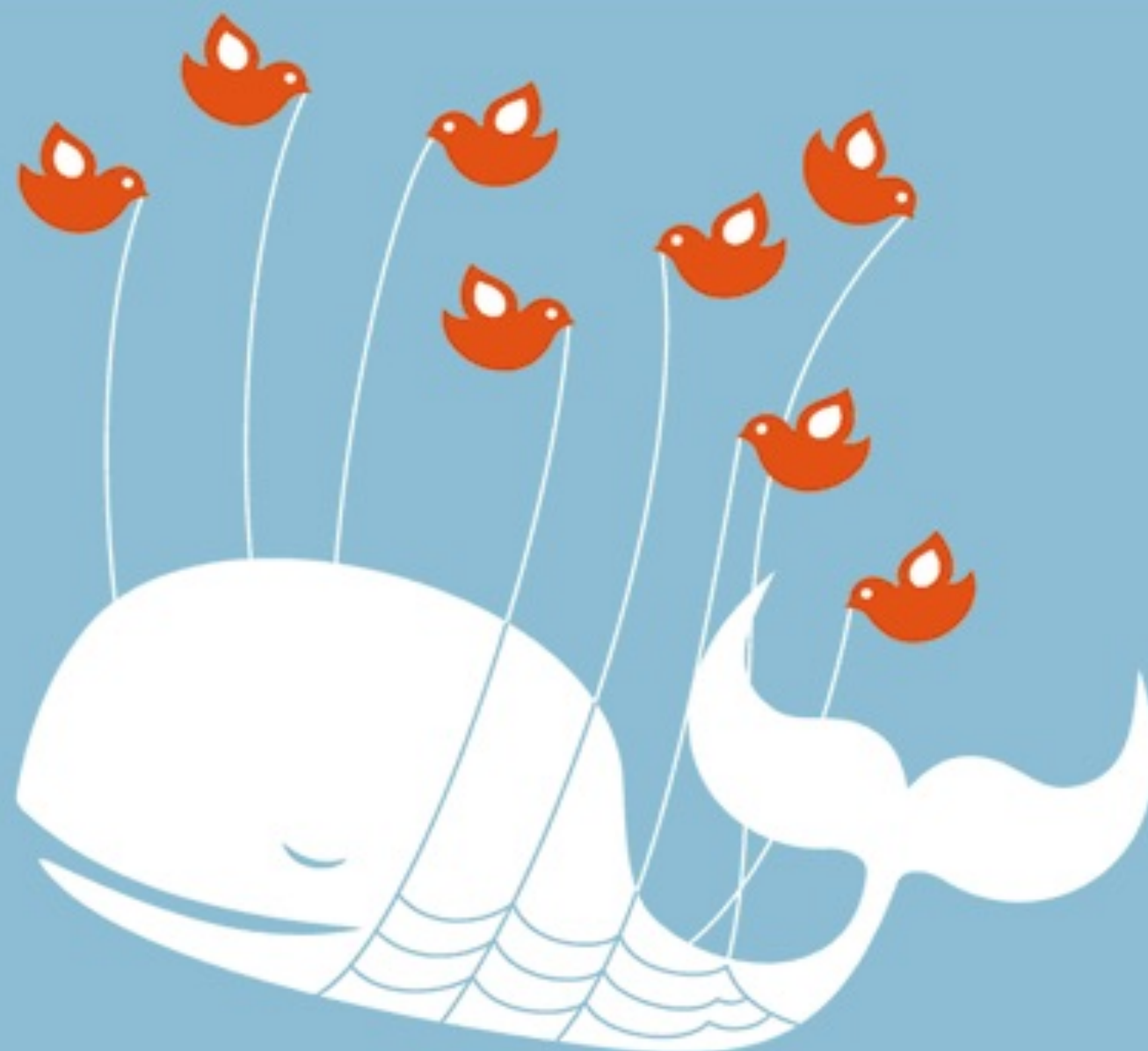
# Twitter History

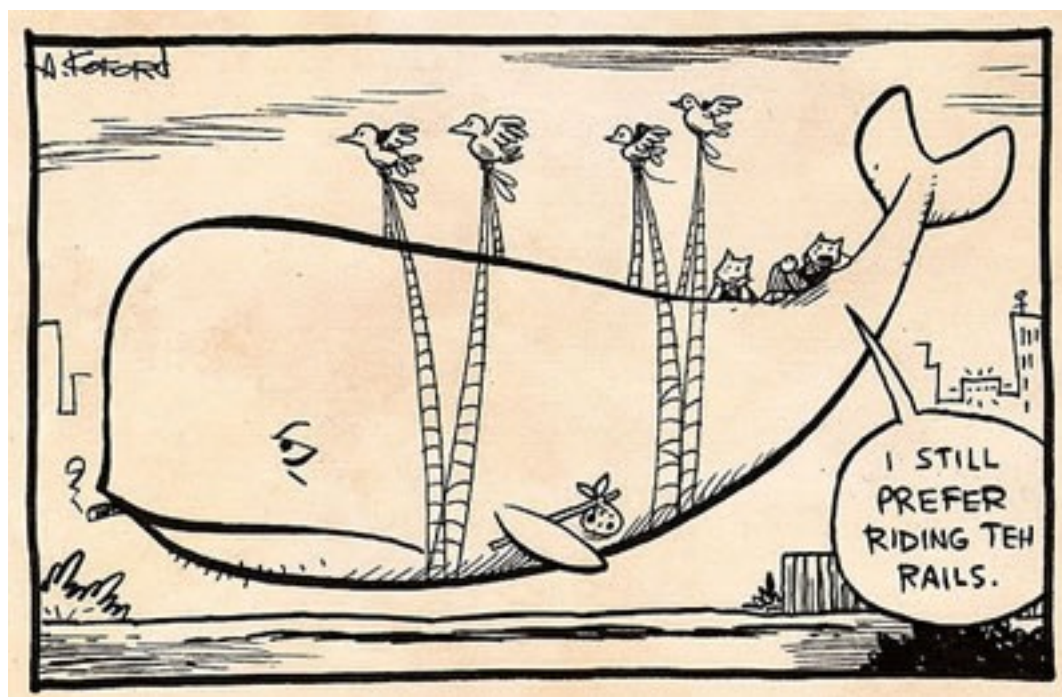
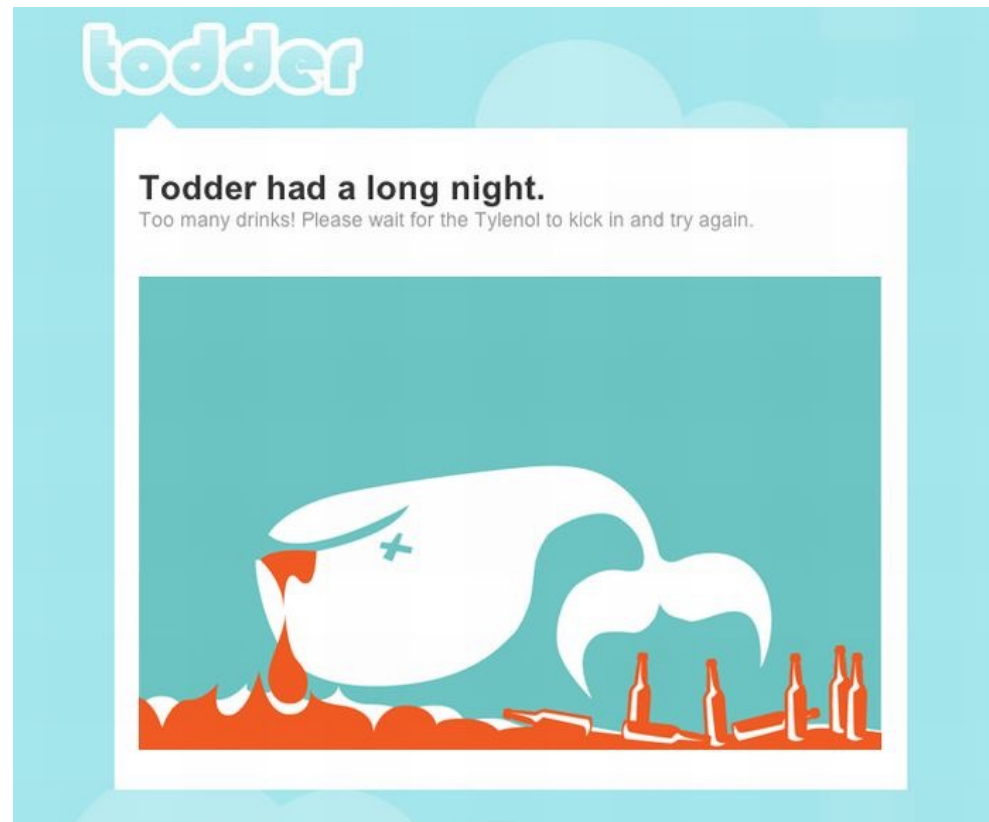
- IPO in 2013 Q4
- market value \$24b, revenue \$435m, net loss \$162m in 2015 Q1
- CEO Dick Costolo resigned July 1st, 2015



# Twitter is over capacity.

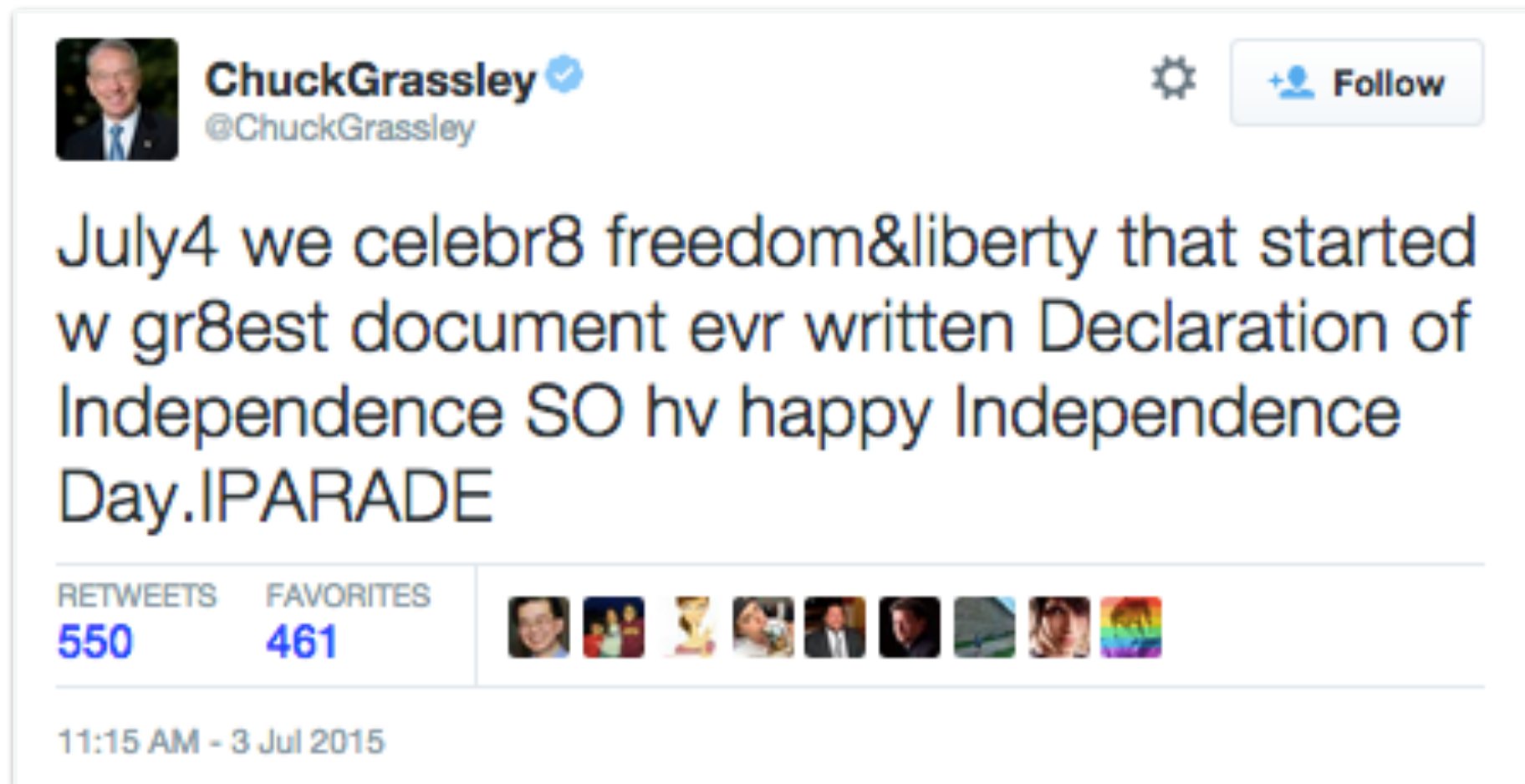
Please wait a moment and try again. For more information, check out [Twitter Status](#) »

[English](#)[Deutsch](#)[Español](#)[Français](#)[Italiano](#)[日本語](#)

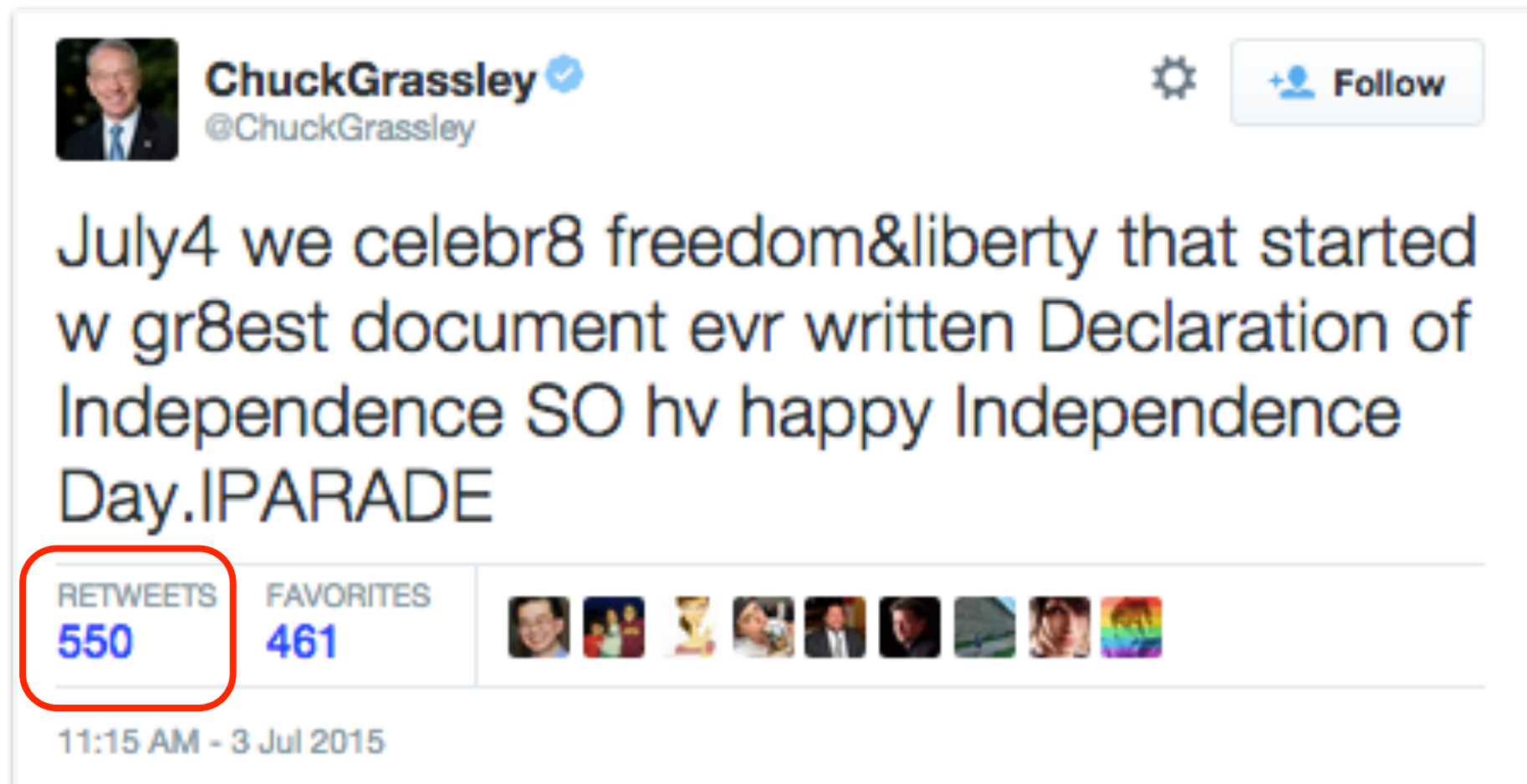




# Tweets



# ReTweets

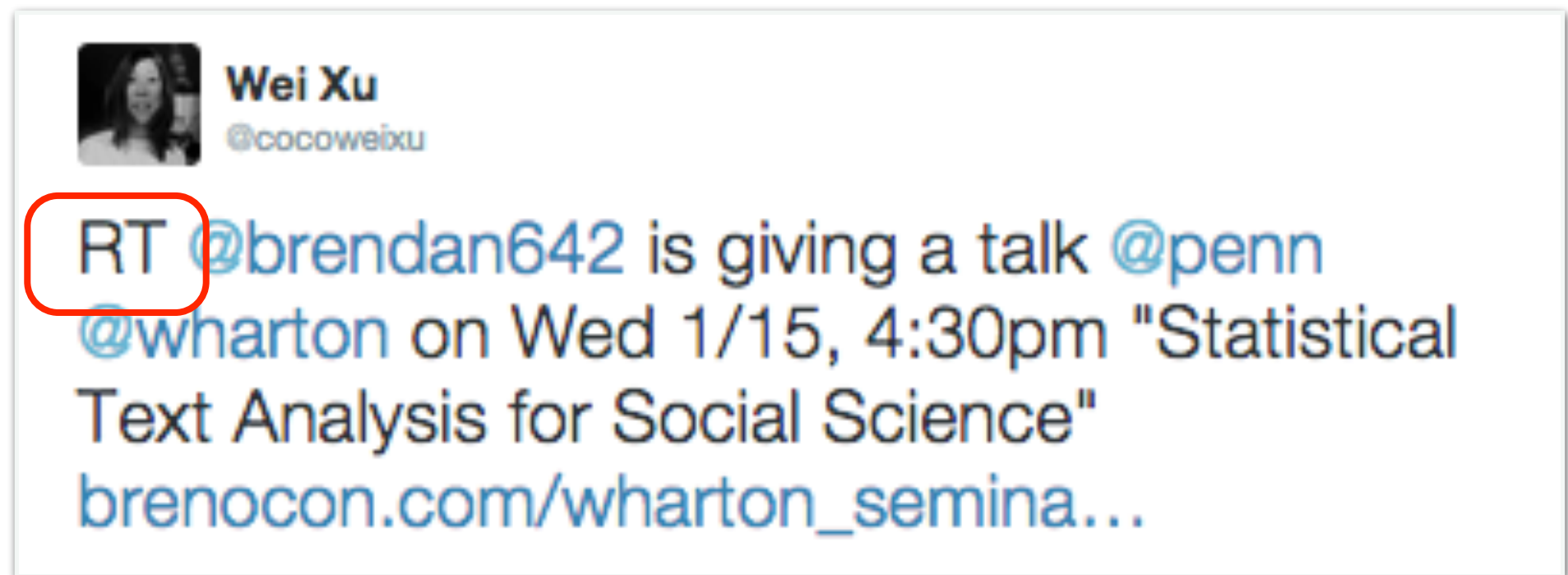


a re-posting of someone else's Tweet

Disclaimer:

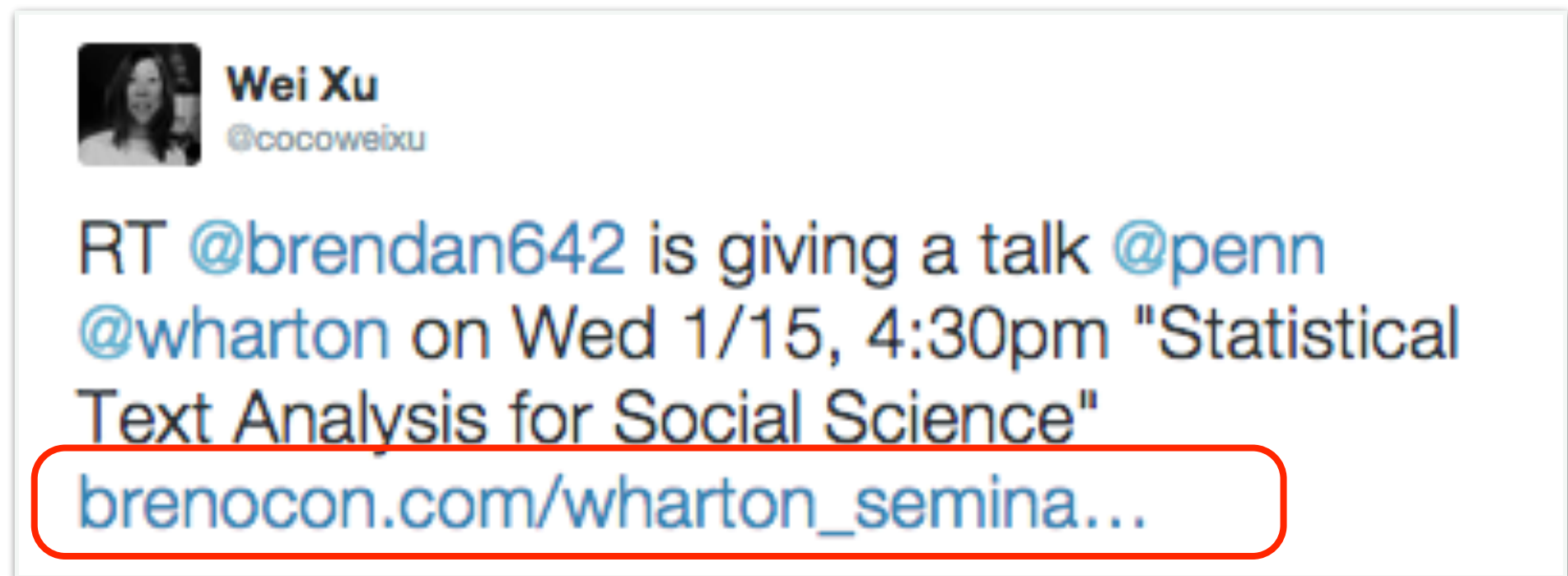
I am going to use many of my own tweets as examples in the slides ...

# ReTweets



- not an official Twitter feature
- often signifies quoting another user
- sometimes creates problems for data analytics

# Embedded Links



- shortened for display

# Embedded Links



- can provide extra external information for text processing



# Mentions

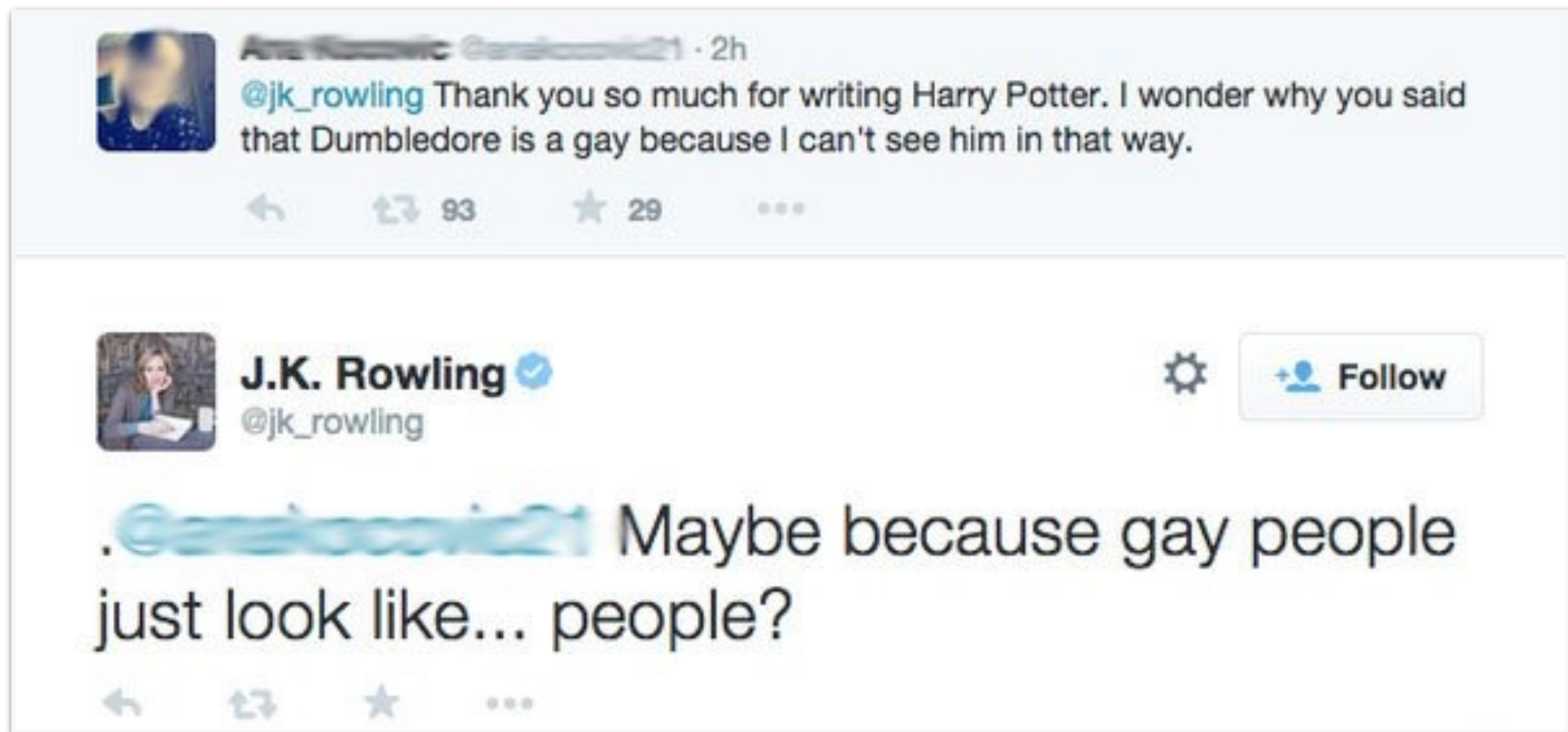


**Wei Xu**  
@cocoweixu

RT [@brendan642](#) is giving a talk [@penn](#)  
[@wharton](#) on Wed 1/15, 4:30pm "Statistical  
Text Analysis for Social Science"  
[brenocon.com/wharton\\_semina...](http://brenocon.com/wharton_semina...)

- user's @username anywhere in the body of the Tweet

# Replies/Conversations



- Tweet starts with a @username



# Replies/Conversations



- can have multi-round conversations

# Images



**Wei Xu**  
@cocoweixu

I wrote an ultimate Twitter API tutorial:  
[socialmedia-class.org/twittertutorial...](http://socialmedia-class.org/twittertutorial...)  
[#datascience](#) [#nlproc](#) [@twitterapi](#)

[Social Media & Text Analytics](#) [Syllabus](#) [Twitter API Tutorial](#) [Homework Assignments](#) ▾



Twitter's 404 error page --  
the Fail Whale

## Twitter API tutorial

by [Wei Xu](#) (July 1, 2015) [Follow @cocoweixu](#)

### 1. Getting Twitter API keys

To start with, you will need to have a Twitter account and obtain credentials from the Twitter developer site to access the Twitter API, following these steps:

- Create a Twitter user account if you do not already have one.
- Go to <https://apps.twitter.com/> and log in with your Twitter user account.
- Click "Create New App"

RETWEETS

45

LIKES

79

11:55 AM - 2 Jul 2015

45 79

# Hashtags



**Wei Xu**  
@cocoweixu

I wrote an ultimate Twitter API tutorial:  
[socialmedia-class.org/twittertutorial...](http://socialmedia-class.org/twittertutorial...)  
[#datascience](#) [#nlproc](#) @twitterapi

Social Media & Text Analytics

Syllabus

Twitter API Tutorial

Homework Assignments ▾



Twitter's 404 error page --  
the Fail Whale

## Twitter API tutorial

by [Wei Xu](#) (July 1, 2015)

[Follow @cocoweixu](#)

### 1. Getting Twitter API keys

To start with, you will need to have a Twitter account and obtain credentials from the Twitter developer site to access the Twitter API, following these steps:

- Create a Twitter user account if you do not already have one.
- Go to <https://apps.twitter.com/> and log in with your Twitter user account.
- Click "Create New App"

RETWEETS

45

LIKES

79



11:55 AM - 2 Jul 2015



45



79



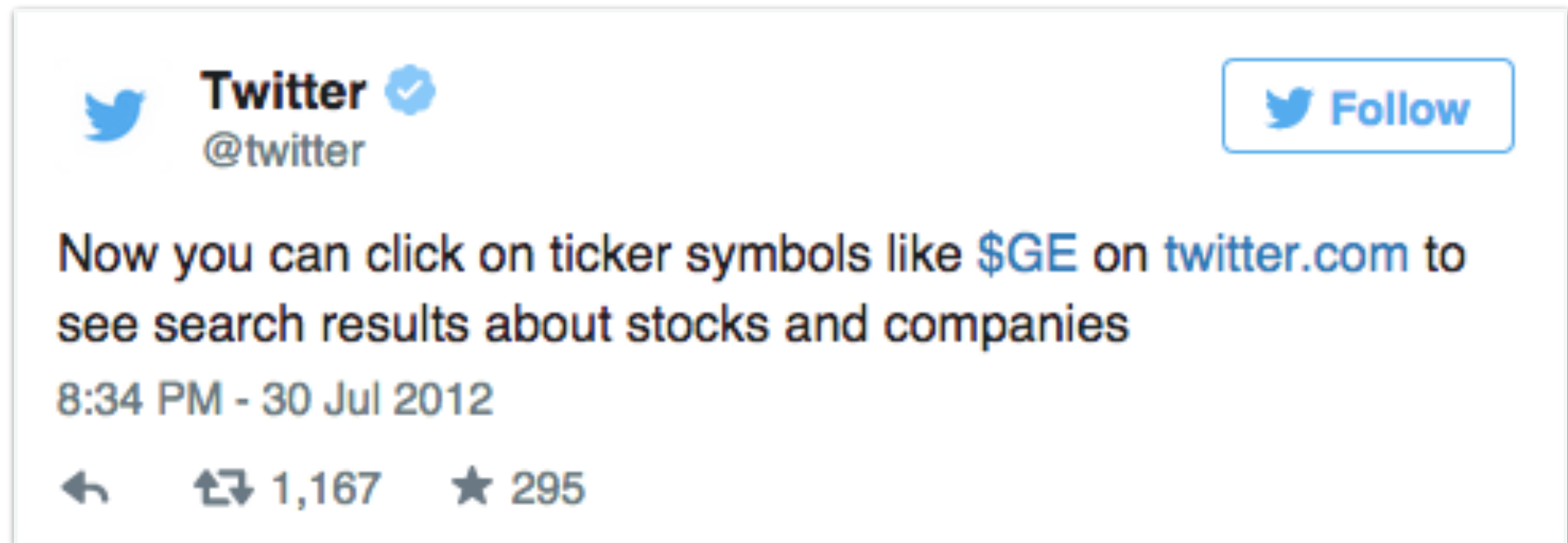




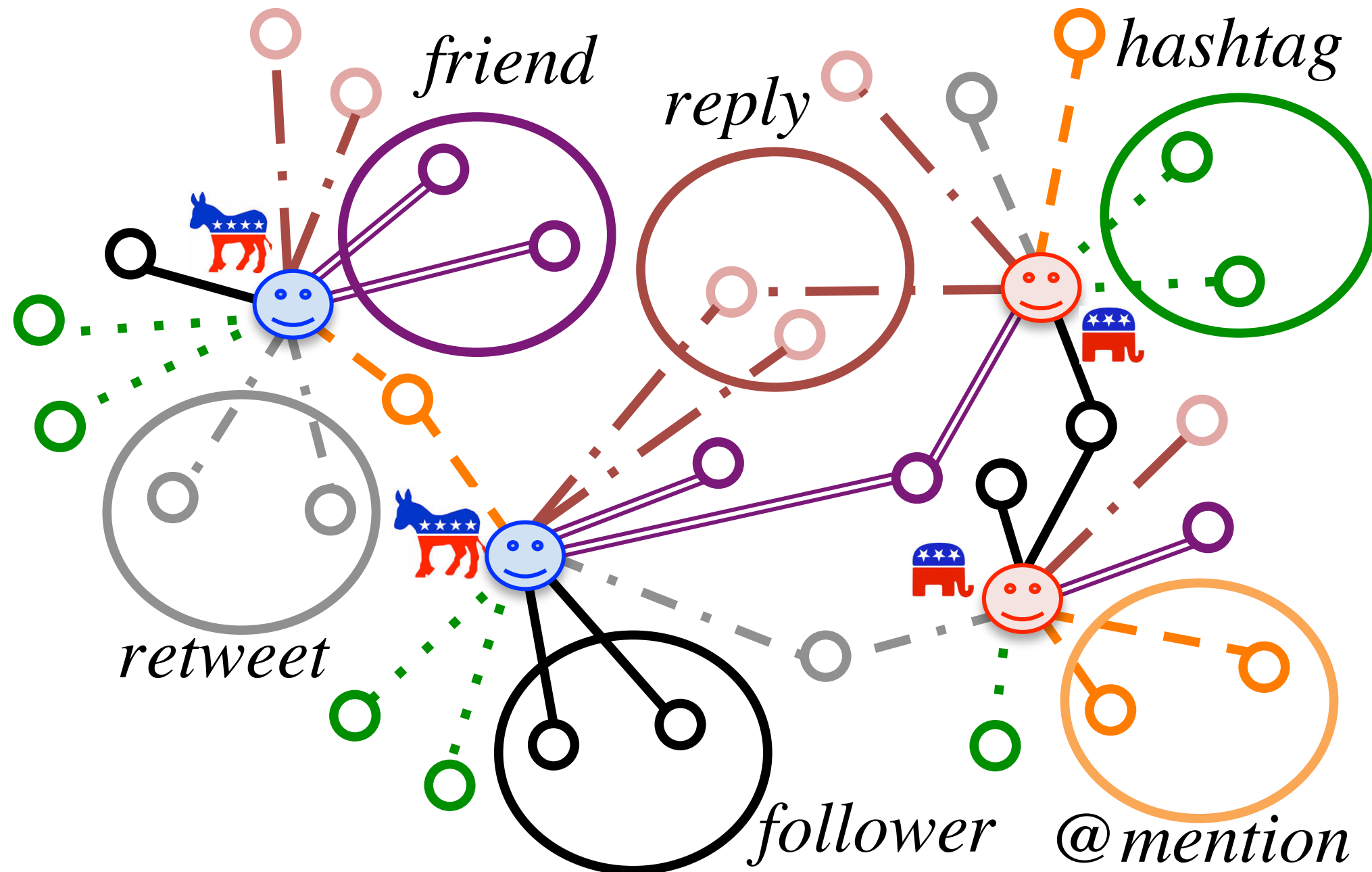
hashtags are powerful



# Cashtags



# Twitter's Social Graph



Source: Volkova, Van Durme, Yarowsky, Bachrach  
"Tutorial on Social Media Predictive Analytics" NAACL 2015

# Twitter API

# What is an API?

**A**pplication **P**rogramming **I**nterface

API is a set of protocols that specify how software programs communicate with each other.



# What is an API?

## **Without API:**

An app finds the current weather in London by opening <http://www.weather.com/> and reading the webpage like a human does, interpreting the content.

## **With API:**

An app finds the current weather in London by sending a message to the [weather.com](#) API (in a structured format like XML). The [weather.com](#) API then replies with a structured response.

# Two Most Popular APIs

Streaming API	REST API
a sample of public tweets and events as they published on Twitter (can specify search terms or users)	<ul style="list-style-type: none"><li>- search</li><li>- trends</li><li>- read author profile and follower data</li><li>- post / modify</li></ul>
<b>only</b> real-time data	historical data up to a week
continuous net connection	one-time request
no limit	rate limit (varies for different requests)

# OAuth

- Twitter uses OAuth to provide authorized access to its API.
- which means, to start with needs:
  - a Twitter account
  - OAuth access tokens from [apps.twitter.com](https://apps.twitter.com)

OAuth settings	
Your application's OAuth settings. Keep the "Consumer secret" a secret. This key should never be human	
Access level	Read-only <a href="#">About the application permission model</a>
Consumer key	1234567890
Consumer secret	NZsJqxVPe4IP1XebbXtAXpLYrQZcg4RIfCjuXbzjAk4

# Python Twitter Tools

The screenshot shows the PyPI (Python Package Index) page for the 'twitter' package, version 1.17.1. The browser address bar shows the URL 'https://pypi.python.org/pypi/twitter'. The page features the Python logo and a search bar. The main content area displays the package name 'twitter 1.17.1' and a description: 'An API and command-line toolset for Twitter (twitter.com)'. A green 'Downloads ↓' button is visible. Below the description, there are links for build status and coverage status. The left sidebar contains a 'PACKAGE INDEX' menu with links to browse packages, submission, classifiers, and more. The right sidebar shows a 'Not Logged In' status with links for login, register, and other user actions.

Python Software Foundation [US] <https://pypi.python.org/pypi/twitter>

python™

» Package Index > twitter > 1.17.1

**PACKAGE INDEX** >>

- Browse packages
- Package submission
- List trove classifiers
- List packages
- RSS (latest 40 updates)
- RSS (newest 40 packages)
- Python 3 Packages
- PyPI Tutorial
- PyPI Security
- PyPI Support
- PyPI Bug Reports
- PyPI Discussion
- PyPI Developer Info

**ABOUT** >>

**NEWS** >>

## twitter 1.17.1

*An API and command-line toolset for Twitter (twitter.com)*

Python Twitter Tools  
=====

[\[!\[Build Status\]\(https://travis-ci.org/sixohsix/twitter.svg\)\]](https://travis-ci.org/sixohsix/twitter.svg)  
(<https://travis-ci.org/sixohsix/twitter>) [\[!\[Coverage Status\]\(https://coveralls.io/repos/sixohsix/twitter/badge.png?branch=master\)\]](https://coveralls.io/repos/sixohsix/twitter/badge.png?branch=master)  
(<https://coveralls.io/r/sixohsix/twitter?branch=master>)

The Minimalist Twitter API for Python is a Python API for Twitter, everyone's favorite Web 2.0 Facebook-style status updater for people on the go.

**Downloads ↓**

**Not Logged In**

- [Login](#)
- [Register](#)
- [Lost Login](#)
- [Use Open](#)
- [Login with](#)

**Status**

[Nothing to](#)

# Streaming API

```
# Import the necessary package to process data in JSON format
try:
    import json
except ImportError:
    import simplejson as json

# Import the necessary methods from "twitter" library
from twitter import Twitter, OAuth, TwitterHTTPError, TwitterStream

# Variables that contains the user credentials to access Twitter API
ACCESS_TOKEN = 'YOUR ACCESS TOKEN'
ACCESS_SECRET = 'YOUR ACCESS TOKEN SECRET'
CONSUMER_KEY = 'YOUR API KEY'
CONSUMER_SECRET = 'ENTER YOUR API SECRET'

oauth = OAuth(ACCESS_TOKEN, ACCESS_SECRET, CONSUMER_KEY, CONSUMER_SECRET)

# Initiate the connection to Twitter Streaming API
twitter_stream = TwitterStream(auth=oauth)

# Get a sample of the public data following through Twitter
iterator = twitter_stream.statuses.sample()
```

OAuth →

connection →

# JSON

## **JavaScript Object Notation**

JSON is a minimal, readable format for structuring data.



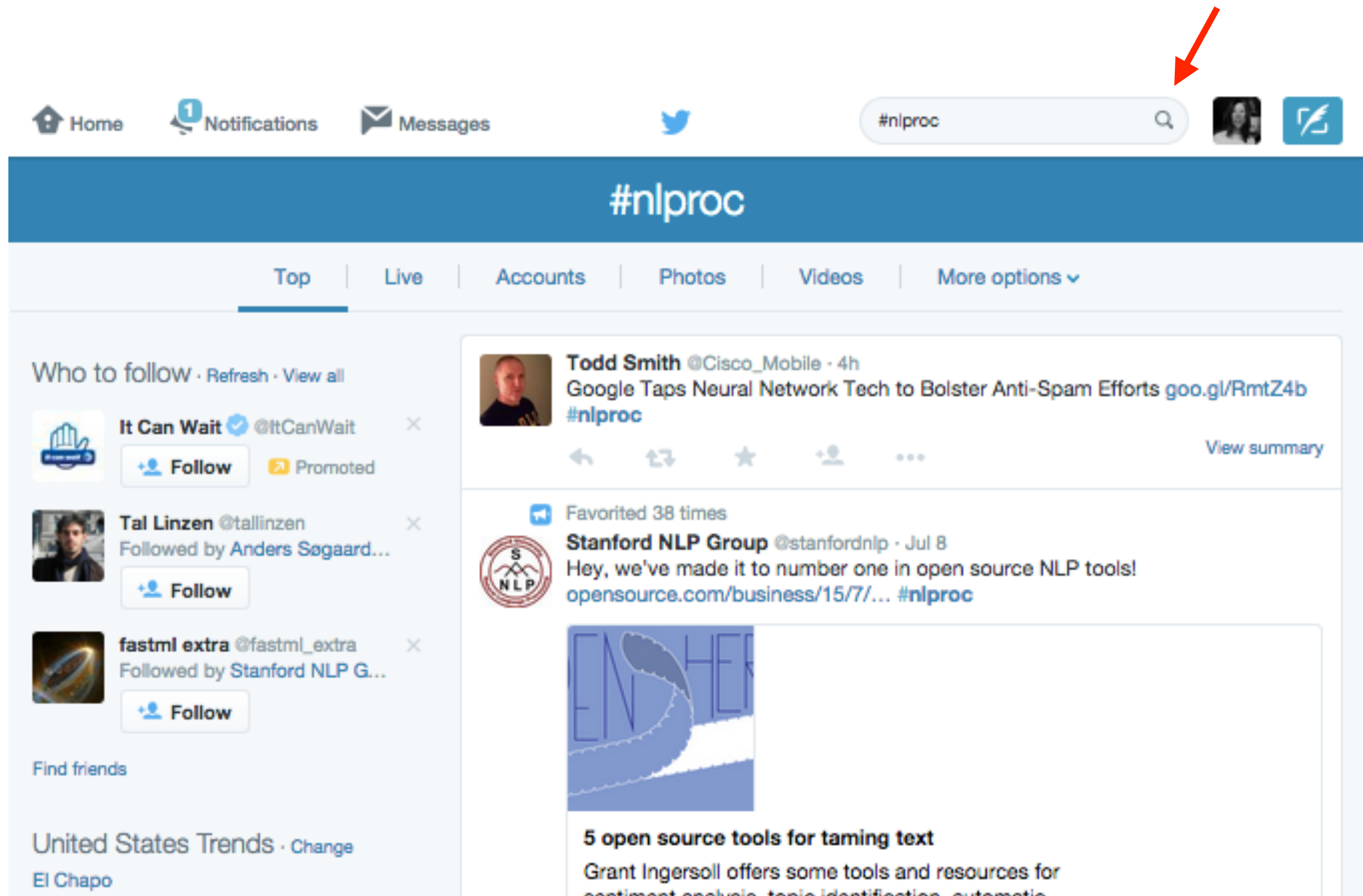
# A Tweet in JSON



#CFP Workshop on Noisy User-generated Text at ACL - Beijing 31 July 2015. Papers due: 11 May 2015. [noisy-text.github.io](http://noisy-text.github.io)  
#NLProc #WNUT15

```
{
  "favorited": false,
  "contributors": null,
  "truncated": false,
  "text": "#CFP Workshop on Noisy User-generated Text at ACL - Beijing 31 July 2015. Papers due: 11 May 2015. http://t.co/rcygyEowqH #NLProc #WNUT15",
  "possibly_sensitive": false,
  "in_reply_to_status_id": null,
  "user": {
    "follow_request_sent": null,
    "profile_use_background_image": true,
    "default_profile_image": false,
    "id": 237918251,
    "verified": false,
    "profile_image_url_https": "https://pbs.twimg.com/profile_images/527088456967544832/Dn"
```

# Search



The image shows a Twitter search results page for the hashtag #nlproc. At the top, there is a navigation bar with icons for Home, Notifications, Messages, and the Twitter logo. A search bar on the right contains the text "#nlproc" and a magnifying glass icon, which is highlighted by a red arrow. Below the navigation bar is a blue header with the hashtag "#nlproc". Underneath the header is a filter bar with tabs for "Top", "Live", "Accounts", "Photos", "Videos", and "More options". The main content area is divided into two columns. The left column, titled "Who to follow", lists three accounts: "It Can Wait" (verified), "Tal Linzen", and "fastml extra". The right column displays a list of tweets. The first tweet is from Todd Smith (@Cisco\_Mobile) about Google Taps Neural Network Tech. The second tweet is from the Stanford NLP Group (@stanfordnlp) about open source NLP tools, featuring a link to an open source website and a thumbnail image with the text "EN HER".

Home Notifications Messages

#nlproc

#nlproc

Top Live Accounts Photos Videos More options

Who to follow · Refresh · View all

It Can Wait @ItCanWait  
Follow Promoted

Tal Linzen @tallinzen  
Followed by Anders Søgaard...  
Follow

fastml extra @fastml\_extra  
Followed by Stanford NLP G...  
Follow

Find friends

United States Trends · Change  
El Chapo

Todd Smith @Cisco\_Mobile · 4h  
Google Taps Neural Network Tech to Bolster Anti-Spam Efforts [goo.gl/RmtZ4b](https://goo.gl/RmtZ4b)  
#nlproc

Favorited 38 times

Stanford NLP Group @stanfordnlp · Jul 8  
Hey, we've made it to number one in open source NLP tools!  
[opensource.com/business/15/7/...](https://opensource.com/business/15/7/...) #nlproc

EN HER

5 open source tools for taming text  
Grant Ingersoll offers some tools and resources for  
sentiment analysis, topic identification, automatic



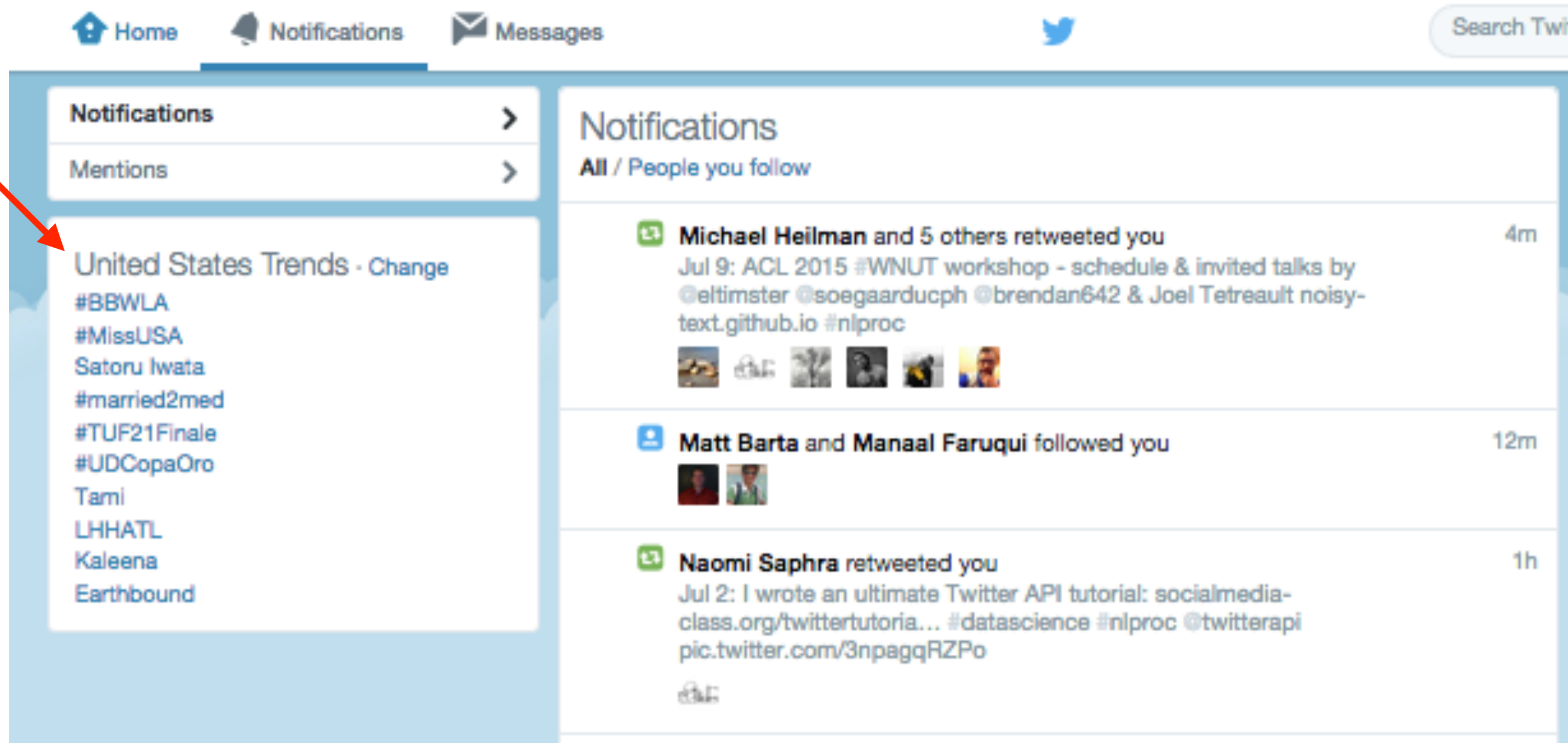
# Search API

```
# Initiate the connection to Twitter REST API  
twitter = Twitter(auth=oauth)
```

```
# Search for latest tweets about "#nlproc"  
twitter.search.tweets(q='#nlproc')
```

```
twitter.search.tweets(q='#nlproc', result_type='recent', lang='en', count=10)
```

# Trends



The image shows a screenshot of the Twitter web interface. At the top, there are navigation links for Home, Notifications, and Messages, along with a search bar. The left sidebar contains a 'Notifications' section with a sub-section for 'Mentions'. Below this, a red arrow points to the 'United States Trends' section, which lists various trending topics and users. The main content area on the right displays a 'Notifications' feed for 'All / People you follow'. This feed includes three notifications: a retweet by Michael Heilman and 5 others, a follow by Matt Barta and Manaal Faruqui, and a retweet by Naomi Saphra.

Home Notifications Messages Search Twitter

**Notifications** >  
Mentions >

**United States Trends · Change**  
#BBWLA  
#MissUSA  
Satoru Iwata  
#married2med  
#TUF21Finale  
#UDCopaOro  
Tami  
LHHATL  
Kaleena  
Earthbound

**Notifications**  
All / People you follow

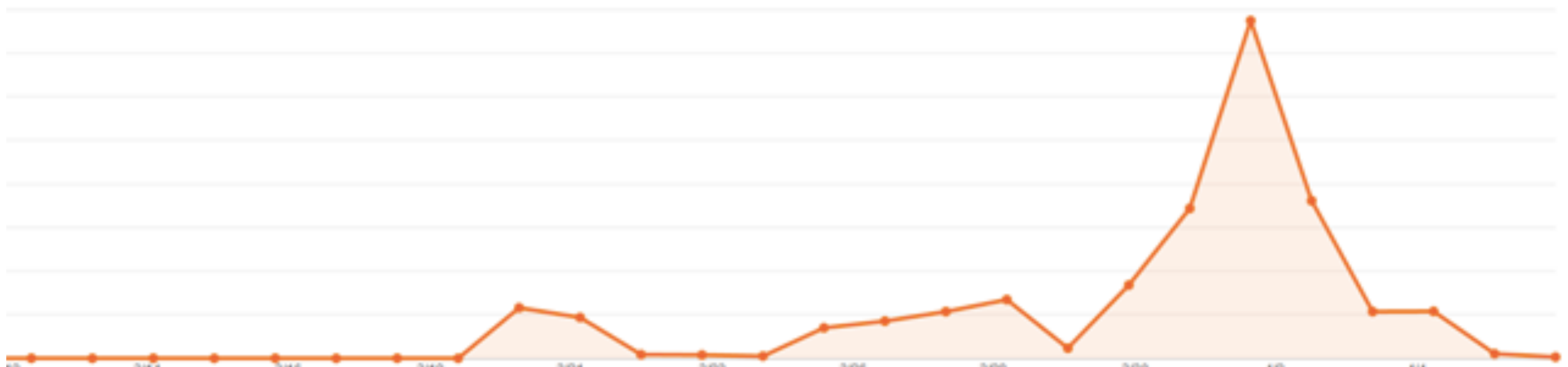
**Michael Heilman** and 5 others retweeted you 4m  
Jul 9: ACL 2015 #WNUT workshop - schedule & invited talks by @elitimster @soegaarducph @brendan642 & Joel Tetreault noisy-text.github.io #nlproc

**Matt Barta** and **Manaal Faruqui** followed you 12m

**Naomi Saphra** retweeted you 1h  
Jul 2: I wrote an ultimate Twitter API tutorial: socialmedia-class.org/twittertutoria... #datascience #nlproc @twitterapi pic.twitter.com/3npagqRZPo

# Trends

trending topics are determined by an unpublished algorithm, which finds words, phrases and hashtags that have had a sharp increase in popularity, as opposed to overall volume.



# Trends API

Where On Earth ID



```
# Get all (it's always 10) trending topics in San Francisco (its WOEID is 2487956)
sfo_trends = twitter.trends.place(_id = 2487956)
```

```
{
  "created_at": "2015-07-01T22:09:55Z",
  "trends": [
    {
      "url": "http://twitter.com/search?q=%23LiesIveToldMyParents",
      "query": "%23LiesIveToldMyParents",
      "name": "#LiesIveToldMyParents",
      "promoted_content": null
    },
    {
      "url": "http://twitter.com/search?q=%22Kevin+Love%22",
      "query": "%22Kevin+Love%22",
      "name": "Kevin Love",
      "promoted_content": null
    },
    ... [and another 8 trends omitted here to save space]
```

# In-class Exercise

- Become a Twitter User (if you haven't)
- Learn to use Twitter API
- Work on homework #1

[Social Media & Text Analytics](#) [Syllabus](#) [Twitter API Tutorial](#) [Homework Assignments ▾](#)



Twitter's 404 error page -- the Fail Whale

## Twitter API tutorial

by [Wei Xu](#) (July 1, 2015) [Follow @cocoweixu](#)

### 1. Getting Twitter API keys

To start with, you will need to have a Twitter account and obtain credentials from the Twitter developer site to access the Twitter API, following these steps:

- Create a Twitter user account if you do not already have one.
- Go to <https://apps.twitter.com/> and log in with your Twitter user account.
- Click "Create New App"

# Homework #1 is out

## Due in a week

The screenshot shows a web browser window with the URL <https://osu.instructure.com/courses/9542/assignments>. The page header displays a hamburger menu icon, the course ID **AU16 5539**, and the section **Assignments**. On the left, a dark sidebar contains icons and labels for **Account**, **Dashboard**, and **Courses**. The main content area features a search bar labeled "Search for Assignment" and a list of navigation links: **Autumn 2016**, **Home**, **Assignments** (highlighted), **Grades**, **People**, **Modules**, **Collaborations**, and **Announcements**. The assignment list shows "Assignment #1: Twitter's Language Mix" with a due date of "Sep 8 at 11:59pm" and a value of "20 pts".