Measurement and Collection of Social Network Data

Pawan Goyal

CSE, IITKGP

July 22, 2014

Social Network Site (SNS)

"A web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system." (Boyd and Elison, 2007, p. 211)

What People Do in Social Media

- People make “friends” with others and build social relationships, connections and communities.
- People ask and answer one another.
- People create, publish or distribute information in the form of text, photos, video, audio, or tweets.
- People share bookmarks, presentation slides, or other files.
- People provide feedback on or rate others’ information.
- People create social tags or folksonomies.

Social Networking Websites

Name	Description/Focus	Date launched	Registered users	Registratio
Google+	General	28 June 2011	1,600,000,000[135]	Open to people 13 a [136]
Facebook	General: Photos, Videos, Blogs, Apps.	February 2004	1,280,000,000[83]	Open to people 13 a
Qzone	General. In Simplified Chinese; caters for mainland China users		480,000,000[245][246]	Open to the general
Sina Weibo	Social microblogging site in Mainland China.	14 August 2009	300,000,000[260]	Open
Formspring	social Q&A website	November 2009	290,000,000[100]	Registration post re questions
Habbo	General for teens. Over 31 communities worldwide. Chat room and user profiles.	August 2000	268,000,000[139][140][141]	Open to people 13 a
Vkontakte	General, including music upload, listening and search. Popular in Russia and former Soviet republics.	September 2006	249,409,900[307]	Open
LinkedIn	Business and professional networking	May 2003	200,000,000[182]	Open to people 18 a
Renren	Significant site in China. Was known as 校内 (Xiaonei) until August 2009.		160,000,000[250]	Open
Instagram	A photo and video sharing site	October 2010	150,000,000[160]	Open to anyone 13
Bebo	General	July 2005	117,000,000[16]	Open to people 13 a
Tagged	General.	October 2004	100,000,000[281]	Open
Orkut	General. Owned by Google Inc. Popular in India and Brazil.[227]	22 January 2004	100,000,000[228]	Open to people 18 a (Google login)
Netlog	General. Popular in Europe, Turkey, the Arab World and Canada's Québec province. Formerly known as Facebox and Redbox.[214]		95,000,000[215]	Open to people 13 a
Twitter	General. Micro-blogging, RSS, updates	15 July 2006	93,800,000[301]	Open to all ages [30
Friendster	General. Popular in Southeast Asia. No longer popular in the western world	2002	90,000,000[110]	Open to people 16 a
hi5	General. Popular in Nepal, Mongolia, Thailand, Romania, Jamaica, Central Africa, Portugal and Latin America. Not very popular in the USA.	2003	80,000,000[143]	Open to people 13 a

Reference for this lecture

Larger Cover

Mining the Social Web, 2nd Edition

Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More

By Matthew A. Russell

Publisher: O'Reilly Media
Final Release Date: October 2013
Pages: 448

4.6

Read **11** Reviews | Write a Review

How can you tap into the wealth of social web data to discover who's making connections with whom, what they're talking about, and where they're located? With this expanded and thoroughly revised edition, you'll learn how to acquire, analyze, and summarize data from all corners of the social web, including Facebook, Twitter,...

Full description

Tools Used

IPhython Notebook

May use the following url: `http://smash.psych.nyu.edu/courses/spring12/modeling/ipythonhints.html`

Twitter

A microblogging service, allowing people to communicate with short, 140-character messages that roughly correspond to thoughts or ideas.

Twitter

A microblogging service, allowing people to communicate with short, 140-character messages that roughly correspond to thoughts or ideas.

Twitter's relationship model

Twitter

A microblogging service, allowing people to communicate with short, 140-character messages that roughly correspond to thoughts or ideas.

Twitter's relationship model

Allows you to keep up with the latest happenings of any other user, even though the other user may not choose to follow you back or even know that you exist.

Twitter

A microblogging service, allowing people to communicate with short, 140-character messages that roughly correspond to thoughts or ideas.

Twitter's relationship model

Allows you to keep up with the latest happenings of any other user, even though the other user may not choose to follow you back or even know that you exist.

Interest Graphs

A way of modeling connections between people and their arbitrary interests.

Fundamental Twitter Terminology

Tweets are the essence of Twitter. In addition to the textual content, we get:

Tweet entities

User mentions, hashtags, URLs and media

Places

Locations in the real world, attached to a tweet

Fundamental Twitter Terminology

@ptwobrussell is writing @SocialWebMining, 2nd Ed. from his home office in Franklin, TN. Be #social: http://on.fb.me/16WJAf9

Tweet entities

- User mentions: @ptwobrussell, @SocialWebMining
- hashtag: #social
- URL: http://on.fb.me/16WJAf9
- Place: Franklin, Tennesee, in the tweet and a location metadata, where the tweet is authored

Fundamental Twitter Terminology

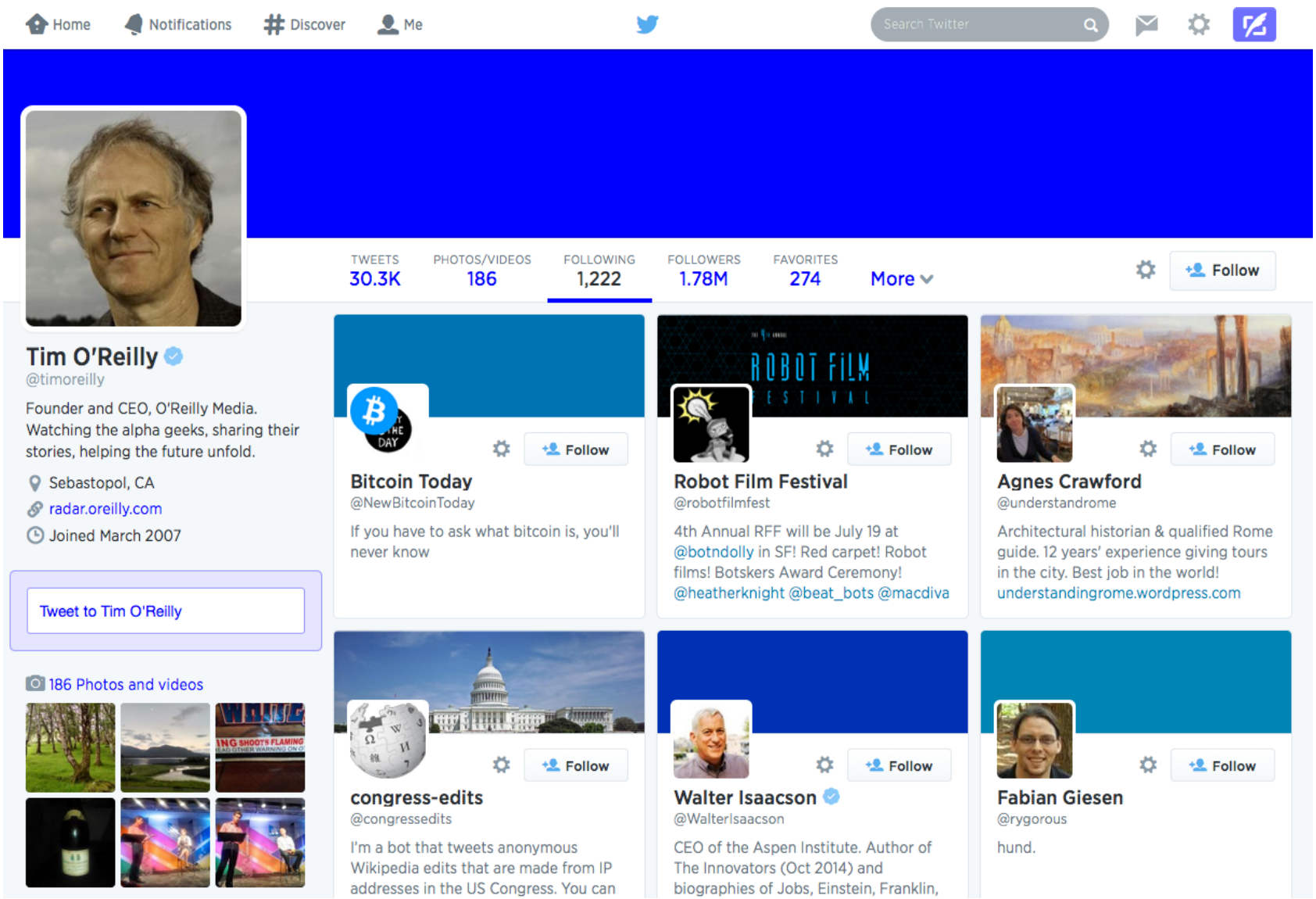
Timelines

Chronologically sorted collections of tweets

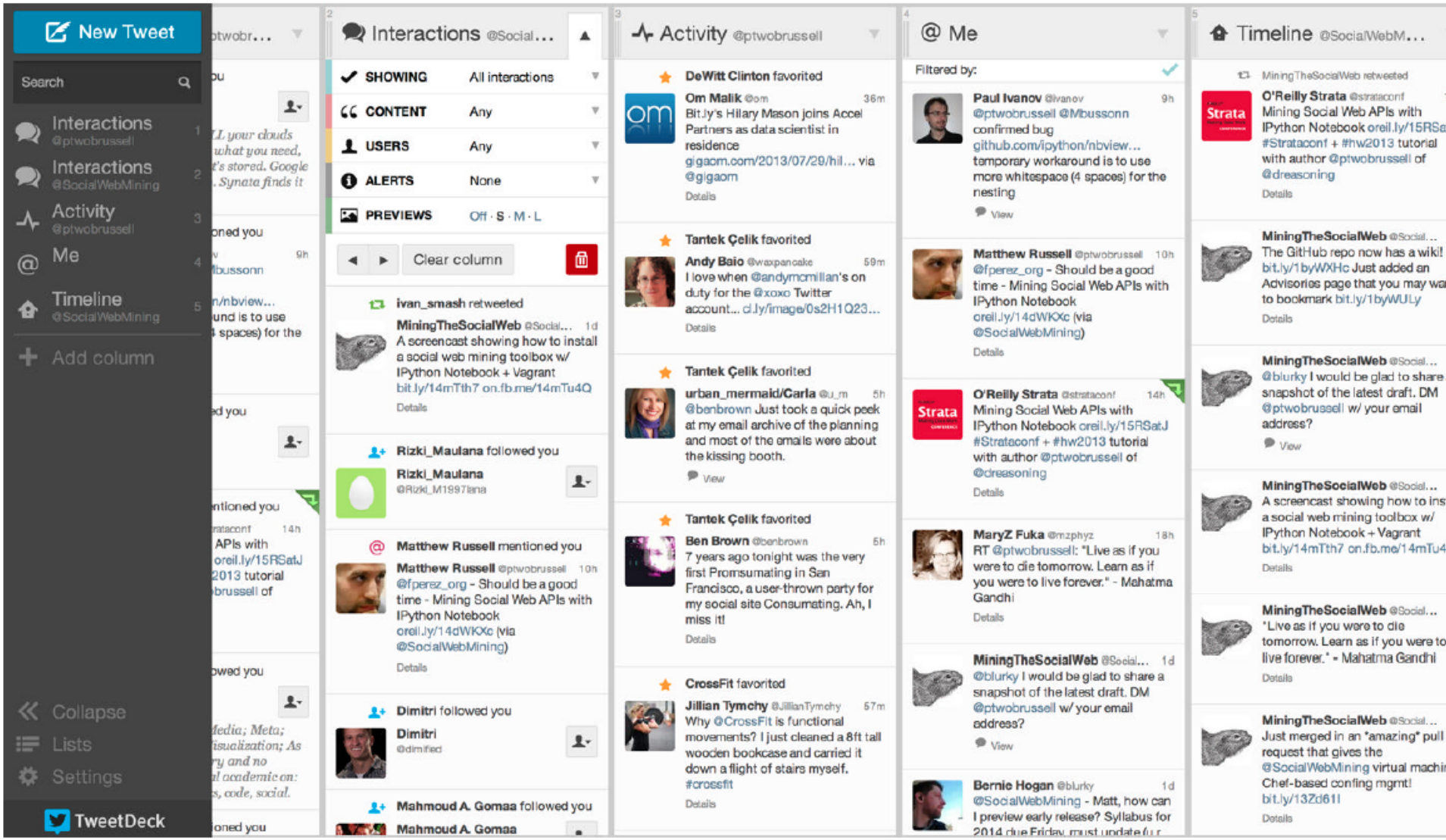
- **home timeline**: view that you see when you log into your account and look at all the tweets from users that you are following:
  `https://twitter.com`
- **user timeline**: collection of tweets only from a certain user:
  `https://twitter.com/timoreilly`
- **User's home timeline:** can be accessed with the additional *following* suffix, appended to the URL. :
  `https://twitter.com/timoreilly/following`

user timeline

user home timeline



TweetDeck: a highly customizable user interface



Streams

Samples of public tweets flowing through Twitter in realtime

Public firehose

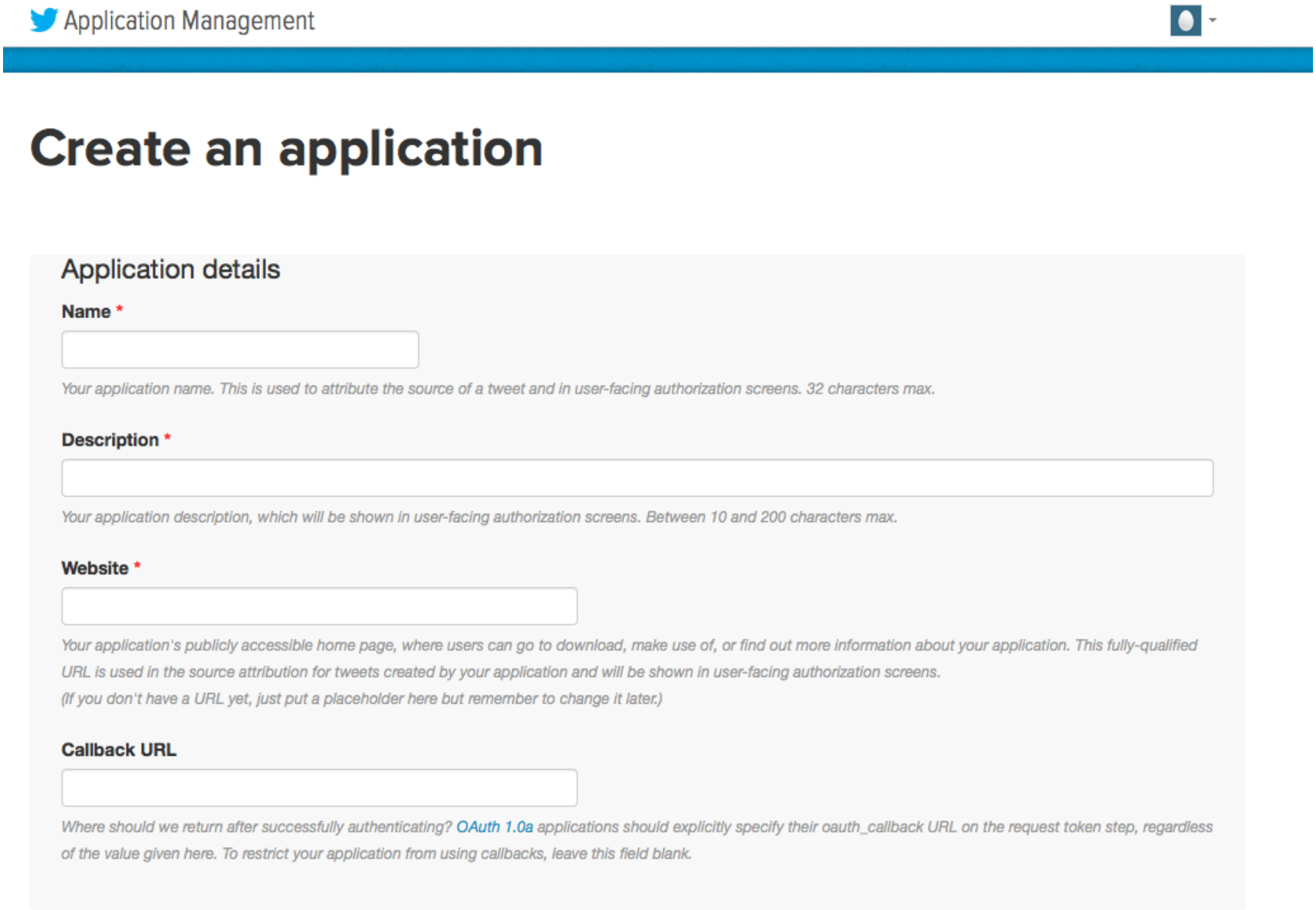
Known to peak at hundreds of thousands of tweets per minute during events with particularly wide interest.

Public timeline

A small random sample of the public timeline is available, that provides filterable access to enough public data for API developers

Creating Twitter API connection

Create an application at `https://dev.twitter.com/apps`.



Key pieces of information

Create a new Twitter application to get OAuth credentials and API access

- Consumer key
- Consumer secret
- access token
- access token secret

Exploring Twitter data

`https://github.com/ptwobrussell/Mining-the-Social-Web-2nd-Edition`

Requires : IPython Notebook

Things to try out

- Exploring Trending Topics
- Searching for tweets using a query
- Analyzing he 140 characters

Exploring Twitter data

Computing Lexical Diversity

Number of unique tokens in the text, divided by the total number of tokens in the text

Can be a primitive statistic for

- How broad or narrow the subject matter is that an individual or group discusses
- Comparing different groups/individuals
- Comparing across time periods

Accessing Facebook user's account data

Register an application as the entry point into the Facebook developer platform.

Data accessible

Whatever the user has explicitly authorized it to access.
As a developer, the application would be like any of your Facebook friends, in that you are ultimately in control of what the application can access.

Accessing Facebook user's account data

Register an application as the entry point into the Facebook developer platform.

Data accessible

Whatever the user has explicitly authorized it to access.
As a developer, the application would be like any of your Facebook friends, in that you are ultimately in control of what the application can access.

Developer Tools

Graph API Explorer app for querying the Social Graph

Accessing Facebook user's account data

Register an application as the entry point into the Facebook developer platform.

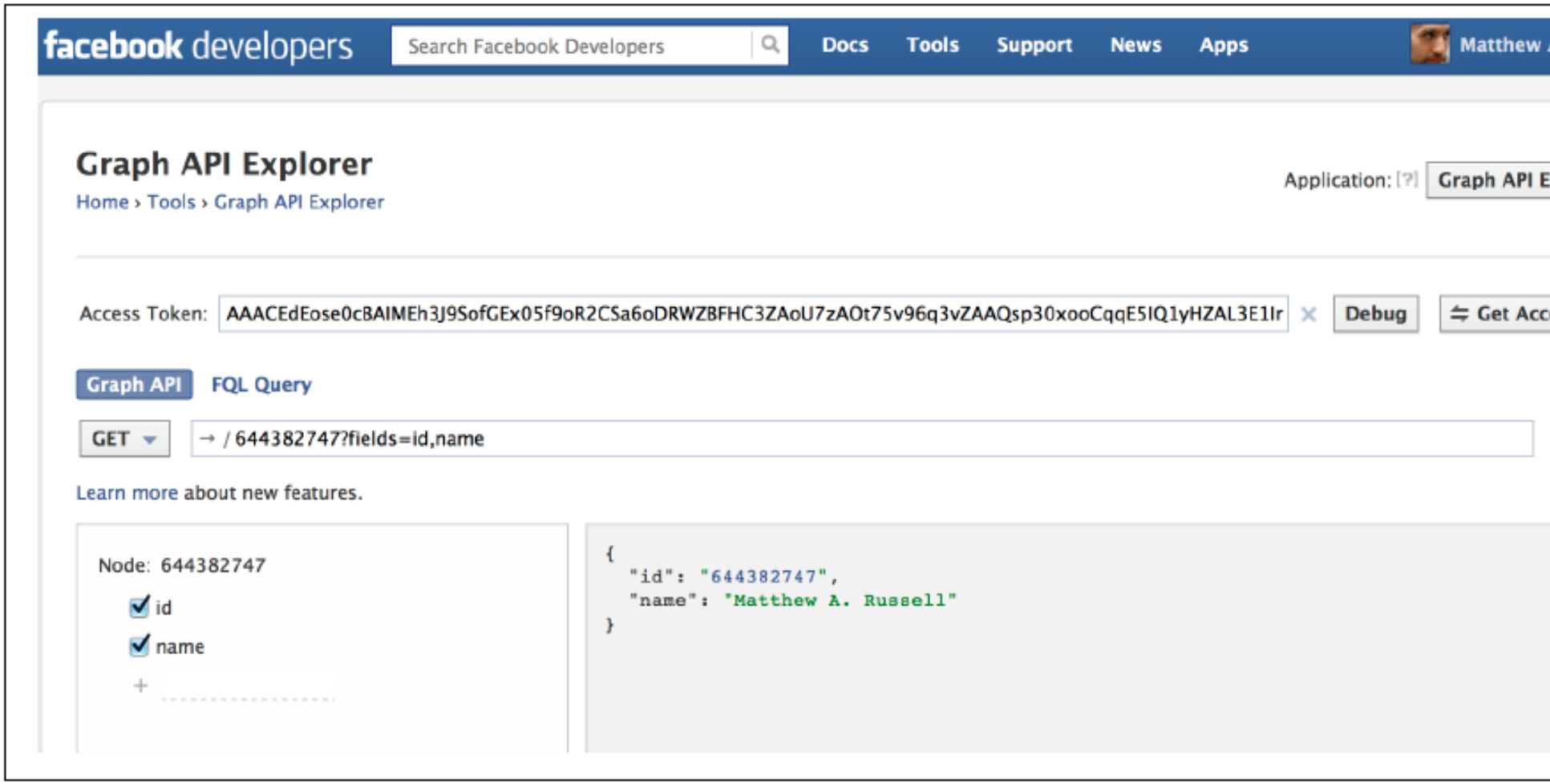
Data accessible

Whatever the user has explicitly authorized it to access.
As a developer, the application would be like any of your Facebook friends, in that you are ultimately in control of what the application can access.

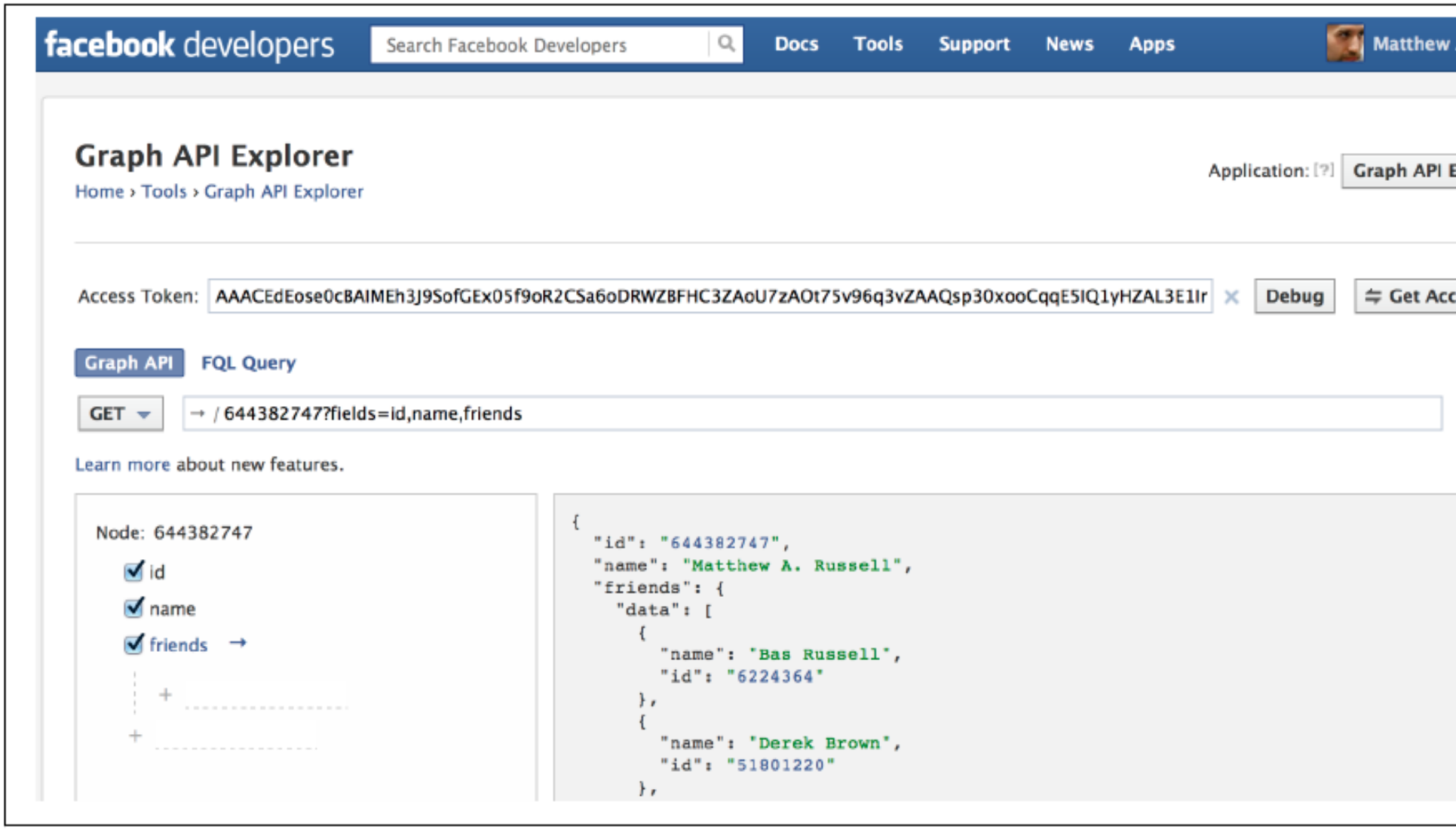
Developer Tools

Graph API Explorer app for querying the Social Graph
You can translate your queries into Python code for automation

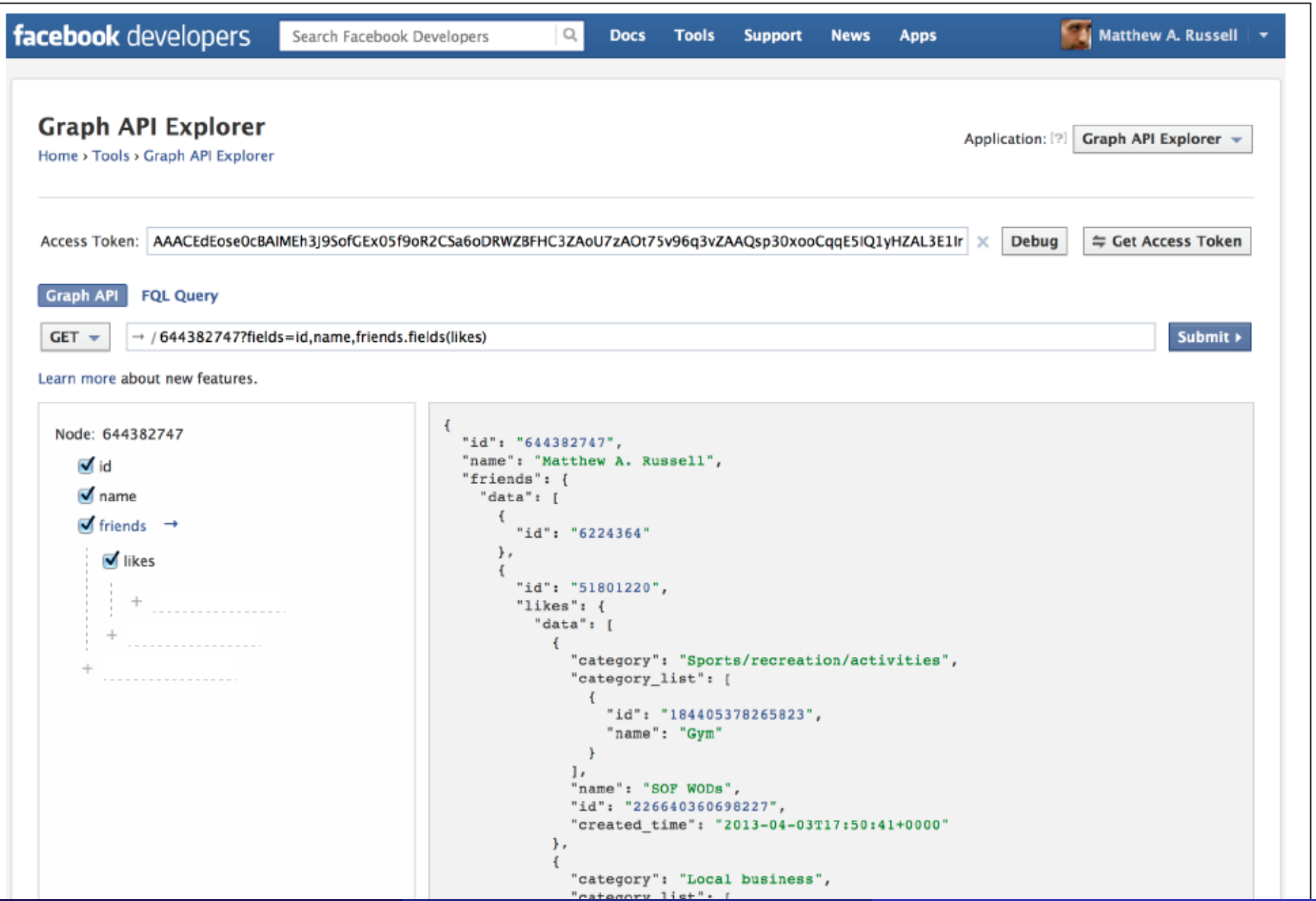
Graph API query



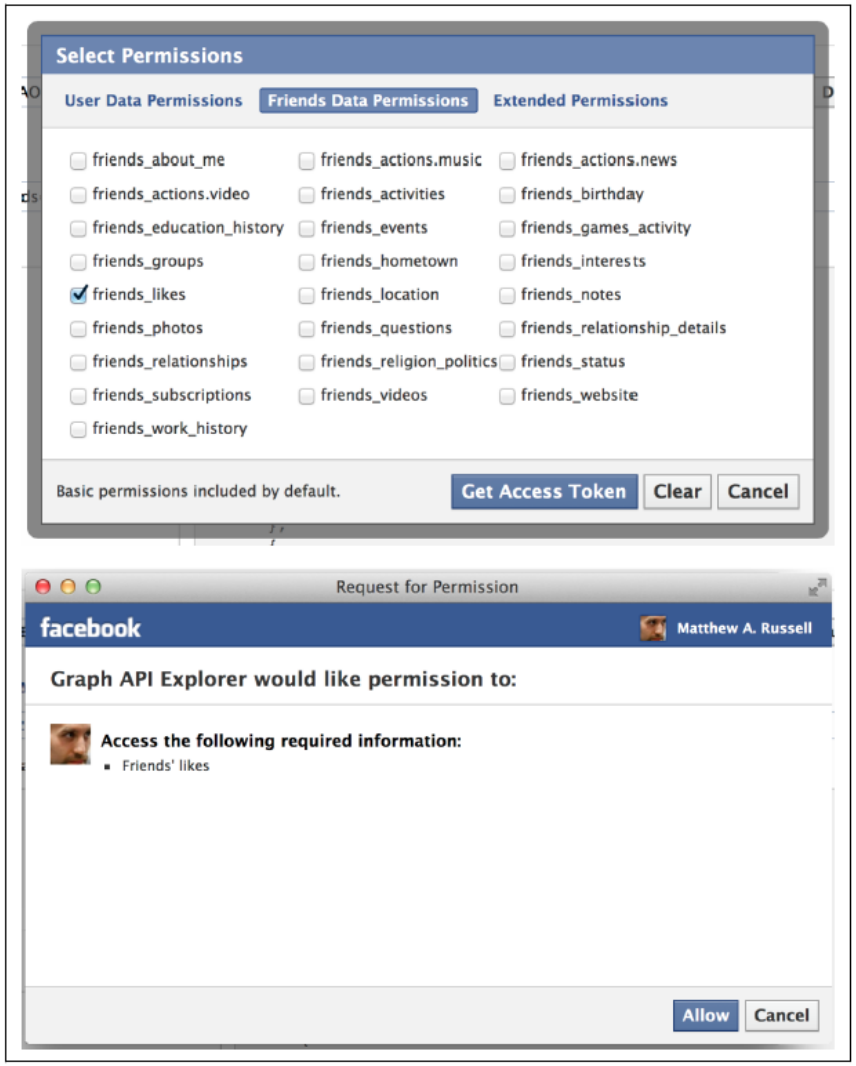
Graph API query



Graph API query



Graph API query



Analyzing Facebook Pages

```
# Get an instance of Mining the Social Web
# Using the page name also works if you know it.
# e.g. 'MiningTheSocialWeb' or 'CrossFit'
mtsw_id = '146803958708175'
pp(g.get_object(mtsw_id))
```

Sample output for the query reveals the data that backs the object's Facebook page, as shown here:


```
{
 "category": "Book",
 "username": "MiningTheSocialWeb",
 "about": "Analyzing Data from Facebook, Twitter, LinkedIn, and Other Social...",
 "talking_about_count": 22,
 "description": "Facebook, Twitter, and LinkedIn generate a tremendous ...",
 "company_overview": "Like It here on Facebook!\n\nFollow @SocialWebMining...",
 "release_date": "January 2011",
 "can_post": true,
 "cover": {
  "source": "https://sphotos-b.xx.fbcdn.net/...",
  "cover_id": 474206292634605,
  "offset_x": -41,
  "offset_y": 0
 },
 "mission": "Teaches you how to...\n\n* Get a straightforward synopsis of ...",
 "name": "Mining the Social Web",
 "founded": "January 2011",
 "website": "http://amzn.to/d1Ci8A",
 "link": "http://www.facebook.com/MiningTheSocialWeb",
 "likes": 911,
 "were_here_count": 0,
 "general_info": "Analyzing Data from Facebook, Twitter, LinkedIn, ...",
 "id": "146803958708175",
 "is_published": true
}
```

Further Exploration

Things to try out

- Analyzing things your friends “like”: top ten things, categories
- Common likes between an ego and its friendship