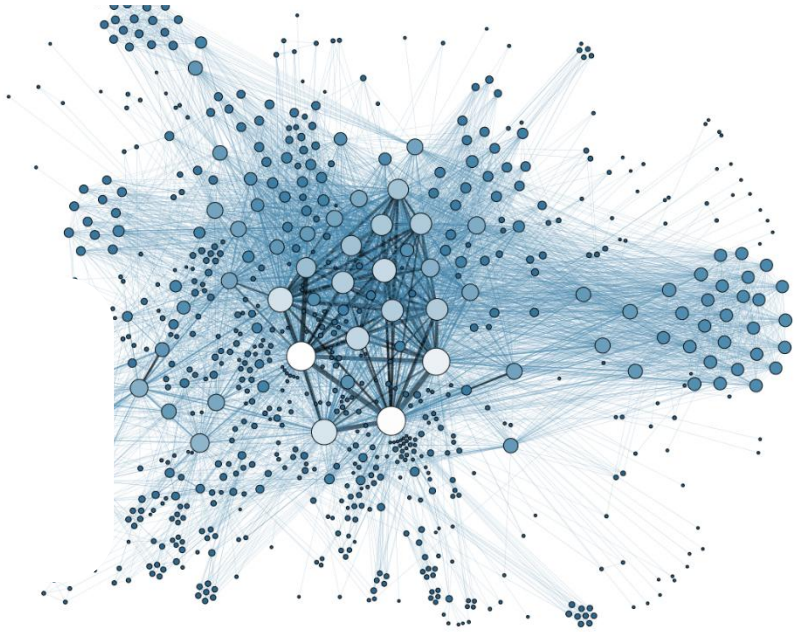
Social Media Mining

Introduction

Dr. Shahid Mahmood Awan
github.com/shahidmawan
shahidmawan@gmail.com



What is data mining?

"Data mining is the efficient discovery of valuable, non-obvious information from a large collection of data."

– Joseph P. Bigus

SOCIAL MEDIA

Social media is defined as a group of Internet-based applications that allow the creation and exchanges of user-generated content.

Social media gives users an easy-to-use way to communicate and network with each other on an unprecedented scale.

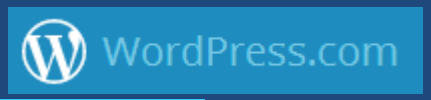
Facebook, the social networking site, recorded more than 845 million active users as of December 2011.

Social Media Landscape 2015

Classification of social media

9 different types of social media:

- Online social networking
- Blogging
- Micro-blogging
- Wikis
- Social news
- Social bookmarking
- Media sharing
- Opinion, reviews, and ratings
- Answers



Types of Social Media

- Online Social Networking
- Publishing
  - Blogging
  - Wiki
- Micro blogging
- Social News
- Social Bookmarking
- Media Sharing
  - Video Sharing
  - Photo Sharing
  - Podcast Sharing
- Opinion, Review, and Ratings Websites
- Answers
- Entertainment

What is social media mining?

Social media mining is extracting information from social media.

Primary objectives of the data mining process are to effectively handle large-scale data, extract actionable patterns, and gain insightful knowledge.

Users on Twitter generate over 400 million Tweets everyday

- The study and development of these new techniques are under the purview of social media mining,an emerging discipline under the umbrella of data mining.

- **Social Media Mining is the process of representing, analyzing, and extracting actionable** patterns from social media data.

- Social Media Mining introduces basic concepts and principal algorithms suitable for investigating massive social media data;

- it discusses theories and methodologies from dierent disciplines such as computer science, data mining, machine learning, social network analysis, network science, sociology, ethnography, statistics, optimization, and mathematics.

- It encompasses the tools to formally represent, measure, model, and mine meaningful patterns from large-scale social media data.

What is social media mining?

Data mining of social media can expand researchers' capability of understanding new phenomena to provide better services and develop innovative opportunities.

Mining social media is a growing multidisciplinary area where researchers of different backgrounds can make important contributions that matter for social media research and development.

The reasons for growth of social media mining

Social media growth is driven by these:

(1) How can a user be heard?
(2) Which source of information should a user use?
(3) How can user experience be improved?

Answers to these questions are hidden in the social media data.

The amount of data!

For example, Facebook and Twitter report Web data from approximately 149 million and 90 million unique U.S. visitors per month, respectively.

According to the video sharing site YouTube, more than 4 billion videos are viewed per day, and 60 hours of videos are uploaded every minute.

The picture sharing site Flickr, as of August 2011, hosts more than 6 billion photo images.

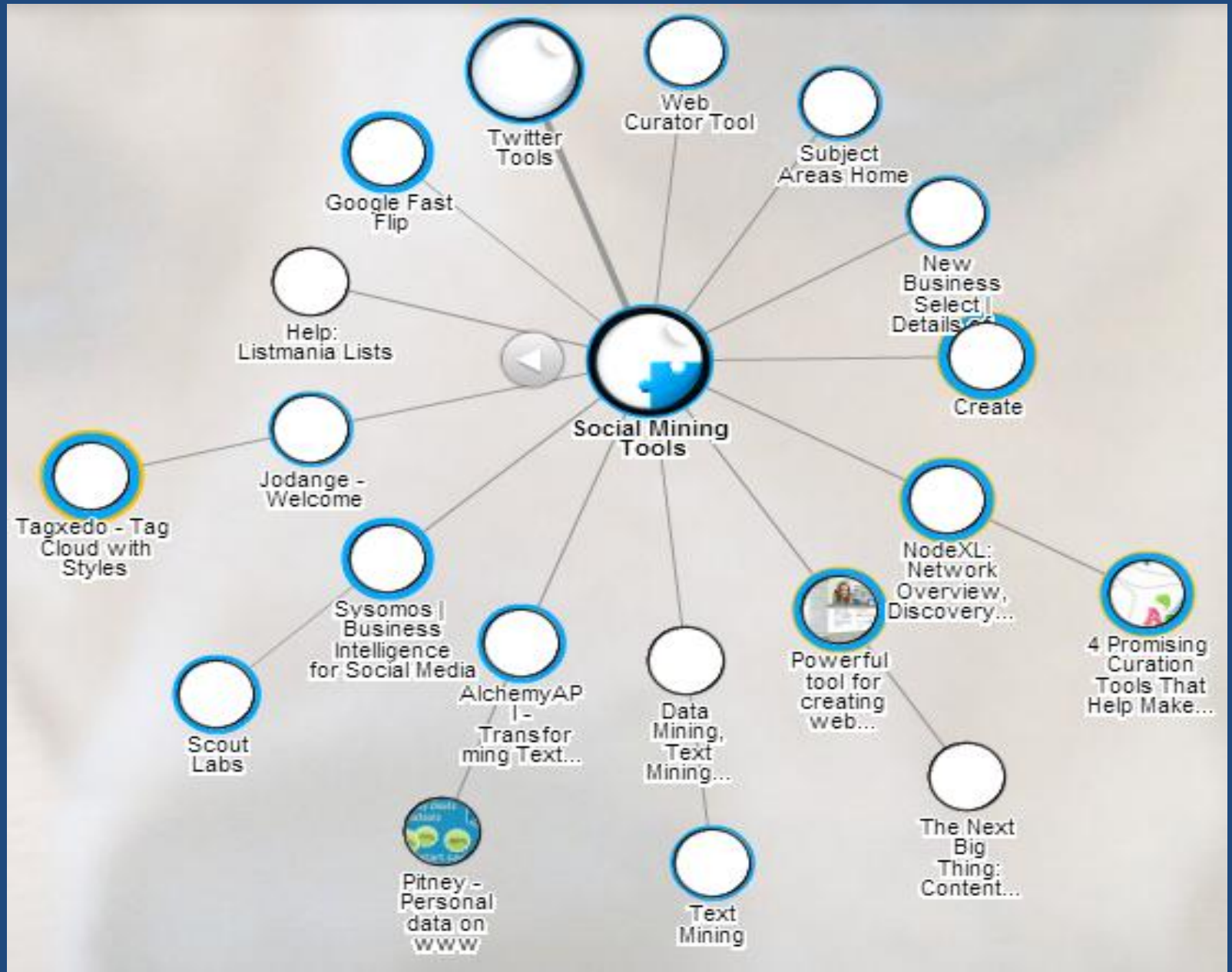
Web-based, collaborative, and multilingual Wikipedia hosts over 20 million articles attracting over 365 million readers.

Challenges in social media mining

Social media data are
vast,
noisy,
distributed,
unstructured,
dynamic.

These characteristics pose challenges to data mining tasks to invent new efficient techniques and algorithms.

Tools used for social media mining



List of common tools

Twitter tools
Cloud4Trends
TweetTracker

What is use of data mining in social media mining?

Social Media data is everywhere.
There is an overload.
Information Overload (blogs, photos, videos,bookmarks)
Interaction Overload (friends, taggers, followers, commenters)

How to extract data from this chaos?

What is use of data mining in social media mining?

Social media captures 'pulse of humanity'.
-Can directly study opinions and behaviors of millions of users to gain insight into:

-Human behavior
-Market analytics
-Product sentiments

Applications of social media mining

Personalization
Targeted marketing
Suggesting choices

Community Analysis
Sentiment Analysis and Opinion Mining
Social Recommendation
Influence Modeling

Research Issues in social media mining

- Community Analysis
- Sentiment Analysis and Opinion Mining
- Social Recommendation
- Influence Modeling
- Information Diffusion and Provenance
- Privacy, Security and Trust

References

- Data Warehousing Fundamentals- Paulraj Ponniah
- Mining Social Media: A Brief Introduction Pritam Gundecha, Huan Liu, Arizona State University
- Shamanth Kumar, Fred Morstatter, and Huan Liu. "Twitter Data Analytics", Springer 2013

Analyze Facebook with R

Facebook with R

- Basics of extracting Facebook data using R & Facebook API.
- **Rfacebook** Package: Provides an interface to the Facebook API.
- **Rfacebook** package in R provides functions that allow R to access Facebook's API to get information about posts, comments, likes, group that mention specific keywords & much more.
- Install "Rfacebook" package from **CRAN** : *install.packages("Rfacebook")*

https://bigdataenthusiast.wordpress.com/2016/03/19/mining-facebook-data-using-r-facebook-api/

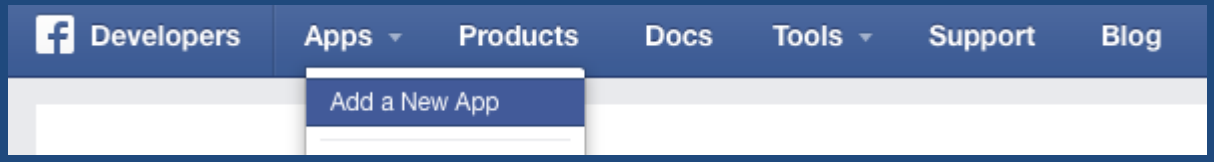
Step 1: Registering an Application with Facebook.

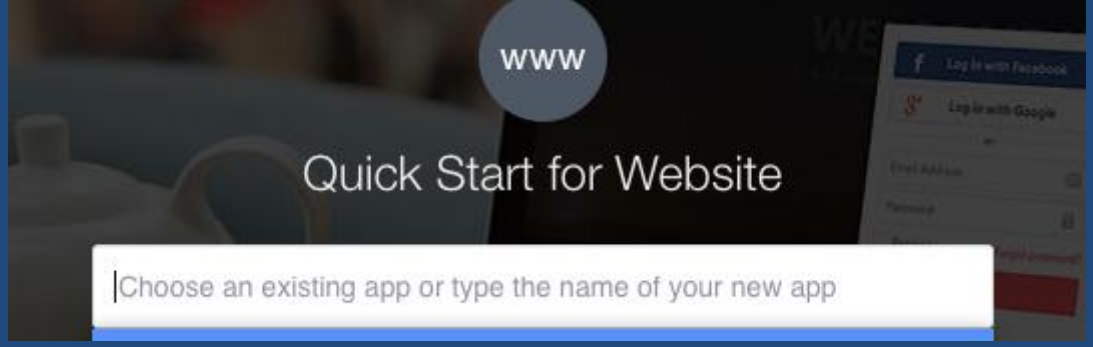
- If you already have an account with Facebook, go to FacebookDeveloper and register.

- Click “Register Now” button. After you register as a Facebook developer, you can register a new application.

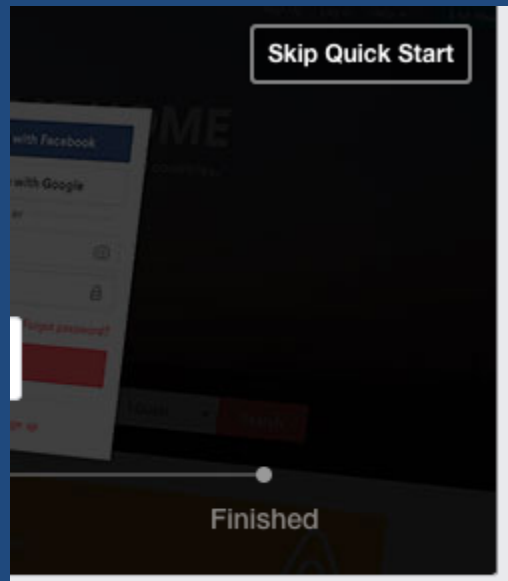
Register a new application

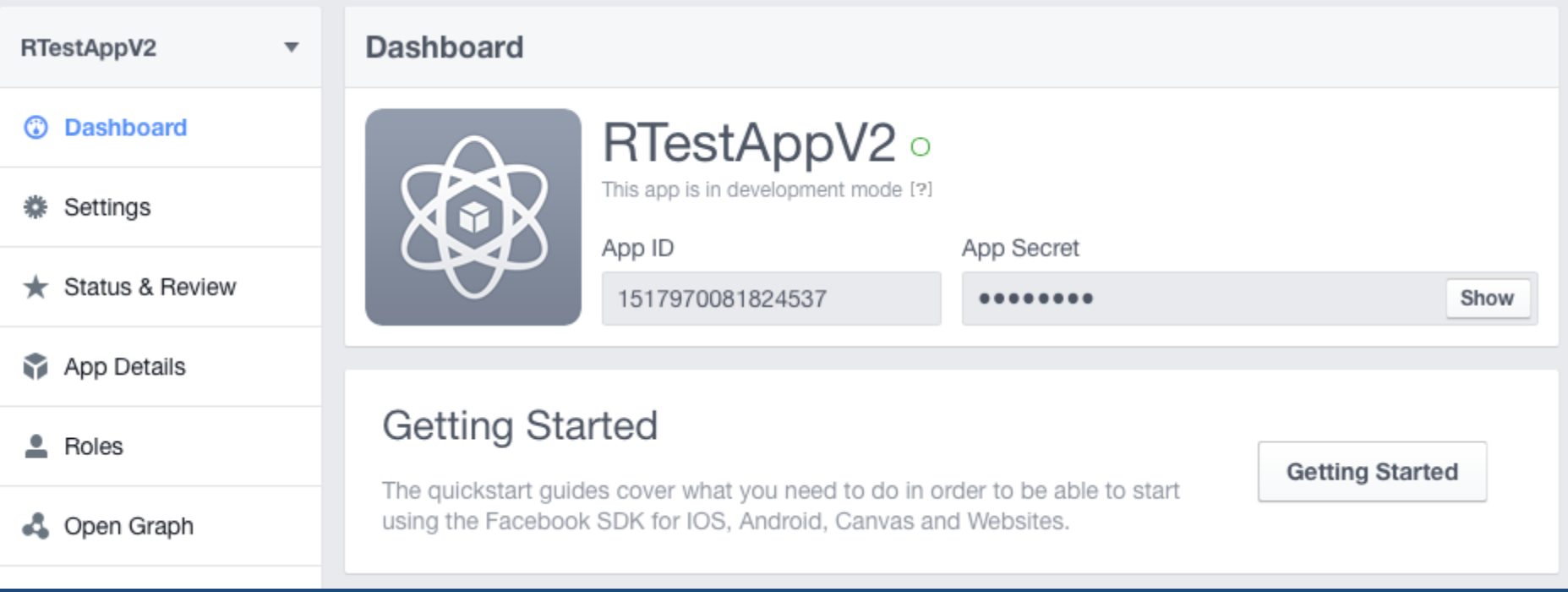
- To create a new app go to https://developers.facebook.com
- From FacebookDeveloper click on Apps at the top of the page to go to the application dashboard.
- Click on “Apps” and choose “**Add a New App**“. In the next window choose “**Website**” and give your app a fancy name.
- Click the button near the top. Once you are done with the verification process, your application is created. Note down the App Id & App Secret.





- After clicking on “**Create a New App ID**“, choose a category for your app in the next window and apply the changes with “**Create App ID**“.
- You can then click on “**Skip Quick Start**” to get directly to the settings of your app.





http://thinktostart.com/analyzing-facebook-with-r/

R

- First we need to install the packages **devtools** and **Rfacebook** from github as this is currently the most recent version.

- install.packages("devtools")
- library(devtools)
- install_github("Rfacebook", "pablobarbera", subdir="Rfacebook")

Connect R session with test app

- Connect R session with test app and authenticate it to Facebook Profile for Data Mining.
- Rfacebook offers a very easy function for that.
- Just copy your app id and your app secret from your app settings on the Facebook developer page.

- *Note: I use an app for the authentication. You can also you OAuth tokens, but they will expire after 2 hours and you have to get a new one.*

fbAuth

- require("Rfacebook")
- fbOAuth(app_id, app_secret, extended_permissions = TRUE)
- fb_oauth <- fbOAuth(app_id="123456789", app_secret="1A2B3C4D",extended_permissions = **TRUE**)
- Of course you have to insert the app-id and the app-secret of your app you just created

- The console will then print you the message:

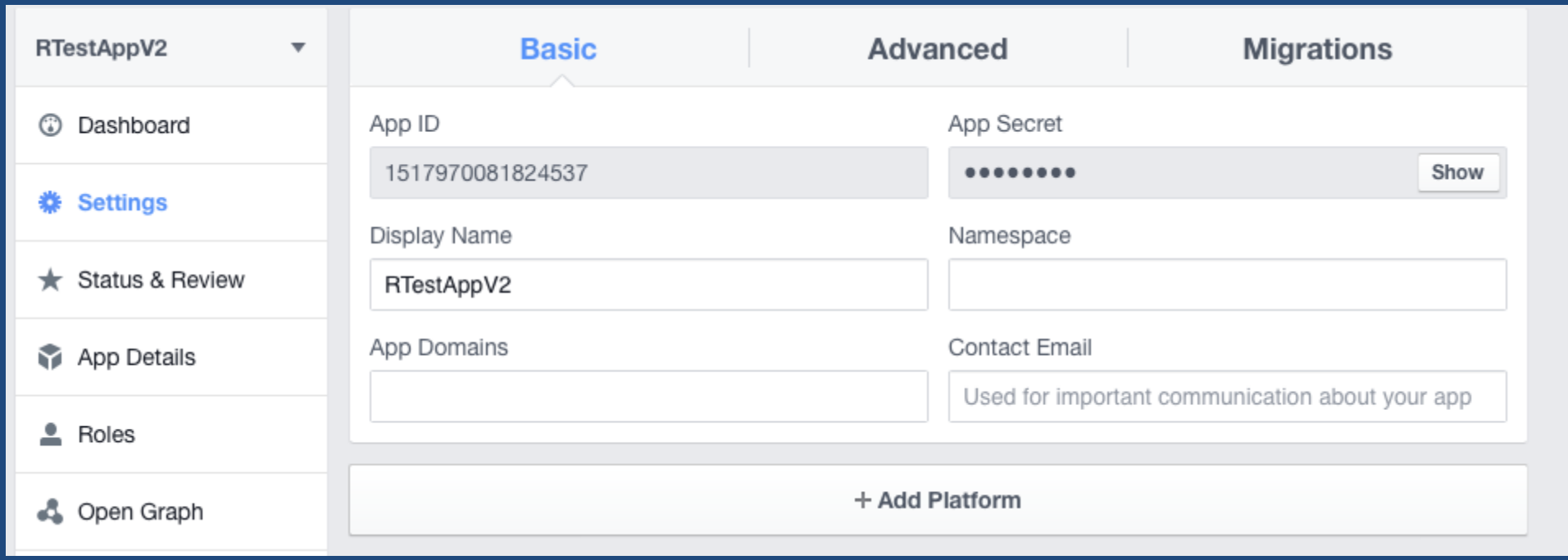
```
Copy and paste into Site URL on Facebook App Settings: http://localhost:1410/
When done, press any key to continue...
```

- Copy the URL and go to the settings of your Facebook app.

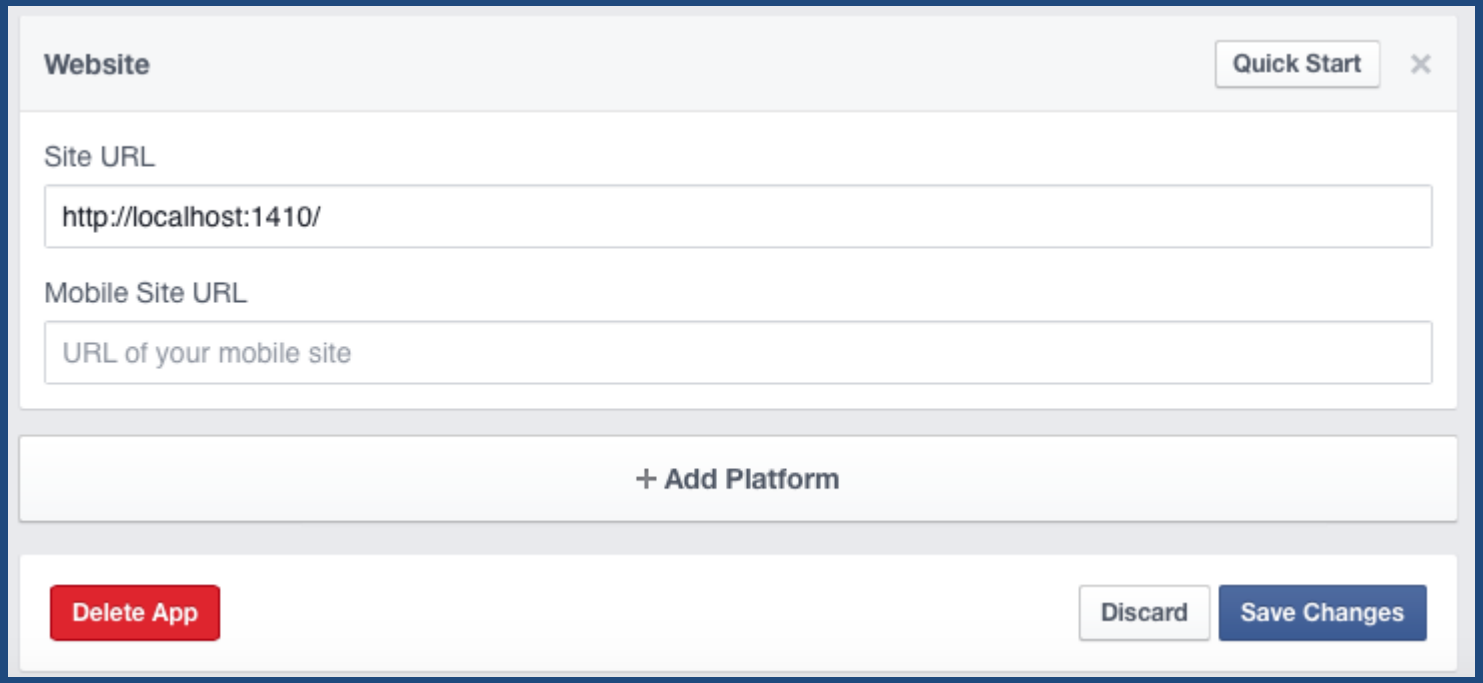
- Click on the settings tab on the left side and then choose “**+ Add Platform**“.

```
Restarting R session...

> library("Rfacebook")
> my_oauth <- fbOAuth(app_id="[redacted]", app_secret="[redacted]")
Copy and paste into Site URL on Facebook App Settings: http://localhost:1410
When done, press any key to continue...
Waiting for authentication in browser...
Press Esc/Ctrl + C to abort
Authentication complete.
Authentication successful.
>
```


RTestAppV2
Dashboard
Settings
Status & Review
App Details
Roles
Open Graph
Basic
Advanced
Migrations
App ID
1517970081824537
App Secret
Show
Display Name
RTestAppV2
Namespace
App Domains
Contact Email
Used for important communication about your app
+ Add Platform

- Then add the URL in the field “**Site URL**” and save the changes.



- Go back to your R session and hit enter. Then a browser window should open you have to allow the app to access your Facebook account.
- If everything worked the browser should show the message

```
Authentication complete. Please close this page and return to R.
```

- And your R console will confirm it with

```
Authentication complete.
Authentication successful.
```

Analyze Facebook with R!

- The getUsers function returns public information about one or more Facebook user. If we use “me” as the username argument, it will return our own profile info.

- me <- getUsers("me",token=fb_oauth)

- An advantage of the new API version is that you can get more than 100 likes. You can get the things you liked with:

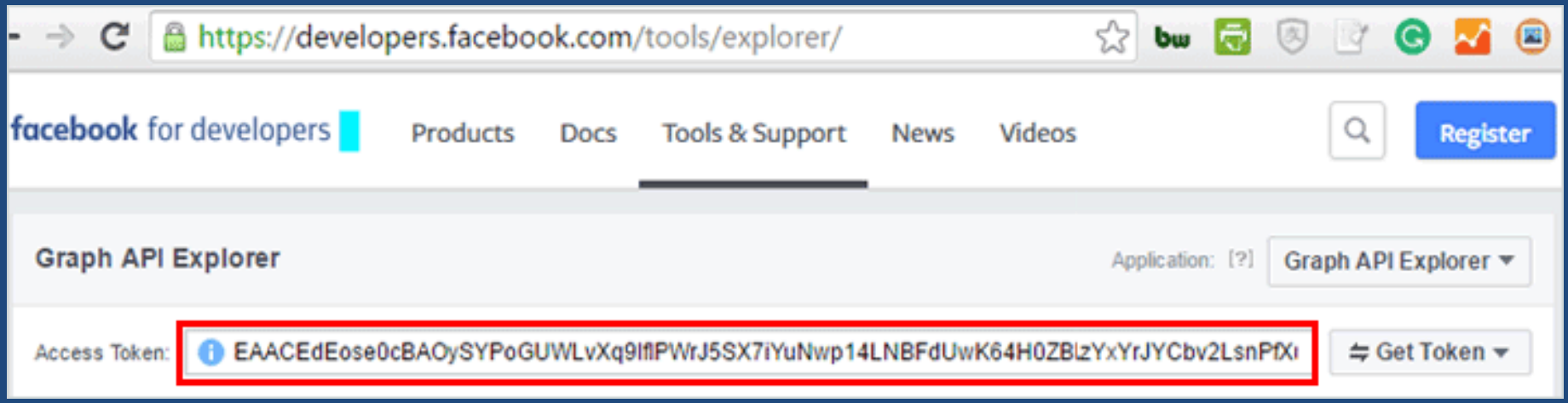
- my_likes <- getLikes(user="me", token=fb_oauth)

Demo

- https://www.youtube.com/watch?v=WP-cbfC3j6A

Get access Token from the Facebook Graph API explorer

- https://developers.facebook.com/tools/explorer/



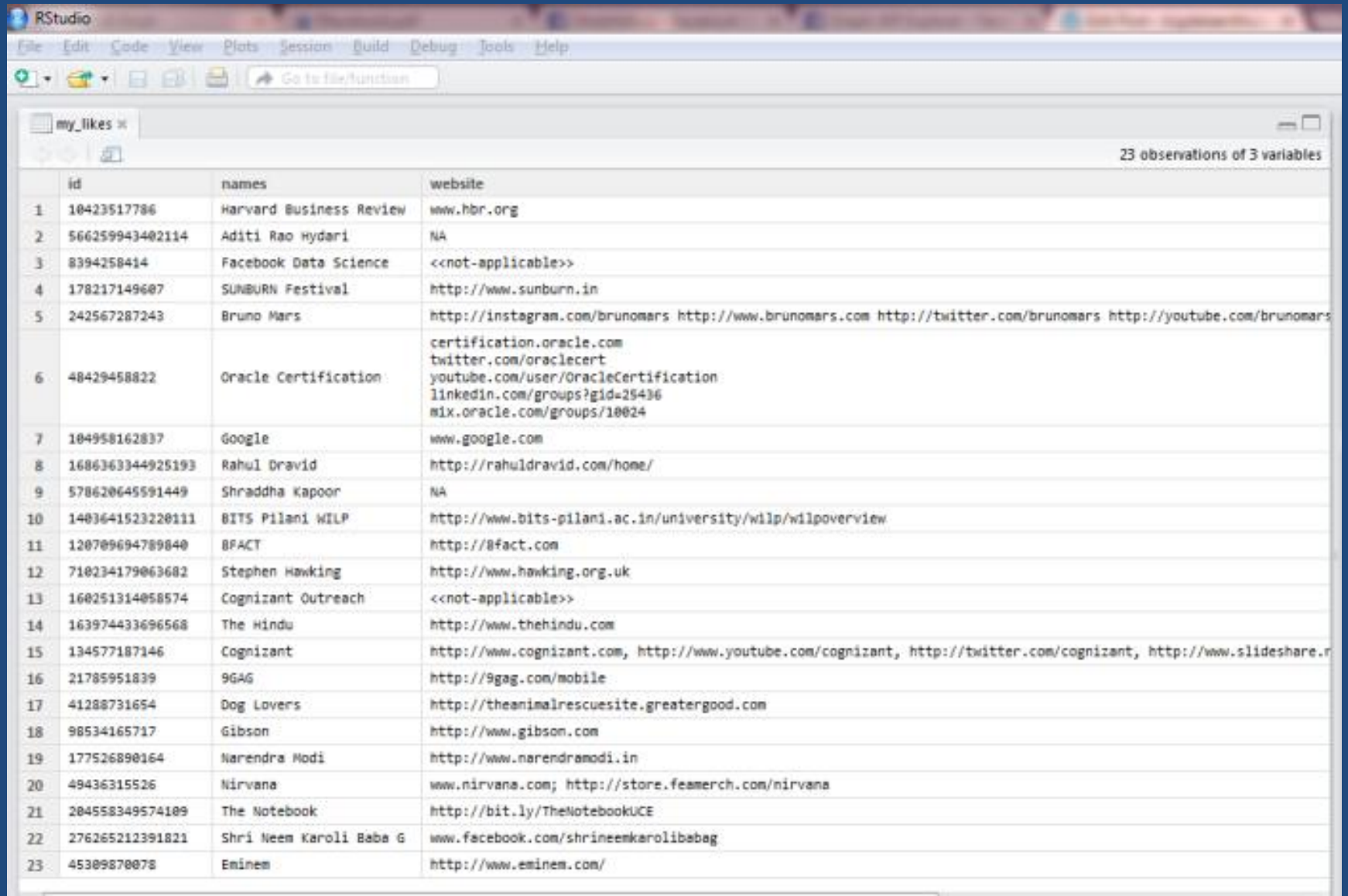
function getLikes

- **getLikes(user, n = n , token)**: Extract list of liked pages of a Facebook user with page id.
- **Arguments:** user: user name/ID , n: Number of liked pages to return for user.
- Use below command in R to get likes.

```
> my_likes <- getLikes(user="me", token = my_oauth)
```

- Here is the my result data set “my_likes”. With three variables: Id, names & website of pages.

```
> View(my_likes)
```



RStudio

File Edit Code View Plots Session Build Debug Tools Help

my_likes

23 observations of 3 variables

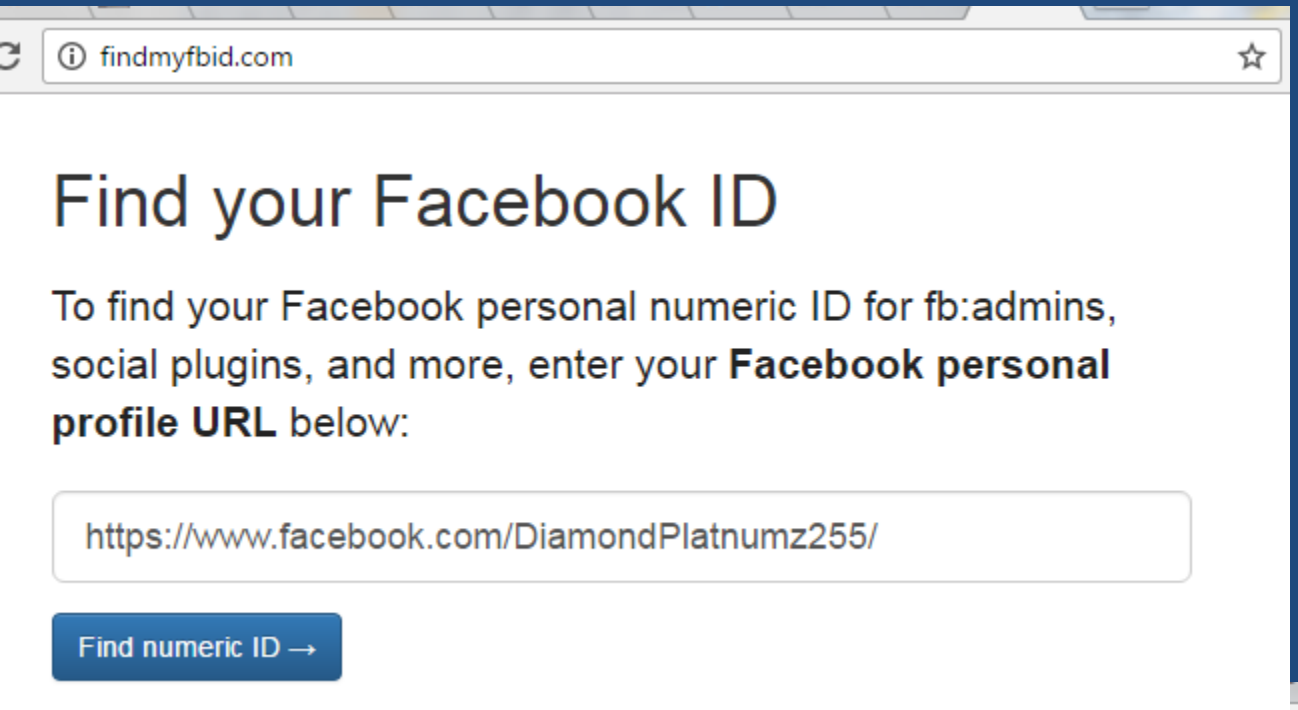
	id	names	website
1	10423517786	Harvard Business Review	www.hbr.org
2	566259943402114	Aditi Rao Hydari	NA
3	8394258414	Facebook Data Science	<<not-applicable>>
4	178217149607	SUNBURN Festival	http://www.sunburn.in
5	242567287243	Bruno Mars	http://instagram.com/brunomars http://www.brunomars.com http://twitter.com/brunomars http://youtube.com/brunomars
6	48429458822	Oracle Certification	certification.oracle.com twitter.com/oraclecert youtube.com/user/OracleCertification linkedin.com/groups?gid=25436 mix.oracle.com/groups/10024
7	104958162837	Google	www.google.com
8	1686363344925193	Rahul Dravid	http://rahuldravid.com/home/
9	578620645591449	Shraddha Kapoor	NA
10	1403641523220111	BITS Pilani WILP	http://www.bits-pilani.ac.in/university/wilp/wilpoverview
11	120709694789840	8FACT	http://8fact.com
12	710234179063682	Stephen Hawking	http://www.hawking.org.uk
13	160251314058574	Cognizant Outreach	<<not-applicable>>
14	163974433696568	The Hindu	http://www.thehindu.com
15	134577187146	Cognizant	http://www.cognizant.com, http://www.youtube.com/cognizant, http://twitter.com/cognizant, http://www.slideshare.r
16	21785951839	9GAG	http://9gag.com/mobile
17	41288731654	Dog Lovers	http://theanimalrescuesite.greatergood.com
18	98534165717	Gibson	http://www.gibson.com
19	177526890164	Narendra Modi	http://www.narendramodi.in
20	49436315526	Nirvana	www.nirvana.com; http://store.feamerch.com/nirvana
21	204558349574109	The Notebook	http://bit.ly/TheNotebookUCE
22	276265212391821	Shri Neem Karoli Baba G	www.facebook.com/shrineemkarolibabag
23	45309870078	Eminem	http://www.eminem.com/

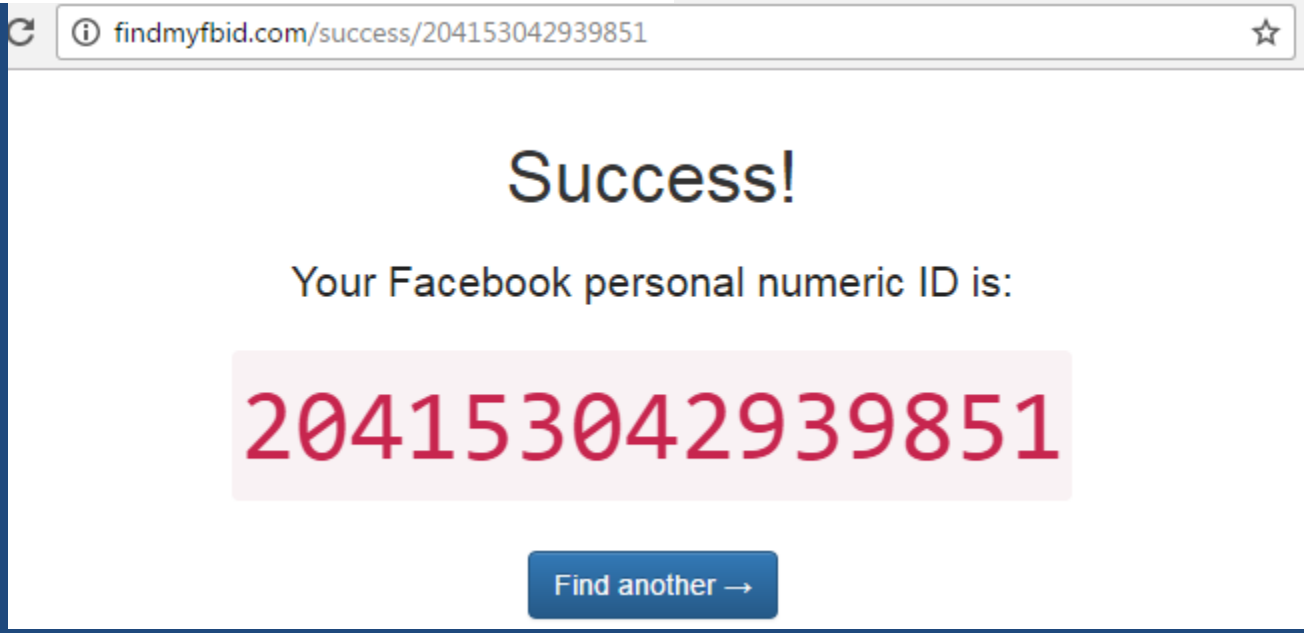
Get your competitor's Facebook fan page ID from the Graph API

- https://www.facebook.com/DiamondPlatnumz255/

- Put it in
- http://findmyfbid.com/

https://bigdataenthusiast.wordpress.com/2016/03/19/mining-facebook-data-using-r-facebook-api/


findmyfbid.com
Find your Facebook ID
To find your Facebook personal numeric ID for fb:admins, social plugins, and more, enter your **Facebook personal profile URL** below:
https://www.facebook.com/DiamondPlatnumz255/
Find numeric ID →


findmyfbid.com/success/204153042939851
Success!
Your Facebook personal numeric ID is:
204153042939851
Find another →

Lengthy Way

- Next Time + JSON and CSV Parsing

Get your competitor's Facebook fan page ID from the Graph API

- After getting the page id use below command in R to get posts.

```
> getpagedata = getPage (177526890164, token = my_oauth, n = 10)
```

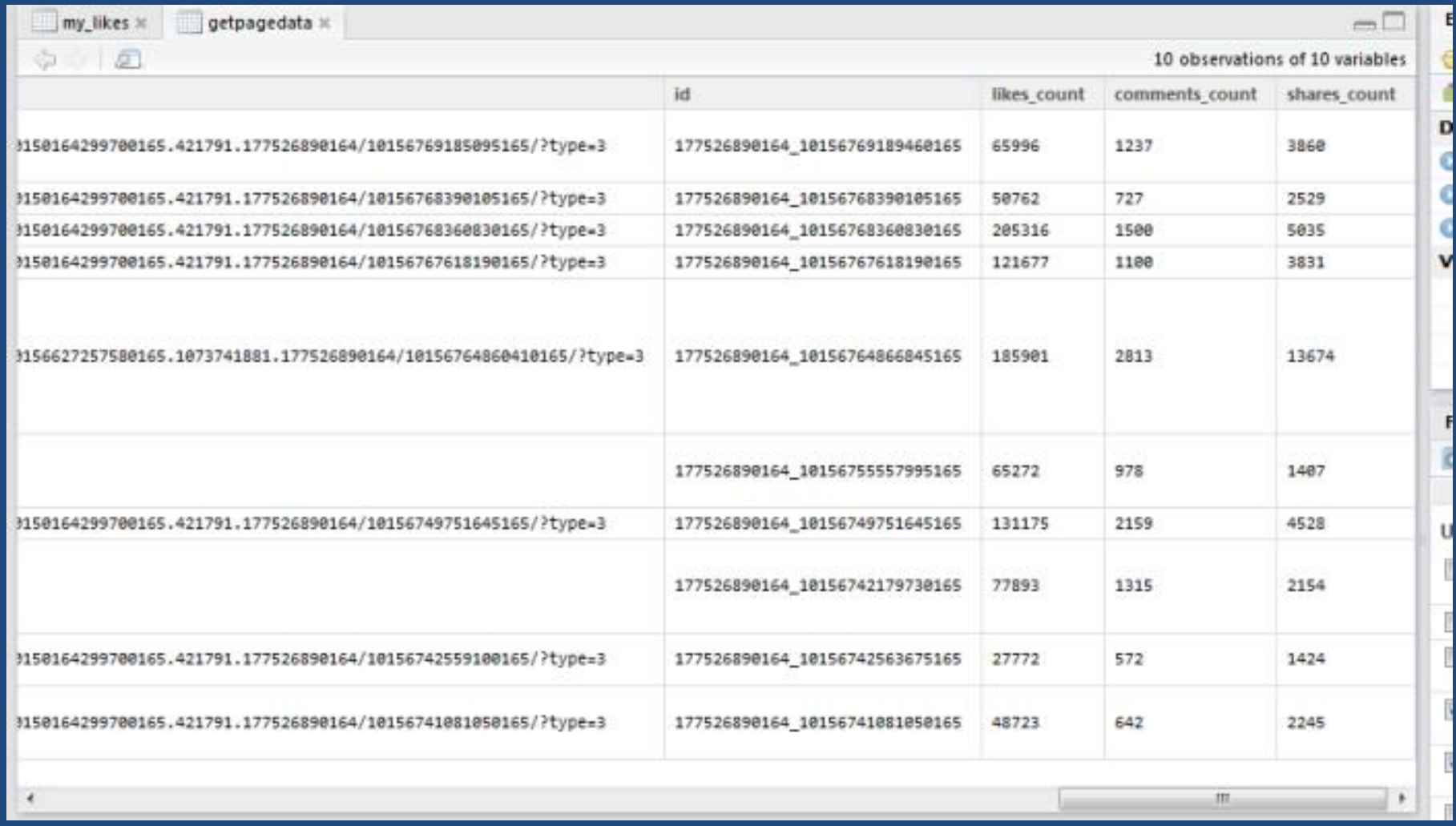
https://bigdataenthusiast.wordpress.com/2016/03/19/mining-facebook-data-using-r-facebook-api/

```
> View(getpagedata)
```

my_likes | getpagedata

10 observations of 10 variables

	from_id	from_name	message
1	177526890164	Narendra Modi	Happy to witness 'Iron Fist 2016' at Pokhran in the Thar Desert, with Rashtrapati ji. Firepower of the Air Force & the skill of the Air Warriors fills our hearts with immense pride. Combat & manc
2	177526890164	Narendra Modi	Reached Jaisalmer. I will be in Pokhran to witness the IAF Fire Power Demonstration ◆Iron Fist 2016.◆ https
3	177526890164	Narendra Modi	Touched to receive a copy of the Gita from Shri Gopal Krishna Goswami Maharaja, Head, ISKCON.
4	177526890164	Narendra Modi	Last evening, had a wonderful meeting with Dr. K.P. Yohannan, Metropolitan Bishop, Believers Church, Kerala.
5	177526890164	Narendra Modi	World Sufi Forum is an extraordinary event, giving the message of peace, tolerance & love. Glad to have atter In a time when terrorism and extremism have become the most destructive force of our times, the message of Su We need to reject any link between terror & religion. Those who spread terror in the name of religion are not Come, together let us challenge violence with kindness & compassion. Let us turn this world into a garden of http://nm-4.com/wbhj
6	177526890164	Narendra Modi	Joined the programme to honour colleagues from the Rajya Sabha who recently retired or are retiring soon. Am sure retiring MPs will continue to work in the fields that interest them & serve the nation. http://nm-4.c
7	177526890164	Narendra Modi	Interacted with a delegation of farmers. Discussed various issues relating to agriculture & farmer welfare.
8	177526890164	Narendra Modi	Projects inaugurated in Hajipur today will make transportation easier & improve economic condition of people Reiterated why India's progress is incomplete without progress of Eastern India & how roads & railways will l http://nm-4.com/s9zi
9	177526890164	Narendra Modi	Delighted to join programme at Patna HC. Congrats to HC for completing a century & best wishes in serving the http://nm-4.com/s6zi
10	177526890164	Narendra Modi	Spoke of India's old relations with IMF & IMF's role in helping in policy making at #AdvancingAsia conference My dream is of a transformed India along with an advanced Asia, where people live with happiness & fulfilment



my_likes | getpagedata

10 observations of 10 variables

	id	likes_count	comments_count	shares_count
0150164299700165.421791.177526890164/10156769185095165/?type=3	177526890164_10156769189460165	65996	1237	3860
0150164299700165.421791.177526890164/10156768390105165/?type=3	177526890164_10156768390105165	50762	727	2529
0150164299700165.421791.177526890164/10156768360830165/?type=3	177526890164_10156768360830165	205316	1500	5035
0150164299700165.421791.177526890164/10156767618190165/?type=3	177526890164_10156767618190165	121677	1100	3831
0156627257580165.1073741881.177526890164/10156764860410165/?type=3	177526890164_10156764866845165	185901	2813	13674
	177526890164_10156755557995165	65272	978	1407
0150164299700165.421791.177526890164/10156749751645165/?type=3	177526890164_10156749751645165	131175	2159	4528
	177526890164_10156742179730165	77893	1315	2154
0150164299700165.421791.177526890164/10156742559100165/?type=3	177526890164_10156742563675165	27772	572	1424
0150164299700165.421791.177526890164/10156741081050165/?type=3	177526890164_10156741081050165	48723	642	2245

function search_groups

- **searchGroup("text",token, n = n):** Find any group with its privacy status & Facebook ID.
- **Arguments:** text: text string, n : Number of groups to return.
- Use below command in R to search groups.

```
> search_groups <- searchGroup("DataMining",token=my_oauth)
> View(search_groups)
```

Here is my result data set “search_groups”. With many observation & 3 Variables.

	name	privacy	id
1	Data Mining and Predictive Analytics News	CLOSED	103235839754287
2	Data Mining / Machine Learning / AI	OPEN	5582633474
3	Analytics, Data Mining, Predictive Modeling, Artificial Intelligence	OPEN	17647299304
4	Big Data, Data Science, Data Mining & Statistics	OPEN	484618291590939
5	Data Mining/Big Data - Social Network Analysis	OPEN	490590987703350
6	Beginning Data Science, Analytics, Machine Learning, Data Mining, R, Python	OPEN	995474220466742
7	Web Scraping and Data Mining	OPEN	540503769381938
8	Data Mining 2013/2014	CLOSED	1478900355664382
9	Data Mining TIF UMG	CLOSED	158538314232927
10	Data Mining 11	CLOSED	647264591963870
11	DATA MINING 3	OPEN	257424714317196
12	Data-Mining	CLOSED	1484850668397040
13	Data Mining	CLOSED	654354121326567
14	DataminingWithShahImranAlam	CLOSED	337017723122281
15	Soft Computing and Data Mining	OPEN	188666561268723
16	<U+0645><U+062C><U+0645><U+0648><U+0639><U+0629> <U+0645><U+062A><U+062E><U+06OT	CLOSED	572493699564509
17	Data Mining ﷺElective 2ﷺ	CLOSED	159537381044025
18	Data mining (unisza)	CLOSED	1533182883585342
19	Data Warehouse and Data Mining batch 2015	OPEN	859564894083270
20	BIT33603 Data Mining	CLOSED	200004710171579
21	Data Mining (Stat 412)	CLOSED	257652777642357
22	Data Mining Bu Nur 2012	CLOSED	196286847139649
23	Data Mining 7E	CLOSED	1512795718966438
24	BITS WILP Advanced Data Mining	CLOSED	949718985110057
25	DATA MINING 5	CLOSED	313148448709071

function getGroup

- **getGroup(ID, token, n = n):** Extract list of posts from a public Facebook page. Whose privacy is open.
- **Arguments:** ID: Group ID , n: Number of posts to return for group.
- **Example**: We have to extract 10 posts from 7th group “Web Scraping and Data mining” present in above image groups.
- Use below command in R to get groups.

```
> groups_post <- getGroup("540503769381938", token=my_oauth, n = 10, since = NULL, until = NULL)
10 posts
> View(groups_post)
>
```

```
> View(groups_post)
>
```

groups_post

	from_id	from_name	message
1	1135178093200197	Anil Bhatia	Hello all, Does any one looking to scrap data from website? Here Aruhat India introduces Datacrops web Data Extraction tool for its users. Datacrops extracts data from websites, social media sites, Business profiles, Reviews si In professional industry many business owners use this tool to strengthen their business whether they have eCommerce site or Online business. This tools helps you to c So if you want to use this tool for your business then just contact on info@aruhat.com Or you can visit our website for more info
2	10205255968272443	Jose Garcia	Join my free course about web scraping and leave a review. Thank you to anyone that joins. If you want an specific topic about web scraping i can create a lecture for https://www.udemy.com/scraping-and-data-mining-from-websites/
3	1052253694821347	Muhammed Ansari	So I am looking for a co-founder and CTO for a new #fintech related start-up anyone interested in taking up the challenge?
4	10205110855865642	Venkata Krishna	This is group is to help people who is looking for Web Scraping Experts. Find a perfect Web Scraper for your requirements. Note: No spam is allowed and no marketing allowed.
5	10205110855865642	Venkata Krishna	This is group is to help people who is looking for Web Scraping Experts. Find a perfect Web Scraper for your requirements. Note: No spam posts, No promotion posts and No marketing posts are allowed. Those posts will be removed from group without notifying .
6	10205110855865642	Venkata Krishna	Invite more people and make this group as center to all scraping related queries:) Happy scraping to all of you
7	10156736017055381	Surya C. Akasam	We are considering to develop web crawler in house which scans & curates the data across the web particular to "anti counterfeiting news". I want to discuss the know how of web crawlers and discussing the feasibility for the same... So share your thoughts and lets get in touch ..Coffee, Lunch or Dinner is on me.
8	1042639152473631	Vidya Sagar S D	hello friends, can you please guide me how can i scrap the webpage and store the data in my data base using php .please guide me step by step .thanks in advance .....
9	10205110855865642	Venkata Krishna	looking for web scraping freelance expert to scrape few websites. Please pm me for more details. This is long term job and continuous.
10	10207059470977853	Richie Enzerillo	I am just starting to hear about web scraping can it be used to gather information such as email addresses and contact info off of social media sites and craigs list e

function searchPages

- **searchPages**(<string>, token, n = n): It Search pages that having a string/keyword.
- **Arguments:** string: any string , n: Number of pages to return
- **Example**: We have to search 10 pages that mention a string “Sports”.

- Use below command in R to search pages.

```
> searchPages <- searchPages("Sports", token = fb_oauth, n = 10)
10 pages
> View(searchPages)
>
```

searchPages

	id	about	category
1	116151972998	Sportskeeda.com is one of the largest all-sports websites based in India. Download our new Android app: http://sportskeeda.co/AndroidLive	News/Media Website
2	1427355644218382	The official Sportskeeda Cricket page. Download our new Android app: http://sportskeeda.co/AndroidLive	Media/News/Publishing
3	105863031495	Sportsnet is a multi-platform network that connects Canadian sports fans to their favourite teams and athletes.	TV Network
4	147262525292598	ESPN's flagship program provides TV's most thorough presentation of sports news, highlights and analysis.	TV Show
5	318816341568454	A multi-brand sporting gear retail destination with over 100 world class brands including Nike,Adidas, Puma,Reebok, Asics, Diamondback, NordicTrack & more.	Outdoor Gear/Sporting Goods

function updateStatus

- updateStatus(“text”, token)
- **Arguments:** text: any string , token

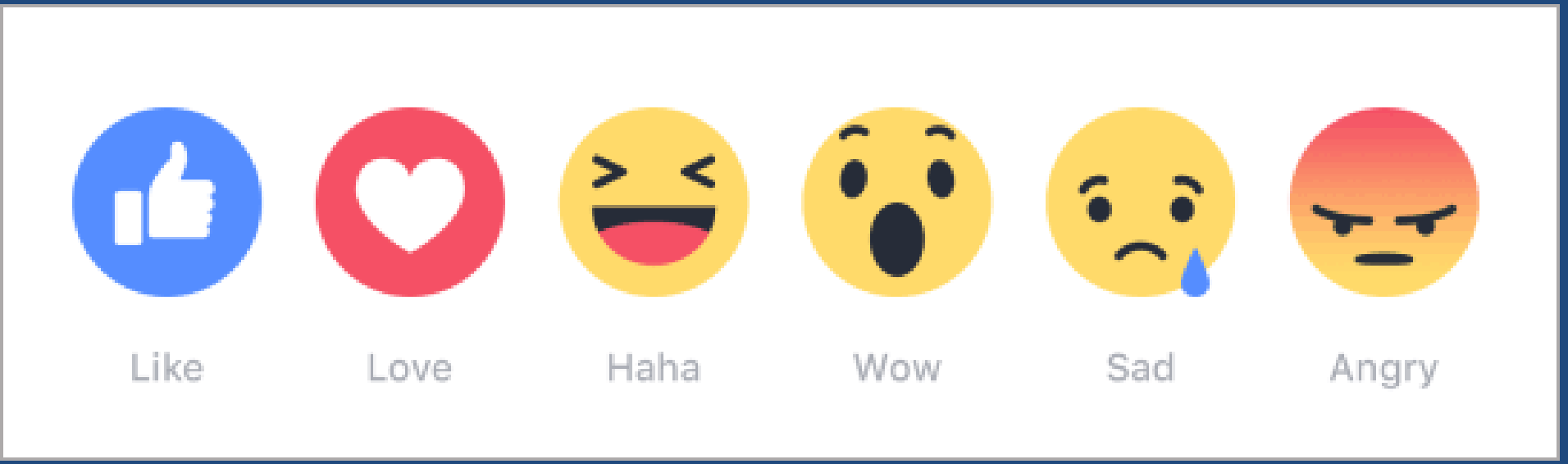
- Use below command in R to update Facebook status.

```
> updateStatus("Updating status using Facebook Graph API and R!", token=my_oauth)
```

Sentiment Analysis in 4 Minutes

- https://www.youtube.com/watch?v=AJVP96tAWxw

Extracting Facebook Reactions for each post into Excel



http://nocodewebscraping.com/how-to-extract-data-from-facebook-page-competitor-analysis/

Sentiment Analysis in R

- https://www.youtube.com/watch?v=2nXwNJO16OE

Facebook graph api : Graph api to extract data from facebook

- https://www.youtube.com/watch?v=qdy_2GS4Yrs

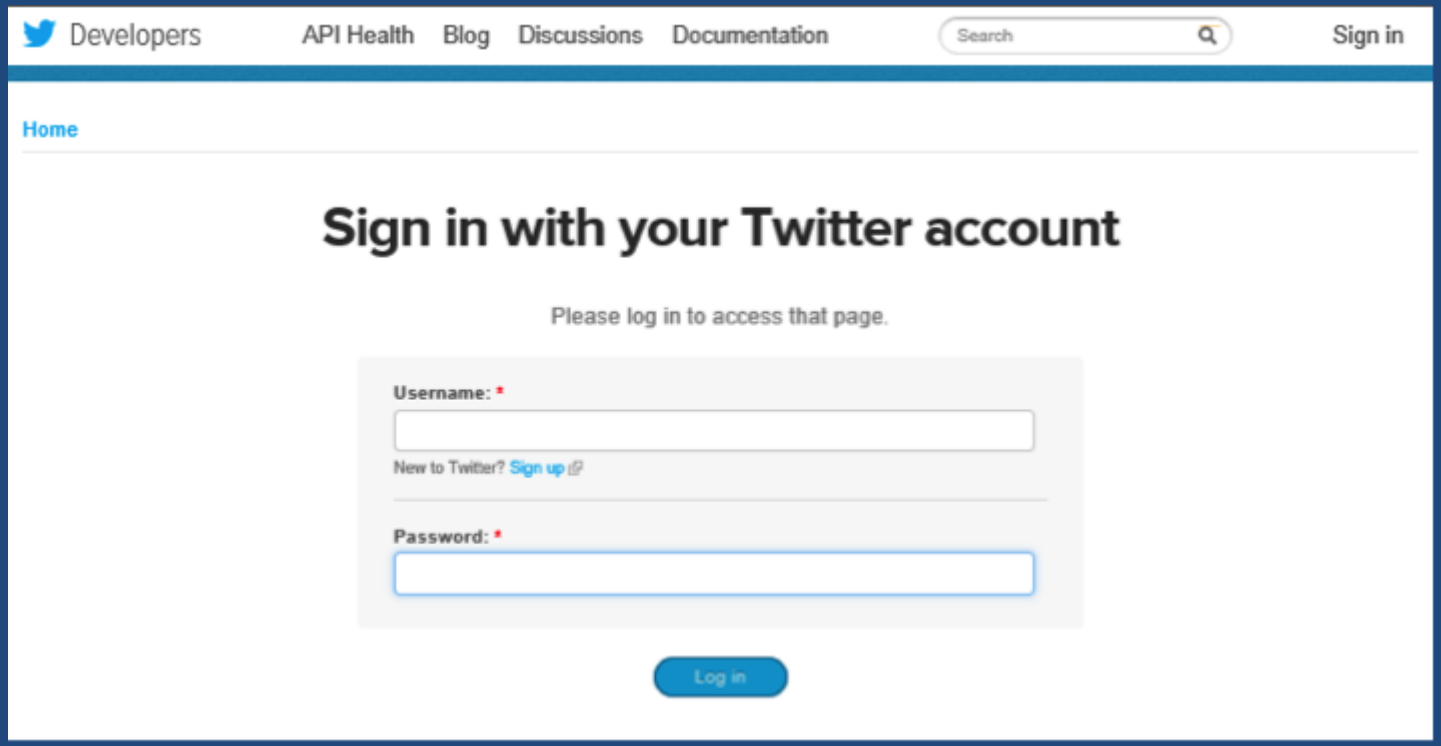
How to extract data from Facebook page: Competitor analysis

- http://nocodewebscraping.com/how-to-extract-data-from-facebook-page-competitor-analysis/

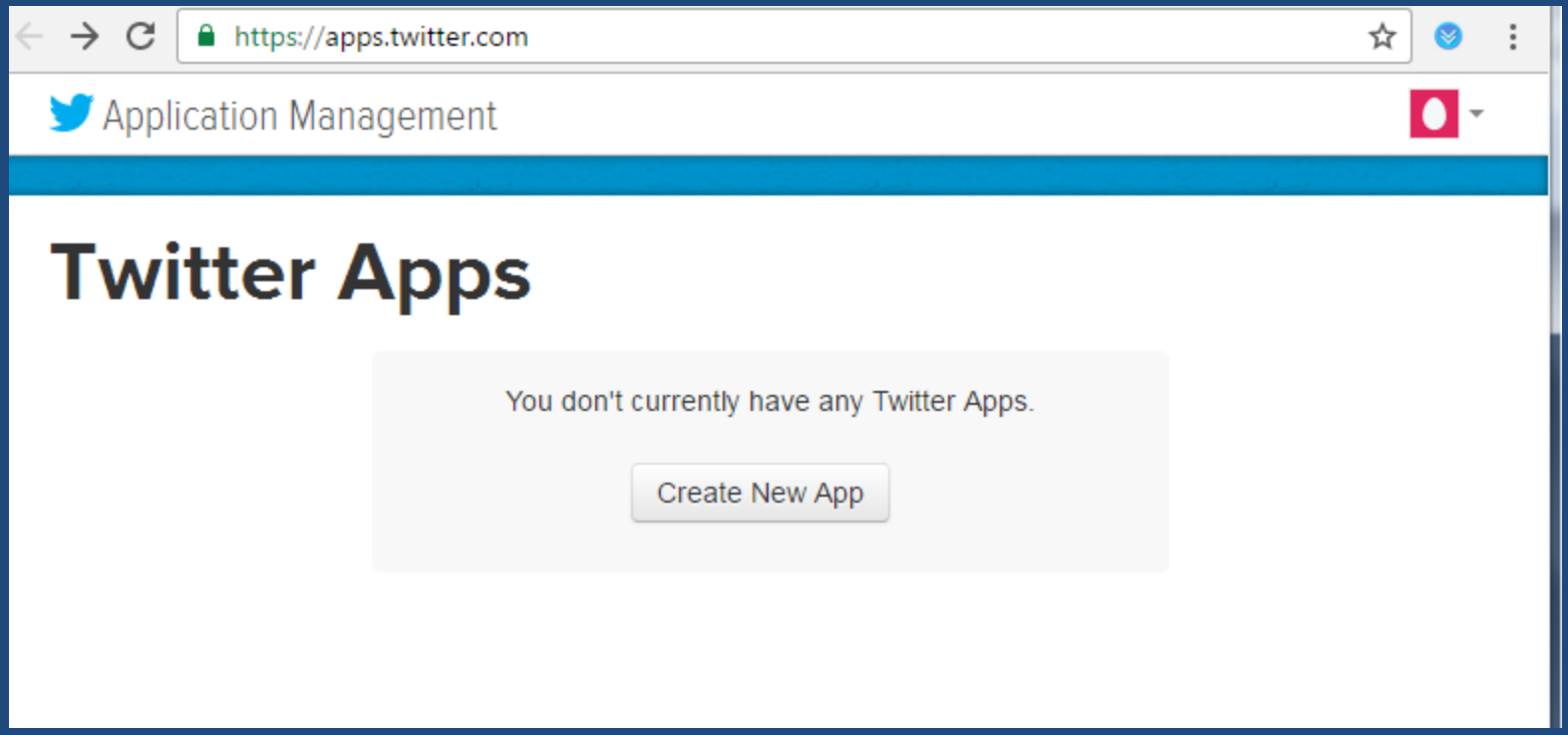
Facebook App Development

- http://www.socialcubix.com/facebook-app-development
- http://www.softwaredeveloper.com/features/develop-facebook-app-072607/
- http://www.emanueleferonato.com/2009/09/20/developing-a-facebook-application-for-absolute-beginners/

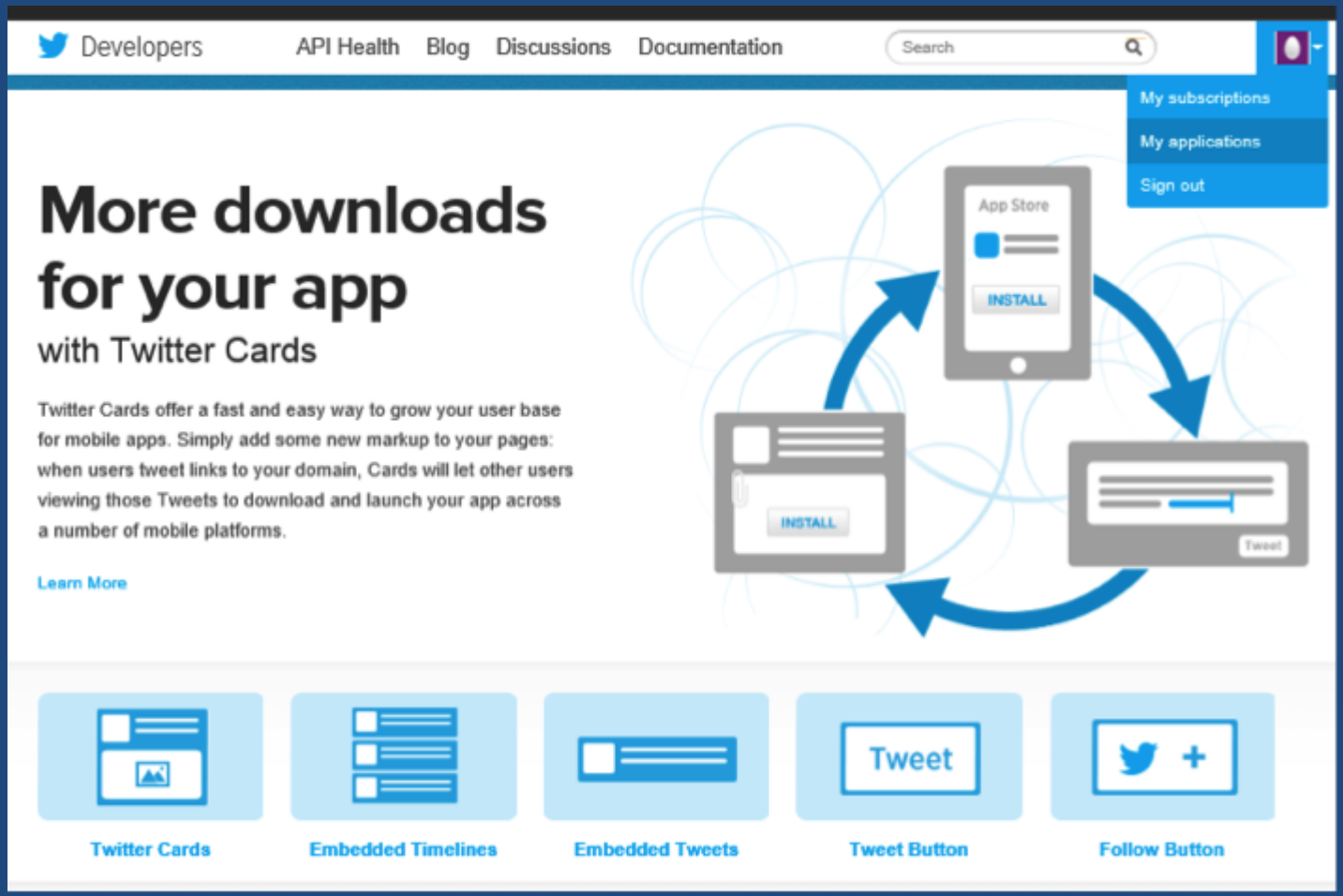
Mining Twitter Data



Mining Twitter Data

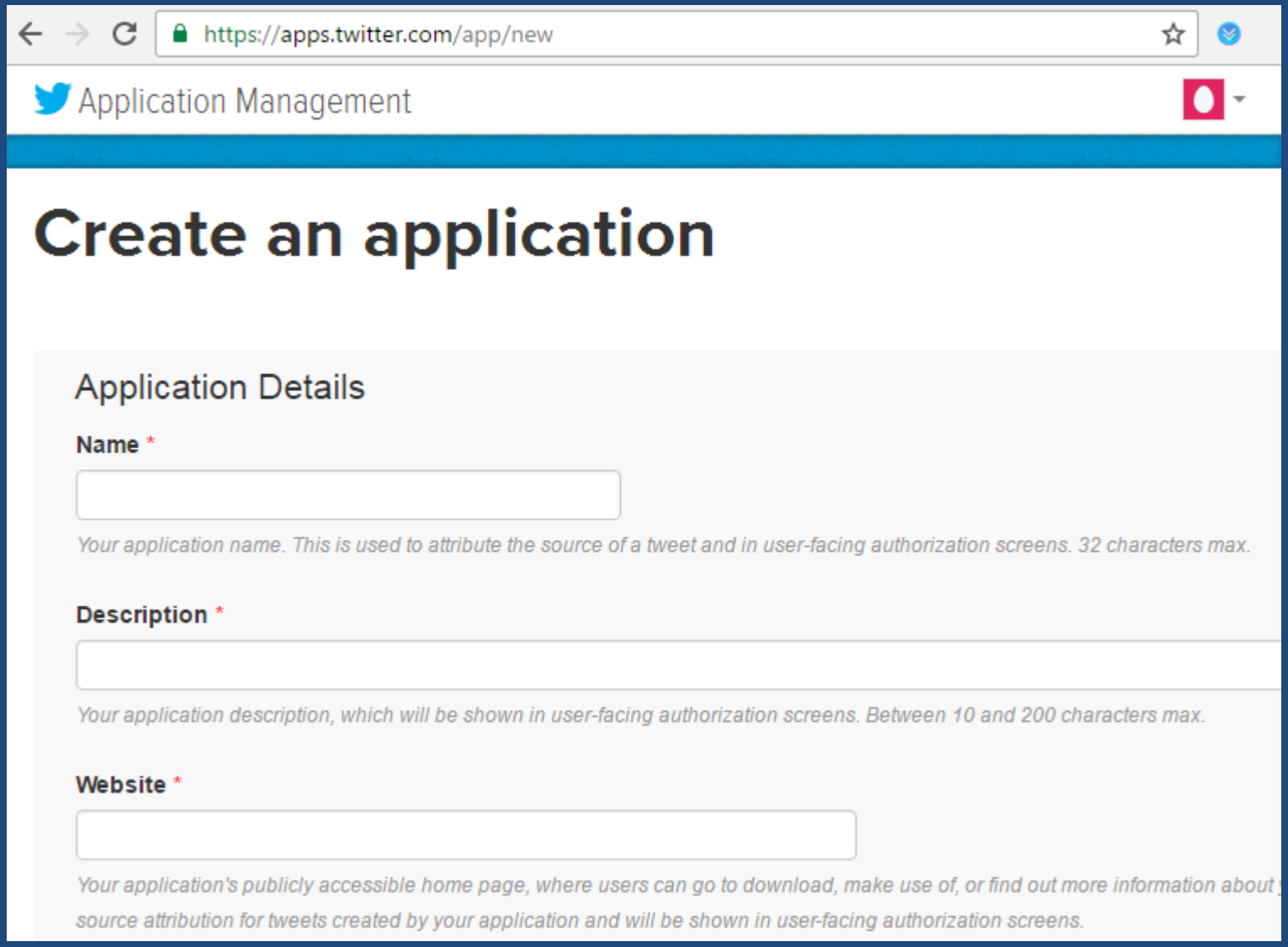


Mining Twitter Data

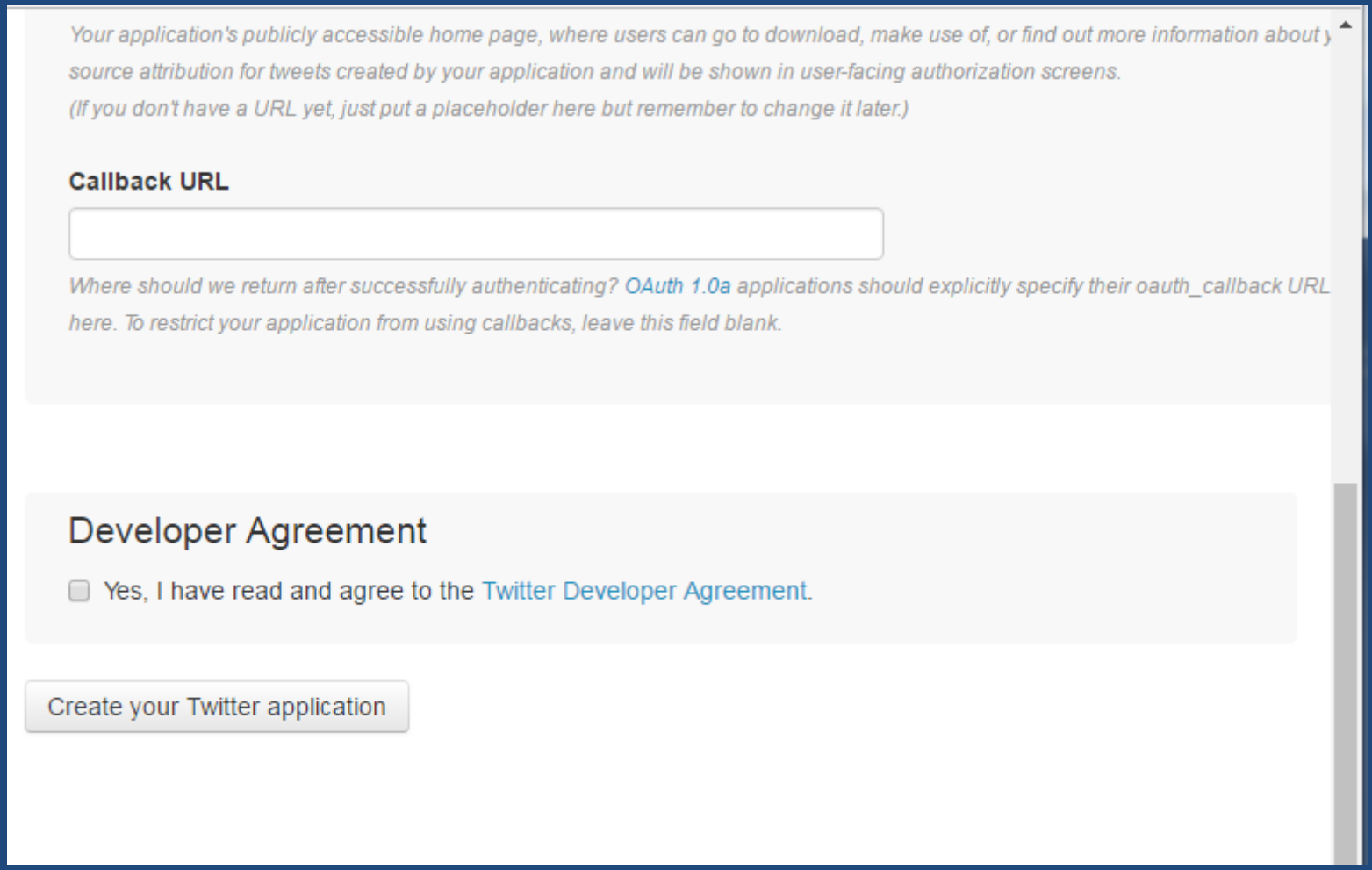


https://www.credera.com/blog/business-intelligence/twitter-analytics-using-r-part-1-extract-tweets/

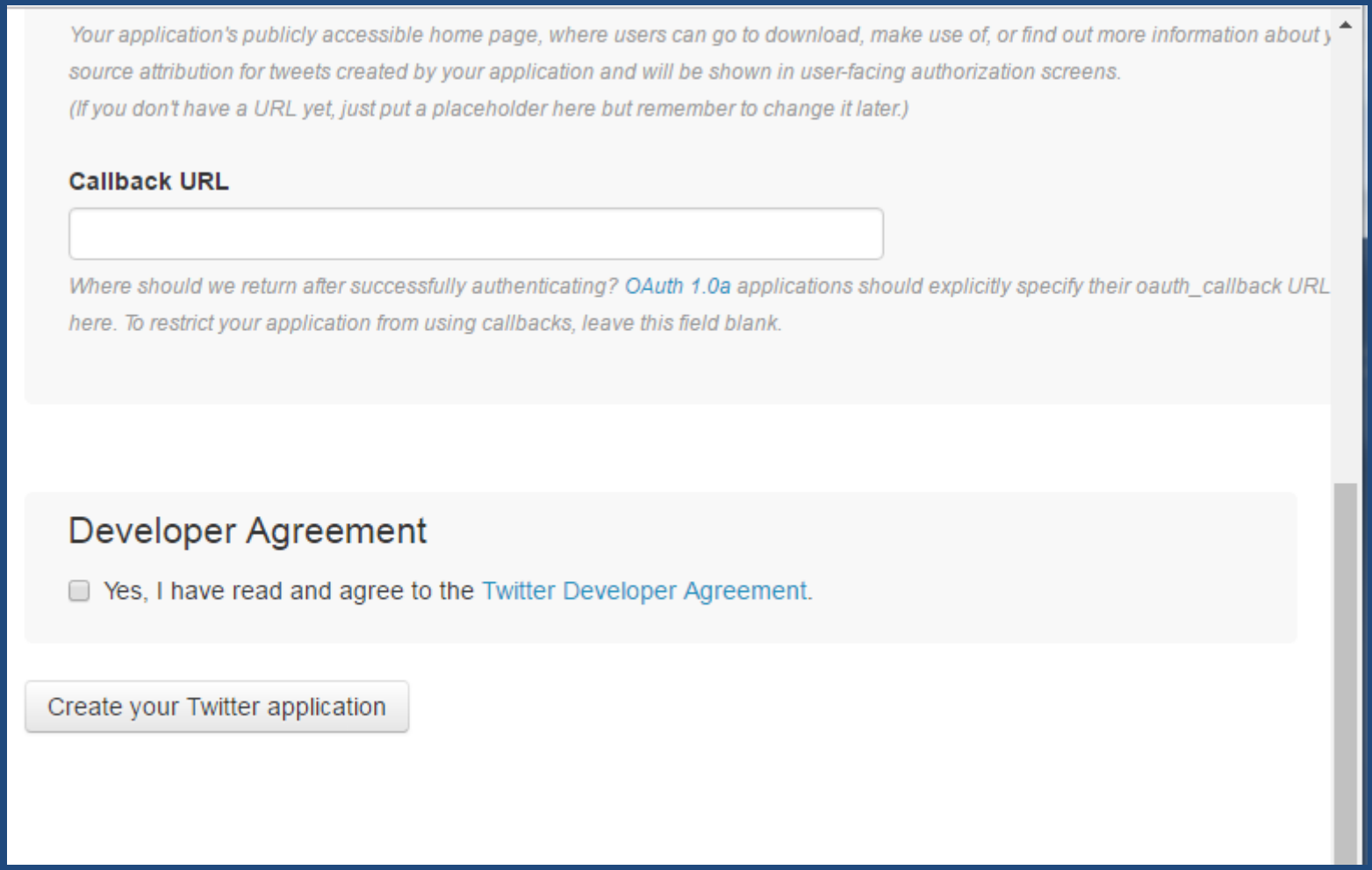
Mining Twitter Data



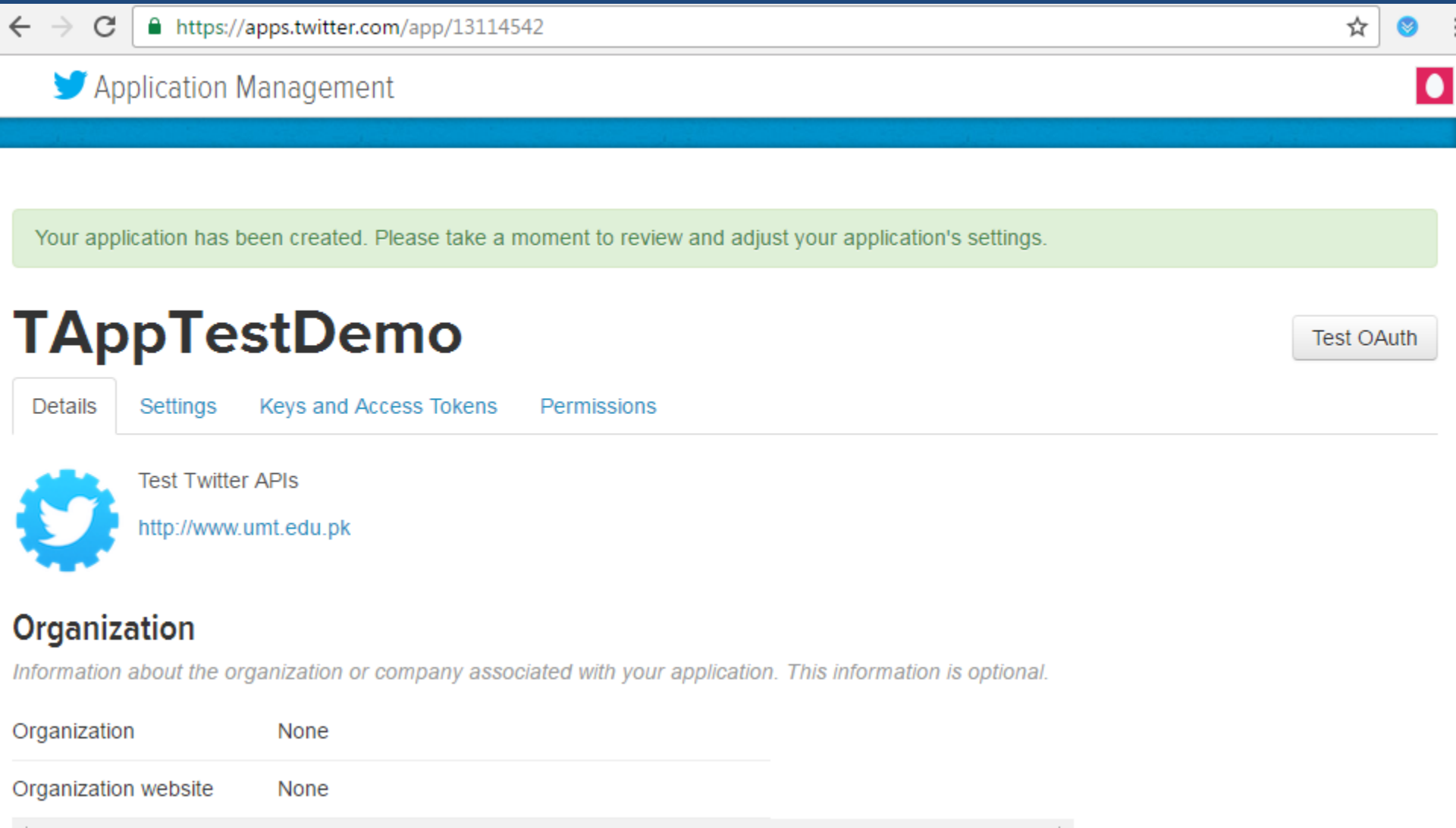
Mining Twitter Data



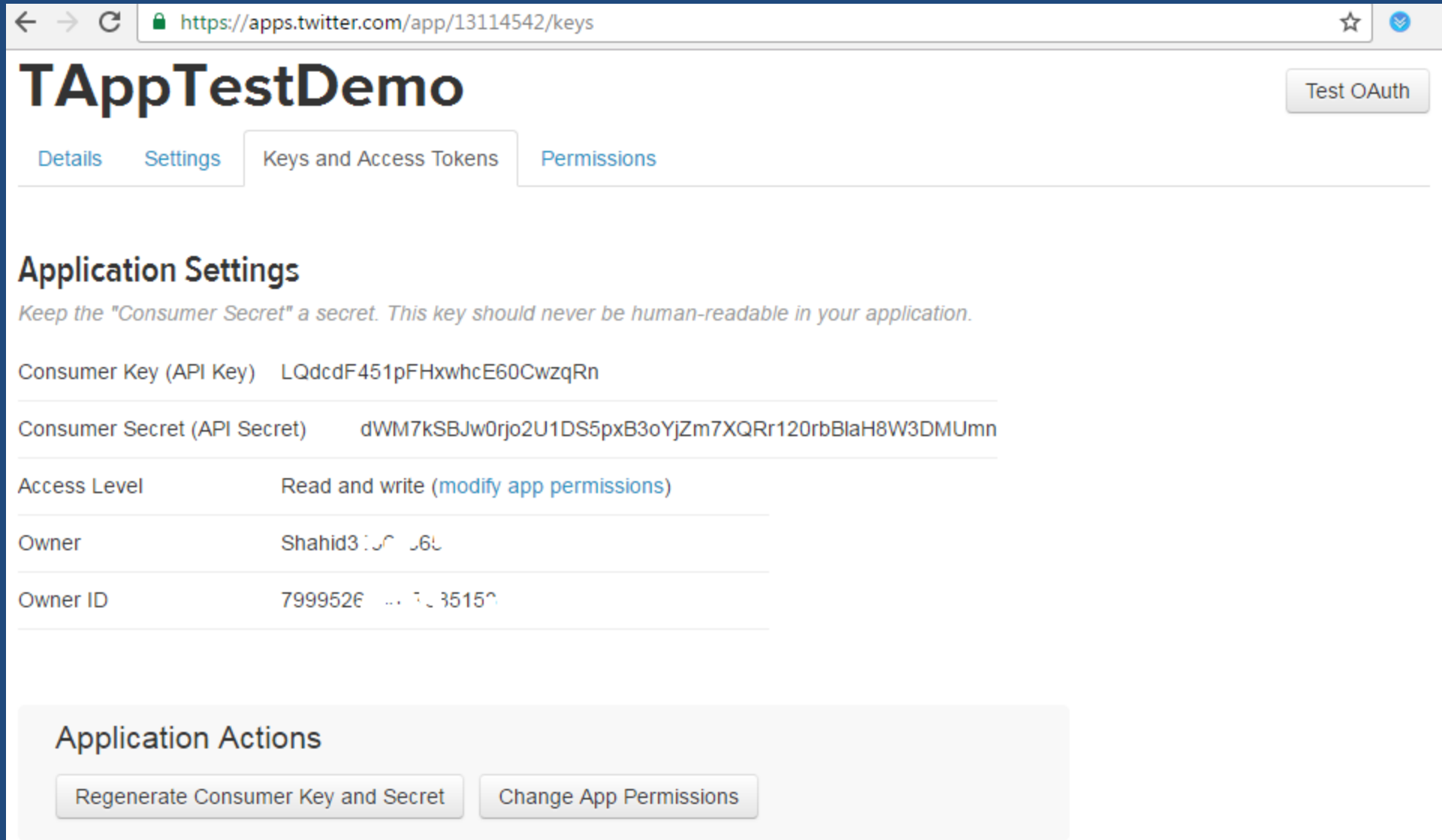
Mining Twitter Data



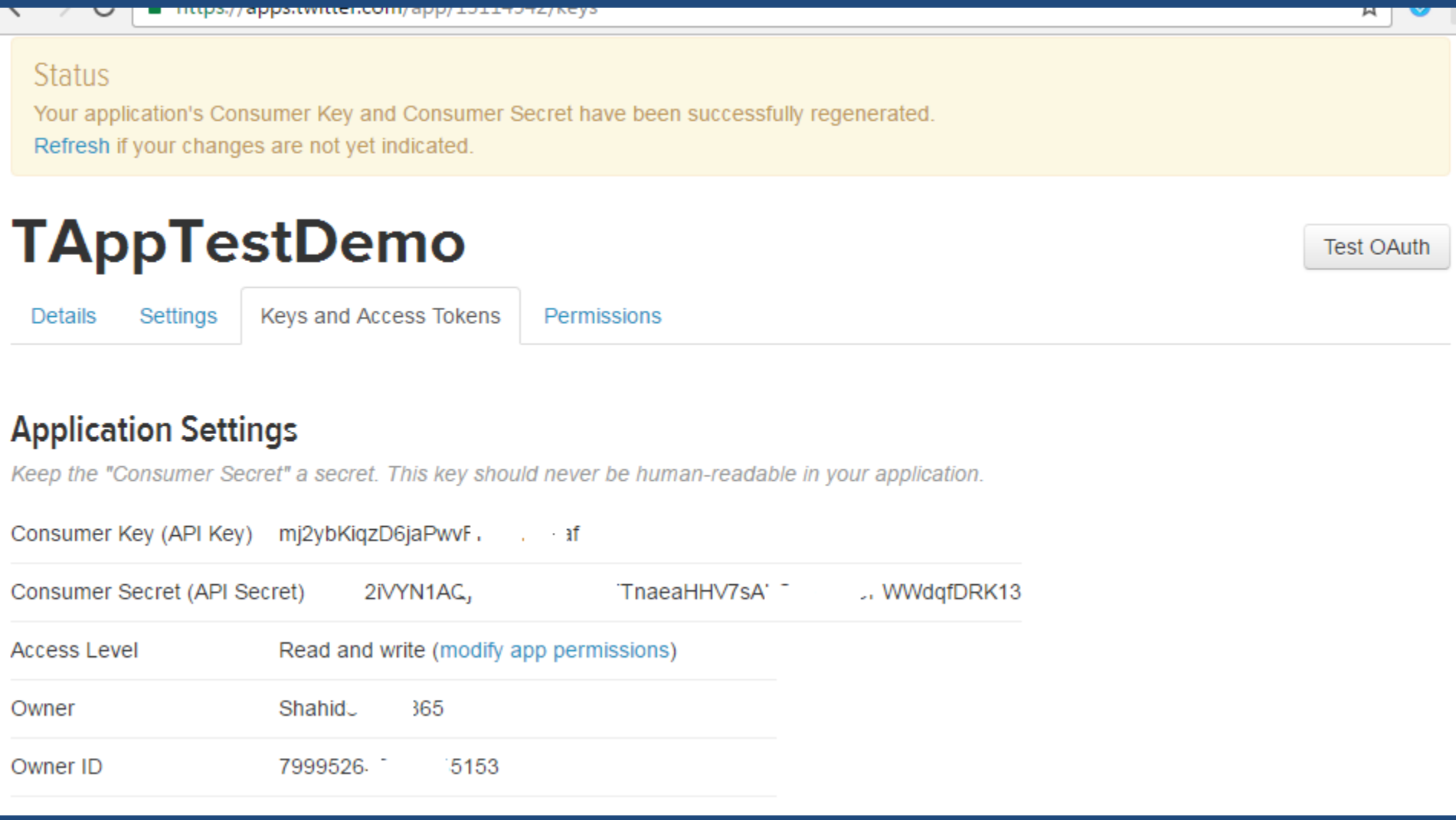
Mining Twitter Data



Mining Twitter Data



Mining Twitter Data



Demo

- https://www.youtube.com/watch?v=lT4Kosc_ers

Links

- https://sites.google.com/site/miningtwitter/basics
- https://sivaanalytics.wordpress.com/2013/10/10/sentiment-analysis-on-twitter-data-using-r-part-i/
- http://andybromberg.com/sentiment-analysis/
- http://www.dataperspective.info/2013/08/sentiment-analysis-using-r.html
- https://www.credera.com/blog/business-intelligence/twitter-analytics-using-r-part-1-extract-tweets/
- https://www.credera.com/blog/business-intelligence/twitter-analytics-using-r-part-2-create-word-cloud/
- https://www.credera.com/blog/technology-insights/open-source-technology-insights/twitter-analytics-using-r-part-3-compare-sentiments/

Step 2: **INSTALL AND LOAD R PACKAGES**

- R comes with a standard set of packages. A number of other packages are available for download and installation. For the purpose of this post, we will need the following packages:
- – ROAuth: *Provides an interface to the OAuth 1.0 specification, allowing users to authenticate via OAuth to the server of their choice.*
- – Twitter: *Provides an interface to the Twitter web API.*
- Let's start by installing and loading all the required packages.
- install.packages("twitteR")
- install.packages("ROAuth")
- library("twitteR")
- library("ROAuth")

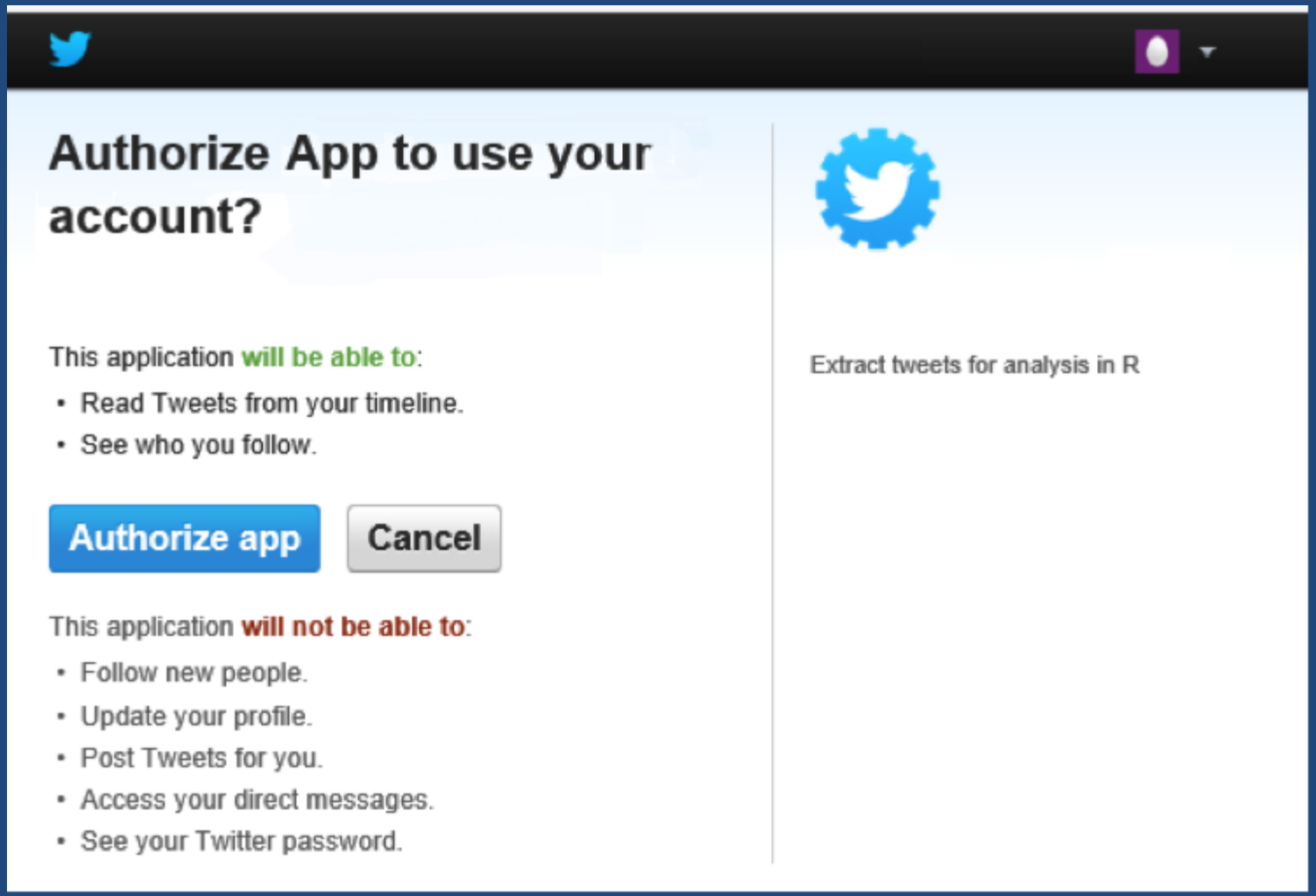
3. CREATE AND STORE TWITTER AUTHENTICATED CREDENTIAL OBJECT

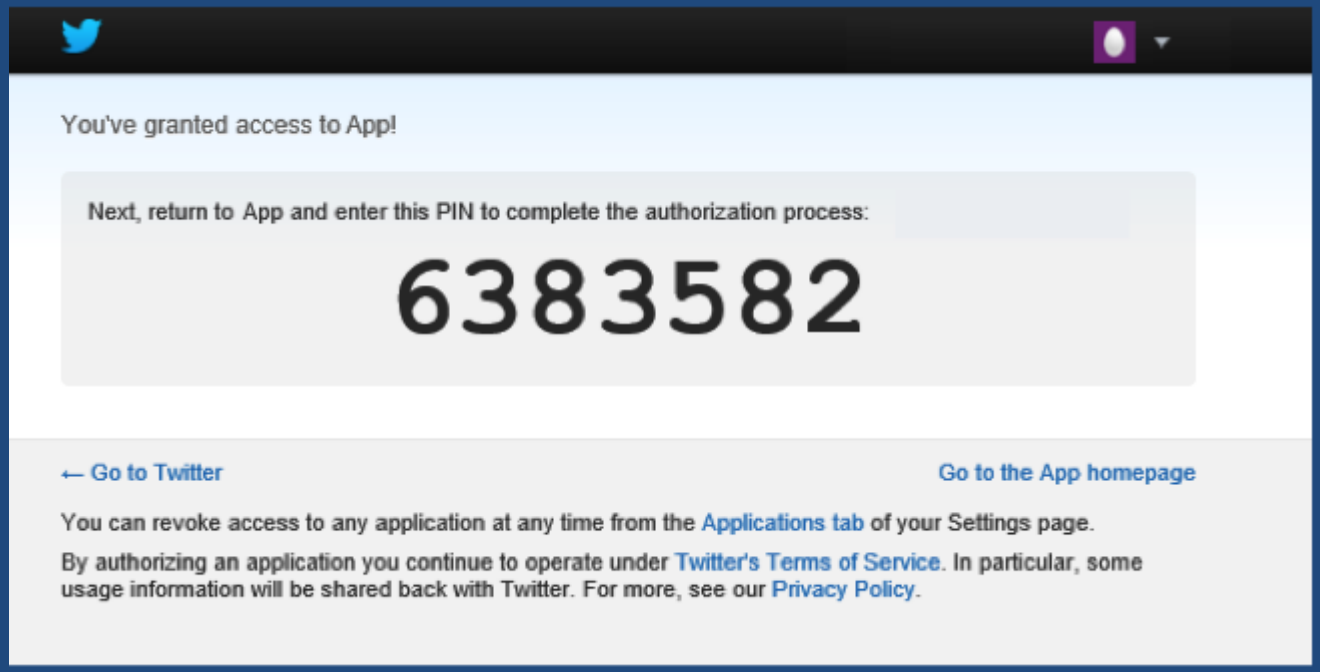
- If you are a Windows user, you need to get “cacert.pem” file. Download the “cacert.pem” file from the specified URL and store it in your working directory. Then create an object “cred” that will save the authenticated object for later sessions and initiate the handshake. This is where you will enter the consumerKey and consumerSecret from the first step. Once the handshake is complete it will direct you to a hyperlink in the console window.

3. CREATE AND STORE TWITTER AUTHENTICATED CREDENTIAL OBJECT

- # Download "cacert.pem" file download.file(url="http://curl.haxx.se/ca/cacert.pem",destfile="cacert.pem") #create an object "cred" that will save the authenticated object that we can use for later sessions cred <- OAuthFactory$new(consumerKey='XXXXXXXXXXXXXXXXX', consumerSecret='XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX', requestURL='https://api.twitter.com/oauth/request_token', accessURL='https://api.twitter.com/oauth/access_token', authURL='https://api.twitter.com/oauth/authorize') # Executing the next step generates an output --> To enable the connection, please direct your web browser to: *<hyperlink>* . Note: You only need to do this part once cred$handshake(cainfo="cacert.pem")

Navigate to the specified link to authorize app and click “Authorize App”.




You've granted access to App!
Next, return to App and enter this PIN to complete the authorization process:
6383582
← Go to Twitter
Go to the App homepage
You can revoke access to any application at any time from the Applications tab of your Settings page.
By authorizing an application you continue to operate under Twitter's Terms of Service. In particular, some usage information will be shared back with Twitter. For more, see our Privacy Policy.

- #save for later use for Windows
- save(cred, file="twitter authentication.Rdata")

4. EXTRACT TWEETS

- Load “twitter authentication.Rdata” file in your session and run registerTwitterOAuth. This should return “TRUE” indicating that all is good and we can proceed. Then we set two variables, one for the search string, which could be a hashtag or user mention, and the second variable is the number of tweets we want to extract for analysis. Use searchTwitter to search Twitter based on the supplied search string and return a list. The “lang” parameter is used below to restrict tweets to the “English” language.

- load("twitter authentication.Rdata") registerTwitterOAuth(cred) search.string <- "#nba" no.of.tweets <- 100 tweets <- searchTwitter(search.string, n=no.of.tweets, cainfo="cacert.pem", lang="en") tweets

Twitter Sentiment Analysis using R

- The implementation of the Review Engine will be as follows:
- Gets Tweets from Twitter
- Clean the data
- Create a Word Cloud
- Create a data dictionary
- Score each tweet.

http://www.dataperspective.info/2013/08/sentiment-analysis-using-r.html

1. EXTRACT TWEETS

- Load the Twitter authentication and extract tweets using #nba.
- load("twitter authentication.Rdata")
  registerTwitterOAuth(cred)

- tweets <- searchTwitter(“#nba”, n=1499, cainfo=”cacert.pem”, lang=”en”)

- tweets.text <- sapply(tweets, function(x) x$getText())

2. CLEAN UP TEXT

- We have already been authenticated and successfully retrieved the text from the tweets using #nba.
- The first step in creating a word cloud is to clean up the text by using lowercase and removing punctuation, usernames, links, etc.
- We are using the function gsub to replace unwanted text. Gsub will replace all occurrences of any given pattern.
- Although there are alternative packages that can perform this operation, we have chosen gsub because of its simplicity and readability.

2. CLEAN UP TEXT

- #convert all text to lower case
- tweets.text <- tolower(tweets.text)

- # Replace blank space (“rt”)
- tweets.text <- gsub("rt", "", tweets.text)

- # Replace @UserName
- tweets.text <- gsub("@\\w+", "", tweets.text)

- # Remove punctuation
- tweets.text <- gsub("[[:punct:]]", "", tweets.text)

2. CLEAN UP TEXT

- # Remove links
- tweets.text <- gsub("http\\w+", "", tweets.text)

- # Remove tabs
- tweets.text <- gsub("[|\t]{2,}", "", tweets.text)

- # Remove blank spaces at the beginning
- tweets.text <- gsub("^ ", "", tweets.text)

- # Remove blank spaces at the end
- tweets.text <- gsub(" $", "", tweets.text)