Twitter Sentiment Analysis

CAB432 – Assessment 2

Authors:
Kevin Duong
N9934731

Luna McLaren
N9720472

Due Date:
27th October 2019

Table of Contents

Introduction

Purpose & Description

Twitter Query Processor is a visual data service primarily based on the social media application 'Twitter', where users can look up a "hashtag" or multiple hashtags as queries in one submission via search box. A "live filter" is applied to reject tweets that are not related to the query/s. This filter works by finding each hashtag from the queries within a message tweet and drawing out those tweets onto the screen. The queries shall remain 'active' until the user decides to click cancel; this will constantly add Twitter message after Twitter message in the feed. If the user cancels the live feed, they may grab a sample summary of the tweets shown in the feed, which is supported by a tag cloud chart for text data analysis.

The main purpose of this application is to analyse and interpret sentimental value polarity for each hashtag/tweet. This reflects the type of words being used, which correlates to social media's response to an event or topic. These words carry significant importance by size and often leaves meaning to what people have to say. The sentiment mean score is given to users who wish to see the feed statistics, summing all sentiment numbers and dividing it by the number of observed values. A low to negative value number indicates that the hashtag is used negatively, while a high number aiming close to 1 or greater shows the hashtag is expressed with positive opinions and attitudes. For a brief walkthrough, see the user guide (Appendix A).

Services Used

Twitter public REST API V1.1

Twitter's REST API allows users to be able to retrieve tweets and related information from Twitter. In the application, queries sent to the server via submission will therefore be responded with results with the help of the API.

https://developer.twitter.com/en/docs

Redis Server

Redis is an open source, RAM based data structure storage service which can be used as a database or a cache. It supports multiple types of data structures and will be used by this application for short term caching.

https://redis.io/

Amazon S3 Service

Amazon S3 is a simple object storage service provided by AWS, which allows for scalability, data availability, security and performance. This will be used for longer term storage and in instances in which the Redis cache is unavailable.

https://aws.amazon.com/s3/

Libraries Used

Natural Node

"Natural" is a natural language processing module for Node JS, specifying in tokenizing, stemming, classification, phonetics, tf-idf, WordNet, string similarity and inflection. It will help serve the user in finding their tweets relating to the submitted queries and calculating sentimental values.

https://github.com/NaturalNode/natural

Stopword

"Stopword" is another natural language processing module for Node JS that filters out common words in a language. Words such as 'the', 'if' and 'what' will be removed which prioritises important words first for visual data analysis.

https://www.npmjs.com/package/stopword

CSS: Bootstrap

A CSS framework to provide responsive front-end web development, containing designed templates for buttons, navigation and other interface components. This template can be used to help replicate Twitter's theme.

https://getbootstrap.com/

amCharts 4: WordCloud

For visual representation, WordCloud is a tag cloud format that is able to present text data in a creative way. It can show the word size based on its frequent usage along with hundreds of other words and their value. A large chunk of tweets and hashtags will be used to formulate the cloud.

https://www.amcharts.com/docs/v4/chart-types/wordcloud/

Twitter for Node.js

An asynchronous library for Node which includes wrappers for accessing the Twitter REST API.

https://www.npmjs.com/package/twitter

Redis for Node.js

A Redis client for Node JS which the application uses to communicate with the Redis server.

https://www.npmjs.com/package/redis

AWS SDK

The official SDK for JavaScript which the application uses to communicate with the Amazon S3 storage service.

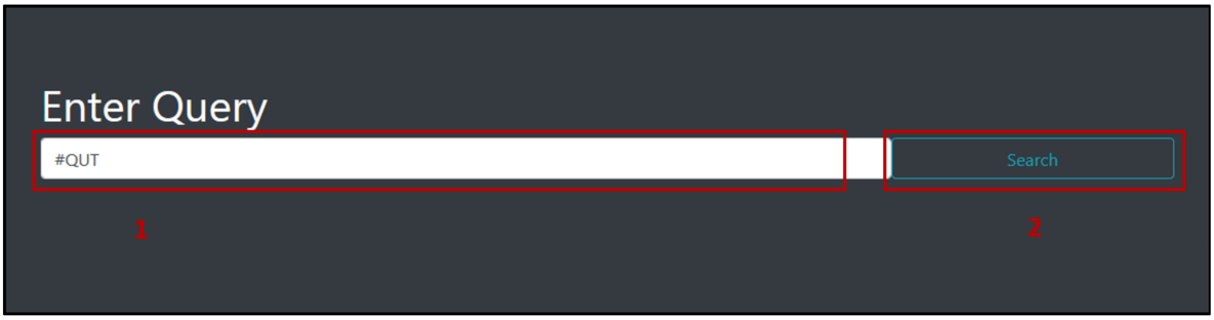
https://www.npmjs.com/package/aws-sdk

Use Cases

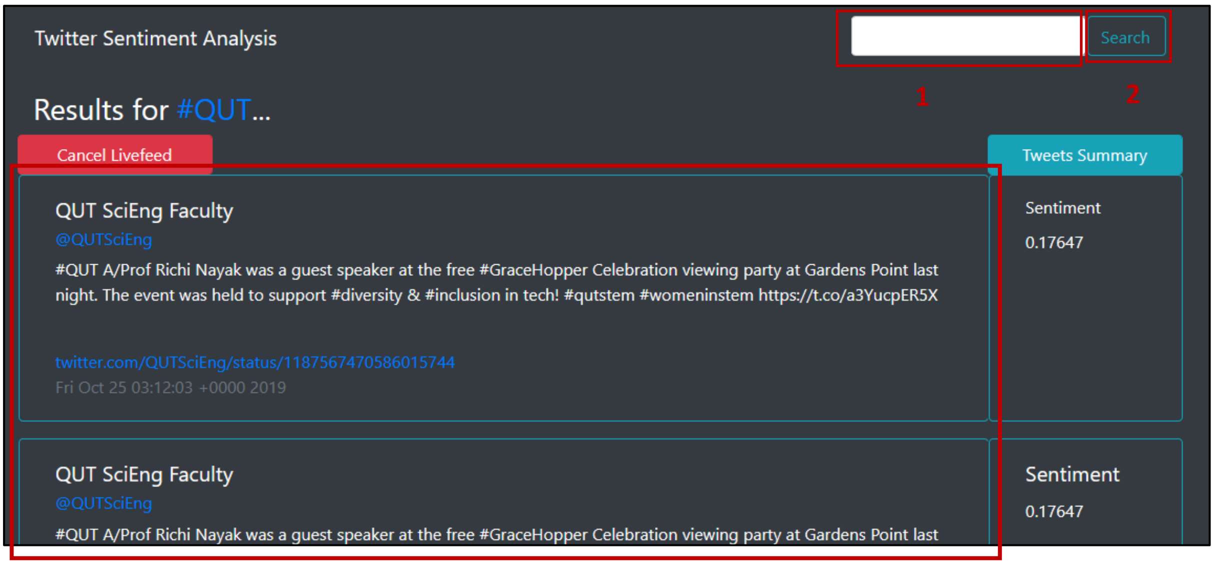
Use 1

As a user I would like to draw out tweets from queries I've submitted.

The user is introduced to the index page and decides to enter a trending hashtag from Twitter. Satisfied with the queries, the user then clicks the search button for submission.



A page full of tweets is shown to the user right after submitting the queries. Here, the user can read the tweets shown in the feed as well as fresh tweets added to the feed. The user can also search for queries again by using the search box on the navigation bar.



Results for #QUT...

QUT SciEng Faculty
@QUTSciEng
#QUT A/Prof Richi Nayak was a guest speaker at the free #GraceHopper Celebration viewing party at Gardens Point last night. The event was held to support #diversity & #inclusion in tech! #qutstem #womeninstem https://t.co/a3YucpER5X
twitter.com/QUTSciEng/status/1187567470586015744
Fri Oct 25 03:12:03 +0000 2019

Sentiment
0.17647

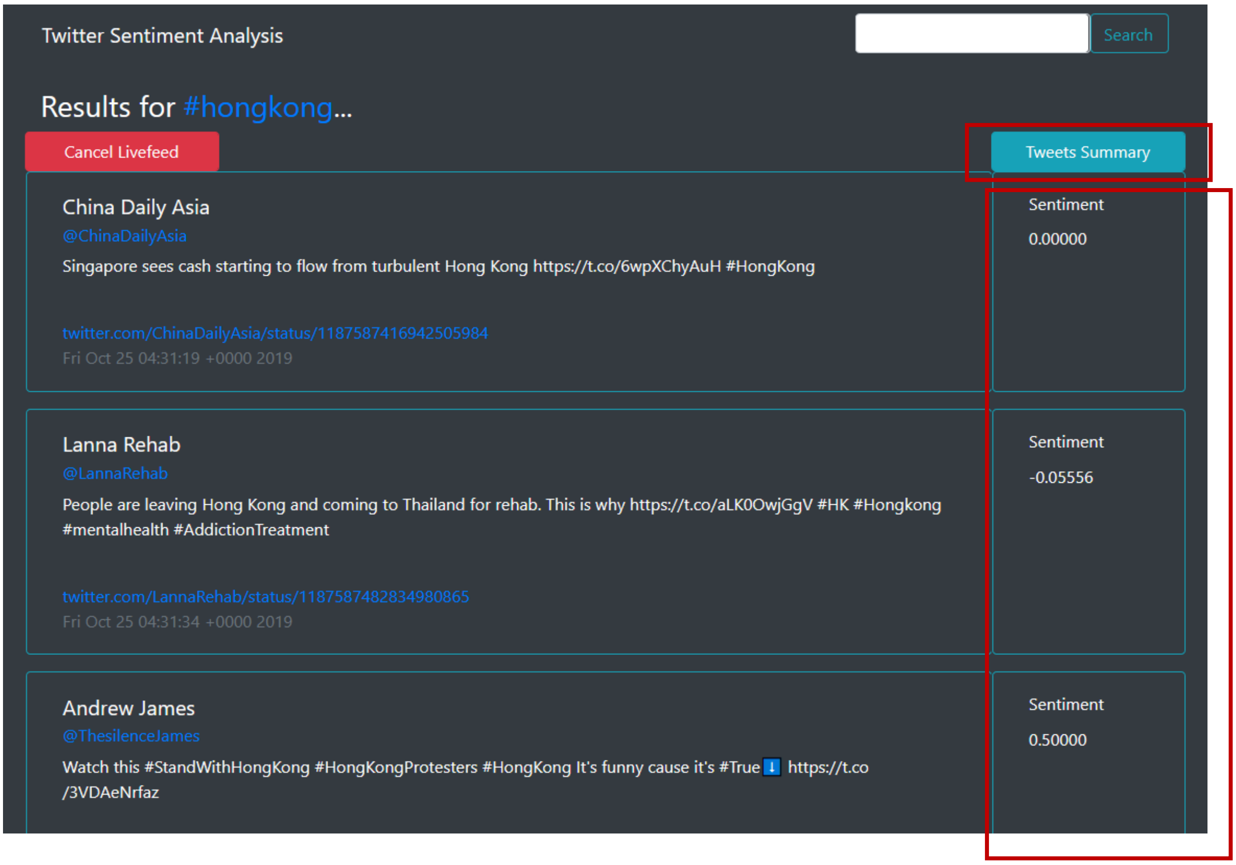
QUT SciEng Faculty
@QUTSciEng
#QUT A/Prof Richi Nayak was a guest speaker at the free #GraceHopper Celebration viewing party at Gardens Point last

Sentiment
0.17647

Use 2

As a user I want to know what people are feeling about the hashtags used.

The user wishes to lookup a trending hashtag and requests its tweets. As the user reads the feed, they can view the sentiment values which correlates to the tweet expressed in the left column. This indicates how well opinions are received with the hashtag being used.



Results for #hongkong...

China Daily Asia
@ChinaDailyAsia
Singapore sees cash starting to flow from turbulent Hong Kong https://t.co/6wpXChyAuH #HongKong
twitter.com/ChinaDailyAsia/status/1187587416942505984
Fri Oct 25 04:31:19 +0000 2019

Sentiment
0.00000

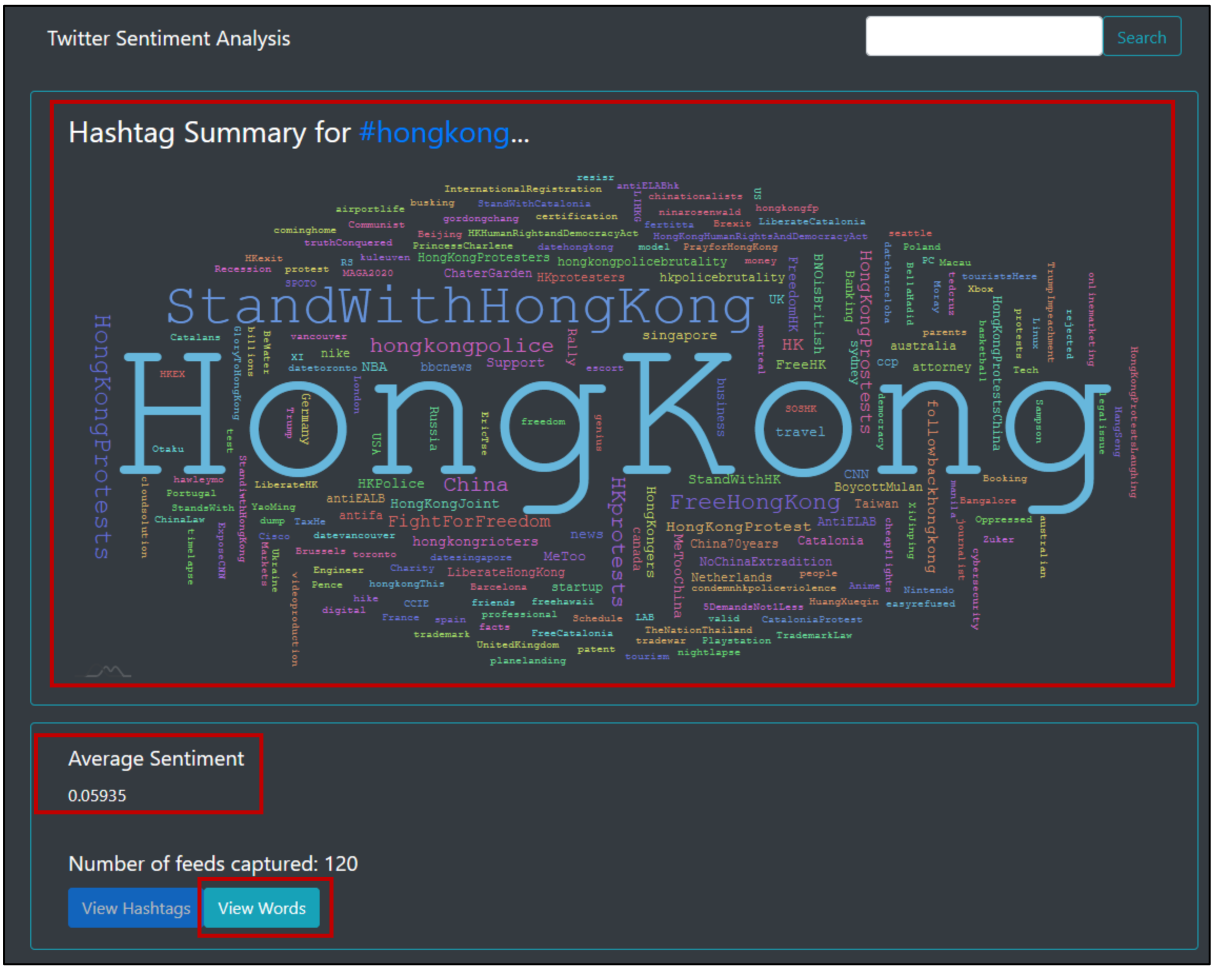
Lanna Rehab
@LannaRehab
People are leaving Hong Kong and coming to Thailand for rehab. This is why https://t.co/aLK0OwjGgV #HK #Hongkong #mentalhealth #AddictionTreatment
twitter.com/LannaRehab/status/1187587482834980865
Fri Oct 25 04:31:34 +0000 2019

Sentiment
-0.05556

Andrew James
@ThesilenceJames
Watch this #StandWithHongKong #HongKongProtesters #HongKong It's funny cause it's #True⬇ https://t.co /3VDAeNrfaz

Sentiment
0.50000

Furthermore, if the user wants to know the overall thoughts and opinions shared around the hashtag, they can click on the 'Tweets Summary' button to get a visual idea of other hashtags used with the query, and keywords and phrases expressed with sentimentalism.


Twitter Sentiment Analysis

Hashtag Summary for #hongkong...

Average Sentiment

0.05935

Number of feeds captured: 120

View Hashtags View Words

Technical Description

Application Architecture

Client vs Server

Server

Tweet Retrieval: The Twitter API is used to retrieve recent tweets by searching a query specified by the user. However, before the application checks Twitter, it retrieves the most recent tweets from the Redis Cache, or Amazon S3 if it is not present in the cache. Once this is complete it checks for any tweets more recent than the latest one in the cache. If the cache and S3 is empty, the application simply checks the Twitter API.

Sentiment Analysis: Once the tweets are fetched, it passes them onto the Natural API to perform sentiment analysis. Once this is complete, the server caches all the results and returns them to user in the form of HTML.

Data Visualisation: The server fetches the latest feeds from the Redis cache and analyses each tweet in the data. Each sentiment value per tweet is added together for the total value and calculated for the average. If a value is 0, then the length of the observed values is decreased for this will not affect the average. The tweets are parsed into strings and filtered for two outputs: the data containing all hashtags in the feeds and the data for words from all tweets. These outputs are passed on to WordCloud to analyse in the client.

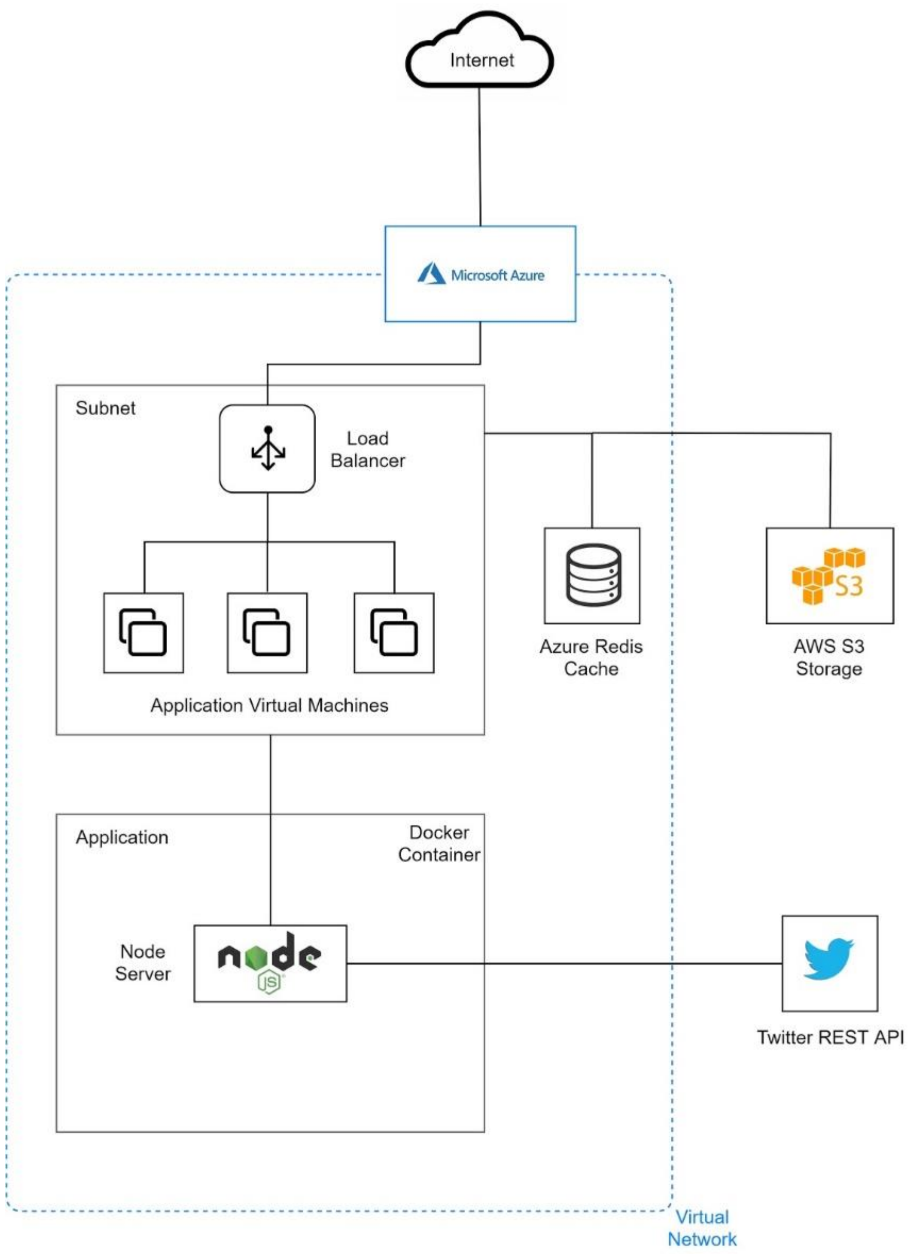
Client

Webpage: The results of the sentiment analysis are shown in the form of a HTML page.

Live Filter: A JavaScript timer fires every 5 seconds, calling a GET request to the server to retrieve all tweets newer than the latest in the page. The response to this is JSON, which is used to dynamically update the page with new tweets.

Data Visualisation: Supported by amCharts 4, the WordCloud retrieves the data passed by the server to load a visual representation of text data for a summary of hashtags and notable keywords expressed in tweets.

Network Diagram



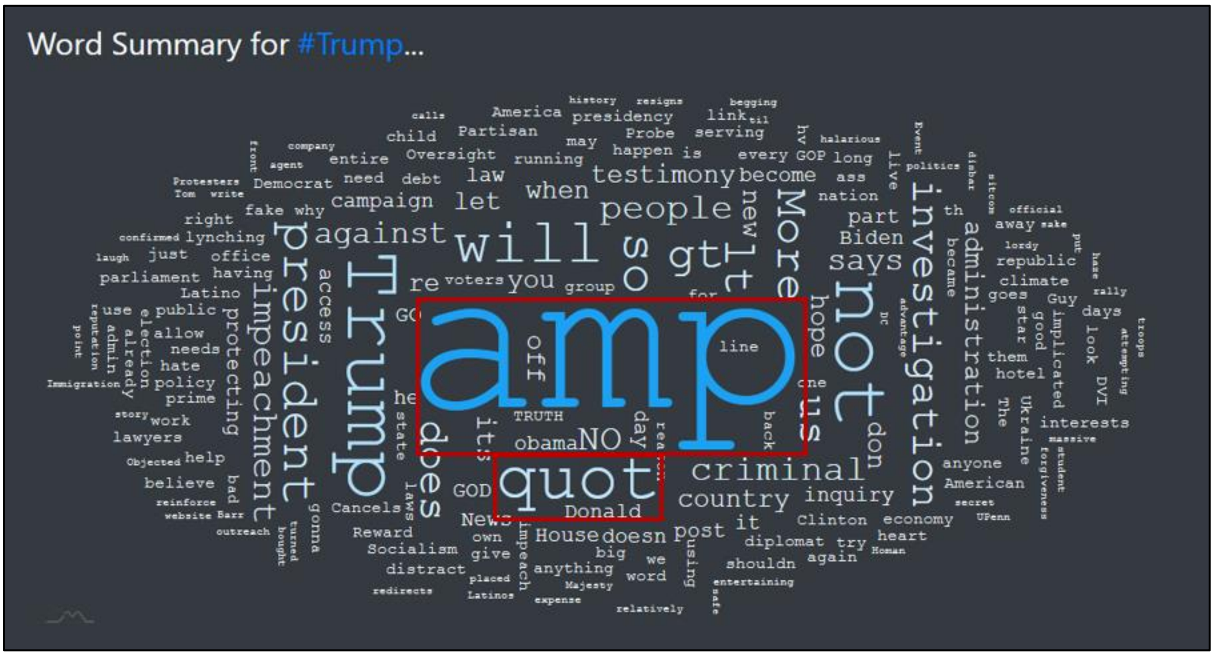
Testing and limitations

Task	Expected Outcome	Result
Search for hashtags	Displays a feed full of tweets	PASS
Enter incorrect query	Prompts an error message	PASS
Update new tweets	Fades in new tweets to add to feed	PASS
Cancel active filter	Stops feed update	PASS
Load tag cloud (hashtags)	Renders WordCloud made out of hashtags	PASS
Load tag cloud (words)	Renders WordCloud made out of keywords	PASS
Click a hashtag in cloud	Redirects user with a query to Twitter results	PASS

Issues

Unexpected Words in Data Visualisation

In most cases when previewing the tag cloud for words used in the summary page, the most frequent words expressed in the data is ‘amp’ and ‘quot’. Even by removing the irrelevant words with regular expression, the text data becomes corrupted and will misspell every word in the cloud. It is assumed that the words are ‘&’ and “ ” and are integral to help structure the data.



Word Summary for #Trump...

Possible extensions

Capture Data

If the user likes to keep an image of the tag clouds displayed, a button is available to click and download the captures. There will also be a time log to show when the tags were created.

Pagination

As the page contains a long list of tweets, a pagination feature would be beneficial to reduce the length of the page when scrolling down the feed.

References

1. Twitter. (n.d). Docs. Retrieved from
   https://developer.twitter.com/en/docs

2. Redis. (n.d). Redis. Retrieved from
   https://redis.io/

3. Amazon. (n.d). Amazon S3. Retrieved from
   https://aws.amazon.com/s3/

4. GitHub. (2019). NaturalNode. Retrieved from
   https://github.com/NaturalNode/natural

5. Npm. (2019). Stopword. Retrieved from
   https://www.npmjs.com/package/stopword

6. Bootstrap. (n.d). Bootstrap. Retrieved from
   https://getbootstrap.com/

7. Amcharts. (n.d). Anayomt of a WordCloud. Retrieved from
   https://www.amcharts.com/docs/v4/chart-types/wordcloud/

8. Npm. (2017). Twitter for Node.js. Retrieved from
   https://www.npmjs.com/package/twitter

9. Npm. (2017). Redis – a node.js redis client. Retrieved from
   https://www.npmjs.com/package/redis

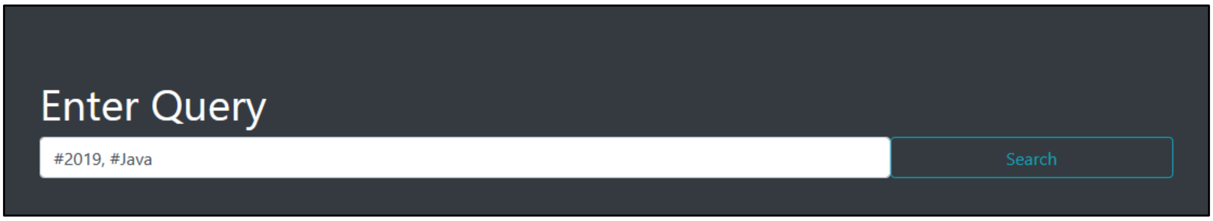
10. Npm. (2019). AWS SDK for JavaScript. Retrieved from
    https://www.npmjs.com/package/aws-sdk

Appendices

Appendix A: Brief User Guide

Getting Started

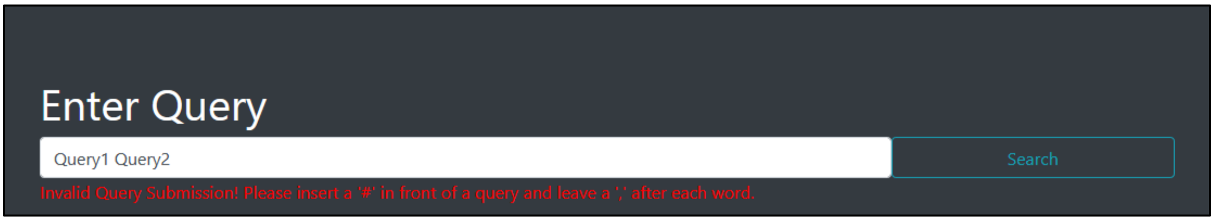
Welcome to Twitter Query Processor, a web application that can lookup hashtags in Twitter to retrieve related tweets and perform sentiment analysis. The first page requires a single or handful of queries to be submitted before results can be achieved.



Enter Query

#2019, #Java

As a format rule, the query must contain a hashtag '#' in front of the word. With the case of having multiple queries, a comma must be inserted between two words. If the formatting is incorrect, an error message will display to indicate how to submit a query.



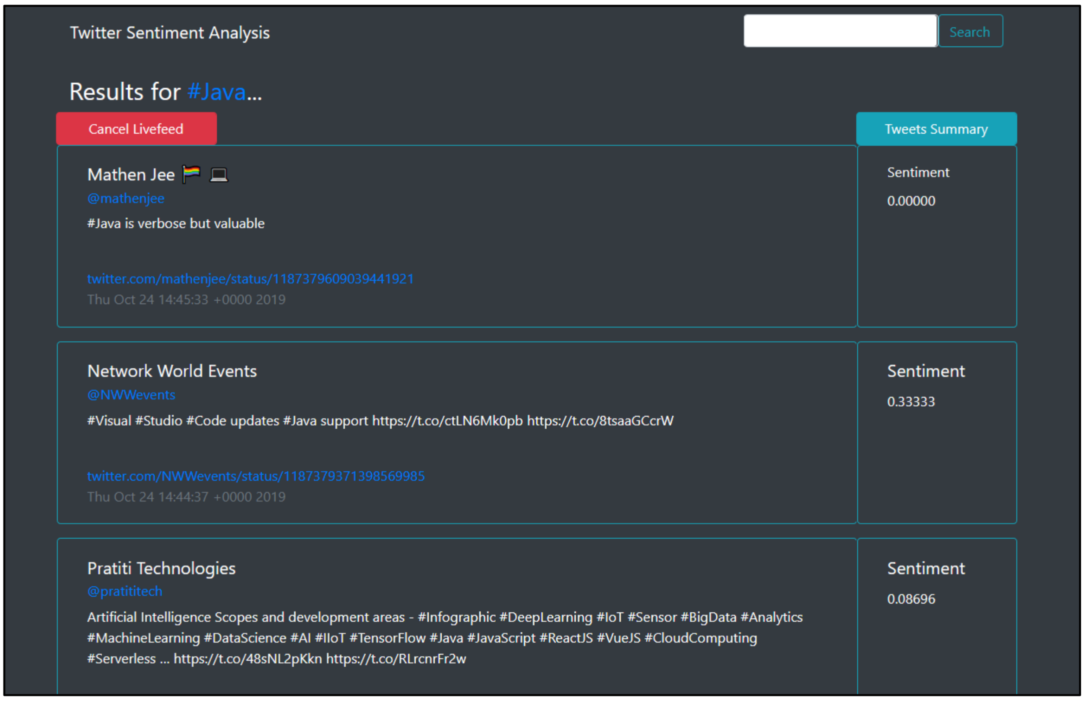
Enter Query

Query1 Query2

Invalid Query Submission! Please insert a '#' in front of a query and leave a ',' after each word.

Twitter Feed

After submitting a query, you are directed to a page full of tweets retrieved from Twitter. The left column represents each tweet, with each tweet separated in rows by table borders. The right column is for sentiment values to its assigned tweet, which determines how the tweet is expressed in a positive or negative manner through numbers.



Twitter Sentiment Analysis

Results for #Java...

Mathen Jee 🏳️‍🌈 💻
@mathenjee
#Java is verbose but valuable

twitter.com/mathenjee/status/1187379609039441921
Thu Oct 24 14:45:33 +0000 2019

Sentiment
0.00000

Network World Events
@NWWevents
#Visual #Studio #Code updates #Java support https://t.co/ctLN6Mk0pb https://t.co/8tsaaGCcrW

twitter.com/NWWevents/status/1187379371398569985
Thu Oct 24 14:44:37 +0000 2019

Sentiment
0.33333

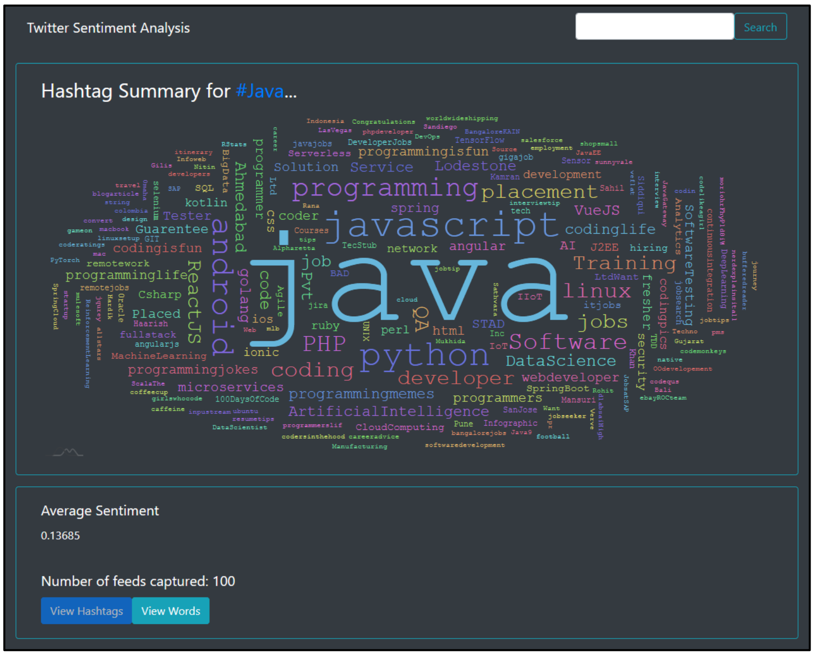
Pratiti Technologies
@pratititech
Artificial Intelligence Scopes and development areas - #Infographic #DeepLearning #IoT #Sensor #BigData #Analytics #MachineLearning #DataScience #AI #IIoT #TensorFlow #Java #JavaScript #ReactJS #VueJS #CloudComputing #Serverless ... https://t.co/48sNL2pKkn https://t.co/RLrcnrFr2w

Sentiment
0.08696

With the active filter enabled, new tweets will be added to the feed until it is cancelled by clicking the ‘Cancel Livefeed’ button. Clicking the ‘Tweets Summary’ button will collect all the tweets in the feed and create a tag cloud presentation.

Tag Cloud Statistics

A cloud-bubbled text is created by analysing the feed on the previous page. In this page, you can observe many relatable hashtags used in conjunction to the hashtag you’ve submitted and even click on any hashtag to lookup results on Twitter. Each word will come in different sizes based on how many times they are used in the feed. The average sentiment value is also available to see in the second row, with the addition of the number of feeds captured by the web application.



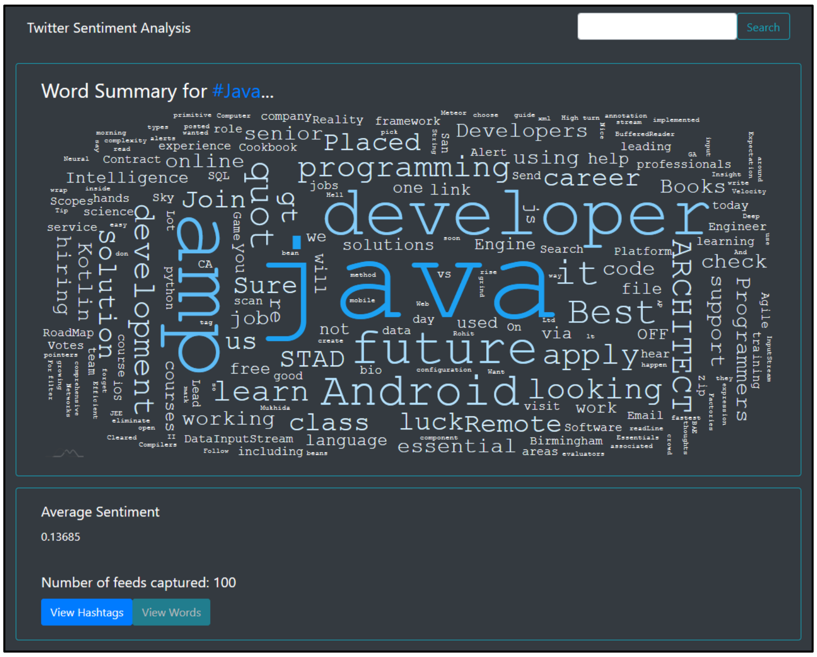
Hashtag Summary for #Java...

Average Sentiment

0.13685

Number of feeds captured: 100

By clicking on the ‘View Words’ button, you can see what kind of words are used in the tweet instead of hashtags. These keywords and phrases have a colour range from white to blue to show how frequent the words have been used.



Word Summary for #Java...

Average Sentiment

0.13685

Number of feeds captured: 100

Appendix B: Deployment Instructions (Local)

Prerequisite Programs

Using the web application requires Redis to cache data into memory on the computer. This can be installed using the command line on Ubuntu's terminal:

```
sudo apt install redis-server
```

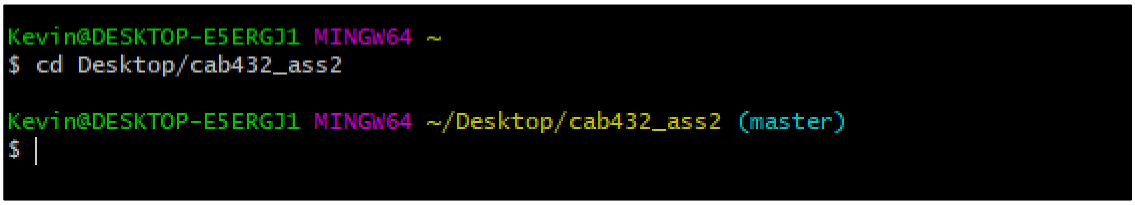
For windows users, it is possible to install Redis by enabling the Windows subsystem Linux on PowerShell. Be sure to run as an administrator to be able to run the script:

```
Enable-WindowsOptionalFeature -Online -FeatureName Microsoft-Windows-
Subsystem-Linux
```

Then, download and install Ubuntu from windows store. Run Ubuntu on your computer and install Redis by using the same command line from the Ubuntu version.

Installing Packages

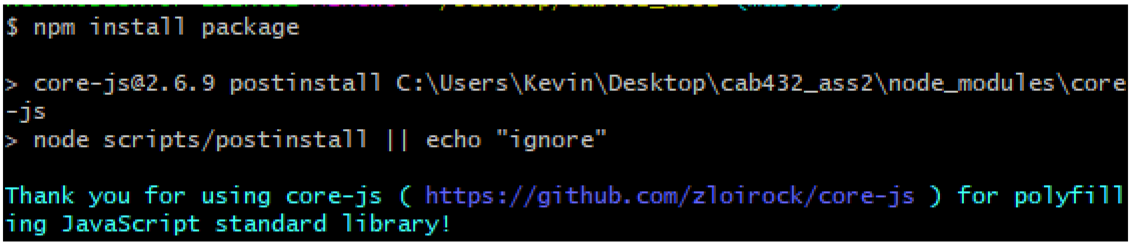
Locate the file directory of the web application using Git bash or an alternative. It is imperative that npm and node.js are installed on the system.



```
Kevin@DESKTOP-E5ERGJ1 MINGW64 ~
$ cd Desktop/cab432_ass2

Kevin@DESKTOP-E5ERGJ1 MINGW64 ~/Desktop/cab432_ass2 (master)
$ |
```

Install the packages in the directory by referencing the .json file called 'package'. This will install the requirements needed to run the server and acquire the features of the application.



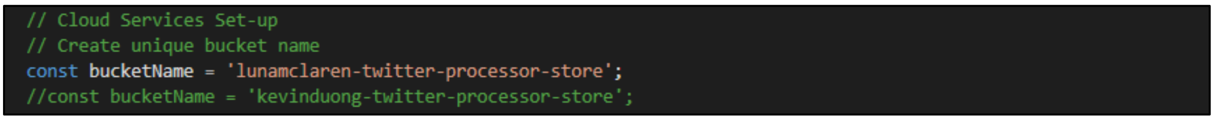
```
$ npm install package

> core-js@2.6.9 postinstall C:\Users\Kevin\Desktop\cab432_ass2\node_modules\core
-js
> node scripts/postinstall || echo "ignore"

Thank you for using core-js ( https://github.com/zloirock/core-js ) for polyfill
ing JavaScript standard library!
```

Running the Server

It is imperative to configure the S3 bucket name to your own and obtain your own token credentials from AWS. Changing the bucket store can be done so in routes/search.js.



```
// Cloud Services Set-up
// Create unique bucket name
const bucketName = 'lunamclaren-twitter-processor-store';
//const bucketName = 'kevinduong-twitter-processor-store';
```

Finally, to launch the application, simply write 'node start' on the bash terminal. You can now access the webpage using the following URL:

http://localhost:3000/

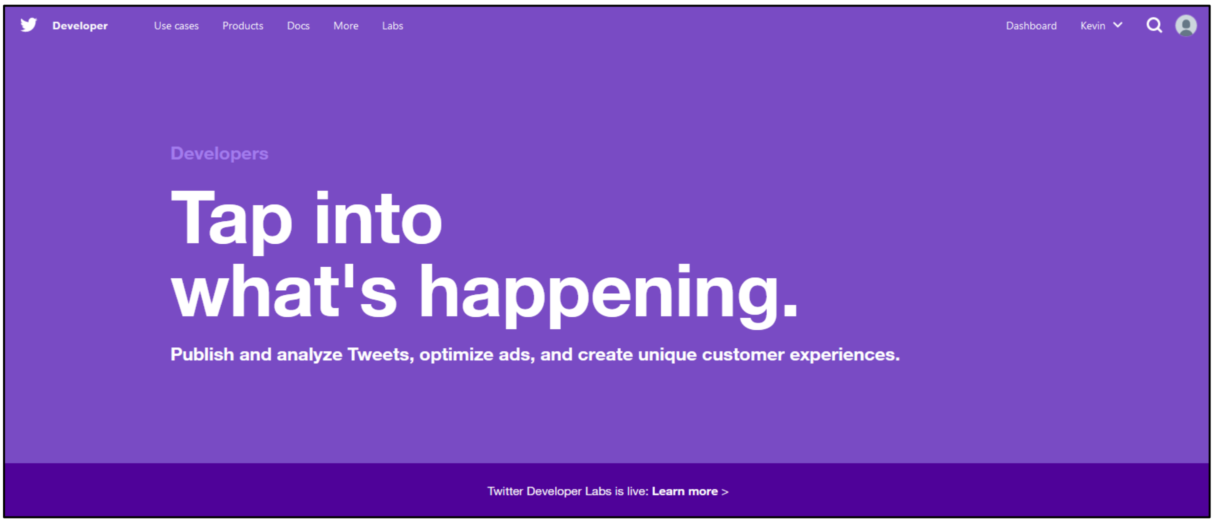
Appendix C: Twitter API Keys and Applications

Registering an Account

Twitter is a main component of building the web application – using their API allows us to retrieve their tweets and produce favourable content to the user. To apply for a key, you must first create a Twitter account and apply for a developer account at the Twitter Developer page:

Twitter Developer Registration Link
https://developer.twitter.com/en/apply-for-access

After following the procedures to register an account, you should have access to the developer page with your own account name labelled on top of the page.



Developers

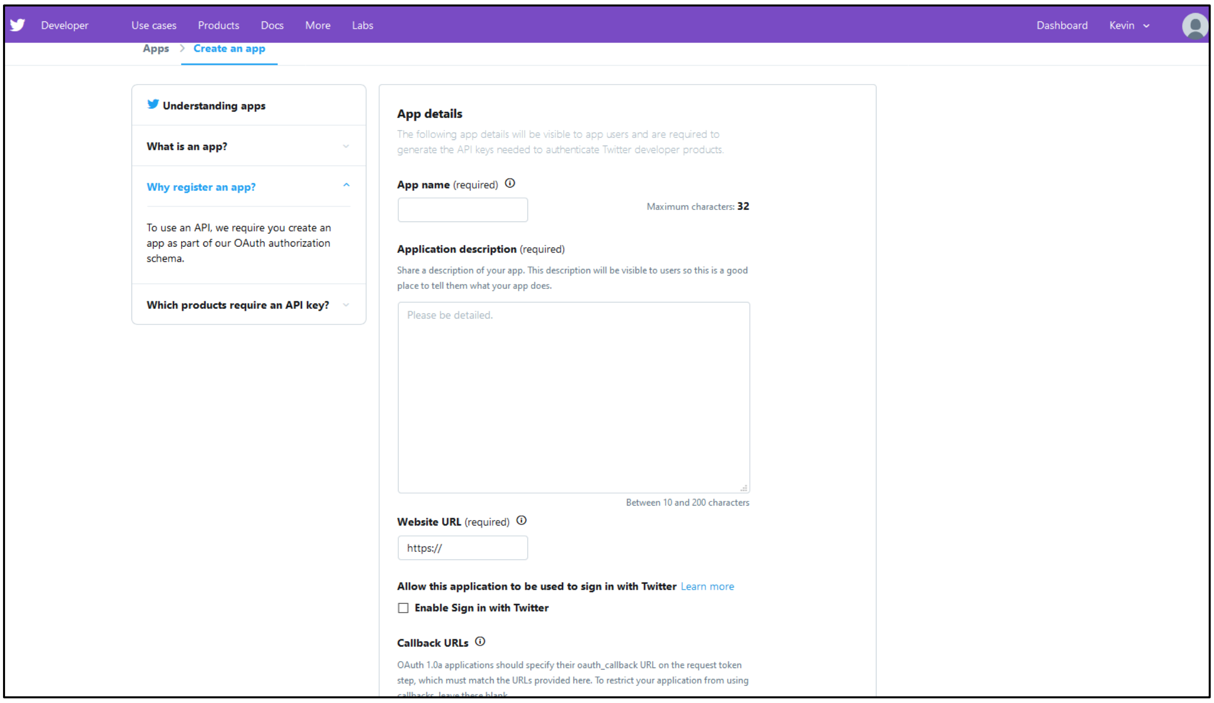
Tap into
what's happening.

Publish and analyze Tweets, optimize ads, and create unique customer experiences.

Twitter Developer Labs is live: **Learn more >**

Applying for Keys and Tokens

Click on your account name, then go to ‘Apps’. Then click on ‘Create an app’ to obtain a key. A form is required to complete as part of Twitter’s OAuth authorisation schema. When it comes to the website URL, you can use your Twitter account profile page if you don’t have a website.



Understanding apps

What is an app?

Why register an app?

To use an API, we require you create an app as part of our OAuth authorization schema.

Which products require an API key?

App details

The following app details will be visible to app users and are required to generate the API keys needed to authenticate Twitter developer products.

App name (required)

Maximum characters: **32**

Application description (required)

Share a description of your app. This description will be visible to users so this is a good place to tell them what your app does.

Please be detailed.

Between 10 and 200 characters

Website URL (required)

https://

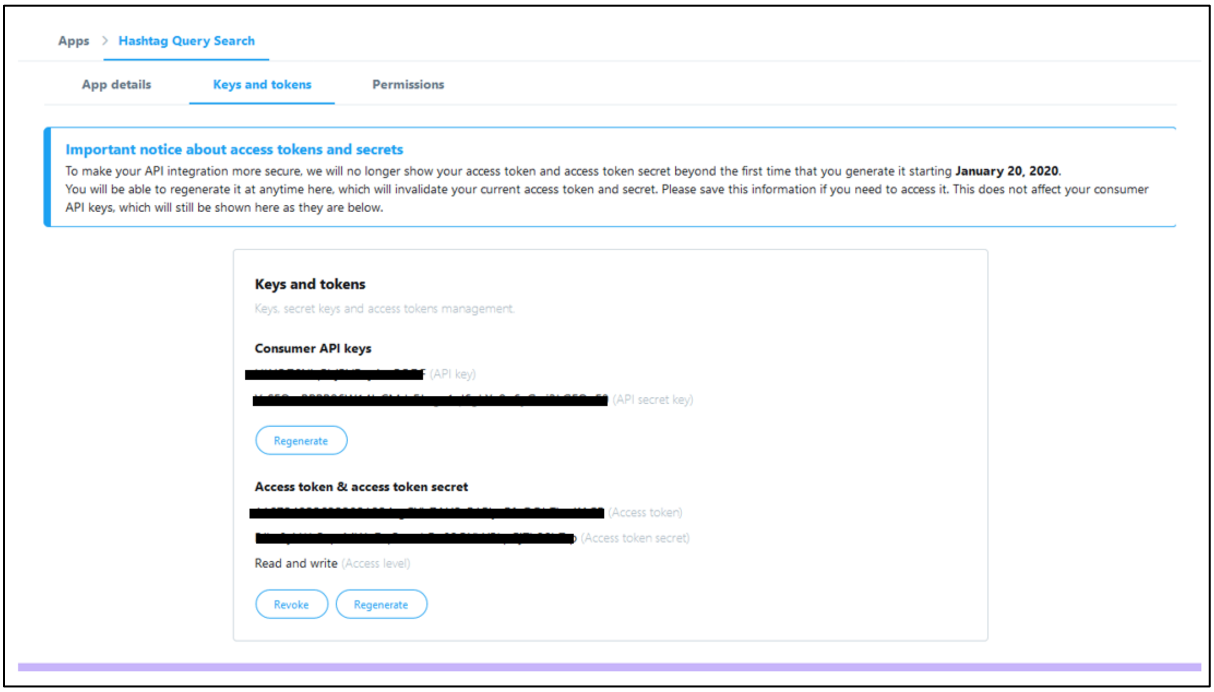
Allow this application to be used to sign in with Twitter Learn more

□ **Enable Sign in with Twitter**

Callback URLs

OAuth 1.0a applications should specify their oauth_callback URL on the request token step, which must match the URLs provided here. To restrict your application from using callbacks, leave these blank.

After completing the form request, Twitter should now grant you two API keys and tokens, both being standard and secret for each type. You can generate the keys anytime to create new ones.



Important notice about access tokens and secrets

To make your API integration more secure, we will no longer show your access token and access token secret beyond the first time that you generate it starting **January 20, 2020**. You will be able to regenerate it at anytime here, which will invalidate your current access token and secret. Please save this information if you need to access it. This does not affect your consumer API keys, which will still be shown here as they are below.

Keys and tokens

Keys, secret keys and access tokens management.

Consumer API keys

[redacted] (API key)

[redacted] (API secret key)

Access token & access token secret

[redacted] (Access token)

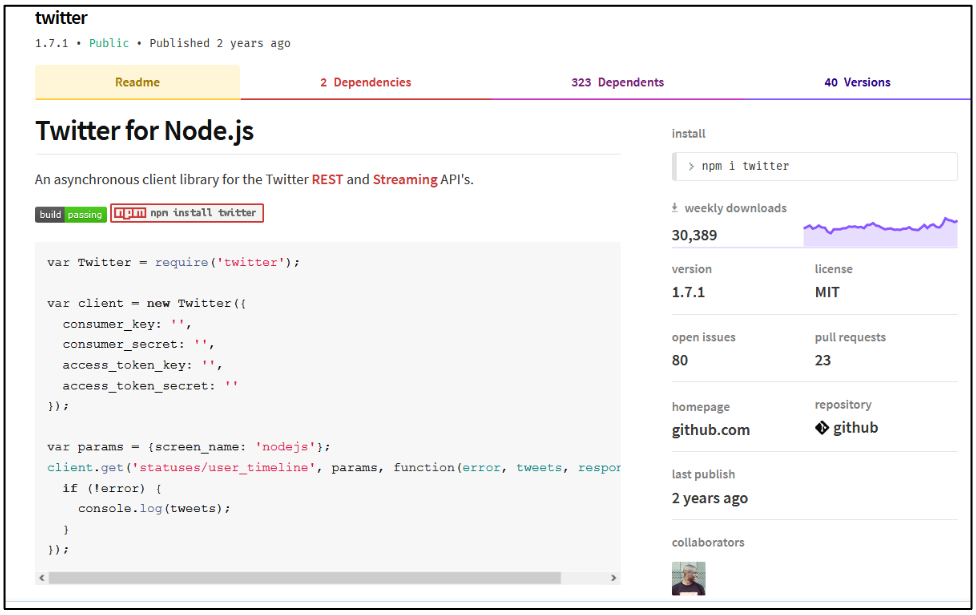
[redacted] (Access token secret)

Read and write (Access level)

Using the API

Npm has provided a module package for Twitter API users at their website, an asynchronous client library that is used in Node.js:

Twitter for Node.js
https://www.npmjs.com/package/twitter



twitter

1.7.1 • Public • Published 2 years ago

Twitter for Node.js

An asynchronous client library for the Twitter REST and Streaming API's.

```
var Twitter = require('twitter');

var client = new Twitter({
  consumer_key: '',
  consumer_secret: '',
  access_token_key: '',
  access_token_secret: ''
});

var params = {screen_name: 'nodejs'};
client.get('statuses/user_timeline', params, function(error, tweets, respor
  if (!error) {
    console.log(tweets);
  }
});
```

install

```
> npm i twitter
```

weekly downloads 30,389	
version 1.7.1	license MIT
open issues 80	pull requests 23
homepage github.com	repository github
last publish 2 years ago	
collaborators	

To learn more on how to use the Twitter API for various uses, check out the official documentation page:

API Reference Index – Twitter
https://developer.twitter.com/en/docs/api-reference-index