

DS 200 Lab2

Gather Real-Time Twitter Data Using Keywords and Hashtags

Aug 28, 2018

Instructor: John Yen

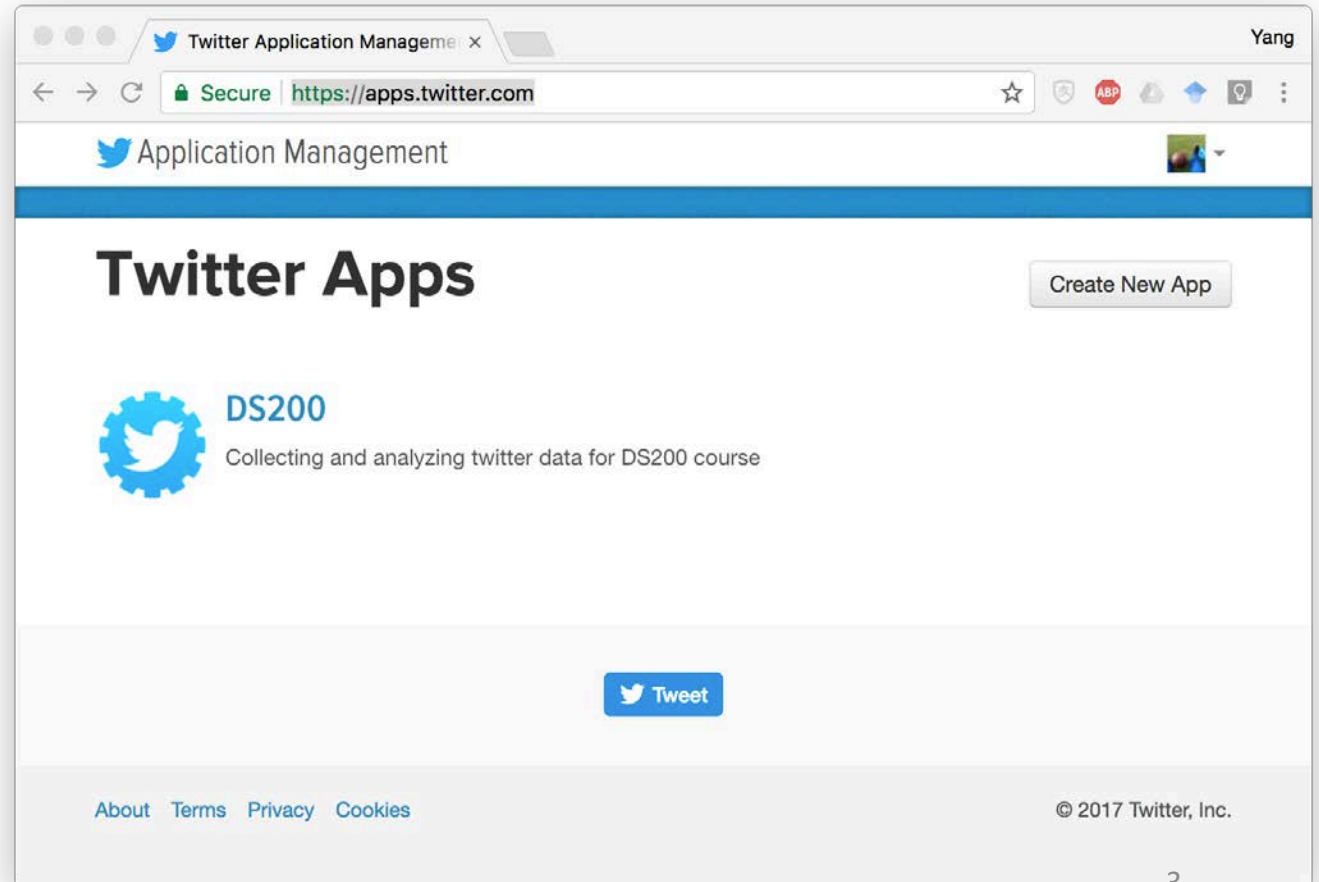
LA: Luewi Lei

Table of content

- Create a Twitter app (Slide 3 -- 9)
- Install tweepy package (Slide 10)
- Edit Python code for Gathering Tweets (Slide 11 -- 13)
- Gather tweets using the pipeline (Slide 14 -- 22)
- Load tweets in CSV file into Jupyter or Excel (Slide 23 -- 25)

Login In to Twitter Developer

- Go to <https://apps.twitter.com/>
- Login with your Twitter account



Apply for a Twitter developer account

 Application Management



Twitter Apps

As of July 2018, you must [apply for a Twitter developer account](#) and be approved before you may create new apps. Once approved, you will be able to create new apps from [developer.twitter.com](#).

For the near future, you can continue to manage existing apps here on [apps.twitter.com](#). However, we will soon retire this site and consolidate all developer tools, API access, and app management within the developer portal at [developer.twitter.com](#). You will be able to access and manage existing apps through that portal when we retire this site.

Click

Apply for a developer account

Select a Twitter user profile to the developer account

The screenshot shows the 'Apply for developer access' page on developer.twitter.com. The page has a purple header with navigation links: Developer, Use cases, Products, Docs, More, and an 'Apply' button with a user icon. The main content area is titled 'Apply for developer access' and includes a progress sidebar on the left, a status section, and a main form area.

STATUS: IN PROGRESS

- ☒ User profile
- ☐ Account details
- ☐ Use case details
- ☐ Terms of service
- ☐ Email verification


Interested in a developer account?

Some of our premium APIs are currently in Beta. By applying, you agree to receive emails from our team requesting feedback on your experience.

Select a user profile to associate

By default, this @username will be the admin of this developer account. If you are creating a developer account on behalf of your organization, you may wish to use your organization's @username as it is most likely to own the Apps you will use to access the API endpoints or warrant special permissions. You'll be able to invite teammates and re-assign roles later within your developer account settings.

Associate your current Twitter @username

 **John Yen**
@JohnCHYen

Continue

[Sign in as a different Twitter @username](#)
[Create new Twitter @username](#)

[Developer policy and terms](#) [Follow @twitterdev](#) [Subscribe to developer news](#)

1a. Use the Current Profile or
1b. Choose a Different Profile

2. Click "Continue"

Request access to a development account for personal use

- Choose a name for your development account (e.g., DS 200)
- Choose “United States” as the country of operation
- Click continue

The screenshot shows the Twitter Developer account application form. On the left, a sidebar lists the steps: STATUS: IN PROGRESS, User profile, Account details (highlighted with a green checkmark), Use case details, Terms of service, and Email verification. Below this is a section titled "Why the questions?" explaining the purpose of the form. The main content area is titled "Add your account details" and contains two radio button options for "Who are you requesting access for?". The second option, "I am requesting access for my own personal use", is selected. Below this is a section titled "Tell us about yourself" with a text input field for "Account name" (containing "DS200") and a dropdown menu for "Primary country of operation" (set to "United States"). A blue "Continue" button is at the bottom right. Blue arrows from the list on the left point to the "Account name" field, the "Primary country of operation" dropdown, and the "Continue" button.

STATUS: IN PROGRESS

- User profile
- Account details
- Use case details
- Terms of service
- Email verification

Why the questions?

We empower freedom of expression by providing a platform that protects the voices of our users — both on Twitter, and via our developer products. To help verify that all uses of Twitter data comply with our policies, we require additional information from developers signing up to use this service. Providing thorough answers will help us understand your use cases and will help expedite the evaluation of your application. Learn more about our restricted use cases.

Add your account details

Who are you requesting access for?

☐ I am requesting access for my organization

I plan to use Twitter's developer platform for projects owned by / in affiliation with a business, organization or institution. Ex: SaaS product, proof of concept, academic research, etc. To enable collaboration, this selection includes additional tools to support team development.

☒ I am requesting access for my own personal use

I plan to use Twitter's developer platform for projects unaffiliated with an existing business, organization or institution. Ex: Side project, hobby, etc. Personal use accounts do not include team development tools.

Tell us about yourself

Account name

e.g., username, project name, etc.

DS200

Primary country of operation

United States

Continue

Developer policy and terms Follow @twitterdev Subscribe to developer news

John Yen - Outlook Web ApDS 200, Section 001: Intro to Twitter DeveloperNew tab

developer.twitter.com/en/apply/usecase

DeveloperUse casesProductsDocsMoreApply

User profile

Account details

Use case details

Terms of service

Email verification

Why the questions?

We empower freedom of expression by providing a platform that protects the voices of our users — both on Twitter, and via our developer products. To help verify that all uses of Twitter data comply with our policies, we require additional information from developers signing up to use this service. Providing thorough answers will help us understand your use cases and will help expedite the evaluation of your application. [Learn more about our restricted use cases.](#)

What use case(s) are you interested in?

Select all that apply

☒ Academic research

☐ Advertising

☒ Audience analysis

☒ Chatbots and automation

☒ Consumer / end-user experience

☒ Engagement and customer service

☒ Publish and curate Tweets

☒ Student project / Learning to code

☒ Topic analysis

☒ Trend and event detection

☐ Other

Describe in your own words what you are building

In English, please describe your product - the more detailed the response, the easier it is to review and approve. Be sure to answer the following:

- What is the purpose of your product or service?
- What will you deliver to your users/customers?
- How do you intend to analyze Tweets, Twitter users, or their content?
- How is Twitter data displayed to users of your end product or service (e.g. will Tweets and content be displayed at row level or in aggregate)?

The purpose of this twitter student project is to build a predictive model for classifying tweets (gathered based on keywords and hashtags) into two or more pre-defined classes (e.g., sentiment analysis regarding a person, an institution; support or against an opinion/policy), applying the model to tweets gathered later to assess accuracy of the model, trend or changes of the sentiment/opinion.

Will your product, service, or analysis make Twitter content or derived information available to a government entity?

In general, schools, colleges, or universities do *not* fall under this category.

☒ No

☐ Yes

7

Read and agree to the Terms of Service

The screenshot shows a web browser window with the URL `developer.twitter.com/en/apply/terms`. The page has a purple header with navigation links: Developer, Use cases, Products, Docs, More, and an 'Apply' button with a user icon. On the left, a sidebar contains 'Terms of service' and 'Email verification'. The main content area is divided into sections: 'E. Twitter for Websites', 'F. Periscope Producer', and 'G. Definitions'. Each section contains a numbered list of terms and conditions. At the bottom, there are two checkboxes for agreement and a 'Submit application' button.

John Yen - Outlook Web Ap | DS 200, Section 001: Intro to | Apply — Twitter Develo | New tab

developer.twitter.com/en/apply/terms

Developer Use cases Products Docs More Apply

Terms of service

Email verification

E. Twitter for Websites

1. If you expect your [embedded Tweets](#) and [embedded timelines](#) to exceed 10 million daily impressions, you must contact us about your Twitter API access, as you may be subject to additional terms.
2. If you use [Twitter for Websites](#) widgets, you must ensure that an end user is provided with clear and comprehensive information about, and consents to, the storing and accessing of cookies or other information on the end user's device as described in Twitter's [cookie use](#) where providing such information and obtaining such consent is required by law.
3. If you use embedded Tweets or embedded timelines, you must provide users legally sufficient notice that fully discloses Twitter's collection and use of data about users' browsing activities on your website, including for interest-based advertising and personalization. You must also obtain legally sufficient consent from users for such collection and use, and provide legally sufficient instructions on how users can opt out of Twitter's interest-based advertising and personalization as described [here](#).
4. If you operate a Service targeted to children under 13, you must opt out of tailoring Twitter in any embedded Tweets or embedded timelines on your Service by setting the opt-out parameter to be true as described [here](#).

F. Periscope Producer

1. You must provide a reasonable user-agent, as described in the Periscope Producer technical documentation, for your Service when accessing the Periscope API.
2. If you expect the number of broadcasts created by your hardware will exceed (10 million) daily broadcasts, you must [contact us](#) about your Twitter API access, as you may be subject to additional terms.
3. You must honor user requests to log out of their Periscope account on your Service.
4. You may not provide tools in your service to allow users to circumvent technological protection measures.

G. Definitions

1. **Twitter Content** - Tweets, Tweet IDs, Direct Messages, Direct Message IDs, Twitter end user profile information, User IDs, Periscope Broadcasts, Periscope Broadcast IDs and any other data and information made available to you through the Twitter API or by any other means authorized by Twitter, and any copies and derivative works thereof.
2. **Developer Site** - Twitter's developer site located at <https://developer.twitter.com>.
3. **Periscope Broadcast** - A user generated live video stream that is available live or on-demand, that is publicly displayed on Twitter Services.
4. **Broadcast ID** - A unique identification number generated for each Periscope Broadcast.
5. **Tweet** - A short-form text and/or multimedia-based posting made on Twitter Services.
6. **Tweet ID** - A unique identification number generated for each Tweet.
7. **Direct Message** - A text and/or multimedia-based posting that is privately sent on the Twitter Service by one end user to another.

☒ By clicking on the box, You indicate that you have read and agree to this Developer Agreement and the Twitter Developer Policy, additionally as it relates to your display of any of the Content, the [Display Requirements](#); as it relates to your use and display of the Twitter Marks, the [Twitter Brand Assets and Guidelines](#); and as it relates to taking automated actions on your account, the [Automation Rules](#). These documents are available in hardcopy upon request to Twitter.

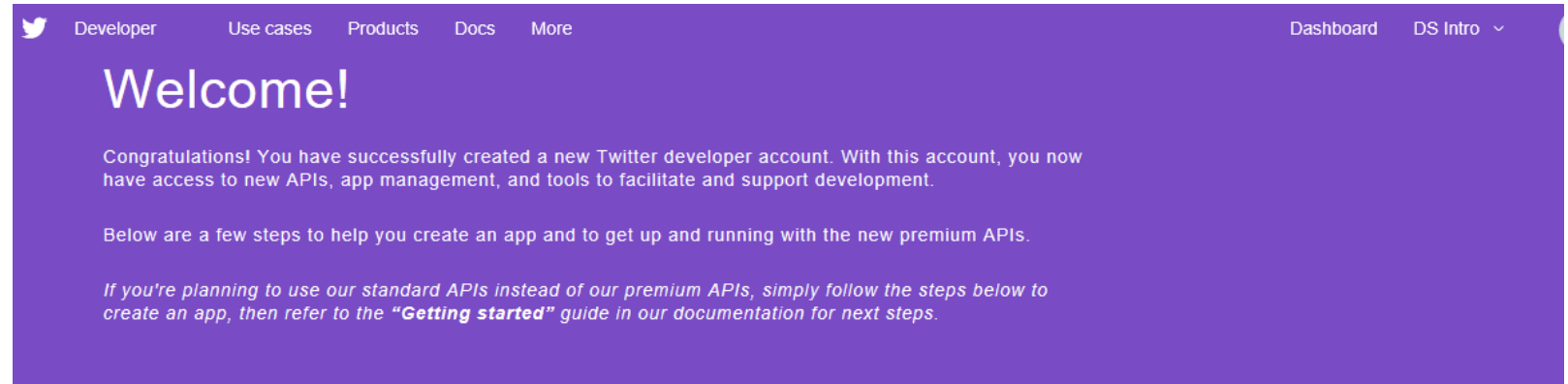
☒ Subscribe to our email list for product updates, developer news, and marketing communications.

Submit application

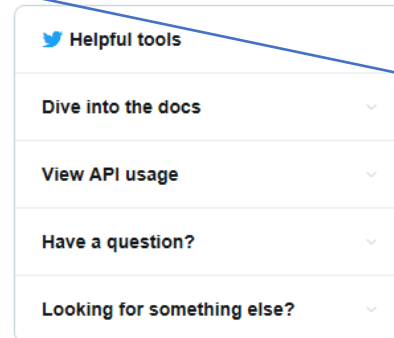
You will receive a confirmation email on your e-mail address associated with the account.

After you click the confirmation link on your email, you can create a Twitter App

Click to
Create a
Twitter App



The screenshot shows the Twitter Developer 'Welcome!' page. The header includes the Twitter logo and navigation links: Developer, Use cases, Products, Docs, More, Dashboard, and DS Intro. The main content area has a purple background with white text. It says 'Welcome!' and 'Congratulations! You have successfully created a new Twitter developer account. With this account, you now have access to new APIs, app management, and tools to facilitate and support development.' Below this, it lists steps to help create an app and get up and running with premium APIs. It also mentions that for standard APIs, users should follow the 'Getting started' guide.



The screenshot shows a sidebar titled 'Helpful tools' with a Twitter logo. It contains four expandable sections: 'Dive into the docs', 'View API usage', 'Have a question?', and 'Looking for something else?'. Each section has a downward arrow icon.

Get started



Create an app

To use an API, we require you create an app as part of our OAuth authorization scheme. Visit the [Apps](#) page of this developer portal to create one. Then, return to this page to complete the next step.



Set up a dev environment

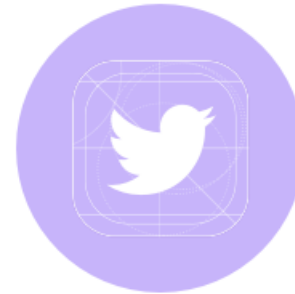
To begin using the new Premium APIs, you need to [set up one or more dev environments](#) for the endpoint — and connect it to an app. Dev environments can be used to isolate usage, rules, rate limits, and more. If you are planning to use our standard APIs, you can skip this step.

Obtain An API key to create a Twitter App



Apps

Click



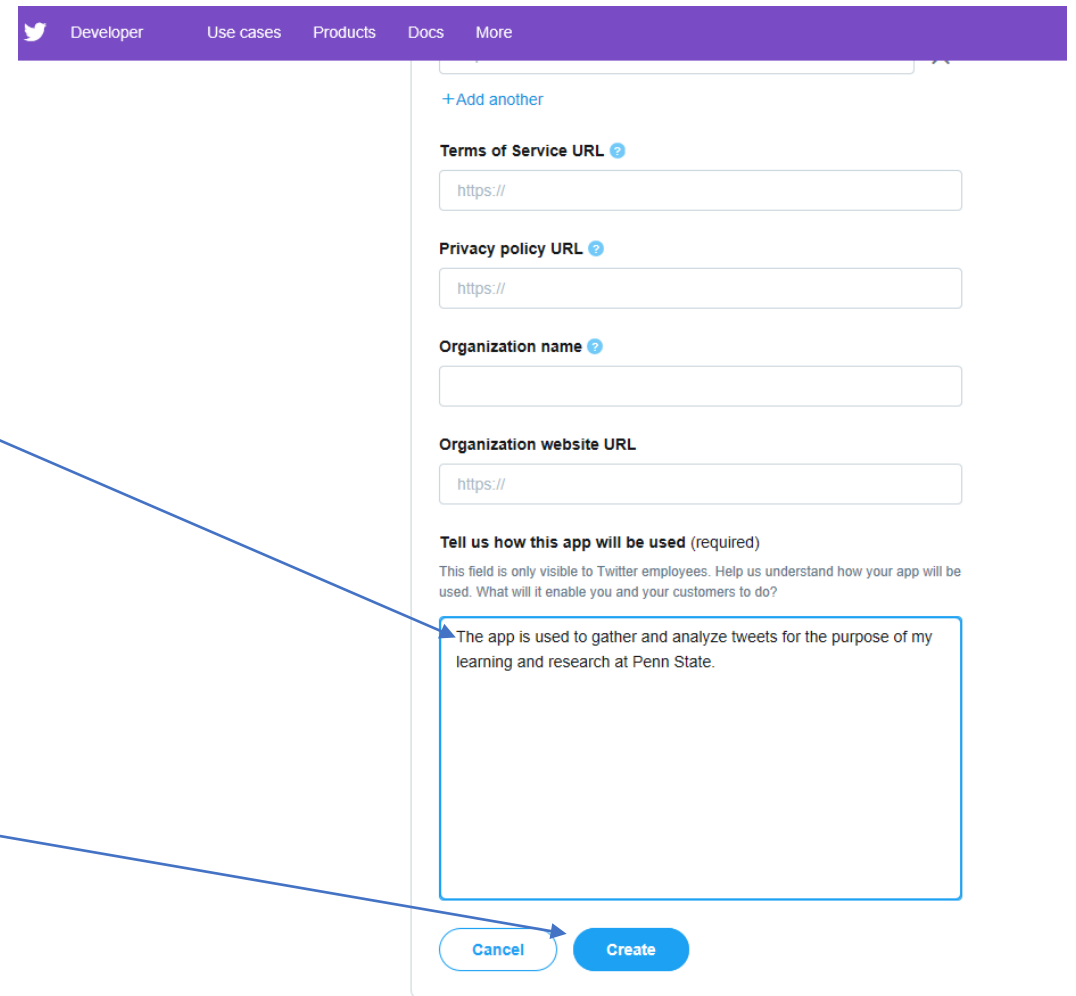
Create your first app

You'll need an app and API key in order to authenticate and integrate with most Twitter developer products. Create an app to get your API key.

Create an app

Create a Twitter app

- Fill in all the necessary information
- “App name”, “Application description”, “Website URL”, and “Tell us how this App will be used” are required
- “Website” could be your Facebook homepage, your personal home page, etc.
- Scroll down and click on “Create”



The screenshot shows the Twitter Developer portal's 'Create a new app' form. The form is titled '+ Add another' and includes several input fields: 'Terms of Service URL' (with a placeholder 'https://'), 'Privacy policy URL' (with a placeholder 'https://'), 'Organization name', and 'Organization website URL' (with a placeholder 'https://'). Below these is a section titled 'Tell us how this app will be used (required)' with a subtext: 'This field is only visible to Twitter employees. Help us understand how your app will be used. What will it enable you and your customers to do?'. A text area contains the response: 'The app is used to gather and analyze tweets for the purpose of my learning and research at Penn State.' At the bottom of the form are two buttons: 'Cancel' and 'Create'. Blue arrows from the text on the left point to the 'Tell us how this app will be used' section and the 'Create' button.

Review Developer Terms

1. Read a summary of the Developer Terms

2. Click “Create”

Privacy policy URL

Review our Developer Terms

As a reminder, you have agreed to our [Developer Agreement and Policy](#). Please be mindful of the following [restricted use cases](#)

- ⊗ **Sensitive Information**
Be careful about using Twitter data to derive or infer potentially sensitive characteristics about Twitter users (ie. health, political or religious affiliation, ethnic origin, sexual orientation, and more).
- ⊗ **Government use and surveillance**
We prohibit the use of Twitter Data and Twitter APIs by any entity for surveillance purposes. Period.
- ⊗ **Automation**
If your app is will be used to post Tweets, follow accounts, or send Direct Messages, you should carefully review the [Automation Rules](#) to ensure you comply with our guidelines and never perform bulk, aggressive, or spammy actions.

[Cancel](#) [Create](#)

Create Keys and Tokens for Tweets Gathering

You will see a web page for your twitter App, showing the App name, description, and URL you provided in the previous step.

Click
"Keys and tokens"

The screenshot shows the Twitter Developer Portal interface. At the top is a purple navigation bar with links: Developer, Use cases, Products, Docs, More, Dashboard, and DS Intro. Below this is a breadcrumb trail: Apps / DS 200 Opinion Classification. A horizontal tab bar contains three tabs: App details (selected), Keys and tokens, and Permissions. The main content area is titled 'App details' with a sub-header 'Details and URLs' and an 'Edit' button. It lists the following information: App icon (Twitter logo), App Name (DS 200 Opinion Classification), Description (Gather tweets related to a topic based on a set of keywords and hashtags. Gather tweets are tagged to create a predictive model regarding whether a tweet is supportive of an opinion or not.), Website URL (https://faculty.ist.psu.edu/yen/), Sign in with Twitter (Disabled), Callback URL (None), Terms of service URL (None), and Privacy policy URL (None). A blue arrow points from the 'Click "Keys and tokens"' text to the 'Keys and tokens' tab.

You will see your Consumer API keys

Developer Use cases Products Docs More

Apps / DS 200 Opinion Classification

App details **Keys and tokens** Permissions

Keys and tokens
Keys, secret keys and access tokens management.

Consumer API keys

[Redacted Consumer API keys]

Regenerate

Access token & access token secret
None

Create

API key
API secret key

Copy these keys for
later use.
Don't share with anyone.

Click "Create" to create
Access token and
Access token secret

Access token & Access token secret

Developer Use cases Products Docs More Dashboard DS Intro

Apps / DS 200 Opinion Classification

App details **Keys and tokens** Permissions

Keys and tokens
Keys, secret keys and access tokens management.

Consumer API keys

[Redacted Consumer API key]

Regenerate

Access token & access token secret

[Redacted Access token & access token secret]

Read and write (Access level)

Revoke Regenerate

You should see your Access
token and
Access token secret
Generated.
Copy them for later
use

Obtain Twitter Keys for Tweets Gathering -- Done

- Now you have four long keys for the following:
 - ***API key***
 - ***API secret key***
 - ***Access token***
 - ***Access token secret***
- These are what you need for using Twitter API (Application Program Interface) to gather tweets

Install *tweepy* package

- Open Command Line (Windows) or Terminal (Mac/Linux)
- Type: **pip install tweepy**
- Use pip3 in case you also have python2 in your system
- For Windows, go to next page
- For Mac, skip next page, go to slide 20.

Copy the Python code for Gathering Tweets (Windows)

- Download the *gather_tweets_pipeline.py* file from Canvas.
- Open **Command Line** in Windows
- Create a directory for the lab using “mkdir” command
 - E.g., “mkdir C:\Users\YourName\MyDocument\DS200lab3” for Windows
- Enter the directory where you download *get_tweets_pipeline.py*
 - E.g., “cd C:\Users\YourName\Downloads” for Windows
- Copy the file to your lab3 directory
 - E.g., “copy get_tweets_pipeline.py C:\Users\YourName\MyDocument\DS200lab3” for Windows

Copy the Python code for Gathering Tweets (Mac/Linux)

- Download the *gather_tweets_pipeline.py* file from Canvas.
- Open **Terminal** in Mac/Linux
- Create a directory for the lab using “mkdir” command
 - E.g., “mkdir ~/DS200lab3” for Mac/Unix (“~” refers to your home directory)
- Enter the directory where you download *get_tweets_pipeline.py*
 - E.g., “cd ~/Downloads” for Mac/Linux
- Copy the file to your lab3 directory
 - E.g., “cp get_tweets_pipeline.py ~/DS200lab3” for Mac/Linux

Download one of the text editors

We need a text editor suitable for editing code.

Text editor possibilities:

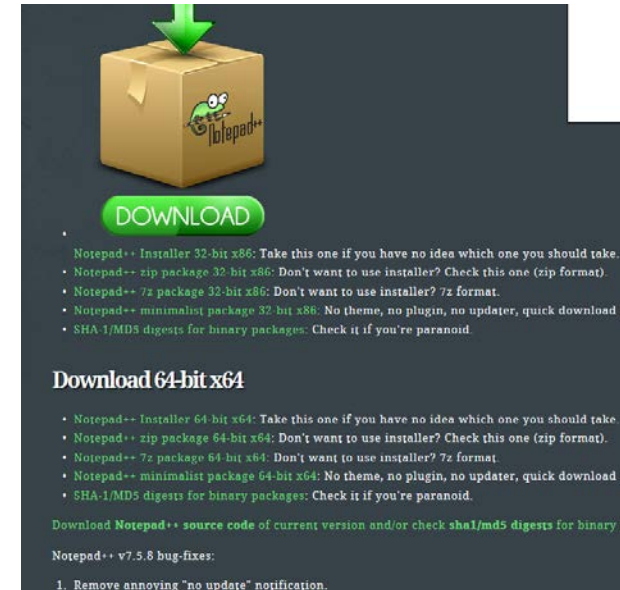
- Windows: Notepad++, sublime text, Atom
- Mac/Linux: Atom, sbulime text, gedit, vim, etc.

Download Notepad++:

<https://notepad-plus-plus.org/download/v7.5.8.html>

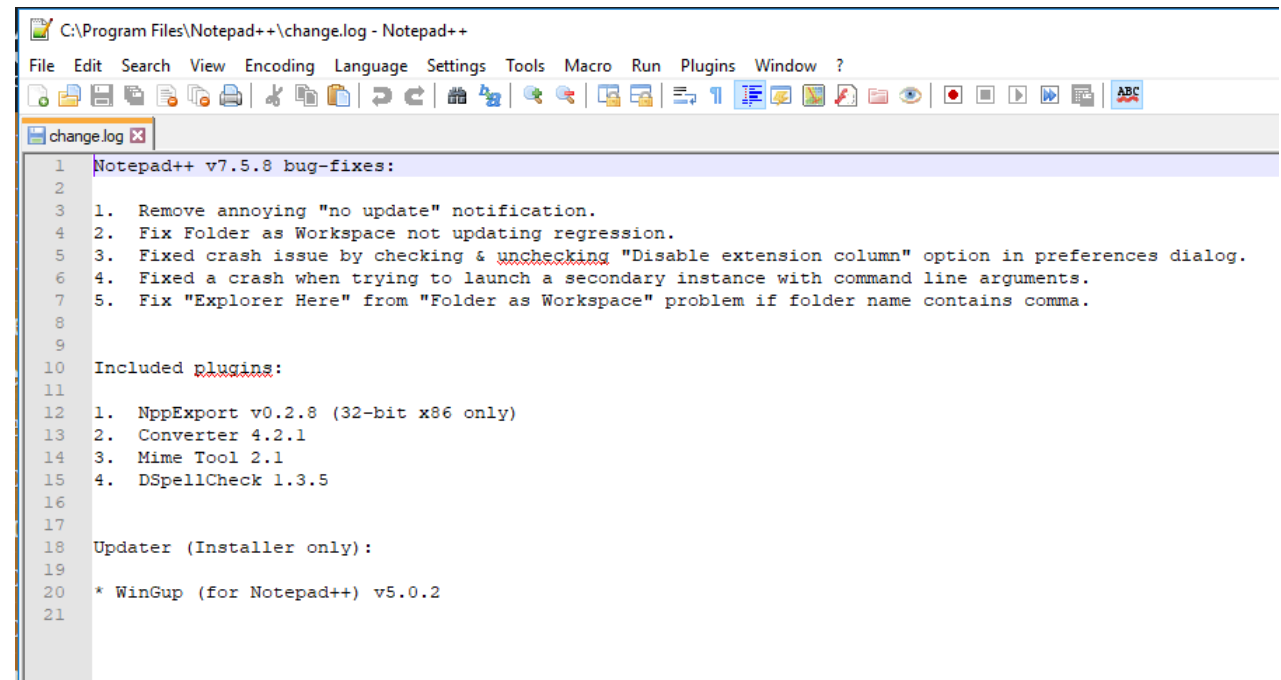
Download Atom:

<http://atom.io>



Edit the Python code using Text Editor

- Open *get_tweets_pipeline.py* file using one of the **text editors** (NOT WORD)
- For Notepad++, select “File”, then “Open”, find the *get_tweets_pipeline.py* file



The screenshot shows the Notepad++ application window with the title bar 'C:\Program Files\Notepad++\change.log - Notepad++'. The menu bar includes File, Edit, Search, View, Encoding, Language, Settings, Tools, Macro, Run, Plugins, Window, and ?. The toolbar contains various icons for file operations and editing. The active window is 'change.log' and the text content is as follows:

```
1 Notepad++ v7.5.8 bug-fixes:
2
3 1. Remove annoying "no update" notification.
4 2. Fix Folder as Workspace not updating regression.
5 3. Fixed crash issue by checking & unchecking "Disable extension column" option in preferences dialog.
6 4. Fixed a crash when trying to launch a secondary instance with command line arguments.
7 5. Fix "Explorer Here" from "Folder as Workspace" problem if folder name contains comma.
8
9
10 Included plugins:
11
12 1. NppExport v0.2.8 (32-bit x86 only)
13 2. Converter 4.2.1
14 3. Mime Tool 2.1
15 4. DSpellCheck 1.3.5
16
17
18 Updater (Installer only):
19
20 * WinGup (for Notepad++) v5.0.2
21
```

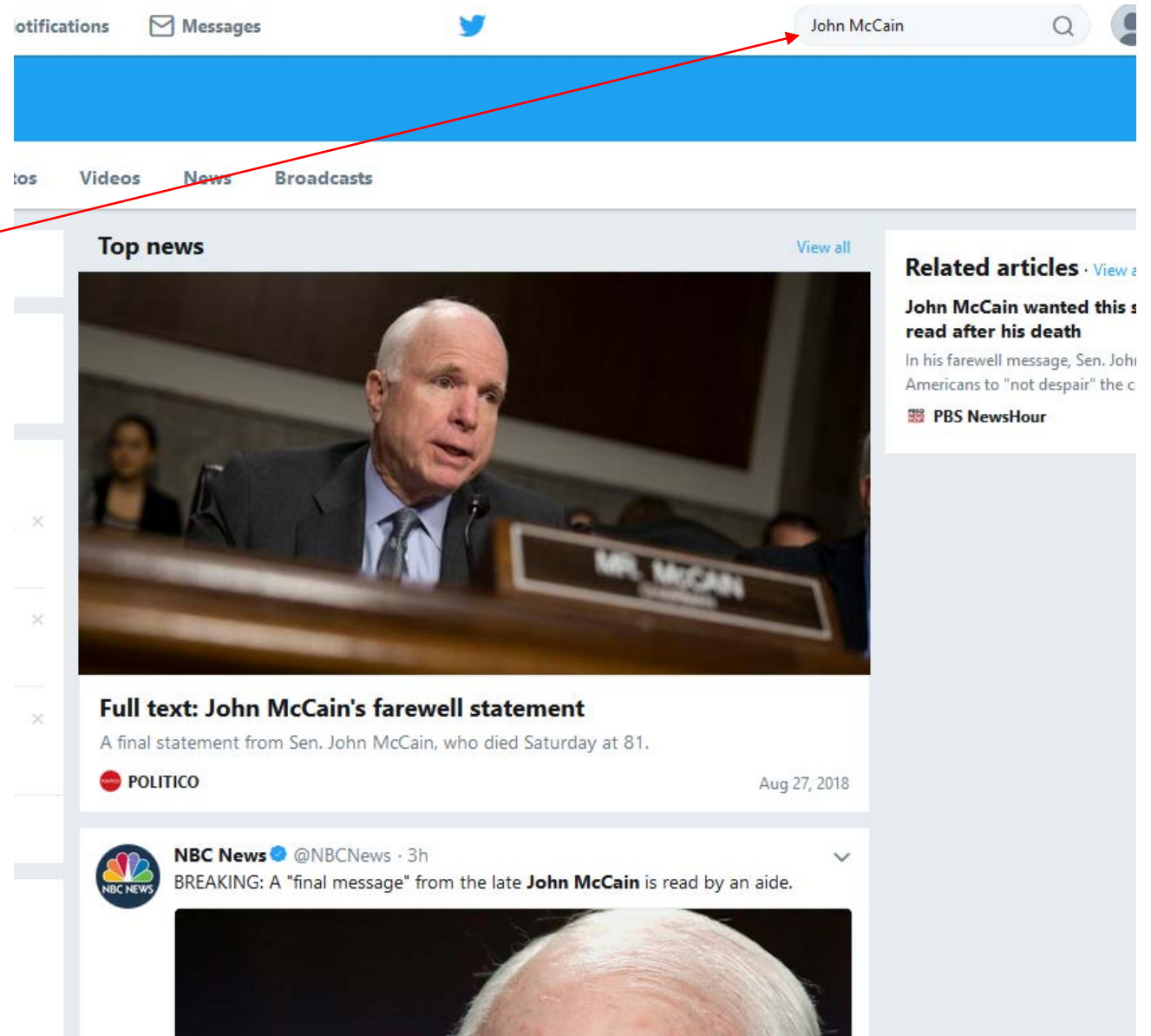
Add API keys to gater_tweets_pipeline.py

- Scroll down to the middle of the file, you will find four assignments for the API keys we gathered earlier.
- Enter API key, API secret key, Access token, Access token secret between the two quotes
- They are strings assigned to the corresponding variables.
- **Save the file**
get_tweets_pipeline.py
(If you don't save, you will lose the changes you made above)

```
98         print(status)
99         return True
100
101
102     # Get the str representation of the current date and time
103     def current_datetime_str():
104         return f'{datetime.datetime.now():%Y-%m-%d_%H-%M-%S}'
105
106
107     # Main
108     def main():
109         # Key and token info needed
110         consumer_key = ''
111         consumer_secret = ''
112         access_token = ''
113         access_secret = ''
114
115         auth = OAuthHandler(consumer_key, consumer_secret)
116         auth.set_access_token(access_token, access_secret)
117         api = tweepy.API(auth)
118
119         # Welcome
120         print('=====')
121         print('Welcome to the user interface of gathering tweets pipeline!')
122         print('You can press "Ctrl+C" at anytime to abort the program.')
123         print('=====')
124         print()
```

How to choose keywords and hashtags?

1. Log in to Twitter.com, use the search box on the upper right.



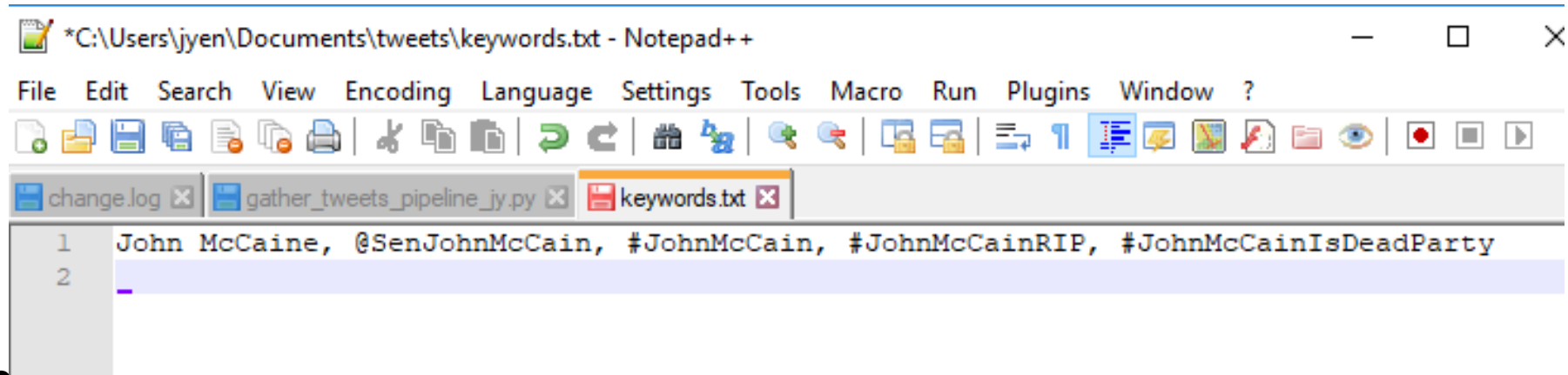
Scroll down to find hashtags, twitterIDs that occur frequently

- Example: @SenJohnMcCain showed up in a few tweets



Create a new file of keywords using a text editor

- Create this file using a Text Editor (File -> New)
- How a key word file (using Notepad++) looks like



```
*C:\Users\jyen\Documents\tweets\keywords.txt - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
change.log x gather_tweets_pipeline_jy.py x keywords.txt x
1 John McCain, @SenJohnMcCain, #JohnMcCain, #JohnMcCainRIP, #JohnMcCainIsDeadParty
2
```

- Hashtags need to be preceded with #
- TweeterIDs need to be preceded with @
- Multiple key words and hashtags are allowed
- Multiple lines are allowed
- Save this keyword file as “keywords.txt” in the same directory as get_tweets_pipeline.py

Start the pipeline for gathering tweets

- Open **Command Line** in Windows, or **Terminal** in Mac/Linux
- Enter the directory (e.g., DS200lab) to which you copied the downloaded files *get_tweets_pipeline.py*
 - E.g., “cd C:\Users\MyDocument\DS200lab” for Windows
 - E.g., “cd ~/DS200lab” for Mac/Linux
- Invoke `get_tweets_pipeline.py` to start the “pipeline” for gathering tweets

`python gather_tweets_pipeline.py` (For Windows)


`python3 gather_tweets_pipeline.py` (For MacOS/Linux)

Step 1

- Choose the way to specify your key words

```
C:\Users\jyen\Documents\tweets>python gather_tweets_pipeline.py
=====
Welcome to the user interface of gathering tweets pipeline!
You can press "Ctrl+C" at anytime to abort the program.
=====

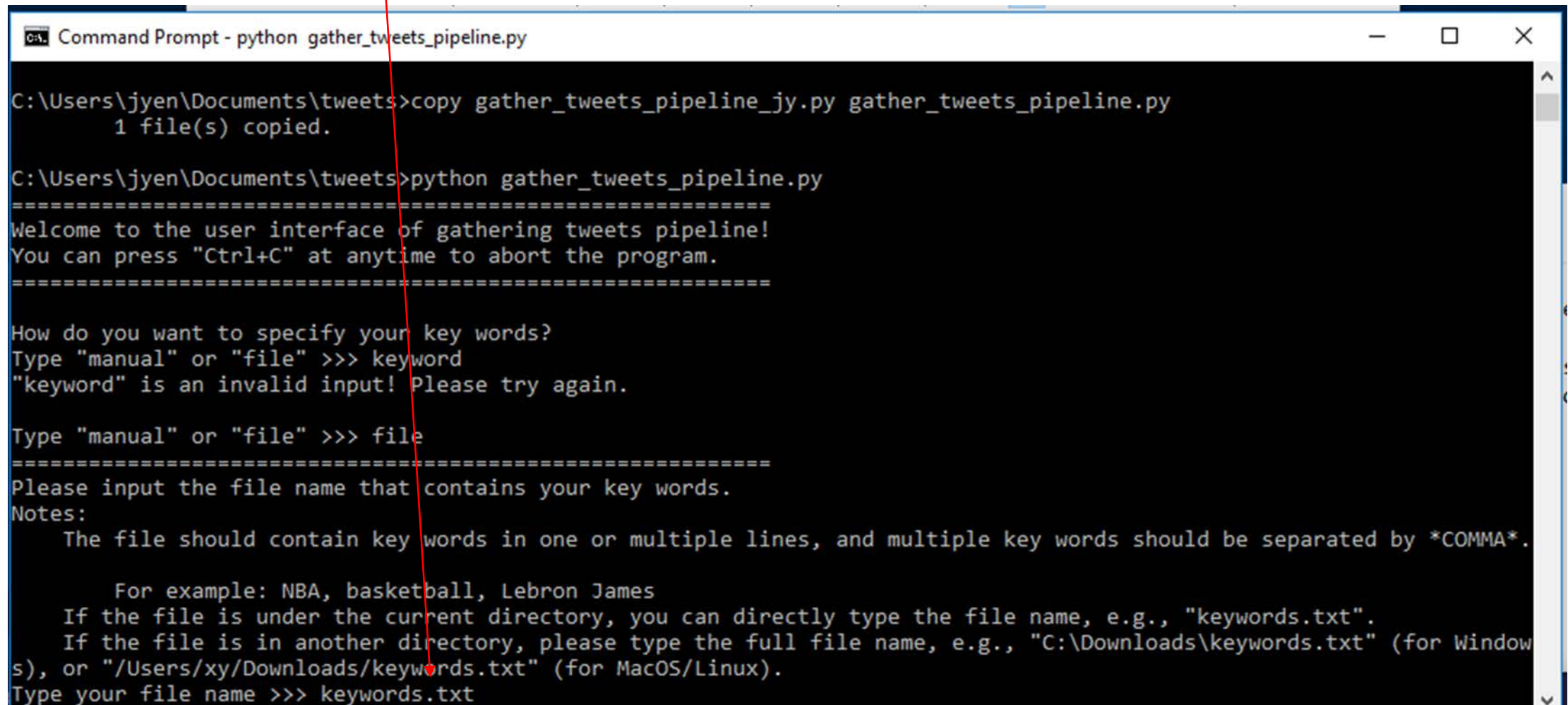
How do you want to specify your key words?
Type "manual" or "file" >>> file
```



- Recommend to enter keywords through a file
- Benefit of using a file: Reuse the keywords and hashtags each time you gather tweets facilitate comparing tweets gathered in different time.

Step 1.1

- Type the file name to the prompt “Type your file name >>>”



```
Command Prompt - python gather_tweets_pipeline.py

C:\Users\jyen\Documents\tweets>copy gather_tweets_pipeline_jy.py gather_tweets_pipeline.py
1 file(s) copied.

C:\Users\jyen\Documents\tweets>python gather_tweets_pipeline.py
=====
Welcome to the user interface of gathering tweets pipeline!
You can press "Ctrl+C" at anytime to abort the program.
=====

How do you want to specify your key words?
Type "manual" or "file" >>> keyword
"keyword" is an invalid input! Please try again.

Type "manual" or "file" >>> file
=====
Please input the file name that contains your key words.
Notes:
    The file should contain key words in one or multiple lines, and multiple key words should be separated by *COMMA*.

    For example: NBA, basketball, LeBron James
    If the file is under the current directory, you can directly type the file name, e.g., "keywords.txt".
    If the file is in another directory, please type the full file name, e.g., "C:\Downloads\keywords.txt" (for Windows), or "/Users/xy/Downloads/keywords.txt" (for MacOS/Linux).
Type your file name >>> keywords.txt
```

Step 1.2

- Alternatively, you can input key words manually

```
~/GitHub/tweepy-exp(master) » python3 gather_tweets_pipeline.py
=====
Welcome to the user interface of gathering tweets pipeline!
You can press "Ctrl+C" at anytime to abort the program.
=====

How do you want to specify your key words?
Type "manual" or "file" >>> manual
=====

Please input your key words (separated by comma), and hit <ENTER> when done.
Type the key words >>> 
```

- Type in your key words and/or hashtags, separated by comma

Step 1.2

- Manual input of key words and hashtags

```
~/GitHub/tweepy-exp(master) » python3 gather_tweets_pipeline.py
=====
Welcome to the user interface of gathering tweets pipeline!
You can press "Ctrl+C" at anytime to abort the program.
=====

How do you want to specify your key words?
Type "manual" or "file" >>> manual
=====

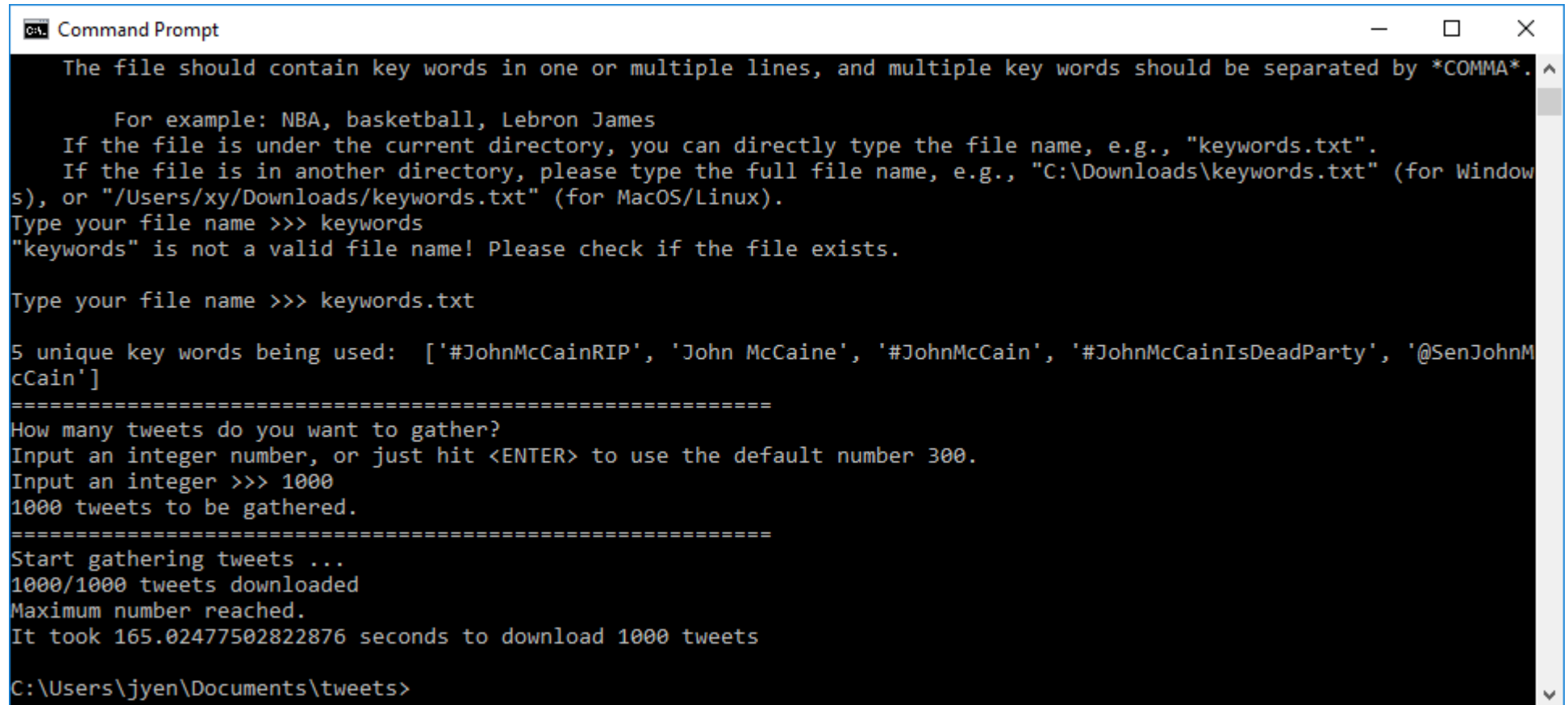
Please input your key words (separated by comma), and hit <ENTER> when done.
Type the key words >>> LeBron James, NBA, basketball

3 unique key words being used: ['basketball', 'NBA', 'Lebron James']
=====

How many tweets do you want to gather?
Input an integer number, or just hit <ENTER> to use the default number 300.
Input an integer >>> 
```

Step 2

- Decide the number of tweets you want to gather this time (recommend: 1000)



```
CA: Command Prompt
The file should contain key words in one or multiple lines, and multiple key words should be separated by *COMMA*.
For example: NBA, basketball, LeBron James
If the file is under the current directory, you can directly type the file name, e.g., "keywords.txt".
If the file is in another directory, please type the full file name, e.g., "C:\Downloads\keywords.txt" (for Windows), or "/Users/xy/Downloads/keywords.txt" (for MacOS/Linux).
Type your file name >>> keywords
"keywords" is not a valid file name! Please check if the file exists.

Type your file name >>> keywords.txt

5 unique key words being used: ['#JohnMcCainRIP', 'John McCain', '#JohnMcCain', '#JohnMcCainIsDeadParty', '@SenJohnMcCain']
=====
How many tweets do you want to gather?
Input an integer number, or just hit <ENTER> to use the default number 300.
Input an integer >>> 1000
1000 tweets to be gathered.
=====
Start gathering tweets ...
1000/1000 tweets downloaded
Maximum number reached.
It took 165.02477502822876 seconds to download 1000 tweets

C:\Users\jyen\Documents\tweets>
```

- You can repeat step 1 and 2 to gather additional tweets.

Step 3

- Wait until all tweets are gathered
- If you do not see the command prompt for a while (which means the number of tweets gathered has not reached the maximum you specified), you can choose to press “Ctrl+C” to interrupt. You will still be able to see the tweets gathered so far.
- Three files will be generated in the same directory for each run
 - raw_YYYY-MM-DD-HH-MM-SS.json (raw tweets from Twitter, JSON format)
 - data_YYYY-MM-DD-HH-MM-SS.csv (selected fields from raw tweets, CSV format)
 - text_YYYY-MM-DD-HH-MM-SS.csv (filtered tweets, where special characters in the body of tweets are removed)
 - YYYY-MM-DD is the date you gathered tweets
 - HH-MM-SS is the time your tweets files were generated

Load the CSV twitter file into Jupyter Notebook

Jupyter Untitled Last Checkpoint: 8 minutes ago (autosaved) Python 3 Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

In [1]: `from datascience import *`

In [2]: `d1 = Table.read_table('data_2018-08-27_17-18-26.csv')`

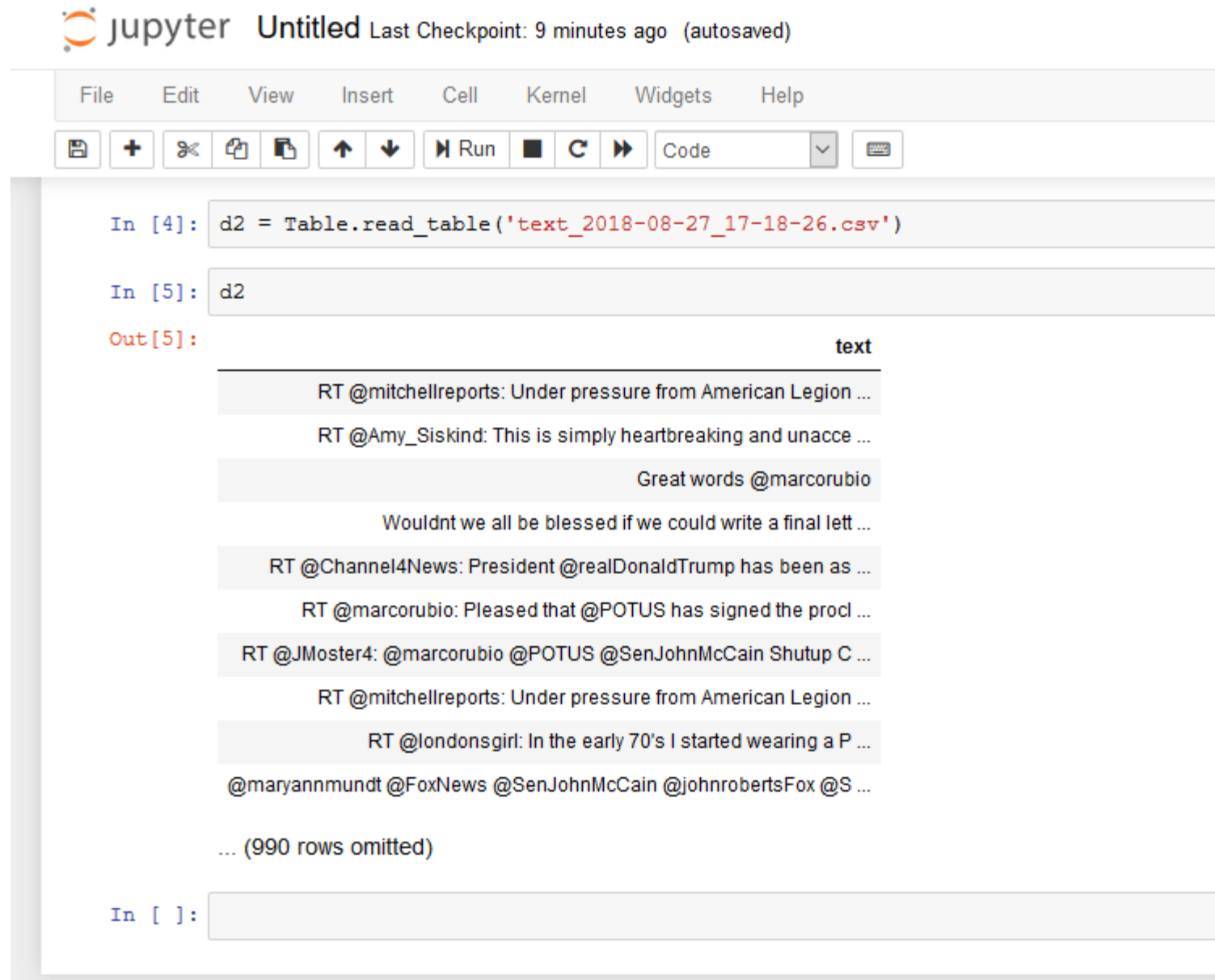
In [3]: `d1`

Out[3]:

user_id	user_name	tweet_time	location	text
1604902290	Catherine Elingburg	Mon Aug 27 21:18:30 +0000 2018	None	RT @mitchellreports: Under pressure from American Legion ...
916846717152010240	CarolynMc	Mon Aug 27 21:18:31 +0000 2018	World Citizen	RT @Amy_Siskind: This is simply heartbreaking and unacce ...
1015272643262582784	Knot M Portant	Mon Aug 27 21:18:31 +0000 2018	None	Great words @marcorubio
722644105931669507	Tasha	Mon Aug 27 21:18:31 +0000 2018	United States	Wouldnt we all be blessed if we could write a final lett ...
2527305086	Dannyboi #FBPE #Stop Brexit #ABTV #peoplesvote	Mon Aug 27 21:18:31 +0000 2018	London UK	RT @Channel4News: President @realDonaldTrump has been as ...
2429369279	DCanes	Mon Aug 27 21:18:31 +0000 2018	None	RT @marcorubio: Pleased that @POTUS has signed the procl ...
917501732577533953	J Moster	Mon Aug 27 21:18:31 +0000 2018	None	RT @JMoster4: @marcorubio @POTUS @SenJohnMcCain Shutup C ...
131019574	D. A. Wood	Mon Aug 27 21:18:31 +0000 2018	None	RT @mitchellreports: Under pressure from American Legion ...
42322155	M@St@rr	Mon Aug 27 21:18:31 +0000 2018	Deep in the heart of Texas	RT @londonsgirl: In the early 70's I started wearing a P ...
87280858	Mike Hill	Mon Aug 27 21:18:32 +0000 2018	Northern VA	@maryannmundt @FoxNews @SenJohnMcCain @johnrobertsFox @S ...

... (990 rows omitted)

Load the CSV file of the filtered tweets in Jupyter Notebook



The screenshot shows a Jupyter Notebook titled "Untitled" with a status bar indicating "Last Checkpoint: 9 minutes ago (autosaved)". The interface includes a menu bar with options: File, Edit, View, Insert, Cell, Kernel, Widgets, and Help. Below the menu is a toolbar with icons for saving, adding cells, undo, redo, copy, paste, and navigation, along with buttons for "Run", "Code", and a dropdown menu.

The notebook contains the following code and output:

```
In [4]: d2 = Table.read_table('text_2018-08-27_17-18-26.csv')
```

```
In [5]: d2
```

Out[5]:

text
RT @mitchellreports: Under pressure from American Legion ...
RT @Amy_Siskind: This is simply heartbreaking and unacce ...
Great words @marcorubio
Wouldnt we all be blessed if we could write a final lett ...
RT @Channel4News: President @realDonaldTrump has been as ...
RT @marcorubio: Pleased that @POTUS has signed the procl ...
RT @JMoster4: @marcorubio @POTUS @SenJohnMcCain Shutup C ...
RT @mitchellreports: Under pressure from American Legion ...
RT @londonsgirl: In the early 70's I started wearing a P ...
@maryannmundt @FoxNews @SenJohnMcCain @johnrobertsFox @S ...
... (990 rows omitted)

```
In [ ]:
```

Lab 2 Submission

- Submit two CSV files containing tweets you gathered (named “data_YYYY-MM-DD-HH-MM-SS.csv” and “text_YYYY-MM-DD-HH-MM-SS.csv”)
- Submit a screenshot of displaying the first ten tweets of these two files in Jupyter notebook.