

Twitter Streaming Query “App”

Usage Guide

This script captures and geolocates live tweets in the U.S. based on designated keywords. For tweets without registered lat/lon coordinates (i.e. most tweets), geolocation is performed by geocoding locations entered by each Twitter user. This is accomplished using the Geonames api, which in turn searches the Global Name Information Service (GNIS) provided by the U.S. Geological Survey. This process is not perfect, since users can enter any location they want (or even “Nowhere”). In an attempt to identify tweets with reliable locations, this script matches user descriptions to city/state or major city names. The focus here is on the United States, but the script can be easily modified by adding administrative units and major cities in another region.

FILES

The app consists of the following two files, which must be placed in the same folder:

- Twitter_Streaming_Query_App.py
- getpip.py

These files can be downloaded from <http://ux1.eiu.edu/~bjkronenfeld/projects/coding.html>.

REQUIREMENTS

Before running the script, you will need to make sure that python is installed and obtain credentials from Twitter and Geonames.

Python Installation

This script was created for Python 2.7.3. This version of python is installed automatically with ArcGIS 10.3.

Twitter Credentials:

1. Create a twitter account.
2. Confirm your e-mail address: <https://support.twitter.com/articles/97942-confirming-your-email-address#>
3. Sign in as a developer: <https://apps.twitter.com/>
4. Click “Create New App” and enter the details as required. You can use any app name you want.
5. Find and copy/paste the following four keys to a text file for reference:
`consumer key, consumer secret, access token key, access token secret`

Geonames Credentials:

1. Create an account: <http://www.geonames.org/login>
2. Click the confirmation link on the confirmation e-mail.
3. Go to <http://www.geonames.org/manageaccount> and click *Click here to enable*

RUNNING THE SCRIPT

This script is not set up with a separate graphical user interface. You will need to use a python IDE to edit and run the script. My preference is PyScripter (<http://sourceforge.net/projects/pyscripter/>). To run the script from PyScripter:

1. Right-click on the file *Twitter_Streaming_Query_App.py* file and select *Edit with PyScripter*.
2. Follow the instructions in the script comments to (a) enter your credentials, (b) modify the search keyword(s), (c) designate a file for the results, and (d) designate the desired running time.
3. Run the script by selecting *Run >> Run* from the main menu (or click the green “play” button).
4. The results will be saved to the designated output file in csv format.

INTERPRETING RESULTS

The output file can be opened in MS Excel, ArcMap, etc. (assuming a .csv file extension). It contains several lat/lon fields:

gps: Associated with tweet. Obtained from numeric “coordinates” field. Assumed to be from user device location.
embed: Associated with user. Discovered in “location” text field, but appearing as lat/lon numbers. Not sure where these are from, but I guess they are produced automatically in some situations when user signs up for a Twitter account.
gnis: Associated with user. City/state discovered in “location” text field. This is then geocoded with Geonames/GNIS. These are obviously entered by the user and so may not be trustworthy.
best: Best of the above, ranked as follows: (1) *gps*, (2) *embed*, (3) *gnis*.

To map all tweets, use the “best” lat/lon coordinates.

TROUBLESHOOTING

The script relies on the Birdy python package to communicate with Twitter. If this package is not already installed, the script will attempt to install it automatically. If this fails you will need to install Birdy yourself. This is not difficult, as Birdy is a registered python package. Here are the basic steps:

1. Download the *get-pip.py* file from <http://pip.readthedocs.org/en/latest/installing.html>
 - This is a basic utility for installing python packages.
2. Copy/paste the above file to your python folder, e.g.: C:\python27\arcgis10.3\
3. Start a command prompt. To find the command prompt, try the following, depending on your flavor of Windows:
 - Start >> Accessories >> Command Prompt
 - Start >> enter "cmd.exe" in the search box
 - Start >> enter "PowerShell" in the search box
4. Enter each of the following commands (you don't need to type comments beginning with " : ")

```
cd c:\python27\arcgis10.2  :: Change directory to your python folder
python get-pip.py          :: Run the get-pip utility using python
cd scripts                :: Change directory to python scripts folder
pip install pyopenssl ndg-httpsclient pyasn1  :: enables secure connections
pip install birdy          :: Use PIP to install Birdy
```

Once Birdy is installed, you should be able to run the Twitter Streaming Query script.

FURTHER INFORMATION

1. Twitter API (<https://dev.twitter.com/overview/documentation>)
2. Birdy (<https://pypi.python.org/pypi/birdy/0.2>)
3. Geonames (<http://www.geonames.org/>)
4. PIP (<http://pip.readthedocs.org/en/latest/installing.html>)