Twitter data collection

SCC.413: Applied Data Mining

NLP – Week 16

In this lab exercise you will interact with the Twitter REST API using Python code to download tweets for future analysis. Data collected via APIs are generally much cleaner than web scraped data, and also structured nicely (here, via JSON) for easy querying.

To collect data from Twitter you need to have a Twitter account with two-factor authentication enabled (i.e. with a mobile phone number linked), and also create an authorised application. If you do not want to do this, you could skip this lab and just use pre-collected data, although it is useful to see how to collect your own data. One option would be to partner with a neighbour and use a single Twitter account.

1 Authorising the Twitter application

To get started, you will need to create an authorised Twitter application. You only need one Twitter application for the module, and you can use the same authorisation credentials for any future work on the module utilising the Twitter API.

To create an authorised application, you need to either have a developer account, or you can be added to an organization team (for education). A team (scc413) has been created for this, and students' Twitter accounts can be added to it.

1.1 Developer account

To be able to create a Twitter application, please choose one of the following three options:

1. If you have a developer account already, please proceed to "Create an app" in Section 1.2.
2. If you would like to have your Twitter account added to the scc413 team (to save applying for a developer account):

(a) Create a Twitter account if you haven't one already.

(b) Your account must have email notifications turned on, as well as two-factor authentication, which requires a mobile phone number being attached to your account.

(c) Send your Twitter handle to Alistair (a.baron@lancaster.ac.uk).

(d) Alistair will add your handle to the developer list of the scc413 team.

(e) You will receive an invite to be added to the team, click the link and follow the instructions.

(f) Once you are on the team, if you visit https://apps.twitter.com, you should have the option to "Create an app", follow the instructions in Section 1.2.

3. You can apply for a developer account yourself, though the authorisation from Twitter for this can take some time.

   (a) Go to https://apps.twitter.com.

   (b) Sign into Twitter.

   (c) Click "Create an app".

   (d) Unless you already have a developer account, you will be asked to create one, follow the instructions, selecting:

   - A "Student" project.
   - For the description of what you are building, fill in **with your own words** (if you all copy and paste this, it will flag as bots and be denied):
     - Collecting tweets for analysis, learning natural language processing skills.
     - Analysing tweets and their textual content, using various natural language processing techniques, such as sentiment analysis and topic modelling.
     - No creation of tweets or content, just the collection of tweets.
     - Tweets will not be displayed to end users, only to be used for learning to use Twitter API.
   - "No" for being available to a government entity.
   - Once you have completed the details, submit the application, and verify the email. You may then have to wait a while for Twitter to approve your application.

1.2 Creating an app

1. Once you have created a developer account, or are part of the scc413 team, you will be able to "Create app" at https://apps.twitter.com.

2. Fill in an app name.

3. If needed, agree to the developer agreement (have a quick read), and create.

4. You will need the *Consumer Key (API Key)*, *Consumer Secret (API Secret)*, *Access Token*, and *Access Token Secret*. Note, you should not share these as they are linked to your account with Twitter, and are rate-limited.

5. In the *Keys and tokens* tab, click "Generate" next to "Access Token & Secret". You will use these in your Python code to authenticate your Twitter app.