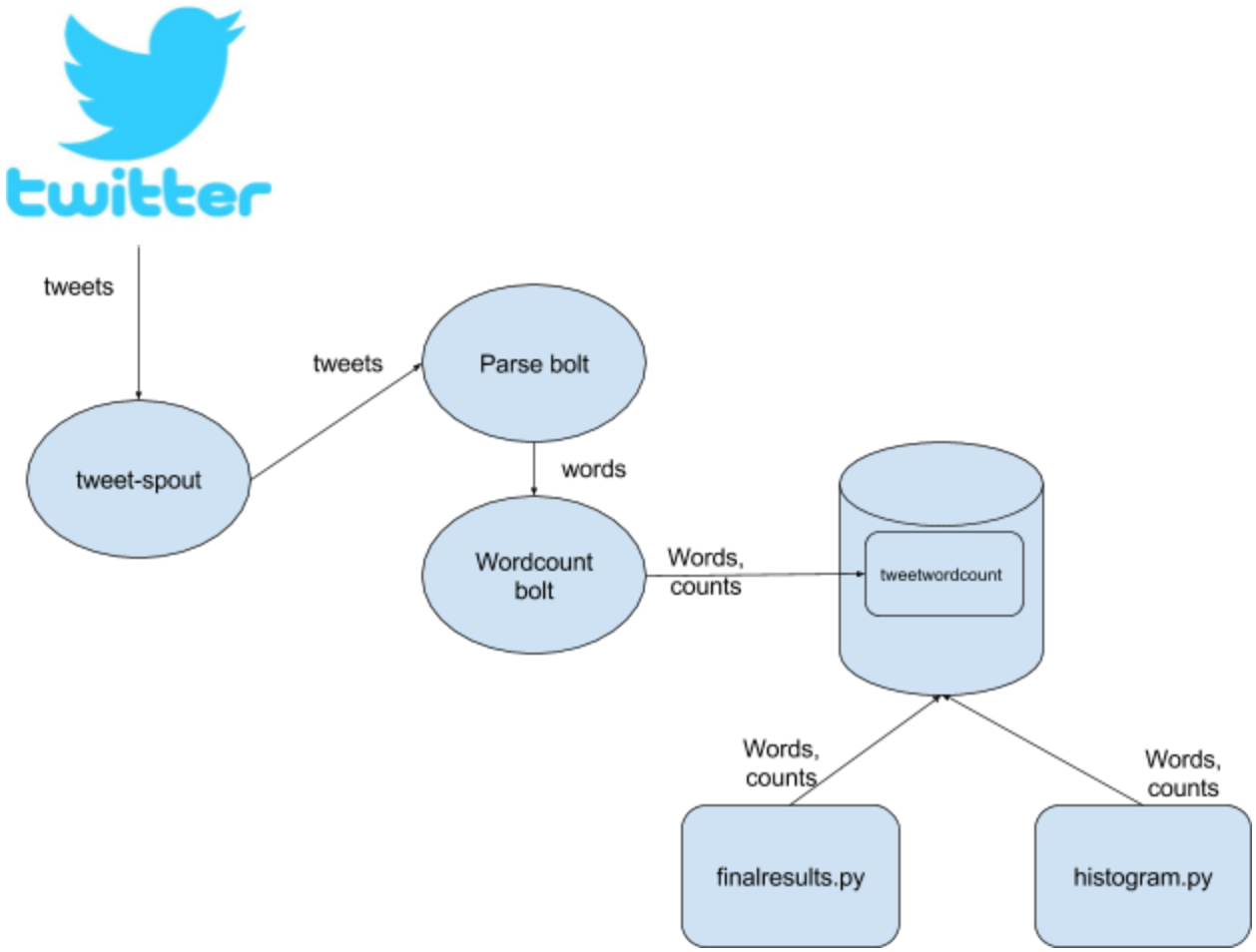
Jay Zuniga w205-3
Exercise 2
8/12/17

Tweet Word Count

This application streams tweets from twitter for a short amount of time, then counts then splits the words in the tweets so they can be counted. The counts of the words will be stored in a postgres database where it can be used for various analysis.

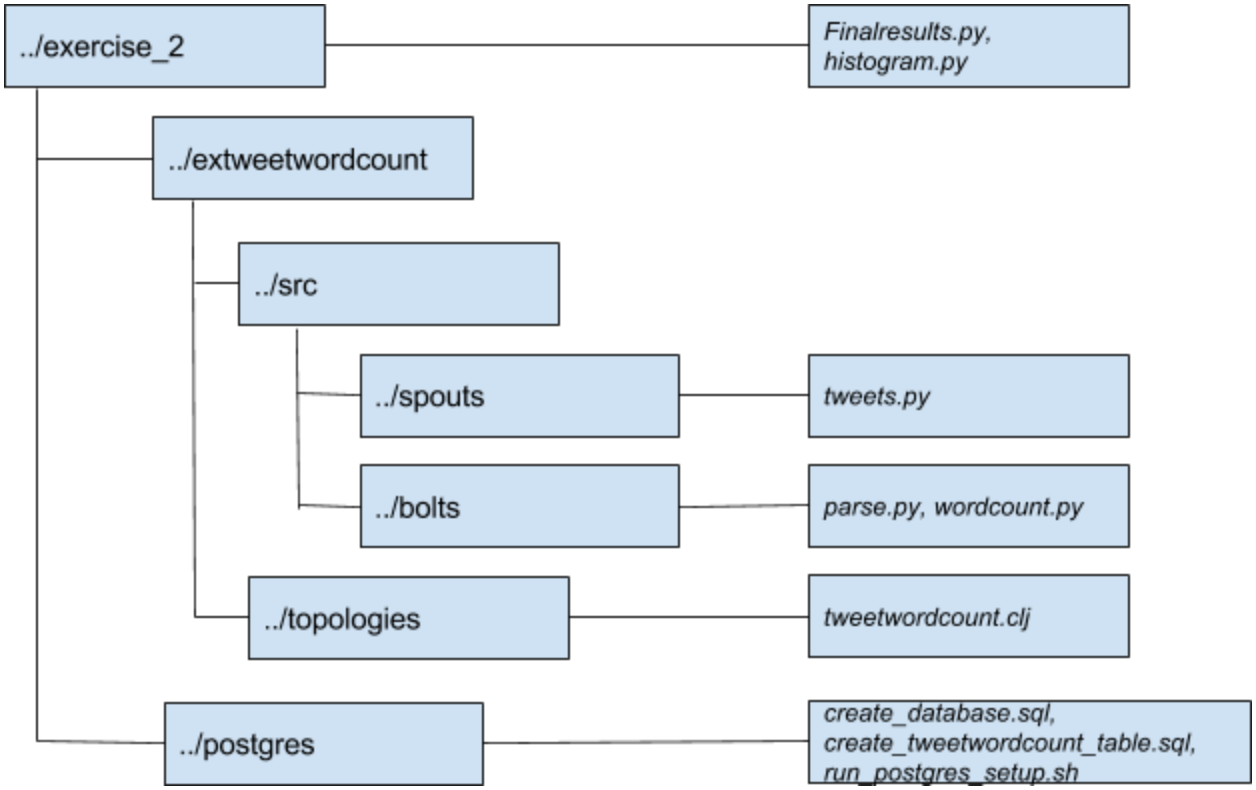
Architecture

The following shows the architecture of our app.



Directory and file structure

The following shows the location of the files.



Directions to Run App

1. Copy Files and Directories

   The configuration I used during my run assumed all the files were under the directory ./exercise_2, however you can use any top directory.

   The following files should be copied using the layout above. In the rest of the document, will refer to the ./exercise_2 (or whatever its replacement is) as $EX2_TOP_DIR.

2. Set up Database

   To set up the database, log in as root (to allow changing to postgres, w205 user) and run the “run_postgres_setup.sh” file passing in the full path to the postgres directory.

   # cd $EX2_TOP_DIR/postgres
   # ./run_postgres_setup.sh $EX2_TOP_DIR

   This step assumes postgres is up.
3. Run Storm

   Before running storm, please enter the twitter credentials in $EX2_TOP_DIR/extweetwordcount/src/spouts/tweets.py.

   $ cd $EX_TOP_DIR/extweetwordcount/
   $ streamparse run -t 60

4. Run scripts

   After Storm has populated the tweetwordcount table, the data can now be used.

   $ cd $EX2_TOP_DIR

   To get number of times a word appears:

   $ python finalresults.py <word>

   To get all the words and their counts:

   $ python finalresults.py

   To get all words within a count range:

   $ python histogram.py <low_value> <high_value>