Extweetwordcount Documentation

1. **Application Description**
   extweetwordcount is a Twitter application which reads live tweets from Twitter, parses each tweet into words, counts numbers of words, and updates word counts in a Postgres table.

2. **Architecture Description**
   The application is divided into 3 main parts:
   a. A spout connected to the Twitter streaming API that pulls tweets and emits them to the parse bolt.
   b. A parse-tweet-bolt that parses the tweets emitted by the spout and extracts individual words out of the received tweet text.
   c. A count-bolt that counts the number of words emitted by the tweet-parse bolt and updates the total counts for each word in the Postgres table (database: tcount, table: tweetwordcount).

3. **Directory/File Structure**
   a. Spout and bolts are saved in src/spouts and src/bolts respectively.
   b. Topology is saved as extweetwordcount.clj in topologies folder, and it refers to files mentioned in (a).
   c. Sample serving scripts to analyze word count data are also included.
      i. finalresults.py: When passed a single word as an argument, finalresults.py returns the total number of word occurrences in the stream. Running finalresults.py without an argument returns all the words in the stream, and their total count of occurrences, sorted alphabetically, one word per line.
      ii. histogram.py: The script gets two integers k1, k2 and returns all the words with a total number of occurrences greater than or equal to k1, and less than or equal to k2.

4. **Other Necessary Information**
   a. All files assume they will be run in Python 3.
   b. Postgres database and table should be set up before. This can be done by running db_setup.py provided.