MR BASED TWITTER ANALYSIS

FILENAME: get_tweets.py

- Get tweets from the twitter API using tweepy python library
- Save the retrieved json tweets to result file

FILENAME:filter_text.py

- From the downloaded tweets extract only the tweet text

FILENAME: mapper.py

- Open the twitter text only file
- Emit the text pairs to be compared to be collected by the reducer
- The key is set as the count of the outer loop
- The value is the text pair to be compared
- The output is emitted to STDOUT

FILENAME: reduce.py

- The emitted tweets are sorted and fed to the appropriate reducer by hadoop based on the key
- The text is compared using difference library (difflib)
- If there is a 90% match then the text pair is displayed