

Fonseca, Menor, Paul

Question: Given a certain number of Twitter users, how can we find out what time of the day is most successful for them to tweet ( in terms of favorites, replies or retweets)?

#### Description

We aim to answer the question through the use of MongoDB and its function MapReduce. The raw data acquired can then be filtered and transformed into useful information which is relevant to the said question. We will look into the number of favorites, replies and retweets and use this as a basis for whether the tweet is successful or not. This will be observed for different times of the day in the span of 1 month. Then, with the information acquired, we aim to pinpoint what time of the day the user should tweet in order to maximize the number of reacts to his or her tweets. Through MapReduce, we will be able to obtain the total average number of favorites, replies, and retweets for specific spans of time given the tweets of 5-10 users. Each user's data will also be presented separately.

#### Procedure

We will use Tweepy for Python ( <http://www.tweepy.org> ) to interact with the Twitter API to obtain the information needed. This requires setting up a Twitter application and installing Tweepy using the pip manager. We will choose a particular Twitter user and use their screen name in a Python script that will collect all tweets made by that user in the past month. These tweets will be outputted as JSON objects that can later be stored in our MongoDB database for manipulation. We aim to collect at least 50,000 tweets from a total of 5 to 10 Twitter users.

#### Scope and Limitations

The data is limited to the times the said user posts. Therefore, the times at which the user is usually asleep (possibly in the midnight and early mornings) will not be part of the possible time of day. This project only aims to give the user a brief overview of the trends based on the time of day of his or her past tweets. Moreover, external variables such as the content of the tweet and possible tags to certain people will not be differentiated by the project and all tweets will be treated the same.