

Project 3.1: CS415 DS Team Proposal for Dashboard for Visualizing Research and Analysis of Twitter and Reddit Data

Chris Wetzel
Binghamton University
Binghamton, New York, United States
cwetzel2@binghamton.edu

Tyler Gabriel
Binghamton University
Binghamton, New York, United States
tgabrie2@binghamton.edu

Paul Maino
Binghamton University
Binghamton, New York, United States
pmaino1@binghamton.edu

ABSTRACT

The researchers discuss the technologies that they wish to use to visualize the results of the Project 2 analysis. The optional queries that may be performed in the proposed application are discussed. Changes that will be needed to integrate existing collections are outlined.

2 SOURCES

- <https://pymongo.readthedocs.io/en/stable/>
- <https://docs.python.org/3/>
- <https://flask.palletsprojects.com/en/1.1.x/>
- <https://angular.io/>
- <https://www.chartjs.org/>

1 METHODOLOGY

In order to visualize the Project 2 analysis of our data, we have decided to create a web application and server. These will be developed with the Angular and Flask frameworks, respectively. In order to display the results in an intuitive manner, we will use Chart.js. The two analyses that we have decided to support are our analysis of sentiment score on four key individuals across subreddits and our analysis of overall sentiment for all tracked individuals on both Twitter and Reddit. Both of these analyses utilized VADER for sentiment scoring and thus can be rapidly updated compared to Google Cloud Platform Natural Language API.

In our Project 2.2 implementation, we created programs to traverse a collection of documents and update missing sentiment scores with their appropriate values. This process was run once to add the metrics to a .csv file. In our Project 3.2 implementation, our program will run constantly and check for missing sentiment values, updating them appropriately. This process will be conducted on collections for Reddit posts and comments as well as tweets.

Likewise, we will also maintain separate collections for the results of each analysis. We will also create a program to recalculate each analysis at a given interval. This will reduce unnecessary calculations and improve the performance of the request. By sacrificing some short term accuracy, the user will have a more responsive experience.

1.1 Average User Statistics

This feature of our application will be composed of two drop-downs that will allow the user to compare the aggregated statistics for two individuals. A four bar bar-chart will be used to visualize and thus compare the two individuals' respective sentiment scores for both Twitter and Reddit. In addition, two tables summarizing the data for each user's respective analysis will be displayed.

1.2 Subreddit Bias

This feature of our application will allow a user to select a subreddit and view the average sentiment scores for all of the tracked individuals. This visualization will provide insight to general ideologies of subreddits and how they may favor certain individuals or groups.