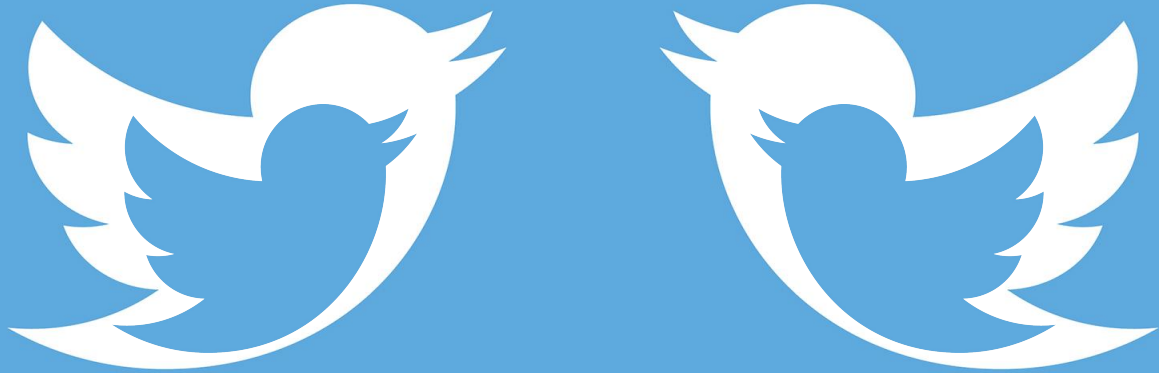


Profil'd



CSCI 49900 - 02

William Wu, Fabio François, Ada Chen, Brandon Troche

Research Idea

- Compare your own tweets with pre-selected Twitter profiles to see how you identify to them.
 - ISIS supporters
 - Donald Trump
 - Hillary Clinton
 - Bernie Sanders
- The expected output will be several charts corresponding to the Twitter profiles that we have available.



Tools

- Spark - Parse data and analyze most commonly used words in the tweets by ISIS sympathizers. Used to find TF-IDF and similarity.
- Python - Create a script to parse tweets to store as JSON objects into our database. Use Tweepy library to dump most recent tweets into a csv (comma separated values) file.
- Heroku - Used to host our webapp.
- MEAN Stack - MongoDB, Express, Angular.js, Node.js. Bootstrap for styling.



MEAN Stack

- MongoDB - database to hold all of the relevant tweets for each Twitter profile. There will be a collection for Trump, Hillary, ISIS and Bernie.
- Express - will be our front end framework
- Angular - will dynamically display information on our site by pulling from our models
- Node - will be used to execute events based on the user command.



Analyzing Data

- The data that the user will be compared to will be stored in a database to be accessed by the web/mobile app
- Sentiment Analysis: Identify keywords and categorize a person's opinions based on their tweets
- Categorize keywords of user and politician tweets and compare them
- Output a percentage for similarity based on a similarity formula.



Web App/Website

- Will allow users to put their Twitter username into a textbox to submit.
- Our Python script will then run and retrieve all the user's tweets.
- The user's tweets are then stored in our database.
- Using Spark, we will submit the information to them for processing.
- The results will be displayed on our web page.



iOS App

- Firebase to store tweets from your twitter profile
- Performs same functionality of the website
- Can possibly pull from the same database in order to prevent redundancy in data
- If active on one's account, can possibly give push notifications when a tweet is sent that is a bit risqué.



Goals

- Provide a fun way of comparing which Twitter personality you are most similar to.
- Possibly provide ways to realize threats when compared to supported Twitter pages.
- Analyzing yourself



Timeline

- Learn the technologies
- Design our website
- Design iOS app
- Implement both website and iOS app with no functionality
- Compile Tweets and create the database
- Get the user Tweets and add to database
- Utilize Spark to compute tf-idf and similarity, store information into MongoDB, or output directly onto the page
- Format the data on our front end platform.



References

ISIS Database - <https://www.kaggle.com/kzaman/how-isis-uses-twitter>

Sentiment Analysis Reference - <https://blog.exploratory.io/sentiment-analysis-with-trump-clinton-sanders-twitter-data-cc978e91960f#.gzdrenvu6>

Tweepy Resource - <https://gist.github.com/yanofsky/5436496>

Project Repo - <https://github.com/Capstone-BAWF/Profil-d>

