Project Progress Report

3/18/16

Tasks and Challenges Identified in Presentation 1 on 2/24/16:

1. Figure out the software stack that best supports the project
2. Implement solid program architecture for handling API requests, script-based NLP, database storage of results, and presentation of final analysis
3. Decide on and code useful analytical functions
4. Build architecture to point that features can be added individually
5. Implement full suite of features
6. Testing, validation, respond to feedback
7. Produce reports/presentations/poster as required

Ongoing Progress:

1. Software stack - COMPLETE

The MEAN stack (MongoDB, ExpressJS, AngularJS, NodeJS) has been implemented.

2. Program Architecture - MOSTLY COMPLETE

A system for handling API requests, NLP processing, database storage, and website presentation of results has been fully implemented. It has also undergone a major revision for order and clarity.

3. Analytical Functions - IN PROGRESS

Early attempts at analytical metadata functions, such as a function to estimate the timezone of a user based on posting activity, have proved difficult. New ideas are possibly needed. Comparison tool for metadata graphing still planned.

4. Architecture set up so that individual features can be added - COMPLETE

Project has moved into the phase where all members of the team can take their own branch of the code, add features, and merge it back in without serious difficulty.

5. Implement full suite of features - IN PROGRESS

Metadata, analysis, NLP, and front-end features have all been implemented. Planned features include subreddit analysis and more NLP.

6. Testing/Validation - IN PROGRESS

Tests have been done to analyze limits of database storage, API calls, etc. Fixes have been implemented for various features after testing. Still much more to come after program is feature complete.

7. Produce Reports/Presentations/Poster - IN PROGRESS

Complete items include the proposal, the first presentation, use case documentation, Gantt charts, and this document. Coming items include the second presentation, the final report, and the poster.

Key Completed Challenges:

1. Setting up the full stack

Getting the full basic implementation of a website frontend connected to a database backend and a framework for analysis tools was a significant undertaking, as a whole host of new tools and unfamiliar languages needed to be learned and used effectively.

2. Setting up NLP analysis

This required integrating the full stack program architecture with the Stanford open source NLP setup, which was no easy feat. It required a fair bit of updating or adjusting plugins, figuring out how to pass data in and out, and so forth.

3. Building a program structure for complete analysis

The key challenge here was making a basic program structure that could be added onto or modified easily for future development. This was difficult due to the need to make multiple sets of API calls, all of which take time to receive responses. Each call must be made, then the program thread freed up to continue processing other portions of the analysis.

Key Future Challenges:

1. Reducing analysis time

There are existing limits on API calls to Reddit which put a reasonably high floor on response times. However, there is likely room for improving time taken by NLP analyses. Metadata

collection, storage, and analysis are not significantly impacting performance currently. Some early improvements for both API calls and NLP have been made, but more work needs to be done. There has also been work done on displaying loading information to users while the analysis runs.

2. Ensuring robustness

There have been surprise bugs, as there always are, in interpreting data from API calls. Old or changing formats for data responses, forms of text that NLP tools have difficulty interpreting, etc. Some bugs have been fixed, and some are likely still lurking out of sight for now.

Looking forward:

Major team member focuses:

Matthew Chatten: Subreddit analysis, project reporting and documentation
Julian Esteban: More NLP analysis, testing
Kanav Tahilramani: More NLP analysis, testing
Sujay Bandarpalle: Front end display of analysis

Next major project task:

Presentation 2, during week of 3/28/16