Working With Open Data – Project Progress Report 2 – April 15, 2014

|  |  |
| --- | --- |
| Team: | Soteria |
| Members: | Evie Phan, Shreyas, Ashley Desouza, Victor Starostenko |

Progress since April 8 Project Proposal Submission:

* Evie – Sketched up wireframes, brainstorm visualization techniques, data exploration using pandas in IPython notebook and used Jaccard Similarity to look for similarities between our datasubsets ( ie. n(AᴖB)/n(AᴗB) ).
* Ashley – Cleaned data, created IPython notebook for data exploration
* Victor – Cleaned data, created Ipython notebook for data exploration
* Shreyas – Study the data for relevant feature extraction, strategize the learning algorithm, Compare similarity measurement algorithms – Jaccard, Cosine to understand which is best suited for the problem.

Ongoing work this week:

* Evie – Continue with front end wireframes, look at different layout options, and continue with data exploration in IPython notebook
* Ashley – Continue with data exploration and data analysis activities. Look at further ways to reduce features.
* Victor – Analysis of relevant attributes. Going to experiment with various techniques to find correlation between attributes.
* Shreyas - Recoding data in IPython data frame, starting the implementation of learning algorithms.

Major obstacles:

* As a group - Figuring how to use the data from the similarities analysis to arrive at an actual conclusion that can be useful for the open source community
* As a group - How to visualize the data to make it understand to a person who doesn't have domain expertise
* As a group - Breadth of data
* Evie – Decide on what a simple and detail layout for the visualization