OSCAR

Open Social Clustering and Analysis for Reddit

The Team

Anna DeBrunner, debrunac@mail.uc.edu

Charles Greenman, greenmcg@mail.uc.edu

Isaac Mathers, matherit@mail.uc.edu

Advisor: Dr. Raj Bhatnagar

Purpose/Goals

- Explore users, communities, and topics on Reddit to discover flags of potentially harmful users.
- Present information and recommendations to users to make them more informed about the users and content they are interacting with online.
- Implement data analysis algorithms and visualizations to summarize available information about a user in an easily digestible way.
- Implement graph analysis algorithms and visualizations to summarize available information about user networks in an easily digestible way.

Project Description

Abstract

The exploitation of social media networks came to light following the 2016 US Presidential election. Since then, the websites we use every day have become more aware of the issue, but it has also become more pervasive. With another election cycle just around the corner, we must arm ourselves with the correct tools and information to ensure we do not fall prey to false stories, targeted beliefs, and general discord. With than in mind, the Open Social Clustering and Analysis for Reddit tool, or OSCAR, hopes to provide a more informed experience on Reddit regarding its users and communities.

.

User Stories

As a user, I want to submit a social media user's profile for analysis to see trends in their posting activity, including frequency of certain topics and controversial posts.
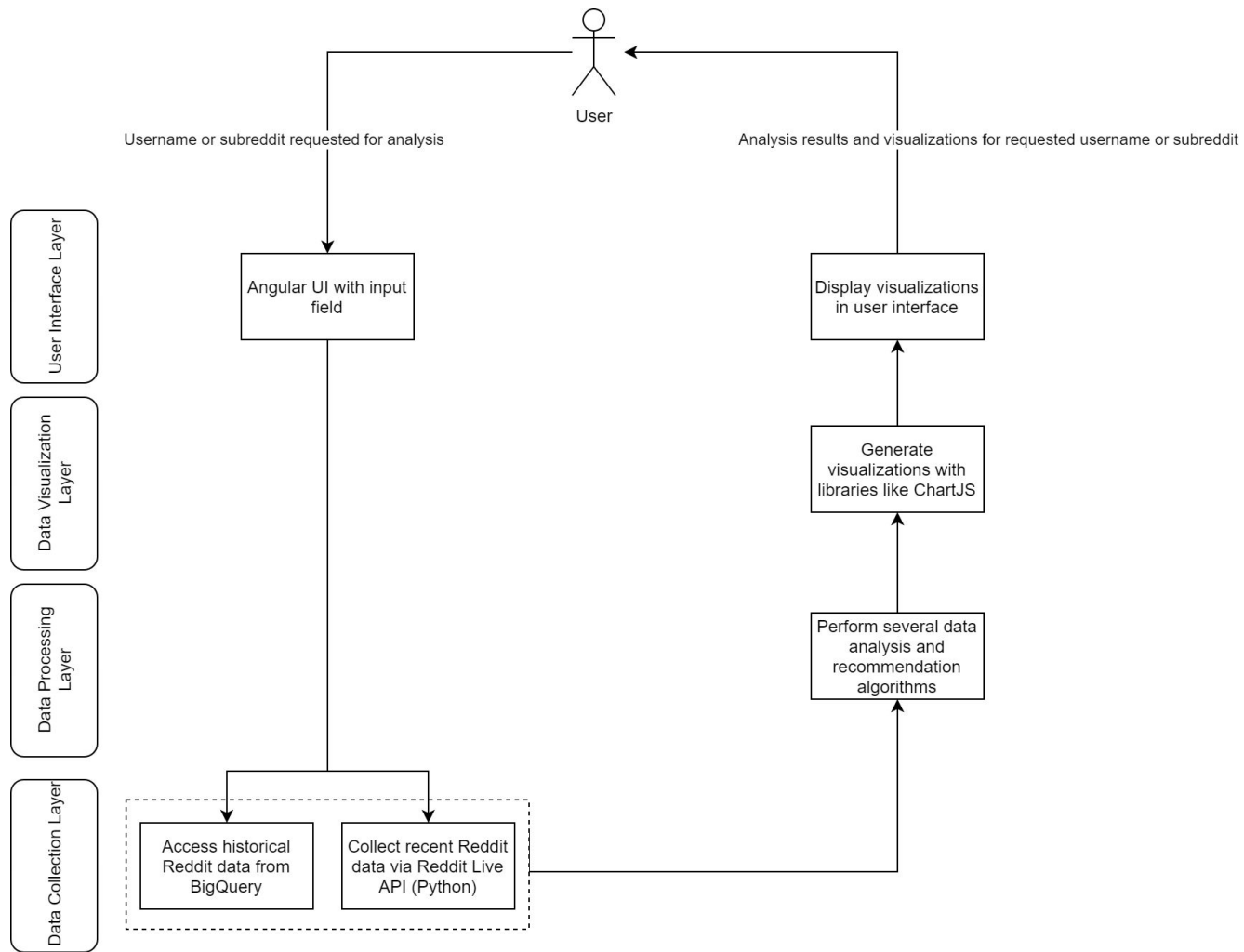
As a user, I want to be able to see data visualizations that summarizes a user's posting activity.

As a user, I want to be able to see data visualizations that compare a user's posting activity about a topic to that of all of Reddit.

As a user, I want to be able to submit a social media user's profile for graph analysis to determine if they are likely to be a member of a sock puppet organization.

As a user, I want to be able to view analysis and data visualizations of a user's interactions with other users, both positive and negative.

As a user, I want to receive a recommendation regarding whether or not a submitted user can be considered trustworthy.

User

Username or subreddit requested for analysis

Analysis results and visualizations for requested username or subreddit

User Interface Layer

Angular UI with input field

Display visualizations in user interface

Data Visualization Layer

Generate visualizations with libraries like ChartJS

Data Processing Layer

Perform several data analysis and recommendation algorithms

Data Collection Layer

Access historical Reddit data from BigQuery

Collect recent Reddit data via Reddit Live API (Python)

Project Constraints

- ○ **Economic** - We are very limited in the amount of computing power available to us since we don't have unlimited resources or the money to buy them.
- ○ **Legal** - Reddit has strict rules about how its API can be used legally. We need to make sure we comply with them.
- ○ **Diversity and Cultural** - We will only be focusing on English speakers for this project to maintain simplicity.
- ○ **Social** - We need to make our solution as usable as possible by the average person to have the maximum positive social impact.

Project Progress

Current State

- Completed download and storage of Reddit activity for the past 3 months
- Completed a preliminary package to interact with Reddit Live API data per user
- Significant design legwork
- Selecting tools and technologies
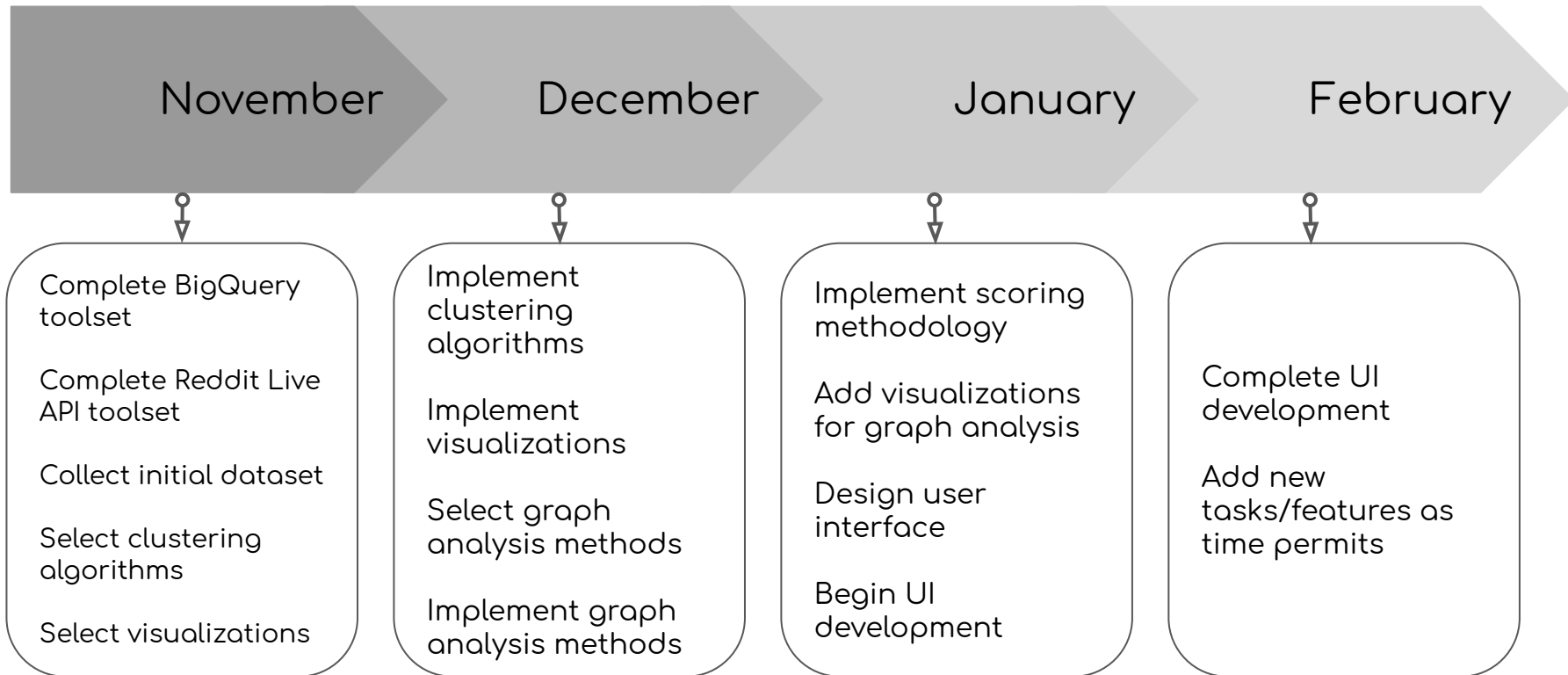
Expected Accomplishments

- ○ Selecting clustering and data visualization methods
- ○ Implementing clustering and data visualization functionality
- ○ Beginning social graph analysis
- ○ Explore UI possibilities

Timeline and Division of Work

Timeline

November	December	January	February

November
- Complete BigQuery toolset
- Complete Reddit Live API toolset
- Collect initial dataset
- Select clustering algorithms
- Select visualizations

December
- Implement clustering algorithms
- Implement visualizations
- Select graph analysis methods
- Implement graph analysis methods

January
- Implement scoring methodology
- Add visualizations for graph analysis
- Design user interface
- Begin UI development

February
- Complete UI development
- Add new tasks/features as time permits

Responsibilities and Demo

- Anna is mainly responsible for tasks regarding visualization and the user interface.
- Charles is mainly responsible for tasks regarding obtaining the initial dataset and social graph analysis.
- Isaac is mainly responsible for tasks regarding clustering and scoring implementation.
- Our demo will consist of our user interface pulled up on a laptop that people can use in real time.

Thanks for watching!