**RedReaper**
Julian Esteban, Kanav Tahlilramani, Matthew Chatten, Sujay Bandarpalle
{jme136, kt295, mwc70, sb1073}@scarletmail.rutgers.edu
Advisor: Dr. Shantenu Jha

# Project Overview

RedReaper is a social media analysis tool built specifically for Reddit\(^1\), a community website organized into “subreddits” dedicated to individual topics, where users vote on, comment on, and post what interests them.

Submissions, comments, and user histories are all publicly available to anyone, so RedReaper makes use of Reddit’s open API to gather this data on requested users or subreddits. That data is then organized, parsed, and analyzed in order to build a full report.

## Motivations and Objectives

- **Motivations**
  - Make interesting, informative data and analysis available
  - Show users what can be revealed about them and their online communities through natural language processing and collated public data

- **Objectives**
  - Build accurate, readable feature outputs for analysis
  - Provide wide variety of results, based on both content and metadata retrieved

## Challenges

- Deciding on best framework for project
  - Chose to build a website using the MEAN stack

- Figuring out best approach for natural language processing
  - Chose to use Stanford open source NLP tools\(^2\) for our own custom analyses

- Make improvements to performance
  - Chose to use ample processing power, reduce database storage, and increase efficiency of API calls and NLP processing

# Acknowledgement

We would like to thank our advisor, Dr. Jha, and all those who create, maintain, and document the open source tools and extensions that make projects like this possible. Many thanks.

# Methodology

## General Software Practices

RedReaper runs on the MEAN stack, a Javascript framework for web applications which uses mongoDB, Express, AngularJS, and NodeJS.

The project was organized on a public Github\(^3\) repository. Special care was taken to ensure that the software is extensible, allowing for individual team members to contribute their own features harmoniously.

In the early stages, the team worked together to outline goals and build the framework of RedReaper. Afterwards, each member concentrated on their area of strength, like front end display or NLP analysis.

## General UML Sequence Diagram

Pictured above: some RedReaper outputs for Reddit user “CBtheDB”

Includes user info, metadata summaries, extracted personal information, and NLP analysis of comments

# Results

## Top 5 Subreddit Average Scores

- Picked by users
- High positive
- Low negative
- Neutral

# References

