Alexander Swanson

February 10, 2018

Professor Lisa Dion

CS 124 – Data Structures and Algorithms

Project 2

Project 2 Report

Dear CS 124 Grader,

For over seven months now, I have been working on a project fundamentally aimed at evaluating and influencing online public opinion. There are many platforms to work on this problem – *Facebook*, *Twitter*, *Instagram*, etc. However, attempting to access all of these platforms immediately is not a practical approach. Thus, working with the social, information aggregation forum *Reddit*, a large amount of data has been collected to begin evaluating public opinion. A portion of this data has been selected for use with this project in order to complete fundamental analysis.

In this second portion of work for *CS 124 – Data Structures and Algorithms*, a class was designed to represent a row of data from the original dataset. The data is composed of general metadata of Reddit comments posted to submissions discussing world news events. An instance of this class, "*RedditElement*," includes the following fields:

ID

The unique ID of the Reddit *comment* object.

Subreddit Name: Prefixed

The common identifier of a categorical forum on Reddit.

Ups

The comment's total resulting *upvotes*.

Downs

The comment's total resulting *downvotes*.

Score

The overall score, a function of the comment's upvotes and downvotes.

Category

The general topic category of the comment's text body.

Sentiment Score

The score of the sentiment, rational to the sentiment magnitude. That is, the type of opinion expressed in the comment text body.

Sentiment Magnitude

The magnitude of the sentiment, rational to the sentiment score. That is, the weight of the opinion expressed in the comment text body.

Created

The UTC unit measure of the date the comment was created.

This newer project has been completed with the dataset used for the *Project 1* – however, some data fields determined to be irrelevant were removed.

Beside functionality for setting and getting data field values for an instance of RedditElement, the class also allows for the formatted output of all the statistics of the object with the *info*() method. Moreover, using the *sentiment_score* data field, the *measure_happiness*() function performs a sum of every instance of *sentiment_score* for each RedditElement object and calculates the arithmetic mean, yielding a naïve measurement of the "happiness." These calculations are not, however, the most complex for the program. Such is, rather, the data input function, which exhibits a complexity of $O(N^2)$, as it contains a nested *while* loop.

Sincerely,

Alexander Swanson

University of Vermont