Clueless Idiots

CPTS 415

Dr. Srinivasulu (Srini) Badri

September 11th, 2022

Project Statement for Milestone 1

Clueless Idiots

Ernad Ljaljic, Matthew Jones, Boris Bugingo, Mathew Bauer, Derek Olson

1. Problem Statement:

a. The project that the ‘Clueless Idiots’ group will be proceeding with is the “YouTube Analyzer”. The input to the problem would be previously gathered data from the YouTube videos. The output would be the analyzed data from the videos. The data would be analyzed and the YouTube analyzer would support different functions including network aggregation (degree distribution and categorized statistics, search, and influence analysis).

b. An interesting problem to solve would be extracting the most influential videos using influence analysis and PageRank algorithms. The videos with high influence could help gather data and information about the people’s tendencies and what videos they watch in order to reach the high influence video. One application of utilizing this information is targeted advertisements. For example, knowing that a certain set of videos leads to the high influence video can indicate a similar audience, hence the ads would be displayed for that audience.

c. ‘Clueless Idiots’ would like to successfully implement the three mentioned functions for the Youtube Analyzer, as well as evaluate the efficacy of the PageRank algorithm.

2. Team:

a.

* Matthew Jones
* Ernad Ljaljic
* Boris Bugingo
* Mathew Bauer
* Derek Olson

b.

* Matthew - Basic experience working with SQL databases (PostgreSql). Experience with development in Python, C, C++. Familiarity with graph algorithms and their implementation/design.
* Ernad - Some experience working with MySQL, PostgresSQL, and SQLAlchemy library for Python. Development experience with Python, C, Java, Javascript.
* Boris - some experience working with SQL databases (Google Cloud Platform/ PostgreSql), Python, C/C++
* Mathew - I have some experience with python mostly in the building out of websites and using it as a backend language to help us read from a database. I also have experience with mostly java and some javascript. I also understand github daily well.
* Derek - I have a lot of professional experience writing Javascript/Typescript, the VueJS and ReactJS librariers and a lot of MSSQL/PostgreSQL. I also have familiarity with setting up and using MongoDB. I also have a lot of experience using Python and GitHub as well. I also have experience with implementing search ranking algorithms such as BM25 and have previous knowledge about how page rank works.

c.

In practice we expect that group members will take it upon themselves to complete sub-portions of whatever is necessary for the project. Matthew has set up a GitHub repository to contribute to, and we are in the process of setting up accounts and permissions for the team. The repository, and accompanying contribution data, will be made available to the professor for review.

3. Dataset and Tools: Pending

a. Find the dataset using the following link <https://netsg.cs.sfu.ca/youtubedata/>. The data has been gathered by using youtube crawlers. There are multiple datasets gathered at different points in time. The datasets will be evaluated and the most fitting dataset for the project will be chosen.

b. The current choice is using PostgreSQL with Python and SQLAlchemy library.

4. Project Progress and Contributions:

a. Multiple datasets for this project have been provided by the instructor. The chosen dataset will be parsed and integrated in PostgreSQL. The datasets hold information such as video ID, uploader, age, category, length, views, rate, ratings, comments, and related IDs.

b.

| Team member | Tasks completed | Approx. number of hours |
| --- | --- | --- |
| Ernad Ljaljic | Helped prepare the document, and have done research necessary for the Youtube analyzer project. Met with the team to discuss next steps. | 3 |
| Matthew Jones | Document preparation, research, and the creation and management of the GitHub Repository. Research into proper integration of SQL solutions in a Python workflow (our chosen programming language). | 2.5 |
| Mathew Bauer | Researched Youtube analyzer and possible ways to implement. Also looked to see if there was any example like this out there. | 2.5 |
| Boris Bugingo | Done research on the Youtube analyzer project. Revisited SQL and how to connect with Python | 2.5 |
| Derek Olson | Completed research on the Youtube analyzer and started to develop a potential relational diagram. Reviewed some PostgreSQL - SQLAlchemy specific syntax. | 2.5 |

c. Current plan for Milestone 2 is to choose the dataset to work with. Once the dataset is chosen, develop a Python parser that can extract the information from the chosen dataset, and insert the data into PostgreSQL locally to begin work on the primary project objectives.