Steve Nowicki

CS 410

Project Proposal

10/20/2022

**What are the names and NetIDs of all your team members? Who is the captain? The captain will have more administrative duties than team members.**

Name: Steve Nowicki

NetID: snowick2

**What is your free topic? Please give a detailed description. What is the task? Why is it important or interesting? What is your planned approach? What tools, systems or datasets are involved? What is the expected outcome? How are you going to evaluate your work?**

My topic is the Classification of Movie plots. The project will use <https://www.omdbapi.com/> to retrieve plot summaries about movies. Once retrieved, movie plots will be analyzed, and an attempt will be made to categorize the movie into a genre. Words from the plot summaries will need to be compared to collections of words that are known to be associated with certain movie genres. To create collections of words that are known to be associated with certain genres, an initial human classification of plot summaries can be performed – or perhaps datasets that contain words associated with certain genres can be found on the web.

The basic expected outcome will be that given an input movie title, the tool will classify the movie.

If this basic outcome is completed, and there is still time, a user interface could be built to help a person search. For example, a user could enter a genre, and the tool could show them movies from that genre. Alternatively, the user could enter a search query and the tool could attempt to find movies that have plot descriptions similar to the query.

**Which programming language do you plan to use?**

Either Python or JavaScript will be used.

**Please justify that the workload of your topic is at least 20\*N hours, N being the total number of students in your team. You may list the main tasks to be completed, and the estimated time cost for each task.**

Create a basic program to interact with the API – 4 hrs

Implement classification code – 8 hrs

Test/document code – 4 hrs

Create presentation – 4 hrs