CS410 Project Proposal

1. 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.

Teague Mitchell, thm2

2. 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 to analyze articles from different publications to help determine the political bias of different articles. I plan to build off of a previous CS410 project: https://github.com/abedoyan/CS410_CourseProject.git. However, instead of using movie scripts to determine a genre, I want to use different publications to help build a metric of political bias. I will source political articles from well known publications with a generally agreed upon bias and use those as training examples. Then I will find more independent articles and see how good the article is at determining bias compared to my own judgment. Time allowing, I will add additional metrics for more specific political aspects like which president or policy the article is in support of. Similar to the Genre Analyzer, I will mainly be using Python and the LSI method to determine similarity. I expect that political bias will be fairly difficult to determine but more specific angles like presidential support may be easier to determine given the potential specificity of an article.

3. Which programming language do you plan to use?

Mainly Python, I believe.

4. 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.

Doing research on political bias and what articles will make good training examples and then compiling a list and weighing each will probably take 3-5 hours since politics can be complex to analyze. Then finding articles that would be suitable for analysis will probably take another few hours. Writing the code will take a long time due to my inexperience but building off a previous project should make it easier, so I expect that will probably take 5-10 hours. Evaluating parameters and the functionality of the code and making adjustments will then probably take a few hours as well. From there if I'm on the lower estimates of the time it would take me I will try to add the additional measurements as I said, which should take a few hours per measurement depending on how difficult it is to find suitable articles.