Wyatt Blair

EMT-678-WS: Big Data Technologies

11/25/2024

**Final Project Proposal – Literature Analysis**

**TOPIC:** Literature sentiment analysis across time.

**PLAN:** Using the Standardized Project Gutenberg Corpus, which is available for free on GitHub, I plan to perform a sentiment analysis across 50,000 books using Natural Language Processing tools. I would like to produce some charts indicating how “positive” or “negative” sentiments have changed over time for several topics such as technology, politics, social problems, etc. I will need to process the data so everything is organized chronologically and to ensure that older language/ outdated terms are handled properly in modern NLP tools.

**SOURCES:**

* <https://github.com/pgcorpus/gutenberg>