**Project Proposal**

**Team Captain:** Zachariah Dick, NetID: zdick2 (One-person team)

**Topic:** Improving a System: MeTA Toolkit

**Sub-Topic:** Add text mining functions to the MeTA toolkit

**Main Proposal:** I propose implementing the aspect segmentation and Latent Rating Regression model (LRR) in C++ from the paper “latent aspect rating analysis on review text data: a rating regression approach” mentioned in the project topics document. This functionality will be developed with the intent to be easily ported into the MeTA codebase. Due to time constraints and the fact that I am a one person team I do not intend on fully integrating this functionality into the meta repository, but instead applying an MIT license to this repository and leaving it public to allow for future integration into the MeTA project.

**Task List Estimate:**

* Setup development environment based on that of the meta-toolkit/meta.
* Review the structure of the MeTA repository to gather needed namespaces, data structures, testing setup, and other necessary information to ensure compatibility in my setup.
* Dissect and Analyze the mentioned paper to develop the pseudo code for the implementation and fully understand the algorithm.
* Gather testing datasets and process them using python tools that have been used in MPs. Use metapy wherever possible to ensure compatibility with MeTA. The link to the dataset used in the paper is no longer valid so I will need to put together my own test data.
* Write the functions needed to do the implementation. Attempt to maintain the same pattern used for functions in MeTA repository for ease of portability.
* Write automated tests for the functionality developed.
* Continuous development of documentation.
* Create demo presentation.

**Deliverables:** The deliverables for this project are the source code implementing functionality for latent aspect rating analysis, an automated test suite to demonstrate the functionality, usage documentation, and any other project requirements listed in the project overview.

**Resources to be Used:**

Hongning Wang, Yue Lu, and ChengXiang Zhai, Latent aspect rating analysis on review text data: a rating regression approach. In Proceedings of ACM KDD 2010, pp. 783-792, 2010. DOI=10.1145/1835804.1835903

meta-toolkit. https://github.com/meta-toolkit/meta