IST736 Text Mining
HW3

Exploration of an Interesting Text Corpus

In this assignment, you have the freedom to find an existing text corpus, or create a new text corpus of your interest. If you are creating a new corpus, make sure it is not too large and thus too time-consuming to create.

Then you will vectorize the text corpus using any tool that you are comfortable with: Weka, R, Python, etc. Explain the decisions you made during the vectorization process, e.g., did you merge lower- and uppercase? Then you will explore the text vectors and see if you can find anything interesting. The lectures showed some examples of comparative analysis and trend analysis. But you have the freedom to define what would be interesting patterns as long as you can explain it in a sensible way.

Follow the HW1 requirement on formatting and grading rubrics.