Katharyn Loweth

Text as Data

Prof. Asimovic

March 13, 2025

Text as Data Final Project Proposal

The pursuit of “government efficiency” in the new Trump administration has adversely impacted many federal agencies and the DC-MD-VA area at large. Although the phrase has is now commonly found in federal government news because of the new Department of Government Efficiency (DOGE), the desire and goal to evaluate government efforts to improve efficiency has been a focus of multiple administrations. As someone who previously worked in the research evaluation space and a current federal employee, I believe that the context of government efficiency– and by extension government evaluation and use of data – is quickly changing and substantially different now than it was during the Biden and first Trump administration.

In my final project, I aim to answer the following research questions:

1. What is the context around federal government efficiency efforts and the tools used to support efficiency (evaluation and data)?
2. How has it changed in official federal documents and in media coverage of federal government since 2017?
   1. How is the context of government efficiency different during the following time periods: first Trump administration (2017-2020), Biden administration (2021 – November 2024), second Trump administration (November 2024 – present)? (For purposes of this project I am grouping the time period after the November 2024 election with the second Trump administration.)

Using two key sources – the New York Times and the Federal Register – I will look at texts with direct references to these key words/phrases such as: evaluation, government evaluation, efficiency, government efficiency, data, and data-driven efforts. I will select texts that were released over the time period of January 20, 2017 (the first day of the first Trump administration) through now. Using both the New York Times and the Federal Register provides two different views of the context; the Federal Register provides context around the formal uses of the words/phrases and the NYTs demonstrates how the words are being interpreted by media sources. I will compile New York Times articles through the New York Times API and compile official federal government documents (published by the Executive Office of the President) from the Federal Register API. I will use filters within the API such as keywords and date ranges to narrow the query results.

To analyze the data, I will use two text analytic methods: keywords in context (KWIC) and word embeddings. These two methods, which both focus on a window of words around a specific target word, will provide different insights into the context of these specific keywords in these documents. KWIC will extract the occurrences of these words/phrases from the documents whereas word embeddings will determine the similarity of these target words/phrases to the surrounding words over time. Both methods also enable me to compare the subgroup the data into the three time periods to compare.

My definition of success for this project would be that I have gained more insight into how evaluation is changing. Having worked for several years on federal surveys and being a Data Science for Public Policy student, I have a personal sense of changes in the field, such as use of AI to inform data-driven decisions, but the results of this project may provide me a source of validation. I also consider success as that I am able to more deeply learn these analytic techniques so that I can confidently explain them in a future job interview.