Simplifying Government Documents

Shay Merchant	Cason White	Dillon Roberts	Jacob Craiglow
Merchants18@students.ecu.edu	whitecas18@students.ecu.edu	robertsd13@students.ecu.edu	craiglowj16@students.ecu.edu

1. Introduction

The general audience for most government documents are politicians. This often leads to proposals, speeches, Senate and Congress bills being wrought with confusion for the everyday person. Besides the technical nature of these documents, the texts can be extremely lengthy, and deceptively titled. A couple examples of this deception, whether intended or not, is the *Patriot Act and Citizen's United.* The *Patriot Act* expanded government surveillance privileges, and *Citizen's United* lifted the cap on what corporations could donate to politicians. Our goal is to remove the confusion of these documents and make a program which can summarize these documents in a more digestible format.

2. Related work

Milad Moradi's group worked on a similar program. A link to his work can be found here, https://www.sciencedirect.com/science/article/pii/S1532046418302156?via%3Dihub. Rather than being focused on political documents, Moradi chose to focus on biomedical texts. Like us, Moradi outlines how the challenge they had to overcome was the different diseases and subcategories of biomedical texts which can affect the way a text is formatted. Moradi's approach was to tackle summarization in 4 steps. Preprocessing, topic extraction, sentence clustering and summary generation. The pros of this approach is that it is concise and

relatively simple to implement. The cons could be that a simple approach may take out some important language and context out of the original text to summarize it.

Anna Kazantseva, and Stan Szpakowicz also had similar work with summarizing short stories. Their work can be found at https://www.aclweb.org/anthology/J10-1003.pdf. Their team took a unique approach, whereas they made sure their reader knew certain details about the literature they were reading. The pro to this choice is that you are ensuring that your reader has some level of understanding at the end of the summary, even if the summary is less than ideal. The con to this approach is that a reader may have to have an understanding of the document before it's summarizes, which could potentially contradict the point of a summary.

3. Approach

Our approach will be to take key terms and their derivatives and search for them in a document. These words will be popular words which are repeatedly found in government documents that we find to have higher significance than others. We will use these words and focus our attention on them. Context of these words will be taken from analyzing the words surrounding them.