Troubleshooting

Result Optimizations

Application is filtering useful sentences from corpus for computations or failing to filter non-useful sentences

- There are several static values that can be set for application usage which determine how exactly sentences are filtered. If more/fewer sentences are determined to need filtered, these values can easily be modified.

Computing results is taking too long

- The analysis done by this application is extremely computation heavy and will likely take a significant amount of time to run. If the computer used for running the program does not have relatively strong performance, our team recommends running the program on Colab and have provided Colab files on the repository. Another option is to modify the number of documents being analyzed for the current run iteration.

The number of clusters seems suboptimal

- Methods within the code are provided to determine the optimal cluster size and may be used to benchmark the relative performance of the current amount of clusters that is being used. The number of clusters may then easily be modified if it is determined to be suboptimal.

Heatmaps are difficult to read/interpret

- Values within the code related to the color scale of the heatmap and numerical scale to determine the shading may be modified.

Critical Errors

Application is failing to read input data

- The input data read by the program is expected to be in the form of a CSV. Double-check the number of documents the program is expecting is not greater than the total number of documents within the imputed corpus.

Application crashes during execution

- The program is extremely computation-heavy and consumer-grade devices may experience difficulty running the code. Our team recommends running the program on Colab and have provided Colab files on the repository.

-