Project Title: Exploring Textual Data through Interactive Visualizations

Deadline is Thursday 7th December 1159 pm

Demos of Project will be conducted on 8th December

Team of 2 students are allowed and even cross-section students allowed.

Datasets: Students are provided with the option to choose from two rich text datasets:

1. Enron Email Dataset
2. Cornell Movie Dialogs Dataset

Requirements:

1. **Data Exploration:**
    - Utilize Python and libraries like NLTK to perform basic Natural Language Processing (NLP) tasks such as Named Entity Recognition (NER) on the chosen dataset. While formal NLP training is not mandatory, students are encouraged to learn and apply these techniques for deeper insights.
2. **Web Integration:**
    - Implement a backend using Flask or Django to seamlessly transfer processed data from Python to the web interface.
3. **Interactive Visualizations:**
    - Develop a web-based interface using D3.js that hosts a minimum of two visualizations.
        - **Visualization Variety:** Avoid standard charts (e.g., bar charts) and focus on more sophisticated visualizations. You can use any visualization already developed either available on d3 website or elsewhere that contains d3 code. You must transform that visualization to handle your given data as per your requirements. The visualizations must be fully interactive with all basic interactivity functions implemented (i.e. highlighting, selection, zooming etc.)
        - **Graph/Tree Visualization:** One of the visualizations must be based on graphs or trees.
        - **Focus+Context Style:** Implement the visualizations in the Focus+Context style for enhanced user understanding and keeping the context in view all the time.
4. **Brushing and Linking (Dashboarding):**
    - Implement brushing and linking functionalities between the visualizations to enhance user interaction and exploration.

Guidance:

- Provide clear documentation guiding students on NLP techniques, especially using libraries like NLTK in Python.
- Encourage creativity in designing visualizations to convey specific details effectively.
- Emphasize the importance of interactivity and Focus+Context design principles in creating user-friendly visualizations.

Note: Students are encouraged to delve into the world of NLP, explore the chosen dataset, and showcase their skills in both data processing and interactive visualization development. The project aims to foster a holistic understanding of text data analysis and visualization techniques.

Submission:

Submit the files as a zip folder with your roll numbers mentioned in the file name.