Design Phase Documentation

Project Title: Enhancing Text Analytics Data Quality with NLP

Team ID: d24f3a13

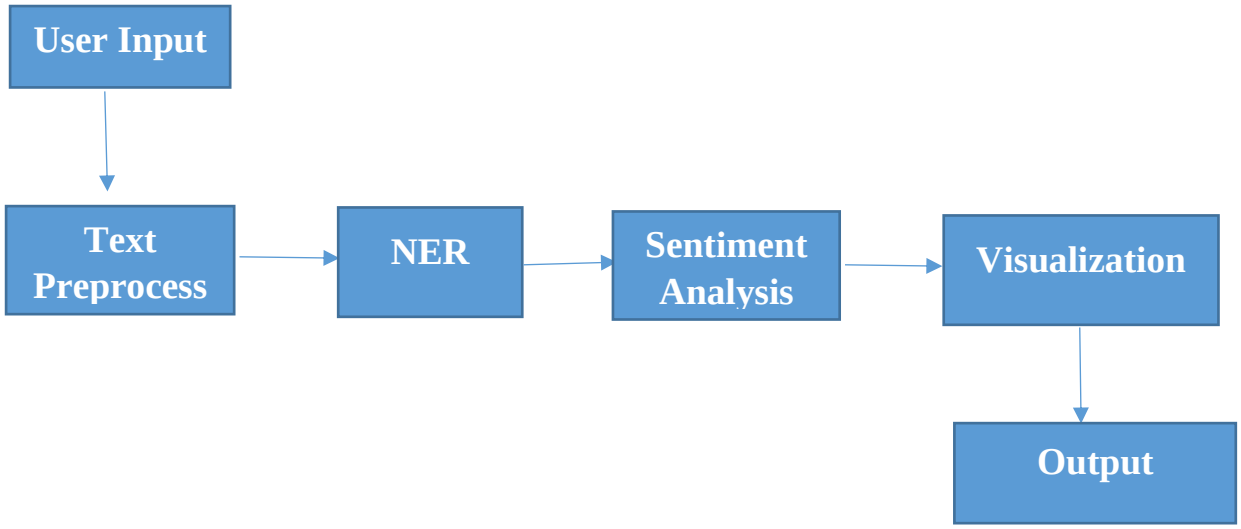
Research Duration: May 1, 2025 to May 10, 2025

Prepared By: Team d24f3a13

1. Primary Objective of the Project

The main goal of this project is to develop a web-based system that can take raw textual input from the user, pre-process it, perform meaningful extraction of entities with the help of Named Entity Recognition (NER), analyze its sentiment, and display the results in the form of dynamic visualizations. The system must be modular, easy to use, and scalable to accommodate possible future features such as multilingual functionality or bigger input data.

2. Data Flow Diagram



3. Module wise design

Frontend

- Technologies: HTML
- Functions:
  - Accepts user input
  - Submits text to backend
  - Displays results

Text Pre-processor

- Technologies: Python, NLTK, spaCy
- Functions:
    - Tokenization
    - Stopword removal
    - Punctuation and case normalization
    - Lemmatization or stemming

NER Module

- Library: spaCy
- Function:
    - Extracts PERSON, ORG, DATE, etc.
    - Returns labelled entities to be color-coded in output

Sentiment Analysis

- Libraries: VADER (NLTK), or TextBlob
- Function:
    - Classifies sentiment polarity: positive, neutral, or negative
    - Outputs a compound score

Visualization

- Libraries: Matplotlib, WordCloud
- Output:
    - Sentiment bar chart
    - Word cloud of frequent terms

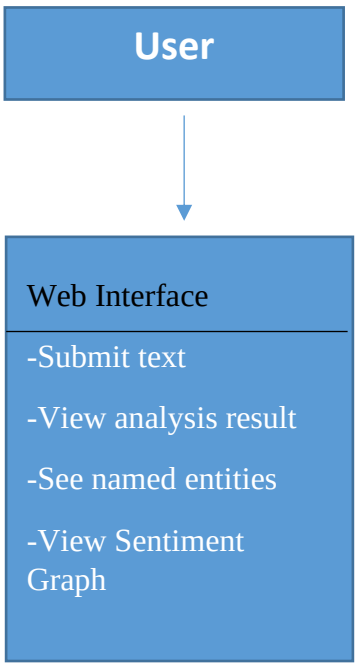
4. Technology Stack

Layer	Technology
Frontend	HTML
Backend	Python
NLP	NLTK, spaCy, TextBlob
Visualization	Matplotlib

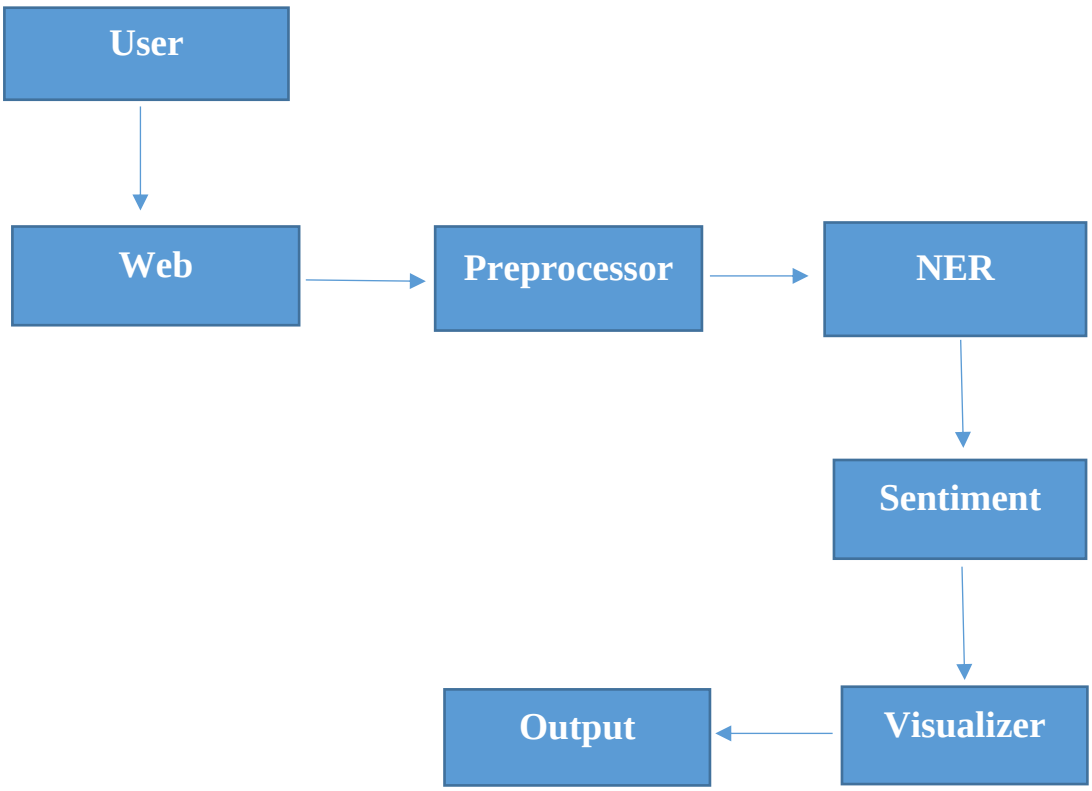
5. Testing Approach

Module	Test Focus
Preprocessor	Edge cases like punctuation only input, emojis
NER	Entity accuracy on structured inputs
Sentiment Analyzer	Positive/Negative sentence classification
Visualizer	Display clarity, empty data cases

6. Use case diagram



7. Sequence Diagram



8. Outcome of Design Phase

The Design Phase offered an in-depth and implementable system for the construction of the text enhancement system that relies on NLP. With defined modules, appropriate technologies, and user-driven planning, the system is now prepared for a systematic and effective implementation during the next phase of development.