Vidyavardhini's College of Engineering and Technology
Department of Artificial Intelligence & Data Science

Experiment No.1
Study various applications of NLP and Formulate the Problem Statement for Mini Project based on chosen real world NLP applications
Date of Performance:
Date of Submission:

Aim: Study various applications of NLP, including Named Entity Recognition and Entity Linkage, and their potential impact in the medical domain.

Objective: Implement and evaluate an NLP model for precise identification and linking of medical entities to facilitate comprehensive analysis of complex relationships in the medical dataset.

Theory:

1. **Healthcare**: Facilitating precise medical record management, improving patient diagnosis, and enabling personalized treatment plans.
2. **Biomedical Research**: Enhancing the understanding of disease patterns, accelerating drug discovery, and supporting the development of advanced treatment methodologies.
3. **Public Health:** Assisting in the analysis of health trends, enabling effective disease surveillance, and supporting the implementation of targeted public health interventions.
4. **Pharmaceutical Industry**: Streamlining drug safety monitoring, expediting pharmacovigilance processes, and aiding in the identification of adverse drug reactions.
5. **Health Informatics**: Enhancing the efficiency of electronic health record systems, facilitating seamless information retrieval, and enabling comprehensive health data analysis for research and policy development.

Conclusion:

Here is the summary of the pros and cons of each paper:

1. NELasso: Group-Sparse Modeling for Characterizing Relations among Named Entities in News Articles:

Pros:

- Introduces a novel system for learning semantic relations among named entities in news articles.
- Uses a sparse group lasso approach for capturing the context and dynamics of relations.
- Rigorous evaluation demonstrates the effectiveness of the system.

Cons:

- Relies on the assumption of reliable detection of named entities from text.
- The predefined word groups may not capture all possible nuances of relations.
- Scalability and efficiency issues are not addressed.

2. Strings and Things: A Semantic Search Engine for news quotes using Named Entity Recognition:

Pros:

- Presents a useful prototype of a semantic search engine for news quotes.
- Integration of spaCy's NER model with Elasticsearch enhances information retrieval.
- Detailed description of data sources, technologies, and evaluation scenarios.

Cons:

- Relies on the GDELT Global Quotation Graph, which may have limitations in accuracy and diversity.
- Uses a default spaCy NER model without discussing its optimality for the domain and language.

From the review of literature, it is evident that the application of NLP techniques in various domains, such as news analysis, entity recognition, and semantic search, is promising. However, there are notable challenges concerning data reliability, model optimization, scalability, and the need for comprehensive comparisons with existing methods and tools.