Megaminds IT Assignment

Name: Shivang Rustagi

Phone Number: 7006935960

Email id: shivangrustagi04@gmail.com

Course and Year: B.TECH(AI-DS) 2nd Year(4th SEM)

Title of the Project: Exploring Natural Language Processing in Model-To-Model Transformations

Name of the Organisation: Megaminds IT Services

Novelty of Work

The provided code demonstrates a natural language processing (NLP) approach to extracting various types of phrases and abbreviations from input text. The primary novelty lies in the implementation of functions that identify specific linguistic patterns and structures within sentences. The code accomplishes the following tasks:

1. **Relation Phrase Extraction:**

   - It identifies verb phrases, conjunctive phrases, disjunctive phrases, and noun phrases based on the part-of-speech (POS) tags of words in the input text.

   - Verb phrases are recognized when a noun (NN) is followed by a verb (VB).

   - Conjunctive phrases are identified when words like 'and' or 'or' connect phrases.

   - Disjunctive phrases are similar to conjunctive phrases but specifically for the word 'or.'

   - Noun phrases are recognized as sequences of consecutive nouns.

2. **Abbreviation Detection:**

   - It detects abbreviations using a regular expression pattern that matches uppercase letters and digits.

   - Detected abbreviations are added to a set for further analysis.

3. **Metric Calculation:**

   - The code includes a function to calculate precision, recall, F1 score, and accuracy, comparing the extracted phrases and abbreviations with expected values.

4. **User Interaction:**

   - The script allows user input, making it interactive and suitable for testing different sentences.

5. **Example Usage:**

- The code provides an example of usage where the user enters a sentence, and the extracted phrases and abbreviations are compared against expected values.

6. **Structured Output:**

   - The extracted phrases and abbreviations are organized into dictionaries and sets, facilitating easy access to specific types of linguistic elements.

7. **Modularity:**

   - The code is modular, with functions for different tasks, promoting reusability and maintainability.

8. **Educational Value:**

   - The code serves as an educational tool for understanding basic NLP concepts, POS tagging, and pattern recognition in natural language.

In summary, the novelty of this work lies in its ability to identify and extract specific linguistic elements from text, providing a foundation for more advanced natural language processing applications. The combination of functionality, modularity, and educational value makes it a valuable resource for both learning and practical implementation in NLP projects.