NLP Exp: 02

Date :

Post Lab Questions

A1 In Python (specifically NLTK), a corpus is used as a large, structured collection of text data (like movie reviews etc.) to train NLP models, perform statistical analysis & test algorithms for patterns, sentiment or frequency

A2 A good corpus is large, structured, representative and often annotated

A3 Tokenization, Stop Word Removal, Stemming, Lemmatization, Part of Speech (~~POS~~ (POS) Tagging, Frequency Distribution etc.

Conclusion :

In this experiment, we successfully analyzed unstructured text about Lionel Messi using Python & NLTK. Techniques like Tokenization and Stop-Word Removal filtered out noise, while Frequency Distribution identified core themes.

The o/p correctly highlights "Messi" & "Football" as top keywords proving the code effectively extracts insights from raw data.

Rainbow