

NLP Assignment 2

1. Mention and define the NLTK steps.

- Tokenization: Segmenting text into words or sentences.
- Stopword Removal: Eliminating common and less meaningful words.
- Stemming and Lemmatization: Reducing words to their base or root form.
- Part-of-Speech Tagging: Assigning grammatical tags to words.
- Named Entity Recognition: Identifying and categorizing named entities.
- Syntax and Parsing: Analysing the grammatical structure of sentences.
- Semantics: Understanding the meaning of words and sentences.
- Pragmatics: Analysing language in context and its intended meaning.

2. What different types of segmentation are there?

Sentence Segmentation: Dividing text into sentences.

Word Segmentation: Dividing text into words or tokens.

3. Explain the method of recognising named entities.

Utilizes machine learning models to identify and categorize entities in text, such as names of persons, organizations, locations, expressions of times, quantities, monetary values, percentages, etc.

4. Make a list of the elements of NLP.

- Tokenization
- Stopword Removal
- Stemming and Lemmatization
- Part-of-Speech Tagging
- Named Entity Recognition
- Syntax and Parsing
- Semantics
- Pragmatics

5. Give an example of pragmatic analysis.

Analysing an utterance like "It's cold in here" requires understanding the context and the intended meaning. If said while pointing at a broken window, it implies a different meaning compared to saying it with a smile in a comfortable room.

6. Explain the morphological and lexical analysis procedure.

Morphological Analysis:

Involves breaking words into morphemes (the smallest units of meaning) to understand their structure and form.

Lexical Analysis:

Involves analysing words (lexemes) to understand their meaning, usage, and how they relate to other words in the language.