

21CSE356T
Natural Language Processing
Unit-1
Question Bank

Q1: Define NLP and explain its different levels.

Q2: How can regex be applied to identify sentence in parsing natural language text?

Q3: Discuss about how the morphological analysis handles for highly inflected languages?

Q4: Explain the implications of tokenization for NLP tasks and discuss about various approaches?

Q5: Differentiate between stemming and lemmatization.

Q6: What are the different stemming and lemmatization approaches?

Q7: Explain the TF-IDF Feature Extraction concept. Compute the TF-IDF for the term "Computer" in a document where:

- i. The term frequency (TF) of "Computer" is 5.
- ii. The total number of documents is 10,000.
- iii. The term "NLP" appears in 100 documents.

Q8: How does POS tagging help in understanding the structure of a sentence? Provide an example of a sentence with POS tags. Discuss its various approaches.

Q9: Explain the following along with various approaches:

- a) NER
- b) POS Tagging
- c) N-Gram Modelling

Q10: Explain the different varieties of smoothing approach.