NLP Introduction and Text Preprocessing

Assignment Questions

NLP Introduction and Text Preprocessing

1. What is the primary goal of Natural Language Processing (NLP)?
2. What does "tokenization" refer to in text processing?
3. What is the difference between lemmatization and stemming?
4. What is the role of regular expressions (regex) in text processing?
5. What is Word2Vec and how does it represent words in a vector space?
6. How does frequency distribution help in text analysis?
7. Why is text normalization important in NLP?
8. What is the difference between sentence tokenization and word tokenization?
9. What are co-occurrence vectors in NLP?
10. What is the significance of lemmatization in improving NLP tasks?
11. What is the primary use of word embeddings in NLP?
12. What is an annotator in NLP?
13. What are the key steps in text processing before applying machine learning models?
14. What is the history of NLP and how has it evolved?
15. Why is sentence processing important in NLP?
16. How do word embeddings improve the understanding of language semantics in NLP?
17. How does the frequency distribution of words help in text classification?
18. What are the advantages of using regex in text cleaning?
19. What is the difference between word2vec and doc2vec?
20. Why is understanding text normalization important in NLP?
21. How does word count help in text analysis?
22. How does lemmatization help in NLP tasks like search engines and chatbots?
23. What is the purpose of using Doc2Vec in text processing?
24. What is the importance of sentence processing in NLP?
25. What is text normalization, and what are the common techniques used in it?
26. Why is word tokenization important in NLP?
27. How does sentence tokenization differ from word tokenization in NLP?
28. What is the primary purpose of text processing in NLP?
29. What are the key challenges in NLP?
30. How do co-occurrence vectors represent relationships between words?
31. What is the role of frequency distribution in text analysis?
32. What is the impact of word embeddings on NLP tasks?
33. What is the purpose of using lemmatization in text preprocessing?

Practical

1. How can you perform word tokenization using NLTK.
2. How can you perform sentence tokenization using NLTK.
3. How can you remove stopwords from a sentence?
4. How can you perform stemming on a word?
5. How can you perform lemmatization on a word?
6. How can you normalize a text by converting it to lowercase and removing punctuation?
7. How can you create a co-occurrence matrix for words in a corpus?
8. How can you apply a regular expression to extract all email addresses from a text?
9. How can you perform word embedding using Word2Vec?
10. How can you use Doc2Vec to embed documents?
11. How can you perform part-of-speech tagging?
12. How can you find the similarity between two sentences using cosine similarity?
13. How can you extract named entities from a sentence?
14. How can you split a large document into smaller chunks of text?
15. How can you calculate the TF-IDF (Term Frequency - Inverse Document Frequency) for a set of documents?
16. How can you apply tokenization, stopword removal, and stemming in one go?
17. How can you visualize the frequency distribution of words in a sentence?