Useful NLP Libraries and Networks Assignment

1. What is NLTK
answer: NLTK (Natural Language Toolkit) is a Python library used for NLP tasks like text processing, tokenization, stemming, and sentiment analysis.

2. What is SpaCy and how does it differ from NLTK
answer: SpaCy is an industrial-grade NLP library optimized for speed and production use, while NLTK is a research-focused toolkit with extensive educational resources.

3. What is the purpose of TextBlob in NLP
answer: TextBlob provides simple APIs for NLP tasks such as POS tagging, sentiment analysis, and noun phrase extraction.

4. What is Stanford NLP
answer: Stanford NLP is a comprehensive NLP library developed by Stanford University for tasks like dependency parsing and named entity recognition.

5. Explain what Recurrent Neural Networks (RNN) are
answer: RNNs are neural networks designed for sequential data, maintaining hidden states that capture information from previous steps in a sequence.

6. What is the main advantage of using LSTM over RNN
answer: LSTMs solve the vanishing gradient problem, allowing models to learn long-term dependencies effectively.

7. What are Bi-directional LSTMs, and how do they differ from standard LSTMs
answer: Bi-directional LSTMs process data in both forward and backward directions, capturing more contextual information.

8. What is the purpose of a Stacked LSTM
answer: Stacked LSTMs combine multiple LSTM layers to increase learning capacity and model complexity.

9. How does a GRU (Gated Recurrent Unit) differ from an LSTM
answer: GRUs have fewer parameters and gates, making them computationally simpler than LSTMs while achieving similar performance.

10. What are the key features of NLTK's tokenization process
answer: NLTK offers tokenization methods for splitting text into words or sentences, supporting regex-based and model-based tokenization.

11. How do you perform named entity recognition (NER) using SpaCy
answer: Run nlp(text) and extract entities using doc.ents in SpaCy to identify named entities.

12. What is Word2Vec and how does it represent words
answer: Word2Vec represents words as dense vectors by learning from their context in text, capturing semantic similarities.

13. Explain the difference between Bag of Words (BoW) and Word2Vec
answer: BoW counts word occurrences without capturing meaning, while Word2Vec generates context-based embeddings.

14. How does TextBlob handle sentiment analysis
answer: TextBlob assigns polarity and subjectivity scores to determine the sentiment of a text.

15. How would you implement text preprocessing using NLTK
answer: Use NLTK functions for tokenization, stopword removal, stemming, and lemmatization to preprocess text.

16. How do you train a custom NER model using SpaCy
answer: Create labeled training data, set up pipeline components, and fine-tune using the nlp.update() method.

17. What is the role of the attention mechanism in LSTMs and GRUs
answer: Attention allows models to focus on specific parts of an input sequence, improving performance in tasks like translation.

18. What is the difference between tokenization and lemmatization in NLP
answer: Tokenization splits text into units, while lemmatization reduces words to their base forms.

19. How do you perform text normalization in NLP
answer: Convert text to lowercase, remove special characters, and apply stemming or lemmatization.

20. What is the purpose of frequency distribution in NLP
answer: Frequency distribution helps identify how often words occur in a corpus, revealing patterns and insights.

21. What are co-occurrence vectors in NLP
answer: Co-occurrence vectors capture relationships between words by measuring how often they appear together in a context.

22. How is Word2Vec used to find the relationship between words
answer: Word2Vec embeds words in a vector space where semantic similarity is reflected by proximity.

23. How does a Bi-LSTM improve NLP tasks compared to a regular LSTM
answer: Bi-LSTM captures both past and future context, enhancing the model's understanding of sequences.

24. What is the difference between a GRU and an LSTM in terms of gate structures
answer: GRU uses two gates (reset and update), while LSTM uses three gates (input, output, forget).

25. How does Stanford NLP’s dependency parsing work
answer: It identifies grammatical relationships between words by constructing dependency trees.

26. How does tokenization affect downstream NLP tasks
answer: Accurate tokenization ensures proper input representation, improving model accuracy in subsequent tasks.

27. What are some common applications of NLP
answer: Sentiment analysis, machine translation, chatbots, text classification, and information retrieval.

28. What are stopwords and why are they removed in NLP
answer: Stopwords are frequent words like "the" and "is" that are removed to reduce noise in text analysis.

29. How can you implement word embeddings using Word2Vec in Python
answer: Use the gensim library to train and apply Word2Vec embeddings for text.

30. How does SpaCy handle lemmatization
answer: SpaCy uses the language model's morphological analysis to reduce words to their base forms.

31. What is the significance of RNNs in NLP tasks
answer: RNNs are significant for processing sequential data in tasks like text generation and speech recognition.

32. How does word embedding improve the performance of NLP models
answer: Word embeddings capture semantic relationships between words, enhancing task performance.

33. How does a Stacked LSTM differ from a single LSTM
answer: Stacked LSTM has multiple LSTM layers, increasing model complexity and learning capacity.

34. What are the key differences between RNN, LSTM, and GRU
answer: RNNs struggle with long-term dependencies; LSTMs solve this with gating mechanisms; GRUs are simplified LSTMs.

35. Why is the attention mechanism important in sequence-to-sequence models
answer: Attention helps models focus on the most relevant parts of input sequences, enhancing performance.