Chapter 1 NLP Test

1. What is the primary goal of Natural Language Processing (NLP)?

A) To teach machines how to speak like humans

B) To enable computers to understand, interpret, and generate human language

C) To process and analyze human language for tasks like translation and sentiment analysis

D) To replace human translators with AI

2. Which of the following is NOT a preprocessing step in NLP?

A) Tokenization

B) Stop-word removal

C) Image segmentation

D) Lemmatization

3. What is the difference between stemming and lemmatization?

A) Stemming adds prefixes, while lemmatization removes suffixes

B) Stemming trims word endings, while lemmatization uses grammar rules to find root words

C) Lemmatization is faster but less accurate than stemming

D) Stemming is used for speech recognition, while lemmatization is for text

4. Which technique counts word frequencies in a document?

A) TF-IDF

B) Bag-of-Words (BoW)

C) Word2Vec

D) POS Tagging

5. What does TF-IDF stand for?

A) Text Frequency – Inverse Document Frequency

B) Term Frequency – Inverse Document Frequency

C) Token Frequency – Inverse Data Frequency

D) Term Frequency – Independent Document Factor

6. Which of the following is a real-world application of NLP?

A) Weather forecasting

B) Sentiment analysis of customer reviews

C) Image recognition in self-driving cars

D) Predicting stock prices

7. What is the main challenge of polysemy in NLP?

A) Words with multiple spellings

B) Words with multiple meanings depending on context

C) Words that are rare in a dataset

D) Words that are grammatically incorrect

8. Which NLP task involves identifying names, locations, and organizations in text?

A) Sentiment analysis

B) Named Entity Recognition (NER)

C) Machine translation

D) Text summarization

9. What is the purpose of stop-word removal?

A) To correct spelling errors

B) To eliminate common words (e.g., "the," "is") that add little meaning

C) To translate text into another language

D) To improve image recognition accuracy

10. Which algorithm is commonly used for spam detection in emails?

A) Decision Trees

B) Naïve Bayes

C) K-Means Clustering

D) Support Vector Machines (SVM)

11. What is the key advantage of word embeddings (e.g., Word2Vec)?

A) They reduce computational power requirements

B) They capture semantic relationships between words

C) They work only for English language text

D) They require no preprocessing

12. Which of the following is a deep learning model used in NLP?

A) Linear Regression

B) Recurrent Neural Networks (RNNs)

C) K-Nearest Neighbors (KNN)

D) Random Forest

13. What is the purpose of transfer learning in NLP?

A) To train models from scratch for every new task

B) To fine-tune pre-trained models (e.g., BERT) for specific tasks

C) To replace human annotators in labeling data

D) To convert speech to text

14. Which dataset is commonly used for sentiment analysis?

A) CoNLL-2003

B) IMDb Reviews

C) WMT (Machine Translation)

D) Penn Treebank

15. What is the main challenge of bias in NLP models?

A) Models are too slow to train

B) Models may produce unfair or discriminatory outputs due to biased training data

C) Models cannot handle multilingual text

D) Models require too much labeled data

16. Which Python library is widely used for NLP tasks like tokenization and POS tagging?

A) TensorFlow

B) NLTK (Natural Language Toolkit)

C) PyTorch

D) Scikit-learn

17. What is the primary use of the CoNLL-2003 dataset?

A) Sentiment analysis

B) Named Entity Recognition (NER)

C) Machine translation

D) Text summarization

18. Which technique helps resolve ambiguity in NLP?

A) Removing all punctuation

B) Using contextual embeddings (e.g., BERT)

C) Ignoring rare words

D) Converting text to lowercase

19. What is the role of tokenization in NLP?

A) Translating text into another language

B) Splitting text into words or subwords (tokens)

C) Correcting grammatical errors

D) Generating summaries of long documents

20. Which of the following is an example of a pretrained language model?

A) K-Means

B) GPT-3

C) Random Forest

D) Logistic Regression

21. What is the purpose of the Bag-of-Words (BoW) model?

A) To translate text between languages

B) To represent text as word frequencies

C) To correct spelling mistakes

D) To generate poetry

22. Which NLP task involves predicting the next word in a sentence?

A) Sentiment analysis

B) Language modeling

C) Named Entity Recognition (NER)

D) Speech recognition

23. What is a major limitation of rule-based NLP systems?

A) They require no training data

B) They struggle with ambiguity and variability in human language

C) They are too fast for real-time applications

D) They only work for structured databases

24. Which of the following is NOT a real-world application of NLP?

A) Chatbots for customer service

B) Google Translate

C) Diagnosing diseases from X-ray images

D) Detecting hate speech on social media

25. What is the main advantage of using NLTK in NLP projects?

A) It supports only deep learning models

B) It provides easy-to-use tools for text preprocessing and analysis

C) It requires no programming knowledge

D) It is designed exclusively for speech recognition