NLP Chapter 2

1. What is the primary goal of text cleaning in NLP?

A) To translate text into another language

B) To standardize text by removing noise (e.g., punctuation, case variations)

C) To generate summaries of long documents

D) To convert speech to text

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

A) Tokenization

B) Image segmentation

C) Stop-word removal

D) Lemmatization

3. What does tokenization achieve in NLP?

A) Translates text into numerical vectors

B) Splits text into words or sentences (tokens)

C) Identifies the sentiment of a sentence

D) Corrects grammatical errors

4. Which technique reduces "running," "ran," and "runner" to the root word "run"?

A) Tokenization

B) Stemming

C) POS tagging

D) Named Entity Recognition (NER)

5. What is the key difference between stemming and lemmatization?

A) Stemming uses grammar rules, while lemmatization does not

B) Stemming cuts word endings, while lemmatization uses dictionaries to find root words

C) Lemmatization is faster but less accurate

D) Stemming only works for verbs

6. Which Python library is commonly used for tokenization and POS tagging?

A) TensorFlow

B) NLTK (Natural Language Toolkit)

C) PyTorch

D) Pandas

7. What does POS tagging stand for?

A) Part-of-Speech tagging

B) Identifying grammatical roles of words (e.g., noun, verb)

C) Both A and B

D) Converting text to lowercase

8. In the sentence "Delhi is the capital of India," what is the POS tag for "capital"?

A) VB (Verb)

B) NN (Noun)

C) JJ (Adjective)

D) IN (Preposition)

9. What is Named Entity Recognition (NER) used for?

A) Classifying text sentiment

B) Identifying real-world entities like names, dates, and locations

C) Translating languages

D) Splitting text into sentences

10. Which entity type would "15th August" be classified as in NER?

A) Person

B) Date

C) Organization

D) Location

11. 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

D) To identify named entities

12. Which of the following is a real-world application of NER?

A) Spam detection

B) Extracting names and locations from news articles

C) Speech recognition

D) Image classification

13. What is chunking in NLP?

A) Translating text into another language

B) Grouping words into meaningful phrases (e.g., noun phrases)

C) Correcting grammatical errors

D) Generating word embeddings

14. Which preprocessing step converts "Hello!" to "hello"?

A) Tokenization

B) Lowercasing

C) Stemming

D) Lemmatization

15. What does TF-IDF stand for?

A) Text Frequency – Inverse Data Frequency

B) Term Frequency – Inverse Document Frequency

C) Token Frequency – Independent Document Factor

D) Term Frequency – Integrated Document Frequency

16. Which technique converts text into numerical vectors for machine learning?

A) Tokenization

B) Vectorization (e.g., Bag-of-Words, Word2Vec)

C) POS tagging

D) Chunking

17. What is the output of word tokenization for "NLP is fun"?

A) ["NLP", "is fun"]

B) ["NLP", "is", "fun"]

C) ["NLP is fun"]

D) ["N", "L", "P", "i", "s", "f", "u", "n"]

18. Which of the following is a rule-based approach to chunking?

A) Defining patterns like "adjective + noun = noun phrase"

B) Training a neural network on labeled data

C) Using Word2Vec embeddings

D) Removing stop words

19. What is the first step in syntactic analysis?

A) Stemming

B) Tokenization

C) Sentiment analysis

D) Translation

20. Which NLP task involves identifying "who," "where," and "when" in text?

A) Sentiment analysis

B) Named Entity Recognition (NER)

C) Text summarization

D) Language detection

21. What is the primary challenge of stemming?

A) It requires labeled data

B) It may produce incorrect root forms (e.g., "universities" → "univers")

C) It only works for English

D) It cannot handle verbs

22. Which preprocessing step removes punctuation like commas and periods?

A) Lemmatization

B) Text cleaning

C) POS tagging

D) Chunking

23. What is the purpose of dependency parsing?

A) To translate languages

B) To analyze grammatical relationships between words in a sentence

C) To classify text sentiment

D) To detect spam emails

24. Which of the following is an example of a noun phrase (NP) in chunking?

A) "is running"

B) "The quick brown fox"

C) "under the table"

D) "very quickly"

25. What is the output of NER for "Apple launched iPhone in 2023"?

A) Organization: Apple; Product: iPhone; Date: 2023

B) Person: Apple; Location: iPhone; Date: 2023

C) Organization: Apple; Location: iPhone; Date: 2023

D) Person: Apple; Product: iPhone; Year: 2023