1. What is Natural Language Processing (NLP) a subfield of?

A. Computer Graphics

B. Artificial Intelligence

C. Database Management

D. Web Development

2. What is the primary goal of NLP?

A. Generate random text

B. Enable machines to understand human language

C. Convert text to images

D. Process numerical data only

3. NLP aims to construct technology for useful functions, diverging from which field that applies computer science to learn language rules?

A. Linguistics

B. Computational Biology

C. Computational Linguistics

D. Cognitive Science

4. Which two overlapping subfields comprise NLP?

A. Natural Language Recognition and Natural Language Display

B. Natural Language Understanding and Natural Language Generation

C. Natural Language Input and Natural Language Output

D. Natural Language Processing and Speech Recognition

5. What does Natural Language Understanding (NLU) primarily involve?

A. Text generation

B. Semantic analysis of text

C. Sound to text conversion

D. Analyzing spoken words

6. What kind of data does NLP require systems to have?

A. Small, unstructured data

B. A corpus (vast linguistic data)

C. Numerical data only

D. Pre-processed data only

7. Before any model can learn from language, what must happen to text data?

A. Stored in raw format

B. Cleaned, simplified, machine-readable

C. Converted to image

D. Manually labeled

8. What is the process of splitting text into words or parts of words called?

A. Lemmatization

B. Stemming

C. Tokenization

D. Segmentation

9. What is the purpose of Stemming and Lemmatization in data preprocessing?

A. Increase word length

B. Reduce words to root form

C. Identify grammar errors

D. Remove stop words

10. What is the goal of Stop Word Removal?

A. Increase data noise

B. Strip frequent, low-meaning words

C. Add common words

D. Convert words to numbers

11. What are features in the context of NLP?

A. Raw text data

B. Numerical attributes extracted from text

C. Random numbers for words

D. Grammatical rules

12. Which feature extraction technique counts each word's frequency in a document?

A. TF-IDF

B. Word Embeddings

C. Bag-of-Words (BoW)

D. GLOVE

13. What do Word Embeddings (like Word2Vec, GloVe) convert words into?

A. Images

B. Vectors (collections of numbers)

C. Boolean values

D. Alphabetic sequences

14. What is the primary function of Language Models?

A. Summarize long texts

B. Predict next word/phrase

C. Perform sentiment analysis

D. Translate languages

15. Pretrained models (BERT, GPT-3) are trained on what kind of data?

A. Small, custom datasets

B. Enormous text collections (e.g., Wikipedia)

C. Numerical tables only

D. User-generated content only

16. NLP plays a fundamental role in comprehending enormous quantities of which type of data?

A. Structured numerical data

B. Unstructured text data

C. Image data only

D. Audio data only

17. Which of the following is NOT explicitly mentioned as a real-world application of NLP in Chapter 1?

A. Virtual Assistants

B. Language Translation

C. Spreadsheet Calculation

D. Sentiment Analysis

18. What is a fundamental challenge in NLP due to human languages?

A. Computational Efficiency

B. Data Redundancy

C. Complexity and Diversity

D. Model Interpretability

19. What can significantly affect NLP model performance if training data is incomplete, biased, or inaccurately labeled?

A. Model speed

B. Model accuracy/reliability

C. Model complexity

D. Model interpretability

20. What is a major limitation for deep learning NLP models regarding computational power?

A. Only basic CPUs

B. Substantial GPU/TPU power

C. Any consumer device

D. No specialized hardware

21. What is a natural characteristic of human language where the same phrase can have different meanings based on context?

A. Consistency

B. Clarity

C. Ambiguity

D. Redundancy

22. What is the primary concern when NLP models learn from non-representative or prejudiced data?

A. Decreased computational cost

B. Increased transparency

C. Bias in outputs

D. Faster training times

23. What is Polysemy in NLP?

A. One word, one meaning

B. One word, multiple meanings

C. Multiple words, same meaning

D. Words with no meaning

24. Which technique involves generating new data by altering existing data (e.g., paraphrasing)?

A. Data Synthesis

B. Crowdsourcing

C. Data Augmentation

D. Data Cleaning

25. What method for NLP data involves employing artificial techniques to create data that simulates actual language use?

A. Data Augmentation

B. Data Synthesis

C. Crowdsourcing

D. Data Validation

26. What type of learning trains a model on labeled datasets?

A. Unsupervised learning

B. Supervised learning

C. Reinforcement learning

D. Semi-supervised learning

27. What type of learning involves data that is not explicitly labeled, with the goal of discovering underlying structures?

A. Supervised learning

B. Unsupervised learning

C. Reinforcement learning

D. Transfer learning

28. Which classic machine learning algorithm makes a "naïve" assumption that features (words) are conditionally independent?

A. Support Vector Machine

B. Logistic Regression

C. Naïve Bayes classifier

D. Decision Tree

29. Which NLP algorithm type has revolutionized how machines understand language by learning hierarchical features automatically from raw text?

A. Traditional statistical models

B. Deep learning models (neural networks)

C. Rule-based systems

D. Hidden Markov Models

30. What was a key breakthrough that powered the success of contemporary NLP models, representing words in high-dimensional, continuous vector spaces?

A. One-hot encoding

B. Word embeddings

C. Manual feature engineering

D. Bag-of-Words

31. What is NLTK (Natural Language Toolkit)?

A. A web framework

B. An open-source Python library for NLP applications

C. A database system

D. A hardware component

32. NLTK provides interfaces to more than 50 of what?

A. Programming languages

B. Corpora and lexical resources

C. Operating systems

D. Hardware devices

33. What is the primary role of NLTK in an educational setting?

A. To deploy production-grade solutions

B. To provide a practical tool for learning fundamental NLP concepts

C. To replace deep learning frameworks

D. To perform complex semantic reasoning only

34. Which command is used to install NLTK using Python's package manager?

A. `install nltk`

B. `python -m nltk install`

C. `pip install nltk`

D. `nltk.setup()`

35. What is the purpose of `nltk.download()` after installing NLTK?

A. To compile the library

B. To download necessary datasets and models

C. To update Python

D. To start the training process

36. Why are NLP datasets important for training machines?

A. They make models slower.

B. They serve as learning content for machines.

C. They reduce the need for AI.

D. They only provide rules for machines.

37. What does linguistic diversity in datasets provide?

A. Uniform language patterns

B. Various languages, dialects, colloquial language

C. Limited contextual usage

D. Simplified grammar rules

38. The IMDb Movie Reviews Dataset is primarily used for which NLP task?

A. Named Entity Recognition

B. Machine Translation

C. Sentiment Analysis

D. Text Summarization

39. The CoNLL-2003 Dataset is a benchmarking dataset for which NLP task?

A. Sentiment Analysis

B. Named Entity Recognition (NER)

C. Language Modeling

D. Question Answering

40. What kind of corpus is the Wikipedia Corpus, heavily applied for unsupervised learning and pretraining large language models?

A. Small, domain-specific

B. High-quality, multilingual

C. Structured, numerical

D. Time-series

41. What is the main purpose of Text Content Processing?

A. To store raw text data

B. To convert text into a machine-understandable format

C. To display text on a screen

D. To translate text between languages

42. What happens in the "Text Cleaning" step of text processing?

A. Splitting text into words

B. Converting text to lowercase, removing punctuation/spaces

C. Identifying grammatical roles

D. Converting words to numbers

43. What is the main goal of Syntactic Analysis (parsing)?

A. To extract keywords

B. To analyze sentence structure according to grammar

C. To identify emotional tone

D. To convert text into numerical format

44. What is Chunking also known as?

A. Deep parsing

B. Shallow parsing

C. Full parsing

D. Semantic parsing

45. What does POS Tagging identify in a sentence?

A. The word's meaning

B. The word's part of speech (noun, verb, etc.)

C. The word's numerical value

D. The word's sentiment

46. What is Named Entity Recognition (NER)?

A. Identifying grammatical errors

B. Classifying text into general topics

C. Identifying and classifying specific entities (people, places, orgs) in text

D. Summarizing documents

47. What is the primary focus of Sentiment Analysis?

A. Identifying factual information

B. Extracting key entities

C. Identifying the emotional tone of text

D. Summarizing documents

48. What approach to sentiment analysis relies on a predefined list of words with sentiment scores?

A. Machine learning-based

B. Hybrid

C. Lexicon-based

D. Deep learning-based

49. What does a sentiment score quantify?

A. The length of the text

B. The emotional tone of a piece of text

C. The number of keywords

D. The grammatical correctness

50. What type of text classification categorizes text into one of several mutually exclusive classes?

A. Binary Classification

B. Multi-Class Classification

C. Multi-Label Classification

D. Hierarchical Classification