

Unit 6: Natural Language Processing

Section A: Basics of NLP

1. Natural Language refers to:
 - a) Programming languages
 - b) Human languages like English, Hindi, Spanish
 - c) Machine codes
 - d) Mathematical formulas

Answer: b) Human languages like English, Hindi, Spanish

2. Which of the following is a feature of natural languages?
 - a) Governed by syntax, lexicon, and semantics
 - b) Always fixed and unchanging
 - c) Cannot be redundant
 - d) Only exists in written form

Answer: a) Governed by syntax, lexicon, and semantics

3. Why is NLP important?
 - a) Converts natural language into binary for computers
 - b) Helps humans write programming code
 - c) Replaces computer languages completely
 - d) Eliminates need for translation

Answer: a) Converts natural language into binary for computers

4. NLP is a subfield of:
 - a) Robotics
 - b) Artificial Intelligence
 - c) Data Mining
 - d) Operating Systems

Answer: b) Artificial Intelligence

5. Which of these is NOT a computer language?
 - a) Python
 - b) Java
 - c) Spanish
 - d) C++

Answer: c) Spanish

Section B: Applications of NLP

6. Auto-generated captions on YouTube use:

- a) Image recognition
- b) NLP speech-to-text conversion
- c) Cryptography
- d) Computer Vision

Answer: b) NLP speech-to-text conversion

7. Google Translate is an example of:

- a) Keyword extraction
- b) Sentiment analysis
- c) Language translation
- d) Text summarization

Answer: c) Language translation

8. Siri and Alexa use NLP mainly for:

- a) Speech recognition and task execution
- b) Image analysis
- c) Data mining
- d) Programming

Answer: a) Speech recognition and task execution

9. Which NLP tool is used to detect customer opinions as positive/negative/neutral?

- a) Sentiment Analysis
- b) Machine Translation
- c) Text Summarization
- d) Tokenization

Answer: a) Sentiment Analysis

10. Text classification helps in:

- a) Dividing text into sentences
- b) Categorizing documents into predefined groups
- c) Translating into other languages
- d) Generating captions

Answer: b) Categorizing documents into predefined groups

Section C: Stages of NLP

11. Which is the first stage of NLP?

- a) Lexical Analysis
- b) Semantic Analysis

- c) Pragmatic Analysis
- d) Discourse Integration

Answer: a) Lexical Analysis

12. Syntactic analysis deals with:

- a) Word meanings
- b) Grammar and sentence structure
- c) Document frequency
- d) Contextual use

Answer: b) Grammar and sentence structure

13. Semantic analysis checks:

- a) Sentence grammar
- b) Sentence meaningfulness
- c) Sentence position in text
- d) Sentence length

Answer: b) Sentence meaningfulness

14. “Hot ice cream” is rejected in which stage?

- a) Lexical Analysis
- b) Syntactic Analysis
- c) Semantic Analysis
- d) Pragmatic Analysis

Answer: c) Semantic Analysis

15. Pragmatic analysis focuses on:

- a) Literal meanings only
- b) Real-world context and intended meaning
- c) Word frequency counts
- d) Grammar correction

Answer: b) Real-world context and intended meaning

16. Discourse integration ensures:

- a) Randomized sentences
- b) Logical connection between sentences
- c) Word frequency analysis
- d) Lowercasing text

Answer: b) Logical connection between sentences

17. Lexicon in NLP means:

- a) Document frequency

- b) Grammar rules
- c) Collection of words/phrases of a language
- d) Token frequency

Answer: c) Collection of words/phrases of a language

18. Which stage forms sentence relationships?

- a) Semantic Analysis
- b) Discourse Integration
- c) Lexical Analysis
- d) Tokenization

Answer: b) Discourse Integration

19. Which stage discards literal word meaning and uses intended meaning?

- a) Lexical
- b) Semantic
- c) Pragmatic
- d) Syntactic

Answer: c) Pragmatic

20. Parsing is another name for:

- a) Syntactic Analysis
- b) Semantic Analysis
- c) Stemming
- d) Lemmatization

Answer: a) Syntactic Analysis

Section D: Text Preprocessing

21. Sentence segmentation means:

- a) Splitting corpus into sentences
- b) Splitting words into characters
- c) Splitting text into numbers
- d) Splitting text into stop words

Answer: a) Splitting corpus into sentences

22. Tokenization divides text into:

- a) Numbers
- b) Tokens (words, numbers, special characters)
- c) Phrases only

d) Paragraphs only

Answer: b) Tokens (words, numbers, special characters)

23. Words like *is*, *the*, *and* are called:

- a) Rare words
- b) Stop words
- c) Meaningful words
- d) Punctuation

Answer: b) Stop words

24. Converting all text to lowercase ensures:

- a) More words in vocabulary
- b) Words treated uniformly
- c) Removal of grammar
- d) Faster translation

Answer: b) Words treated uniformly

25. Stemming reduces words to:

- a) Longer forms
- b) Base/root form (may not be meaningful)
- c) Synonyms
- d) Lemmas only

Answer: b) Base/root form (may not be meaningful)

26. Lemmatization differs from stemming because it:

- a) Produces only shorter words
- b) Produces meaningful words (lemmas)
- c) Ignores grammar
- d) Is always faster

Answer: b) Produces meaningful words (lemmas)

27. In stemming, “studies” → ?

- a) study
- b) studi
- c) student
- d) studied

Answer: b) studi

28. In lemmatization, “studies” → ?

- a) study
- b) studi

- c) student
- d) studied

Answer: a) study

29. Removing special characters is part of:

- a) Syntactic analysis
- b) Text normalization
- c) Lexical analysis
- d) Semantic analysis

Answer: b) Text normalization

30. Corpus refers to:

- a) A collection of documents/text
- b) A single sentence
- c) A keyword
- d) A programming function

Answer: a) A collection of documents/text

Section E: Bag of Words & TF-IDF

31. Bag of Words provides:

- a) Sentence grammar rules
- b) Vocabulary & frequency of words
- c) Synonyms and antonyms
- d) Only rare words

Answer: b) Vocabulary & frequency of words

32. Why is it called a “bag” of words?

- a) Sequence of words doesn't matter
- b) Sequence is most important
- c) Words are hidden
- d) Grammar is checked

Answer: a) Sequence of words doesn't matter

33. Bag of Words gives:

- a) Word counts & unique vocabulary
- b) TFIDF values
- c) Grammar correctness
- d) Contextual meanings

Answer: a) Word counts & unique vocabulary

34. Term Frequency (TF) means:

- a) Number of documents word appears in
- b) Frequency of a word in a document
- c) Ratio of rare words
- d) Total frequency of stop words

Answer: b) Frequency of a word in a document

35. Document Frequency means:

- a) How often a word occurs across all documents
- b) How many characters per sentence
- c) How many stop words removed
- d) How many synonyms used

Answer: a) How often a word occurs across all documents

36. Inverse Document Frequency (IDF) formula:

- a) $\text{Total docs} \div \text{Docs containing word}$
- b) $\text{Docs containing word} \div \text{Total docs}$
- c) $\text{Word frequency} \div \text{Document length}$
- d) $\text{Word length} \div \text{Total tokens}$

Answer: a) $\text{Total docs} \div \text{Docs containing word}$

37. TFIDF is calculated as:

- a) $\text{TF} \times \text{IDF}$
- b) $\text{TF} \div \text{IDF}$
- c) $\text{TF} - \text{IDF}$
- d) $\text{TF} + \text{IDF}$

Answer: a) $\text{TF} \times \text{IDF}$

38. High TFIDF value means:

- a) Common word in all documents
- b) Stop word
- c) Word important for one document but not common in all
- d) Word always ignored

Answer: c) Word important for one document but not common in all

39. Stop words usually have:

- a) High TFIDF
- b) Zero/negligible TFIDF
- c) Random TFIDF

d) Maximum IDF

Answer: b) Zero/negligible TFIDF

40. Rare words usually have:

a) High value in TFIDF

b) No importance

c) Always removed

d) Same as stop words

Answer: a) High value in TFIDF

Section F: Chatbots & Tools

41. A chatbot is designed to:

a) Simulate human conversation

b) Classify images

c) Process numbers

d) Encrypt data

Answer: a) Simulate human conversation

42. Eliza, Mitsuku, and Cleverbot are examples of:

a) Voice assistants

b) Chatbots

c) Corpus datasets

d) Image processors

Answer: b) Chatbots

43. Script-bot chatbots are:

a) AI-powered with learning ability

b) Traditional rule-based chatbots

c) Always voice-based

d) Used only in translation

Answer: b) Traditional rule-based chatbots

44. Smart-bot chatbots are:

a) Always rule-based

b) AI-powered with knowledge base

c) Image classifiers

d) Syntax checkers

Answer: b) AI-powered with knowledge base

45. Example of **Code-based NLP tool**:

- a) MonkeyLearn
- b) Orange Data Mining
- c) NLTK (Python)
- d) Lobe.ai

Answer: c) NLTK (Python)

46. Example of **No-code NLP tool**:

- a) SpaCy
- b) NLTK
- c) MonkeyLearn
- d) C++

Answer: c) MonkeyLearn

47. Orange Data Mining is a:

- a) No-code NLP tool
- b) Rule-based chatbot
- c) Text normalization technique
- d) Programming language

Answer: a) No-code NLP tool

48. MonkeyLearn supports:

- a) Text classification, sentiment analysis
- b) Image detection
- c) Speech recognition
- d) Data encryption

Answer: a) Text classification, sentiment analysis

49. SpaCy is mainly used for:

- a) Tokenization, POS tagging, named entity recognition
- b) Image recognition
- c) Grammar correction
- d) Audio transcription

Answer: a) Tokenization, POS tagging, named entity recognition

50. Which NLP application helps analyze customer feedback emotions?

- a) Sentiment Analysis
- b) Text Classification
- c) Keyword Extraction

d) Translation

Answer: a) Sentiment Analysis