

Natural Language Processing (NLP) Interview Questions

Below are few Frequently Asked NLP Interview Questions

Q1. What is NLP?

Natural Language Processing or **NLP** is an automated way to understand or analyze the natural languages and extract required information from such data by applying machine learning Algorithms.

Q2. List some Components of NLP?

Below are the few major components of NLP.

- **Entity extraction:** It involves segmenting a sentence to identify and extract entities, such as a person (real or fictional), organization, geographies, events, etc.
- **Syntactic analysis:** It refers to the proper ordering of words.
- **Pragmatic analysis:** Pragmatic Analysis is part of the process of extracting information from text.

Q3. List some areas of NLP?

Natural Language Processing can be used for

- Semantic Analysis
- Automatic summarization
- Text classification
- Question Answering

Some real-life example of NLP is **IOS Siri**, the **Google assistant**, **Amazon echo**.

Q4. Define the NLP Terminology?

NLP Terminology is based on the following factors:

- **Weights and Vectors:** TF-IDF, length(TF-IDF, doc), Word Vectors, Google Word Vectors
- **Text Structure:** Part-Of-Speech Tagging, Head of sentence, Named entities
- **Sentiment Analysis:** Sentiment Dictionary, Sentiment Entities, Sentiment Features
- **Text Classification:** Supervised Learning, Train Set, Dev(=Validation) Set, Test Set, Text Features, LDA.
- **Machine Reading:** Entity Extraction, Entity Linking,dbpedia, FRED (lib) / Pikes

Q5. What is the significance of TF-IDF?

tf-idf or **TFIDF** stands for **term** frequency–inverse **document frequency**. In information retrieval TFIDF is a numerical statistic that is intended to reflect how important a word is to a document in a collection or in the collection of a set.

Q6. What is part of speech (POS) tagging?

According to The Stanford Natural Language Processing Group :

A **Part-Of-Speech Tagger** (POS Tagger) is a piece of software that reads text in some language and assigns parts of speech to each word (and other token), such as noun, verb, adjective, etc.

PoS taggers use an algorithm to label terms in text bodies. These taggers make more complex categories than those defined as basic PoS, with tags such as “noun-plural” or even more complex labels. Part-of-speech categorization is taught to school-age children in English grammar, where children perform basic PoS tagging as part of their education.

Q7. What is Lemmatization in NLP?

Lemmatization generally means to do the things properly with the use of vocabulary and morphological analysis of words. In this process, the endings of the words are removed to return the base word, which is also known as Lemma.

Example: boy's = boy, cars= car, colors= color.

So, the main attempt of Lemmatization as well as of stemming is to identify and return the root words of the sentence to explore various additional information.

Q8. What is stemming in NLP?

Q9. What is tokenization in NLP?

Natural Language Processing aims to program computers to process large amounts of natural language data. **Tokenization in NLP** means the method of dividing the text into various tokens. You can think of a token in the form of the word. Just like a word forms into a sentence. It is an important step in NLP to slit the text into minimal units.

Q10. What is latent semantic indexing and where can it be applied?

Latent Semantic Indexing (LSI) also called Latent semantic analysis is a mathematical method that was developed so that the accuracy of retrieving information can be improved. It helps in finding out the hidden(latent) relationship between the words(semantics) by producing a set of various concepts related to the terms of a sentence to improve the information understanding. The technique used for the purpose is

called Singular value decomposition. It is generally useful for working on small sets of static documents.

Q11. What is dependency parsing?

Q12. Differentiate regular grammar and regular expression.

Regular expression is a sequence of characters that define a search pattern, mainly for use in pattern matching with strings, or string matching. It includes the following elements:

Example: A and B are regular expressions then

- The regular expression is A. B (concatenation)
- The regular expression (alternation) is A | B
- The regular expression (Kleene Star) is A*

Regular Grammars

There are 4 tuples in Regular Grammars ($N, \Sigma, P, S \in N$). In this formula, N stands for the non-terminals' sets, Σ means the set of terminals, P is the set of productions to change the start symbol, P has its productions from one of the types and lastly S is the start non-terminal.

Q13. List some tools for training NLP models?

Q14. Describe dependency parsing?

Q15. Explain Named entity recognition (NER)?

Named-entity recognition (NER) is the method of extracting information. It arranges and classifies named entity in the unstructured text in different categories like locations, time expressions, organizations, percentages, and monetary values. NER allows the users to properly understand the subject of the text.

Q16. What is NLTK?

Q17. List some OpenSource Libraries for NLP?

Q18. What is the difference between NLP and NLU?

Difference between NLP and NLU are

Natural Language Processing

NLP is the system that works simultaneously to manage end-to-end conversations between computers and humans.

NLP is related to both humans and machines.

Natural Language Understanding

NLU helps to solve the complicated challenges of Artificial Intelligence.

NLU allows converting the unstructured inputs into structured text for easy understanding by the machines.

Q19. What is the difference between NLP and CI(Conversational Interfaces)?

Difference between NLP and CI(Conversational Interfaces)

Natural Language Processing

NLP is a kind of artificial intelligence technology that allows identifying, understanding and interpreting the request of users in the form of language.

NLP aims to make users understand a particular concept.

Conversational Interfaces

CI is a user interface that mixes voice, chat and another natural language with images, videos or buttons.

Conversational Interface provides only what the users need and not more than that.

Q20. List few differences between AI, Machine Learning, and NLP?

Differences between AI, Machine Learning, and NLP

Artificial Intelligence

It is the technique to create smarter machines

AI includes human intervention

Artificial intelligence is a broader concept than Machine Learning

Machine Learning

Machine Learning is the term used for systems that learn from experience.

Machine Learning purely involves the working of computers and no human intervention.

ML is a narrow concept and is a subset of AI.

Natural Language Processing

This is the set of system that has the ability to understand the language

NLP links both computer and human languages.

Q21. Explain the Masked Language Model?

Masked language modelling is the process in which the output is taken from the corrupted input. This model helps the learners to master the deep representations in downstream tasks. You can predict a word from the other words of the sentence using this model.

Q22. What is latent semantic indexing? Where it is applied.

Q23. What is pragmatic analysis in NLP?

Pragmatic Analysis: It deals with outside word knowledge, which means knowledge that is external to the documents and/or queries. Pragmatics analysis that focuses on what was described is reinterpreted by what it actually meant, deriving the various aspects of language that require real-world knowledge.

Q24. Explain dependency parsing in NLP?

Dependency Parsing is also known as Syntactic Parsing. It is the task of recognizing a sentence and assigning a syntactic structure to it. The most widely used syntactic structure is the parse tree which can be generated using some parsing algorithms. These parse trees are useful in various applications like grammar checking or more importantly it plays a critical role in the semantic analysis stage.

Q25. What is pragmatic ambiguity in NLP?

Pragmatic Ambiguity can be defined as the words which have multiple interpretations. Pragmatic Ambiguity arises when the meaning of words of a sentence is not specific; it concludes different meanings. There are various sentences in which the proper sense is not understood due to the grammar formation of the sentence; this multi interpretation of the sentence gives rise to ambiguity.

For example- "do you want a cup of coffee", the given the word is either an informative question or a formal offer to make a cup coffee.

Q26. What is perplexity in NLP?

The word "perplexed" means "puzzled" or "confused", thus Perplexity in general means the inability to tackle something complicated and a problem that is not specified. Therefore, Perplexity in NLP is a way to determine the extent of uncertainty in predicting some text.

In NLP, perplexity is a way of evaluating language models. Perplexity can be high and low; Low perplexity is ethical because the inability to deal with any complicated problem is less while high perplexity is terrible because the failure to deal with a complicated is high.

Q27. What is ngram in NLP?

N-gram in NLP is simply a sequence of n words, and we also conclude the sentences which appeared more frequently, for example, let us consider the progression of these three words:

- New York (2 gram)
- The Golden Compass (3 gram)
- She was there in the hotel (4 gram)

Now from the above sequence, we can easily conclude that sentence (a) appeared more frequently than the other two sentences, and the last sentence(c) is not seen that often. Now if we assign probability in the occurrence of an n-gram, then it will be advantageous. It would help in making next-word predictions and in spelling error corrections.