NLP Interview Questions for Freshers

1. What is Naive Bayes algorithm, When we can use this algorithm in NLP?

[Naive Bayes algorithm](#) is a collection of classifiers which works on the principles of the Bayes' theorem. This series of NLP model forms a family of algorithms that can be used for a wide range of classification tasks including sentiment prediction, filtering of spam, classifying documents and more.

Naive Bayes algorithm converges faster and requires less training data. Compared to other discriminative models like logistic regression, Naive Bayes model it takes lesser time to train. This algorithm is perfect for use while working with multiple classes and text classification where the data is dynamic and changes frequently.

2. Explain Dependency Parsing in NLP?

Dependency Parsing, also known as Syntactic parsing in NLP is a process of assigning syntactic structure to a sentence and identifying its dependency parses. This process is crucial to understand the correlations between the "head" words in the syntactic structure.

The process of dependency parsing can be a little complex considering how any sentence can have more than one dependency parses. Multiple parse trees are known as ambiguities. Dependency parsing needs to resolve these ambiguities in order to effectively assign a syntactic structure to a sentence.

Dependency parsing can be used in the semantic analysis of a sentence apart from the syntactic structuring.

3. What is text Summarization?

[Text summarization](#) is the process of shortening a long piece of text with its meaning and effect intact. Text summarization intends to create a summary of any given piece of text and outlines the main points of the document. This technique has improved in recent times and is capable of summarizing volumes of text successfully.

Text summarization has proved to a blessing since machines can summarise large volumes of text in no time which would otherwise be really time-consuming. There are two types of text summarization:

- Extraction-based summarization

- Abstraction-based summarization

4. What is NLTK? How is it different from Spacy?

NLTK or Natural Language Toolkit is a series of libraries and programs that are used for symbolic and statistical natural language processing. This toolkit contains some of the most powerful libraries that can work on different ML techniques to break down and understand human language. NLTK is used for Lemmatization, Punctuation, Character count, Tokenization, and Stemming. The difference between NLTK and Spacey are as follows:

- While NLTK has a collection of programs to choose from, Spacey contains only the best-suited algorithm for a problem in its toolkit
- NLTK supports a wider range of languages compared to Spacey (Spacey supports only 7 languages)
- While Spacey has an object-oriented library, NLTK has a string processing library
- Spacey can support word vectors while NLTK cannot

5. What is information extraction?

Information extraction in the context of Natural Language Processing refers to the technique of extracting structured information automatically from unstructured sources to ascribe meaning to it. This can include extracting information regarding attributes of entities, relationship between different entities and more. The various models of information extraction includes:

- Tagger Module
- Relation Extraction Module
- Fact Extraction Module
- Entity Extraction Module
- Sentiment Analysis Module
- Network Graph Module
- Document Classification & Language Modeling Module

6. What is Bag of Words?

Bag of Words is a commonly used model that depends on word frequencies or occurrences to train a classifier. This model creates an occurrence matrix for documents or sentences irrespective of its grammatical structure or word order.

7. What is Pragmatic Ambiguity in NLP?

Pragmatic ambiguity refers to those words which have more than one meaning and their use in any sentence can depend entirely on the context. Pragmatic ambiguity can result in multiple interpretations of the same sentence. More often than not, we come across sentences which have words with multiple meanings, making the sentence open to interpretation. This multiple interpretation causes ambiguity and is known as Pragmatic ambiguity in NLP.

8. What is Masked Language Model?

Masked language models help learners to understand deep representations in downstream tasks by taking an output from the corrupt input. This model is often used to predict the words to be used in a sentence.

9. What is the difference between NLP and CI(Conversational Interface)?

The difference between NLP and CI is as follows:

Natural Language Processing (NLP)	Conversational Interface (CI)
NLP attempts to help machines understand and learn how language concepts work.	CI focuses only on providing users with an interface to interact with.
NLP uses AI technology to identify, understand, and interpret the requests of users through language.	CI uses voice, chat, videos, images and more such conversational aid to create the user interface.

10. What are the best NLP Tools?

Some of the best NLP tools from open sources are:

- SpaCy
- TextBlob
- Textacy
- Natural language Toolkit (NLTK)
- Retext
- NLP.js
- Stanford NLP

- CogcompNLP

11. What is POS tagging?

Parts of speech tagging better known as [POS tagging](#) refers to the process of identifying specific words in a document and group them as part of speech, based on its context. POS tagging is also known as grammatical tagging since it involves understanding grammatical structures and identifying the respective component.

POS tagging is a complicated process since the same word can be different parts of speech depending on the context. The same generic process used for word mapping is quite ineffective for POS tagging because of the same reason.

12. What is NES?

Name entity recognition is more commonly known as NER is the process of identifying specific entities in a text document which are more informative and have a unique context. These often denote places, people, organisations, and more. Even though it seems like these entities are proper nouns, the NER process is far from identifying just the nouns. In fact, NER involves entity chunking or extraction wherein entities are segmented to categorise them under different predefined classes. This step further helps in extracting information.

NLP Interview Questions for Experienced

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum\limits_{i=1}^{n} A_i B_i}{\sqrt{\sum\limits_{i=1}^{n} A_i^2} \sqrt{\sum\limits_{i=1}^{n} B_i^2}},$$

$$\sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

13. Which of the following techniques can be used for keyword normalization in NLP, the process of converting a keyword into its base form?

a. Lemmatization

b. Soundex

c. Cosine Similarity

d. N-grams

Answer: a)

Lemmatization helps to get to the base form of a word, e.g. are playing -> play, eating -> eat, etc..

Other options are meant for different purposes.

14. Which of the following techniques can be used to compute the distance between two word vectors in NLP?

a. Lemmatization

b. Euclidean distance

c. Cosine Similarity

d. N-grams

Answer: b) and c)

Distance between two word vectors can be computed using Cosine similarity and Euclidean Distance. **Cosine Similarity** establishes a cosine angle between the vector of two words. A cosine angle close to each other between two word vectors indicates the words are similar and vice a versa.

E.g. cosine angle between two words "Football" and "Cricket" will be closer to 1 as compared to angle between the words "Football" and "New Delhi"

Python code to implement CosineSimlarity function would look like this

def cosine_similarity(x,y):

```
    return np.dot(x,y)/( np.sqrt(np.dot(x,x)) * np.sqrt(np.dot(y,y)) )
```

```
q1 = wikipedia.page('Strawberry')

q2 = wikipedia.page('Pineapple')

q3 = wikipedia.page('Google')

q4 = wikipedia.page('Microsoft')

cv = CountVectorizer()

X = np.array(cv.fit_transform([q1.content, q2.content, q3.content,
q4.content]).todense())

print ("Strawberry Pineapple Cosine Distance", cosine_similarity(X[0],X[1]))

print ("Strawberry Google Cosine Distance", cosine_similarity(X[0],X[2]))

print ("Pineapple Google Cosine Distance", cosine_similarity(X[1],X[2]))

print ("Google Microsoft Cosine Distance", cosine_similarity(X[2],X[3]))

print ("Pineapple Microsoft Cosine Distance", cosine_similarity(X[1],X[3]))
```

Strawberry Pineapple Cosine Distance 0.8899200413701714

Strawberry Google Cosine Distance 0.7730935582847817

Pineapple Google Cosine Distance 0.789610214147025

Google Microsoft Cosine Distance 0.8110888282851575

Usually Document similarity is measured by how close semantically the content (or words) in the document are to each other. When they are close, the similarity index is close to 1, otherwise near 0.

The **Euclidean distance** between two points is the length of the shortest path connecting them. Usually computed using Pythagoras theorem for a triangle.

15. What are the possible features of a text corpus in NLP?

a. Count of the word in a document

b. Vector notation of the word

c. Part of Speech Tag

d. Basic Dependency Grammar

e. All of the above

Answer: e)

All of the above can be used as features of the text corpus.

16. You created a document term matrix on the input data of 20K documents for a Machine learning model. Which of the following can be used to reduce the dimensions of data?

1. Keyword Normalization
2. Latent Semantic Indexing
3. Latent Dirichlet Allocation

a. only 1

b. 2, 3

c. 1, 3

d. 1, 2, 3

Answer: d)

17. Which of the text parsing techniques can be used for noun phrase detection, verb phrase detection, subject detection, and object detection in NLP.

a. Part of speech tagging

b. Skip Gram and N-Gram extraction

c. Continuous Bag of Words

d. Dependency Parsing and Constituency Parsing

Answer: d)

18. Dissimilarity between words expressed using cosine similarity will have values significantly higher than 0.5

a. True

b. False

Ans: a)

19. Which one of the following are keyword Normalization techniques in NLP

a. Stemming

b. Part of Speech

c. Named entity recognition

d. Lemmatization

Answer: a) and d)

Part of Speech (POS) and Named Entity Recognition(NER) are not keyword Normalization techniques. Named Entity help you extract Organization, Time, Date, City, etc..type of entities from the given sentence, whereas Part of Speech helps you extract Noun, Verb, Pronoun, adjective, etc..from the given sentence tokens.

20. Which of the below are NLP use cases?

a. Detecting objects from an image

b. Facial Recognition

c. Speech Biometric

d. Text Summarization

Ans: d)

a) And b) are Computer Vision use cases, and c) is Speech use case.

Only d) Text Summarization is an NLP use case.

21. In a corpus of N documents, one randomly chosen document contains a total of T terms and the term "hello" appears K times.

What is the correct value for the product of TF (term frequency) and IDF (inverse-document-frequency), if the term "hello" appears in approximately one-third of the total documents?

a. KT * Log(3)

b. T * Log(3) / K

c. K * Log(3) / T

d. Log(3) / KT

Answer: (c)

formula for TF is K/T

formula for IDF is log(total docs / no of docs containing "data")

= log(1 / (⅓))

= log (3)

Hence correct choice is Klog(3)/T

22. In NLP, The algorithm decreases the weight for commonly used words and increases the weight for words that are not used very much in a collection of documents

a. Term Frequency (TF)

b. Inverse Document Frequency (IDF)

c. Word2Vec

d. Latent Dirichlet Allocation (LDA)

Ans: b)

23. In NLP, The process of removing words like "and", "is", "a", "an", "the" from a sentence is called as

a. Stemming

b. Lemmatization

c. Stop word

d. All of the above

Ans: c)

In Lemmatization, all the stop words such as a, an, the, etc.. are removed. One can also define custom stop words for removal.

24. In NLP, The process of converting a sentence or paragraph into tokens is referred to as Stemming

a. True

b. False

Ans: b)

The statement describes the process of tokenization and not stemming, hence it is False.

25. In NLP, Tokens are converted into numbers before giving to any Neural Network

a. True

b. False

Ans: a)

In NLP, all words are converted into a number before feeding to a Neural Network.

26. identify the odd one out

a. nltk

b. scikit learn

c. SpaCy

d. BERT

Ans: d)

All the ones mentioned are NLP libraries except BERT, which is a word embedding

27. TF-IDF helps you to establish?

a. most frequently occurring word in the document

b. most important word in the document

Ans: b)

TF-IDF helps to establish how important a particular word is in the context of the document corpus. TF-IDF takes into account the number of times the word appears in the document and offset by the number of documents that appear in the corpus.

- TF is the frequency of term divided by a total number of terms in the document.
- IDF is obtained by dividing the total number of documents by the number of documents containing the term and then taking the logarithm of that quotient.
- Tf.idf is then the multiplication of two values TF and IDF.

Suppose that we have term count tables of a corpus consisting of only two documents, as listed here

Term	Document 1 Frequency	Document 2 Frequency
This	1	1
is	1	1
a	2	
Sample	1	
another		2
example		3

The calculation of tf–idf for the term "this" is performed as follows:

for "this"

----

tf("this", d1) = 1/5 = 0.2

tf("this", d2) = 1/7 = 0.14

idf("this", D) = log (2/2) =0

hence tf-idf

tfidf("this", d1, D) = 0.2* 0 = 0

tfidf("this", d2, D) = 0.14* 0 = 0

for "example"

----

tf("example", d1) = 0/5 = 0

tf("example", d2) = 3/7 = 0.43

idf("example", D) = log(2/1) = 0.301

tfidf("example", d1, D) = tf("example", d1) * idf("example", D) = 0 * 0.301 = 0

tfidf("example", d2, D) = tf("example", d2) * idf("example", D) = 0.43 * 0.301 = 0.129

In its raw frequency form, TF is just the frequency of the "this" for each document. In each document, the word "this" appears once; but as document 2 has more words, its relative frequency is smaller.

An IDF is constant per corpus, and accounts for the ratio of documents that include the word "this". In this case, we have a corpus of two documents and all of them include the word "this". So TF–IDF is zero for the word "this", which implies that the word is not very informative as it appears in all documents.

The word "example" is more interesting - it occurs three times, but only in the second document.

28. In NLP, The process of identifying people, an organization from a given sentence, paragraph is called

a. Stemming

b. Lemmatization

c. Stop word removal

d. Named entity recognition

Ans: d)

29. Which one of the following is not a pre-processing technique in NLP

a. Stemming and Lemmatization

b. converting to lowercase

c. removing punctuations

d. removal of stop words

e. Sentiment analysis

Ans: e)

Sentiment Analysis is not a pre-processing technique. It is done after pre-processing and is an NLP use case. All other listed ones are used as part of statement pre-processing.

30. In text mining, converting text into tokens and then converting them into an integer or floating-point vectors can be done using

a. CountVectorizer

b. TF-IDF

c. Bag of Words

d. NERs

Ans: a)

CountVectorizer helps do the above, while others are not applicable.

text =["Rahul is an avid writer, he enjoys studying understanding and presenting. He loves to play"]

vectorizer = CountVectorizer()

vectorizer.fit(text)

vector = vectorizer.transform(text)

print(vector.toarray())

output

[[1 1 1 1 2 1 1 1 1 1 1 1 1 1]]

The second section of the interview questions c*overs advanced NLP techniques such as Word2Vec, GloVe word embeddings, and advanced models such as GPT, ELMo, BERT, XLNET based questions, and explanations.*

[sc name="cards-style"][/sc]

Our Most Popular Courses:

> <

[sc name="glcard" course_url="https://www.mygreatlearning.com/pg-program-artificial-intelligence-course" title="PGP - Artificial Intelligence & Machine Learning" image_url="https://d1vwxdpzbgdqj.cloudfront.net/s3-public-images/program/pg-program-online-artificial-intelligence-machine-learning.jpg" time="12 Months" type="Online Mentorship" institute_logo_url="https://d1vwxdpzbgdqj.cloudfront.net/assets/learning-partners/new-logo/utaustin-1-d74dd3fcbc8f8d1a2d9b77e04a20801f1ab1fdbc225a6b198f752d905d6a2bc8.png"][/sc] [sc name="glcard" course_url="https://www.mygreatlearning.com/artificial-intelligence-course-for-managers-leaders" title="PGP - Artificial Intelligence for Leaders" image_url="https://d1vwxdpzbgdqj.cloudfront.net/s3-public-images/program/artificial-intelligence-course-for-managers-leaders.jpg" time="4 Months" type="Online with Live Virtual Classes" institute_logo_url="https://d1vwxdpzbgdqj.cloudfront.net/assets/learning-partners/new-logo/utaustin-1-d74dd3fcbc8f8d1a2d9b77e04a20801f1ab1fdbc225a6b198f752d905d6a2bc8.png"][/sc] [sc name="glcard" course_url="https://www.mygreatlearning.com/pg-program-artificial-intelligence-course-classroom" title="PG Program in Artificial Intelligence and Machine Learning" image_url="https://d1vwxdpzbgdqj.cloudfront.net/s3-public-images/program/pg-program-artificial-intelligence-course.jpg" time="12 Months" type="Weekend Classroom" institute_logo_url="https://d1vwxdpzbgdqj.cloudfront.net/assets/learning-partners/new-logo/utaustin-1-d74dd3fcbc8f8d1a2d9b77e04a20801f1ab1fdbc225a6b198f752d905d6a2bc8.png"][/sc]

31. In NLP, Words represented as vectors are called as Neural Word Embeddings

a. True

b. False

Ans: a)

Word2Vec, GloVe based models build word embedding vectors that are multidimensional.

32. In NLP, Context modeling is supported with which one of the following word embeddings

1. a. Word2Vec
2. b) GloVe
3. c) BERT
4. d) All of the above

Ans: c)

Only BERT (Bidirectional Encoder Representations from Transformer) supports context modelling where the previous and next sentence context is taken into consideration. In Word2Vec, GloVe only word embeddings are considered and previous and next sentence context is not considered.

33. In NLP, Bidirectional context is supported by which of the following embedding

a. Word2Vec

b. BERT

c. GloVe

d. All the above

Ans: b)

Only BERT provides a bidirectional context. The BERT model uses the previous and the next sentence to arrive at the context.Word2Vec and GloVe are word embeddings, they do not provide any context.

34. Which one of the following Word embeddings can be custom trained for a specific subject in NLP

a. Word2Vec

b. BERT

c. GloVe

d. All the above

Ans: b)

BERT allows Transform Learning on the existing pre-trained models and hence can be custom trained for the given specific subject, unlike Word2Vec and GloVe where existing word embeddings can be used, no transfer learning on text is possible.

35. Word embeddings capture multiple dimensions of data and are represented as vectors

a. True

b. False

Ans: a)

36. In NLP, Word embedding vectors help establish distance between two tokens

a. True

b. False

Ans: a)

One can use Cosine similarity to establish distance between two vectors represented through Word Embeddings

37. Language Biases are introduced due to historical data used during training of word embeddings, which one amongst the below is not an example of bias

a. New Delhi is to India, Beijing is to China

b. Man is to Computer, Woman is to Homemaker

Ans: a)

Statement b) is a bias as it buckets Woman into Homemaker, whereas statement a) is not a biased statement.

38. Which of the following will be a better choice to address NLP use cases such as semantic similarity, reading comprehension, and common sense reasoning

a. ELMo

b. Open AI's GPT

c. ULMFit

Ans: b)

Open AI's GPT is able to learn complex pattern in data by using the Transformer models Attention mechanism and hence is more suited for complex use cases such as semantic similarity, reading comprehensions, and common sense reasoning.

39. Transformer architecture was first introduced with?

a. GloVe

b. BERT

c. Open AI's GPT

d. ULMFit

Ans: c)

ULMFit has an LSTM based Language modeling architecture. This got replaced into Transformer architecture with Open AI's GPT

40. Which of the following architecture can be trained faster and needs less amount of training data

a. LSTM based Language Modelling

b. Transformer architecture

Ans: b)

Transformer architectures were supported from GPT onwards and were faster to train and needed less amount of data for training too.

41. Same word can have multiple word embeddings possible with _____?

a. GloVe

b. Word2Vec

c. ELMo

d. nltk

Ans: c)

EMLo word embeddings supports same word with multiple embeddings, this helps in using the same word in a different context and thus captures the context than just meaning of the word unlike in GloVe and Word2Vec. Nltk is not a word embedding.

Word2Vec	Supported word embeddings (one word one embedding)
GloVe	Supported word embeddings (one word one embedding)
ELMo	Trained on a massive dataset to generate word embeddings. Same word multiple word embeddings based on the context it is in using a bidirectional LSTM arch. Two independent LSTM models one is left to right, and another right to left.
ULMFit	Introduced Transfer Learning, LSTM based bidirectional arch. But two independent models one is left to right and another right to left.
GPT, GPT-2	Transformer based architecture with attention models. Sentences are trained - Unidirectional Left to right
BERT (Bidirectional Encoder Representations from Transformers)	• Bidirectional context-based model using Transfer learning, Masking words in sentences, and predicting based on previous and next sentence context. • BERT uses WordPiece embedding with 30K token vocabulary and learned positional embeddings with supported sequence length up to 512 tokens
XLNET (Generalized Autoregressive Pretraining for Language Understanding)	• Improves on BERT by addressing the dependency between masked words. It doesn't mask any word in the sentence, unlike BERT. • Enables Learning bidirectional context by maximizing the expected likelihood over all the permutations
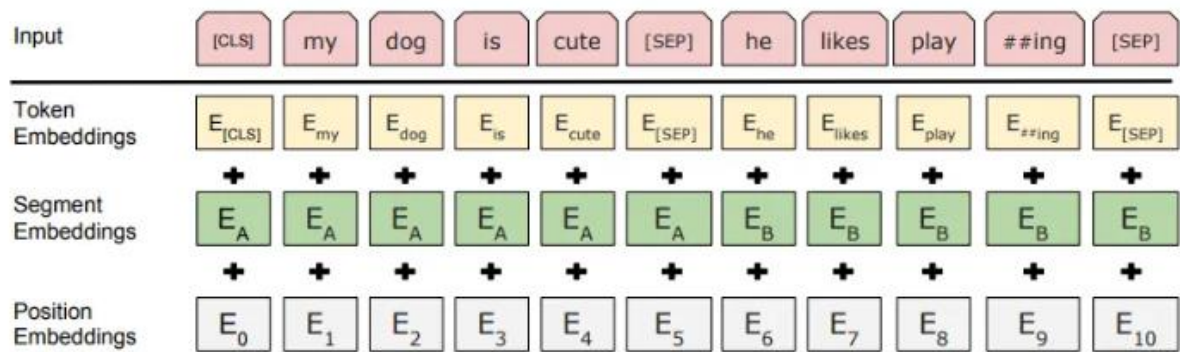
42. For a given token, its input representation is the sum of embedding from the token, segment and position embedding

a. ELMo

b. GPT

c. BERT

d. ULMFit

Ans: c)

BERT uses token, segment and position embedding.

43. Trains two independent LSTM language model left to right and right to left and shallowly concatenates them

a. GPT

b. BERT

c. ULMFit

d. ELMo

Ans: d)

ELMo tries to train two independent LSTM language models (left to right and right to left) and concatenates the results to produce word embedding.

44. Uses unidirectional language model for producing word embedding

a. BERT

b. GPT

c. ELMo

d. Word2Vec

Ans: b)

GPT is a idirectional model and word embedding are produced by training on information flow from left to right. ELMo is bidirectional but shallow. Word2Vec provides simple word embedding.

45. In this architecture, the relationship between all words in a sentence is modelled irrespective of their position. Which architecture is this?

a. OpenAI GPT

b. ELMo

c. BERT

d. ULMFit

Ans: c)

BERT Transformer architecture models the relationship between each word and all other words in the sentence to generate attention scores. These attention scores are later used as weights for a weighted average of all words' representations which is fed into a fully-connected network to generate a new representation.

46. List 10 use cases to be solved using NLP techniques?

- Sentiment Analysis
- Language Translation (English to German, Chinese to English, etc..)
- Document Summarization
- Question Answering
- Sentence Completion
- Attribute extraction (Key information extraction from the documents)

- Chatbot interactions
- Topic classification
- Intent extraction
- Grammar or Sentence correction
- Image captioning
- Document Ranking
- Natural Language inference
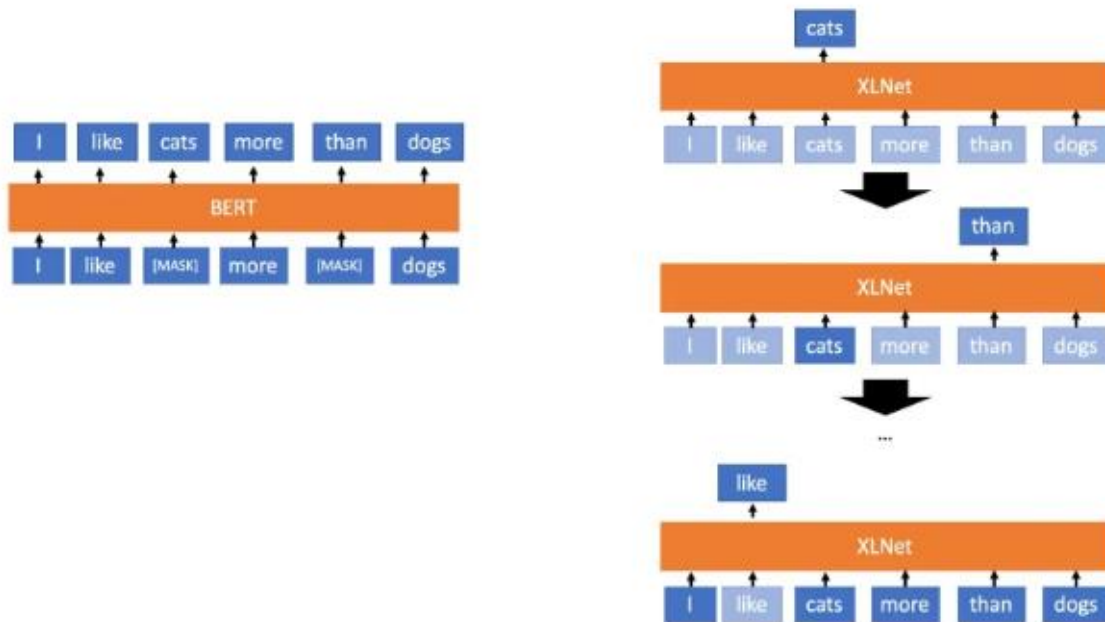
The conceptual difference between BERT and XLNet. Transparent words are masked out so the model cannot rely on them. XLNet learns to predict the words in an arbitrary order but in an autoregressive, sequential manner (not necessarily left-to-right). BERT predicts all masked words simultaneously.

47. Transformer model pays attention to the most important word in Sentence

a. True

b. False

Ans: a) Attention mechanisms in the Transformer model are used to model the relationship between all words and also provide weights to the most important word.

48. Which NLP model gives the best accuracy amongst the following?

a. BERT

b. XLNET

c. GPT-2

d. ELMo

Ans: b) XLNET

XLNET has given best accuracy amongst all the models. It has outperformed BERT on 20 tasks and achieves state of art results on 18 tasks including sentiment analysis, question answering, natural language inference, etc.

49. Permutation Language models is a feature of

a. BERT

b. EMMo

c. GPT

d. XLNET

Ans: d)

XLNET provides permutation-based language modelling and is a key difference from BERT. In permutation language modeling, tokens are predicted in a random manner and not sequential. The order of prediction is not necessarily left to right and can be right to left. The original order of words is not changed but a prediction can be random.

The conceptual difference between BERT and XLNET can be seen from the following diagram.

50. Transformer XL uses relative positional embedding

a. True

b. False

Ans: a)

Instead of embedding having to represent the absolute position of a word, Transformer XL uses an embedding to encode the relative distance between the words. This embedding is used to compute the attention score between any 2 words that could be separated by n words before or after.

There, you have it - all the probable questions for your NLP interview. Now go, give it your best shot. Take up Great Learning's PGP Artificial Intelligence and Machine Learning Course and upskill today. This is an online course with mentorship where you'll be learning from industry experts. You also have the opportunity to gain hands-on learning through real-world projects. Great Learning also offers career assistance, interview preparation, and resume building sessions.

Natural Language Processing FAQ's

1. Why do we need NLP?

One of the main reasons why NLP is necessary is because it helps computers communicate with humans in natural language. It also scales other language-related tasks. Because of NLP, it is possible for computers to hear speech, interpret this speech, measure it and also determine which parts of the speech are important.

2. What must a natural language program decide?

A natural language program must decide what to say and when to say something.

3. Where can NLP be useful?

NLP can be useful in communicating with humans in their own language. It helps improve the efficiency of the machine translation and is useful in emotional analysis too. It can be helpful in sentiment analysis too. It also helps in structuring highly unstructured data. It can be helpful in creating chatbots, Text Summarization and virtual assistants.

4. How to prepare for an NLP Interview?

The best way to prepare for an NLP Interview is to be clear about the basic concepts. Go through blogs that will help you cover all the key aspects and remember the important topics. Learn specifically for the interviews and be confident while answering all the questions.

5. What are the main challenges of NLP?

Breaking sentences into tokens, Parts of speech tagging, Understanding the context, Linking components of a created vocabulary, Extracting semantic meaning are currently some of the main challenges of NLP.

6. Which NLP model gives best accuracy?

Naive Bayes Algorithm has the **highest accuracy** when it comes to NLP models. It gives up to 73% correct predictions.

7. What are the major tasks of NLP?

Translation, named entity recognition, relationship extraction, sentiment analysis, speech recognition, and topic segmentation are few of the major tasks of NLP. Under unstructured data, there can be a lot of untapped information that can help an organization grow.

8. What are stop words in NLP?

Common words that occur in sentences that add weight to the sentence are known as stop words. These stop words act as a bridge and ensure that sentences are grammatically correct. In simple terms, words that are filtered out before processing natural language data is known as a stop word and it is a common pre-processing method.

9. What is stemming in NLP?

The process of obtaining the root word from the given word is known as stemming. All tokens can be cut down to obtain the root word or the stem with the help of efficient and well-generalized rules. It is a rule-based process and is well-known for its simplicity.

10. Why is NLP so hard?

There are several factors that make the process of Natural Language Processing difficult. There are hundreds of natural languages all over the world, words can be ambiguous in their meaning, each natural language has a different script and syntax,

the meaning of words can change depending on the context, and so the process of NLP can be difficult. If you choose to upskill and continue learning, the process will become easier over time.

11. What does a NLP pipeline consist of *?

The overall architecture of an **NLP pipeline consists** of several layers: a user interface; one or several **NLP** models, depending on the use case; a Natural Language Understanding layer to describe the **meaning of** words and sentences; a preprocessing layer; microservices for linking the components together and of course.

12. How many steps of NLP is there?

The five phases of NLP involve lexical (structure) analysis, parsing, semantic analysis, discourse integration, and pragmatic analysis.