**4. What is NLTK?**

NLTK is a Python library, which stands for Natural Language Toolkit. We use NLTK to process data in human-spoken languages. NLTK allows us to apply techniques such as parsing, tokenization, lemmatization, stemming, and more to understand natural languages. It helps in categorizing text, parsing linguistic structure, analyzing documents, etc.

A few of the libraries of the NLTK package that we often use in NLP are:

1. SequentialBackoffTagger
2. DefaultTagger
3. UnigramTagger
4. treebank
5. wordnet
6. FreqDist
7. patterns
8. RegexpTagger
9. backoff\_tagger
10. UnigramTagger, BigramTagger, and TrigramTagger

### ****5. What is Syntactic Analysis?****

Syntactic analysis is a technique of analyzing sentences to extract meaning from them. Using syntactic analysis, a machine can analyze and understand the order of words arranged in a sentence. NLP employs grammar rules of a language that helps in the syntactic analysis of the combination and order of words in documents.

The techniques used for syntactic analysis are as follows

### ****12. What is TF-IDF?****

TFIDF or Term Frequency-Inverse Document Frequency indicates the importance of a word in a set. It helps in information retrieval with numerical statistics. For a specific document, TF-IDF shows a frequency that helps identify the keywords in a document. The major use of TF-IDF in NLP is the extraction of useful information from crucial documents by statistical data. It is ideally used to classify and summarize the text in documents and filter out stop words.

**TF** helps calculate the ratio of the frequency of a term in a document and the total number of terms. Whereas, **IDF** denotes the importance of the term in a document.

The formula for calculating TF-IDF:

**TF(W) = (Frequency of W in a document)/(The total number of terms in the document)**

**IDF(W) = log\_e(The total number of documents/The number of documents having the term W)**

When **TF**\***IDF** is high, the frequency of the term is less and vice versa.

Google uses TF-IDF to decide the index of search results according to the relevancy of pages. The design of the TF-IDF algorithm helps optimize the search results in Google. It helps quality content rank up in search results.

### ****15. What is the difference between NLP and NLU?****

### ****17. What is Pragmatic Analysis?****

Pragmatic analysis is an important task in NLP for interpreting knowledge that is lying outside a given document. The aim of implementing pragmatic analysis is to focus on exploring a different aspect of the document or text in a language. This requires a comprehensive knowledge of the real world. The pragmatic analysis allows software applications for the critical interpretation of the real-world data to know the actual meaning of sentences and words.

**Example**:

Consider this sentence: ‘Do you know what time it is?’

This sentence can either be asked for knowing the time or for yelling at someone to make them note the time. This depends on the context in which we use the sentence.

### ****9. What are unigrams, bigrams, trigrams, and n-grams in NLP?****

When we parse a sentence one word at a time, then it is called a unigram. The sentence parsed two words at a time is a bigram.

When the sentence is parsed three words at a time, then it is a trigram. Similarly, n-gram refers to the parsing of n words at a time.

**Example**: To understand unigrams, bigrams, and trigrams, you can refer to the below diagram:

**20. What are the steps involved in solving an NLP problem?**

Below are the steps involved in solving an NLP problem:

1. Gather the text from the available dataset or by web scraping
2. Apply stemming and lemmatization for text cleaning
3. Apply feature engineering techniques
4. Embed using **word2vec**
5. Train the built model using neural networks or other Machine Learning techniques
6. Evaluate the model’s performance
7. Make appropriate changes in the model
8. Deploy the model

### ****21. What is Feature Extraction in NLP?****

Features or characteristics of a word help in text or document analysis. They also help in sentiment analysis of a text. Feature extraction is one of the techniques that are used by recommendation systems. Reviews such as ‘excellent,’ ‘good,’ or ‘great’ for a movie are positive reviews, recognized by a recommender system. The recommender system also tries to identify the features of the text that help in describing the context of a word or a sentence. Then, it makes a group or category of the words that have some common characteristics. Now, whenever a new word arrives, the system categorizes it as per the labels of such groups.

### ****What are precision and recall?****

### ****Explain Stemming with the help of an example.****

### ****Explain Lemmatization with the help of an example.****

### ****What is Parts-of-speech Tagging?****

**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](https://www.mygreatlearning.com/blog/text-summarization-in-python/) 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

**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](https://www.mygreatlearning.com/blog/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.

### ****11. What is POS tagging?****

Parts of speech tagging better known as [POS tagging](https://www.mygreatlearning.com/blog/pos-tagging/) refer to the process of identifying specific words in a document and grouping 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 general process used for word mapping is quite ineffective for POS tagging because of the same reason.