**NATURAL LANGUAGE PROCESSING INTERVIEW QUESTIONS**

Q:What do you understand by Natural Language Processing?  
A: Natural Language Processing is a field of computer science that deals with communication between computer systems and humans. It is a technique used in Artificial Intelligence and Machine Learning. It is used to create automated software that helps understand human spoken languages to extract useful information from the data it gets in the form of audio. Techniques in NLP allow computer systems to process and interpret data in the form of natural languages.

Q:List any two real-life applications of Natural Language Processing.  
A: Two real-life applications of Natural Language Processing are as follows:

**Google Translate**: Google Translate is one of the famous applications of Natural Language Processing. It helps convert written or spoken sentences into any language. Also, we can find the correct pronunciation and meaning of a word by using Google Translate. It uses advanced techniques of Natural Language Processing to achieve success in translating sentences into various languages.

**Chatbots:** To provide a better customer support service, companies have started using chatbots for 24/7 service. Chatbots helps resolve the basic queries of customers. If a chatbot is not able to resolve any query, then it forwards it to the support team, while still engaging the customer. It helps make customers feel that the customer support team is quickly attending them. With the help of chatbots, companies have become capable of building cordial relations with customers. It is only possible with the help of Natural Language Processing.

Q:What are stopwords?  
A: Stopwords are said to be useless data for a search engine. Words such as articles, prepositions, etc. are considered as stopwords. There are stopwords such as was, were, is, am, the, a, an, how, why, and many more. In Natural Language Processing, we eliminate the stopwords to understand and analyze the meaning of a sentence. The removal of stopwords is one of the most important tasks for search engines. Engineers design the algorithms of search engines in such a way that they ignore the use of stopwords. This helps show the relevant search result for a query.

Q: What is TF-IDF?  
A: TF-IDF 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)**

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.

Q: What is Word embeddings?

##### **A: Word embeddings are numeric vectors that show semantic relationships between words. To see how close the two words are to each other semantically, the cosinus similarity of these word vectors is checked. The closer the value is to 1, the closer the words are semantically.**

Q: What is Short Term Memory Problem?

A: **Short-term memory** is a problem for recurrent neural networks due to vanishing gradient issues. They'll have difficulty transferring information from earlier time steps to later ones if the sequence is lengthy enough. If you're attempting to predict anything from a paragraph of text, RNN’s may leave out essential information at the beginning of sequences. The vanishing gradient problem affects recurrent neural networks during backpropagation.

**The vanishing gradient problem**affects recurrent neural networks during backpropagation. Gradients are values that are used to update the weights of a neural network. When a gradient diminishes as it backpropagates through time, this is known as the vanishing gradient issue. When a gradient value falls below a certain threshold, it no longer contributes much to learning.

Q: What is the function of gates in LSTM and GRU?

A: Internal processes called gates in the LSTM and GRU regulate the flow of information. These gates figure out which data in a sequence should be kept and which should be discarded. They can then send important information along long sequences to create predictions.

Q: What is the function of Cell State in LSTM?

A : **The cell state**carries relevant information throughout the sequence. Thus, previous information isn't forgotten and the model can make more accurate predictions. Therefore, you can think of it as the memory of the network.

Q: What is the function of Forget Gate in LSTM?

A: **The Forget Gate** determines whether or not to retain or discard the information. The sigmoid function is used for this process. The closer the value is to 0, the more likely it will be forgotten, and the closer it is to 1, the more likely it will be retained.

Q: What is the function of Reset Gate in GRU?

A : It is the gate that decides how much of the information from other steps should be forgotten.

Q: What is the working logic of LSTM and GRU ?

A: In NLP, LSTM and GRU determine which class the text belongs to by determining the keywords and patterns in the text. They ignore all other words, phrases, etc.