Interview Questions

Q1) What will happen if we do stemming instead of lemmatization?

Q2) Difference between stemming and lemmatization?

Q3) How to remove punctuation marks?  
\* In this question talk about different ways to remove punctuation mark.

Q4) if you don’t want to remove (‘) apostrophe in sentence how you can do?

Q5) Different libraries and ways to perform stemming, lemmatization, tokenization.

Q) Why we need NLP?

\* We need applications which directly interact with humans like human say something and machine can understand that. That’s why we need NLP. There are many applications came of NLP like Alexa, Siri, etc.

Q) What are the different steps of data pre-processing?

Q) How to remove punctuation marks?

Q) If you are not removing punctuations marks then what is the reason?

Q) How to remove all punctuation marks except semicolon (;) or apostrophe (‘)?

\* You known all the difference ways of performing Data Pre-processing along with codes.

Q) How to remove a particular word?

\* Spend time on Text pre-processing steps and their multiple ways along with code. (Important)

Q) What is TF-IDF and related question are asked from this topic.

\* Learn the formulas of TF-IDF by heart because in interviews questions are asked from TF-IDF.

\* NLP is important topic so do good preparation of it.

\* Explore about NLP on google and you tube, Kaggle. Explore text preparation, feature engineering techniques.

Q) What are the different steps of data pre-processing? A) Data pre-processing involves several steps, including text normalization, tokenization, removing stop words, handling special characters, converting text to lowercase, and lemmatization or stemming.

Q) How to remove punctuation marks? A) Punctuation marks can be removed from text by using regular expressions or built-in string functions in programming languages.

Q) If you are not removing punctuation marks, then what is the reason? A) The decision not to remove punctuation marks may depend on the specific NLP task at hand. In some cases, punctuation can carry important information and semantic meaning, so it may be relevant to retain them.

Q) How to remove all punctuation marks except semicolon (;) or apostrophe (‘)? A) To remove all punctuation marks except semicolon or apostrophe, one can use regular expressions and specify the characters to keep while excluding all others.

Q) How to remove a particular word? A) To remove a specific word from text, one can use string manipulation functions or regular expressions to locate and replace the target word with an empty string.

Q) What is TF-IDF, and what related questions are asked from this topic? A) TF-IDF stands for Term Frequency-Inverse Document Frequency, and it is a numerical representation of the importance of a word in a document relative to a collection of documents. Related questions in interviews might involve understanding its calculation, its use in text representation, and its significance in information retrieval and text analysis tasks.

Q) Learn the formulas of TF-IDF by heart because questions are asked about TF-IDF in interviews. A) It is essential to memorize the formulas of TF-IDF to better explain its concept and use in NLP applications during interviews.

Q) NLP is an important topic, so do good preparation for it. A) NLP is a significant and rapidly growing field, and thorough preparation is crucial to grasp its concepts and techniques effectively.

Q) Explore about NLP on Google, YouTube, Kaggle, and explore text preparation and feature engineering techniques. A) To gain a comprehensive understanding of NLP, it is recommended to explore various resources such as Google, YouTube, and Kaggle, focusing on text preparation and feature engineering techniques.

--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Different methods of Text Preparation: -

1) Lower Casing

(i) String.lower()

2) Removal of Punctuations

(i) string.punctuation() and string.remove()

(ii) Using Regex

(iii) Using spacy library

(iv) Use the replace() method

3) Removal of Stop words

(i) NLTK library

4) Removal of most frequent and rare words

(i) Counter Library

(ii) Using Frequency Distribution Function of NLTK Library

5) Stemming

(i) Porter Stemmer (NLTK Library)

(ii) Snowball Stemmer (NLTK Library)

(iii) Text Blob Library

6) Lemmatization: We can change POS in lemmatization

(i) WordNetLemmatizer of NLTK Library

(ii) Using Text Blob Library

(ii) Using Spacy Library

7) Removal of Emojis

(i) Giving a list of Emojis code and using Regex

8) Removal of Emoticons

(i) Giving a list of Emoticons code and using Regex

9) Converting Emojis and Emoticons into text

(i) Giving a list of Emojis and Emoticons code and using Regex

10) Removal of URL’s

(i) Using Regex

11) Removal of HTML Tags

(i) Using Regex

(ii) Using Beautiful Soup Library

12) Convert Chat Word (Short form to full form)

(i) Create a list and then create user define function

13) Spelling Correction

(i) Using Spell Checker Library

(ii) Using Text Blob Library

14) Tokenization

(i) Using NLTK Library

(ii) Using Text Blob Library

(iii) Using String.Split() Function

(iv) Using SpaCy Library

15) Word Cloud

(i) Using Word Cloud Function of NLTK Library