

Text Preprocessing using NLTK.

NLTK (Natural Language Toolkit) is a powerful Python library for working with human language data. It provides easy-to-use interfaces to over 50 corpora and lexical resources, such as WordNet. Additionally, NLTK includes a variety of text processing libraries for tasks such as tokenization, stemming, tagging, parsing, and more.

Installation:

```
pip install nltk
```

Downloading NLTK Data:

NLTK requires additional data for certain functionalities. To download these data sets, you can use the NLTK data downloader. Here's how to do it:

- Open a Python interpreter or a Jupyter notebook.
- Import NLTK:
`import nltk`
- Download NLTK Data:
`nltk.download()`
- This command will open the NLTK downloader GUI, allowing you to download specific datasets or all available datasets.

- If you prefer to download specific datasets programmatically, you can use commands like:

```
nltk.download('punkt')  
nltk.download('wordnet')  
nltk.download('stopwords')
```

Now you have NLTK installed on your system, along with the necessary data for various NLP tasks. You can start using NLTK to perform text processing, analysis, and other natural language tasks in your Python projects.

Text Preprocessing Steps:-

- Lowercase / Uppercase
- Tokenization
- Removing Special Characters
- Removing Stop words and punctuation
- Stemming