

TEXT SUMMARIZATION



01

SPACY LIBRARY

Provides efficient text processing capabilities, such as tokenization, part-of-speech tagging, and dependency parsing.



02

CLASS STRUCTURE:

- The code defines a Summarizer class that utilizes Spacy for text summarization.
- The class takes an instance of the Spacy NLP model as an input.



03

CLEANING AND SENTENCE SPLITTING:

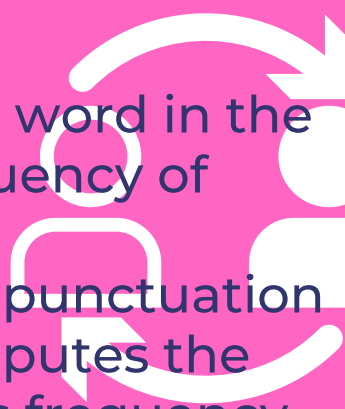
The `_clean_text` method preprocesses the text by removing newlines and splitting it into individual sentences using Spacy's sentence segmentation.



04

WORD WEIGHTS:

- Calculates the weights of each word in the document based on their frequency of occurrence.
- Spacy remove stop words and punctuation from the tokens and then computes the weights as a ratio of the word's frequency to the maximum frequency.



05

SENTENCE WEIGHTS AND SUMMARIZATION:

- `_get_sentence_weights` method calculates sentence weights by averaging word weights.
- Sentences are scored based on pre-obtained word weights.
- `_get_top_n_sentences` method selects highest-weighted sentences and combines them into a summary.
- `summarize_text` method executes summarization, providing word weights, sentence weights, original sentences, and summary as output.

