tokenizers hugging face (can take lemming and stemming into account) it's the idea of subwords (split words into subwords with tags on them).

dropdown menu "chosse tokenization method"

- BERT

lemmatisation racinisation

(not finished yet)

tokenizers GPT

tokenizers GPT copy

\_\_\_\_

create application to calculate sentiment score

POS (Part-Of-Speech) Tagging is a NLP task that involves assigning grammatical categories, such as nouns, verbs, adjectives, etc. to each word in a given text. This process enhances language understanding by providing insight into the syntactic structure and role of words within sentences, enabling more advanced language processing applications.

NER (Named Entitty Recognition) is a NLP task that involves identifying and classifying named entities, such as names of people, organizations, locations, dates, and other specific entities, within a given text. The goal is to extract structured information from unstructured text, enabling machines to understand and process named entities for various applications like information retrieval, text summarization, and question answering.

|  |  |  |
| --- | --- | --- |
| Description | Penn POS Tags | Equivalent SentiMI POS Tag |
| Verb, base form | VB | V |
| Verb, past tense | VBD | V |
| Verb, gerund or present participle | VBG | V |
| Verb, past participle | VBN | V |
| Verb, non-3rd person singular present | VBP | V |
| Verb, 3rd person singular present | VBZ | V |
| Noun, singular or or mass | NN | N |
| Noun, plural | NNS | N |
| Propre noun, singular | NNP | N |
| Proper noun, plural | NNPS | N |
| Adjective | JJ | A |
| Adjective, comparative | JJR | A |
| Adjective, superlative | JJS | A |
| Adverb | RB | R |
| Adverb, comparative | RBR | R |
| Adverb, superlative | RBS | R |