

```
!pip install nltk spacy
```

```
Requirement already satisfied: nltk in /usr/local/lib/python3.12/dist-packages (3.9.1)
Requirement already satisfied: spacy in /usr/local/lib/python3.12/dist-packages (3.8.11)
Requirement already satisfied: click in /usr/local/lib/python3.12/dist-packages (from nltk) (8.3.1)
Requirement already satisfied: joblib in /usr/local/lib/python3.12/dist-packages (from nltk) (1.5.)
Requirement already satisfied: regex>=2021.8.3 in /usr/local/lib/python3.12/dist-packages (from nl
Requirement already satisfied: tqdm in /usr/local/lib/python3.12/dist-packages (from nltk) (4.67.1)
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in /usr/local/lib/python3.12/dist-pac
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in /usr/local/lib/python3.12/dist-pac
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/python3.12/dist-pac
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /usr/local/lib/python3.12/dist-pac
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python3.12/dist-pac
Requirement already satisfied: thinc<8.4.0,>=8.3.4 in /usr/local/lib/python3.12/dist-pac
Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in /usr/local/lib/python3.12/dist-pac
Requirement already satisfied: srslry<3.0.0,>=2.4.3 in /usr/local/lib/python3.12/dist-pac
Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in /usr/local/lib/python3.12/dist-pac
Requirement already satisfied: weasel<0.5.0,>=0.4.2 in /usr/local/lib/python3.12/dist-pac
Requirement already satisfied: typer-slim<1.0.0,>=0.3.0 in /usr/local/lib/python3.12/dist-pac
Requirement already satisfied: numpy>=1.19.0 in /usr/local/lib/python3.12/dist-packages (from sp
Requirement already satisfied: requests<3.0.0,>=2.13.0 in /usr/local/lib/python3.12/dist-pac
Requirement already satisfied: pydantic!=1.8,!=1.8.1,<3.0.0,>=1.7.4 in /usr/local/lib/python3.12/c
Requirement already satisfied: jinja2 in /usr/local/lib/python3.12/dist-packages (from spacy) (3.1
Requirement already satisfied: setuptools in /usr/local/lib/python3.12/dist-packages (from spacy)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.12/dist-packages (from sp
Requirement already satisfied: annotated-types>=0.6.0 in /usr/local/lib/python3.12/dist-pac
Requirement already satisfied: pydantic-core==2.41.4 in /usr/local/lib/python3.12/dist-pac
Requirement already satisfied: typing-extensions>=4.14.1 in /usr/local/lib/python3.12/dist-pac
Requirement already satisfied: typing-inspection>=0.4.2 in /usr/local/lib/python3.12/dist-pac
Requirement already satisfied: charset_normalizer<4,>=2 in /usr/local/lib/python3.12/dist-pac
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.12/dist-pac
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.12/dist-pac
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.12/dist-pac
Requirement already satisfied: blis<1.4.0,>=1.3.0 in /usr/local/lib/python3.12/dist-pac
Requirement already satisfied: confection<1.0.0,>=0.0.1 in /usr/local/lib/python3.12/dist-pac
Requirement already satisfied: cloudpathlib<1.0.0,>=0.7.0 in /usr/local/lib/python3.12/dist-pac
Requirement already satisfied: smart-open<8.0.0,>=5.2.1 in /usr/local/lib/python3.12/dist-pac
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.12/dist-pac
Requirement already satisfied: wrapt in /usr/local/lib/python3.12/dist-pac (from smart-open<8
```

```
import nltk
import spacy
from nltk.tokenize import sent_tokenize, word_tokenize
from nltk.stem import PorterStemmer
from nltk.stem import WordNetLemmatizer

nltk.download('punkt')
nltk.download('wordnet')
nltk.download('omw-1.4')
```

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]  Unzipping tokenizers/punkt.zip.
[nltk_data] Downloading package wordnet to /root/nltk_data...
[nltk_data] Downloading package omw-1.4 to /root/nltk_data...
True
```

```
medical_text = """
Diabetes is a chronic disease that affects how the body processes blood sugar.
If untreated, diabetes may cause heart disease, kidney failure, nerve damage and vision problems.
```

```
Early diagnosis and proper treatment help improve patient outcomes.  
"""
```

```
nltk.download('punkt_tab')  
sentences = sent_tokenize(medical_text)  
sentences  
  
[nltk_data] Downloading package punkt_tab to /root/nltk_data...  
[nltk_data] Package punkt_tab is already up-to-date!  
['\nDiabetes is a chronic disease that affects how the body processes blood sugar.',  
 'If untreated, diabetes may cause heart disease, kidney failure, nerve damage and vision  
problems.',  
 'Early diagnosis and proper treatment help improve patient outcomes.']
```

```
words = word_tokenize(medical_text)  
words
```

```
['Diabetes',  
 'is',  
 'a',  
 'chronic',  
 'disease',  
 'that',  
 'affects',  
 'how',  
 'the',  
 'body',  
 'processes',  
 'blood',  
 'sugar',  
 '.',  
 'If',  
 'untreated',  
 ',',  
 'diabetes',  
 'may',  
 'cause',  
 'heart',  
 'disease',  
 ',',  
 'kidney',  
 'failure',  
 ',',  
 'nerve',  
 'damage',  
 'and',  
 'vision',  
 'problems',  
 '.',  
 'Early',  
 'diagnosis',  
 'and',  
 'proper',  
 'treatment',  
 'help',  
 'improve',  
 'patient',  
 'outcomes',  
 '.']
```

```
nlp = spacy.load("en_core_web_sm")  
doc = nlp(medical_text)
```

```
spacy_tokens = [token.text for token in doc]
spacy_tokens
```

```
['\n',
'Diabetes',
'is',
'a',
'chronic',
'disease',
'that',
'affects',
'how',
'the',
'body',
'processes',
'blood',
'sugar',
'.',
'\n',
'If',
'untreated',
',',
'diabetes',
'may',
'cause',
'heart',
'disease',
',',
'kidney',
'failure',
',',
'nerve',
'damage',
'and',
'vesion',
'problems',
'.',
'\n',
'Early',
'diagnosis',
'and',
'proper',
'treatment',
'help',
'improve',
'patient',
'outcomes',
'.',
'.',
'\n']
```

```
stemmer = PorterStemmer()
```

```
stemmed_words = [stemmer.stem(word) for word in words if word.isalpha()]
stemmed_words
```

```
['diabet',
'is',
'a',
'chronic',
'diseas',
'that',
'affect',
'how',
```

```
'the',
'bodi',
'process',
'blood',
'sugar',
'if',
'untreat',
'diabet',
'may',
'caus',
'heart',
'diseas',
'kidney',
'failur',
'nerv',
'damag',
'and',
'vesion',
'problem',
'earli',
'diagnosi',
'and',
'proper',
'treatment',
'help',
'improv',
'patient',
'outcom']
```

```
lemmatizer = WordNetLemmatizer()

lemmatized_words_nltk = [lemmatizer.lemmatize(word.lower()) for word in words if word.isalpha()]
```

```
['diabetes',
'is',
'a',
'chronic',
'disease',
'that',
'affect',
'how',
'the',
'body',
'process',
'blood',
'sugar',
'if',
'untreated',
'diabetes',
'may',
'cause',
'heart',
'disease',
'kidney',
'failure',
'nerve',
'damage',
'and',
'vesion',
'problem',
'early',
'diagnosis',
'and',
'proper',
```

```
'treatment',
'help',
'improve',
'patient',
'outcome']
```

```
spacy_lemmas = [token.lemma_ for token in doc if token.is_alpha]
spacy_lemmas
```

```
['Diabetes',
'be',
'a',
'chronic',
'disease',
'that',
'affect',
'how',
'the',
'body',
'process',
'blood',
'sugar',
'if',
'untreat',
'diabete',
'may',
'cause',
'heart',
'disease',
'kidney',
'failure',
'nerve',
'damage',
'and',
'vesion',
'problem',
'early',
'diagnosis',
'and',
'proper',
'treatment',
'help',
'improve',
'patient',
'outcome']
```

