

```
!pip install nltk spacy
```

```
Collecting nltk
```

```
  Downloading nltk-0.1-py3-none-any.whl.metadata (1.6 kB)
```

```
Requirement already satisfied: spacy in /usr/local/lib/python3.12/dist-packages (3.8.11)
```

```
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in /usr/local/lib/python3.12/dist-packa
```

```
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in /usr/local/lib/python3.12/dist-packa
```

```
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/python3.12/dist-package
```

```
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /usr/local/lib/python3.12/dist-packages (fr
```

```
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python3.12/dist-packages (f
```

```
Requirement already satisfied: thinc<8.4.0,>=8.3.4 in /usr/local/lib/python3.12/dist-packages (fr
```

```
Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in /usr/local/lib/python3.12/dist-packages (fr
```

```
Requirement already satisfied: srsly<3.0.0,>=2.4.3 in /usr/local/lib/python3.12/dist-packages (fr
```

```
Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in /usr/local/lib/python3.12/dist-packages
```

```
Requirement already satisfied: weasel<0.5.0,>=0.4.2 in /usr/local/lib/python3.12/dist-packages (fr
```

```
Requirement already satisfied: typer-slim<1.0.0,>=0.3.0 in /usr/local/lib/python3.12/dist-packages
```

```
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in /usr/local/lib/python3.12/dist-packages (fr
```

```
Requirement already satisfied: numpy>=1.19.0 in /usr/local/lib/python3.12/dist-packages (from spac
```

```
Requirement already satisfied: requests<3.0.0,>=2.13.0 in /usr/local/lib/python3.12/dist-packages
```

```
Requirement already satisfied: pydantic!=1.8,!1.8.1,<3.0.0,>=1.7.4 in /usr/local/lib/python3.12/d
```

```
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-packages (
```

```
Requirement already satisfied: pydantic-core==2.41.4 in /usr/local/lib/python3.12/dist-packages (f
```

```
Requirement already satisfied: typing-extensions>=4.14.1 in /usr/local/lib/python3.12/dist-package
```

```
Requirement already satisfied: typing-inspection>=0.4.2 in /usr/local/lib/python3.12/dist-packages
```

```
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.12/dist-packages
```

```
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.12/dist-packages (from requ
```

```
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.12/dist-packages (from
```

```
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.12/dist-packages (from
```

```
Requirement already satisfied: blis<1.4.0,>=1.3.0 in /usr/local/lib/python3.12/dist-packages (from
```

```
Requirement already satisfied: confection<1.0.0,>=0.0.1 in /usr/local/lib/python3.12/dist-packages
```

```
Requirement already satisfied: click>=8.0.0 in /usr/local/lib/python3.12/dist-packages (from typer
```

```
Requirement already satisfied: cloudpathlib<1.0.0,>=0.7.0 in /usr/local/lib/python3.12/dist-packag
```

```
Requirement already satisfied: smart-open<8.0.0,>=5.2.1 in /usr/local/lib/python3.12/dist-packages
```

```
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.12/dist-packages (from ji
```

```
Requirement already satisfied: wrapt in /usr/local/lib/python3.12/dist-packages (from smart-open<8
```

```
Downloading nltk-0.1-py3-none-any.whl (15 kB)
```

```
Installing collected packages: nltk
```

```
Successfully installed nltk-0.1
```

```
# Import libraries
```

```
import nltk
```

```
import spacy
```

```
from nltk.corpus import gutenberg
```

```
from nltk.tokenize import sent_tokenize, word_tokenize
```

```
from nltk.stem import PorterStemmer, WordNetLemmatizer
```

```
# Download NLTK resources
```

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

```
nltk.download('gutenberg')
```

```
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 gutenberg to /root/nltk_data...
```

```
[nltk_data] Unzipping corpora/gutenberg.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.
"""
```

```
import nltk
from nltk.tokenize import sent_tokenize
nltk.download('punkt_tab')
sentences = sent_tokenize(medical_text)
sentences
```

```
[nltk_data] Downloading package punkt_tab to /root/nltk_data...
[nltk_data] Unzipping tokenizers/punkt_tab.zip.
['\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',
'.']
```

```
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',
'vision',
'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()]
lemmatized_words_nltk
```

```
['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',  
'vision',  
'problem',  
'early',  
'diagnosis',  
'and',  
'proper',  
'treatment',  
'help',  
'improve',  
'patient',  
'outcome']
```