```
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."""
```

```
# Install NLTK and spaCy libraries
!pip install nltk spacy

# Download a spaCy language model (e.g., en_core_web_sm)
!python -m spacy download en_core_web_sm
```

```
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-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 (frc
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 (frc
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 (frc
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: 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 reque
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: 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
Collecting en-core-web-sm==3.8.0
  Downloading https://github.com/explosion/spacy-models/releases/download/en_core_web_sm-3.8.0/en_
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 12.8/12.8 MB 100.4 MB/s eta 0:00:00
✓ Download and installation successful
You can now load the package via spacy.load('en_core_web_sm')
⚠ Restart to reload dependencies
If you are in a Jupyter or Colab notebook, you may need to restart Python in
order to load all the package's dependencies. You can do this by selecting the
'Restart kernel' or 'Restart runtime' option.
```

```
import nltk

# Download NLTK 'punkt' tokenizer data for sentence tokenization
nltk.download('punkt')
```

```
# Download 'punkt_tab' for comprehensive NLTK tokenization resources
nltk.download('punkt_tab')
```

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package punkt_tab to /root/nltk_data...
[nltk_data]   Package punkt_tab is already up-to-date!
True
```

```
import nltk
import spacy

# Assuming 'medical_text' is already defined in the environment
# If not, you can uncomment and define it here:
# medical_text = "A 45-year-old female presents with persistent migraines and visual disturbances

print("--- NLTK Tokenization ---")
# NLTK Sentence Tokenization
sentences_nltk = nltk.sent_tokenize(medical_text)
print("NLTK Sentences:", sentences_nltk)

# NLTK Word Tokenization (for the first sentence as an example)
words_nltk = [nltk.word_tokenize(sent) for sent in sentences_nltk]
print("NLTK Words (first sentence):", words_nltk[0])

print("\n--- spaCy Tokenization ---")
# Load the spaCy English model
nlp = spacy.load('en_core_web_sm')

# Process the medical text with spaCy
doc = nlp(medical_text)

# spaCy Sentence Tokenization
sentences_spacy = [sent.text for sent in doc.sents]
print("spaCy Sentences:", sentences_spacy)

# spaCy Word Tokenization (including punctuation as tokens)
words_spacy = [token.text for token in doc]
print("spaCy Words:", words_spacy)

# spaCy Word Tokenization (excluding punctuation for a cleaner word list)
words_spacy_no_punct = [token.text for token in doc if not token.is_punct and not token.is_space]
print("spaCy Words (excluding punctuation):", words_spacy_no_punct)
```

```
--- NLTK Tokenization ---
NLTK Sentences: ['Diabetes is a chronic disease that affects how the body processes blood sugar\nI
NLTK Words (first sentence): ['Diabetes', 'is', 'a', 'chronic', 'disease', 'that', 'affects', 'how

--- spaCy Tokenization ---
spaCy Sentences: ['Diabetes is a chronic disease that affects how the body processes blood sugar\n
spaCy Words: ['Diabetes', 'is', 'a', 'chronic', 'disease', 'that', 'affects', 'how', 'the', 'body'
spaCy Words (excluding punctuation): ['Diabetes', 'is', 'a', 'chronic', 'disease', 'that', 'affect
```

```
import nltk
from nltk.stem import PorterStemmer
import spacy

# Ensure medical_text is defined (it should be from previous steps)
# medical_text = "A 45-year-old female presents with persistent migraines and visual disturbances.

print("--- Stemming (NLTK PorterStemmer) ---")
```

```
# Tokenize words for stemming
words_for_stemming = nltk.word_tokenize(medical_text)
porter = PorterStemmer()
stemmed_words = [porter.stem(word) for word in words_for_stemming]
print("Original Words:", words_for_stemming)
print("Stemmed Words:", stemmed_words)

print("\n--- Lemmatization (spaCy) ---")
# Load the spaCy English model (if not already loaded)
try:
    nlp
except NameError:
    nlp = spacy.load('en_core_web_sm')

# Process the medical text with spaCy
doc_for_lemmatization = nlp(medical_text)

# Extract lemmas (excluding punctuation and spaces for cleaner output)
lemmatized_words = [token.lemma_ for token in doc_for_lemmatization if not token.is_punct and not
print("Original Tokens (spaCy):")
for token in doc_for_lemmatization:
    if not token.is_punct and not token.is_space:
        print(f"{token.text:<15} {token.pos_:<10} {token.lemma_:<15}")
print("\nLemmatized Words (spaCy):", lemmatized_words)
```

```
--- Stemming (NLTK PorterStemmer) ---
Original Words: ['Diabetes', 'is', 'a', 'chronic', 'disease', 'that', 'affects', 'how', 'the', 'bo
Stemmed Words: ['diabet', 'is', 'a', 'chronic', 'diseas', 'that', 'affect', 'how', 'the', 'bodi',

--- Lemmatization (spaCy) ---
Original Tokens (spaCy):
Diabetes        NOUN       diabetes
is              AUX        be
a               DET        a
chronic         ADJ        chronic
disease         NOUN       disease
that            PRON       that
affects         VERB       affect
how             SCONJ      how
the             DET        the
body            NOUN       body
processes       VERB       process
blood           NOUN       blood
sugar           NOUN       sugar
If              SCONJ      if
untreated       VERB       untreate
diabetes        VERB       diabete
may             AUX        may
cause           VERB       cause
heart           NOUN       heart
disease         NOUN       disease
kidney          NOUN       kidney
failure         NOUN       failure
nerve           NOUN       nerve
damage          NOUN       damage
and             CCONJ      and
vision          NOUN       vision
problems        NOUN       problem
Early           ADJ        early
diagnosis       NOUN       diagnosis
and             CCONJ      and
proper          ADJ        proper
treatment       NOUN       treatment
```

```
help            NOUN       help
improve         VERB       improve
patient         ADJ        patient
outcomes        NOUN       outcome

Lemmatized Words (spaCy): ['diabetes', 'be', 'a', 'chronic', 'disease', 'that', 'affect', 'how', '
```