```
!pip install nltk spacy

import nltk
import spacy

# Download NLTK data for tokenization, stemming, and lemmatization
nltk.download('punkt')
nltk.download('wordnet')
nltk.download('averaged_perceptron_tagger') # For spaCy's part-of-speech tagging for lemmatization
nltk.download('punkt_tab') # Added: Download punkt_tab resource for NLTK sentence tokenization
nltk.download('averaged_perceptron_tagger_eng') # Added: Download for NLTK POS tagger for lemmatization

# Download spaCy's English model
# You might need to restart the runtime after installing spaCy for this to work correctly on some systems
try:
    spacy.load('en_core_web_sm')
except OSError:
    print('Downloading spaCy model en_core_web_sm...')
    spacy.cli.download('en_core_web_sm')
    print('Download complete. Restart your runtime if you encounter issues.')

# Import specific modules after installation and download
from nltk.tokenize import word_tokenize
from nltk.stem import PorterStemmer, WordNetLemmatizer

# Initialize lemmatizer and stemmer
lemmatizer = WordNetLemmatizer()
stemmer = PorterStemmer()

# Load the spaCy model
nlp = spacy.load('en_core_web_sm')

print("NLTK, spaCy, and required resources are set up.")
```

```
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) (1.5.3)
Requirement already satisfied: regex>=2021.8.3 in /usr/local/lib/python3.12/dist-packages (from nltk) (2025.11.3)
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-packages (from spacy
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in /usr/local/lib/python3.12/dist-packages (from spacy
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/python3.12/dist-packages (from spacy)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /usr/local/lib/python3.12/dist-packages (from spacy) (2.0.1
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python3.12/dist-packages (from spacy) (3.0
Requirement already satisfied: thinc<8.4.0,>=8.3.4 in /usr/local/lib/python3.12/dist-packages (from spacy) (8.3.1
Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in /usr/local/lib/python3.12/dist-packages (from spacy) (1.1.
Requirement already satisfied: srsly<3.0.0,>=2.4.3 in /usr/local/lib/python3.12/dist-packages (from spacy) (2.5.2
Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in /usr/local/lib/python3.12/dist-packages (from spacy) (2
Requirement already satisfied: weasel<0.5.0,>=0.4.2 in /usr/local/lib/python3.12/dist-packages (from spacy) (0.4.
Requirement already satisfied: typer-slim<1.0.0,>=0.3.0 in /usr/local/lib/python3.12/dist-packages (from spacy) (
Requirement already satisfied: numpy>=1.19.0 in /usr/local/lib/python3.12/dist-packages (from spacy) (2.0.2)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in /usr/local/lib/python3.12/dist-packages (from spacy) (2
Requirement already satisfied: pydantic!=1.8,!=1.8.1,<3.0.0,>=1.7.4 in /usr/local/lib/python3.12/dist-packages (f
Requirement already satisfied: jinja2 in /usr/local/lib/python3.12/dist-packages (from spacy) (3.1.6)
Requirement already satisfied: setuptools in /usr/local/lib/python3.12/dist-packages (from spacy) (75.2.0)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.12/dist-packages (from spacy) (25.0)
Requirement already satisfied: annotated-types>=0.6.0 in /usr/local/lib/python3.12/dist-packages (from pydantic!=
Requirement already satisfied: pydantic-core==2.41.4 in /usr/local/lib/python3.12/dist-packages (from pydantic!=1
Requirement already satisfied: typing-extensions>=4.14.1 in /usr/local/lib/python3.12/dist-packages (from pydanti
Requirement already satisfied: typing-inspection>=0.4.2 in /usr/local/lib/python3.12/dist-packages (from pydantic
Requirement already satisfied: charset_normalizer<4,>=2 in /usr/local/lib/python3.12/dist-packages (from requests
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.12/dist-packages (from requests<3.0.0,>=2.1
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.12/dist-packages (from requests<3.0.0
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.12/dist-packages (from requests<3.0.0
Requirement already satisfied: blis<1.4.0,>=1.3.0 in /usr/local/lib/python3.12/dist-packages (from thinc<8.4.0,>=
Requirement already satisfied: confection<1.0.0,>=0.0.1 in /usr/local/lib/python3.12/dist-packages (from thinc<8.
Requirement already satisfied: cloudpathlib<1.0.0,>=0.7.0 in /usr/local/lib/python3.12/dist-packages (from weasel
Requirement already satisfied: smart-open<8.0.0,>=5.2.1 in /usr/local/lib/python3.12/dist-packages (from weasel<0
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.12/dist-packages (from jinja2->spacy) (3
Requirement already satisfied: wrapt in /usr/local/lib/python3.12/dist-packages (from smart-open<8.0.0,>=5.2.1->w
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package wordnet to /root/nltk_data...
[nltk_data]   Package wordnet is already up-to-date!
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data]     /root/nltk_data...
[nltk_data]   Package averaged_perceptron_tagger is already up-to-
[nltk_data]       date!
[nltk_data] Downloading package punkt_tab to /root/nltk_data...
[nltk_data]   Package punkt_tab is already up-to-date!
[nltk_data] Downloading package averaged_perceptron_tagger_eng to
[nltk_data]     /root/nltk_data...
[nltk_data]   Unzipping taggers/averaged_perceptron_tagger_eng.zip.
NLTK, spaCy, and required resources are set up.
```

```
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.
"""

print("Medical text loaded successfully:")
print(medical_text)
```

```
Medical text loaded successfully:

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.
```

```
from nltk.tokenize import sent_tokenize

sentences_nltk = sent_tokenize(medical_text)

print("NLTK Sentence Tokenization:")
for i, sentence in enumerate(sentences_nltk):
    print(f"Sentence {i+1}: {sentence}")
```

```
NLTK Sentence Tokenization:
Sentence 1:
Diabetes is a chronic disease that affects how the body processes blood sugar.
Sentence 2: If untreated, diabetes may cause heart disease, kidney failure, nerve damage and vision problems.
Sentence 3: Early diagnosis and proper treatment help improve patient outcomes.
```

```
print("\n--- NLTK Word Tokenization ---")
words_nltk = word_tokenize(medical_text)
print(words_nltk)

print("\n--- spaCy Word Tokenization ---")
doc_spacy = nlp(medical_text)
words_spacy = [token.text for token in doc_spacy]
print(words_spacy)
```

```

--- NLTK Word Tokenization ---
['The', 'patient', 'presented', 'with', 'acute', 'myocardial', 'infarction', ',', 'confirmed', 'by', 'elevated',

--- spaCy Word Tokenization ---
['The', 'patient', 'presented', 'with', 'acute', 'myocardial', 'infarction', ',', 'confirmed', 'by', 'elevated',
```

```
print("\n--- NLTK Porter Stemming ---")
stemmed_words_nltk = [stemmer.stem(word) for word in words_nltk]

for original, stemmed in zip(words_nltk, stemmed_words_nltk):
    print(f"'{original}' -> '{stemmed}'")
```

```

--- NLTK Porter Stemming ---
'The' -> 'the'
'patient' -> 'patient'
'presented' -> 'present'
'with' -> 'with'
'acute' -> 'acut'
'myocardial' -> 'myocardi'
'infarction' -> 'infarct'
',' -> ','
'confirmed' -> 'confirm'
'by' -> 'by'
'elevated' -> 'elev'
'troponin' -> 'troponin'
'levels' -> 'level'
'and' -> 'and'
'ECG' -> 'ecg'
'changes' -> 'chang'
'.' -> '.'
'A' -> 'a'
'history' -> 'histori'
'of' -> 'of'
'hypertension' -> 'hypertens'
'and' -> 'and'
'hyperlipidemia' -> 'hyperlipidemia'
'was' -> 'wa'
'noted' -> 'note'
'.' -> '.'
'Treatment' -> 'treatment'
'involved' -> 'involv'
```

```
'immediate' -> 'immedi'
'reperfusion' -> 'reperfus'
'therapy' -> 'therapi'
',' -> ','
'antiplatelets' -> 'antiplatelet'
',' -> ','
'and' -> 'and'
'beta-blockers' -> 'beta-block'
'.' -> '.'
'Follow-up' -> 'follow-up'
'is' -> 'is'
'recommended' -> 'recommend'
'for' -> 'for'
'cardiac' -> 'cardiac'
'rehabilitation' -> 'rehabilit'
'.' -> '.'
```

```
from nltk.corpus import wordnet

# Helper function to convert NLTK POS tags to WordNet POS tags for lemmatization
def get_wordnet_pos(word):
    tag = nltk.pos_tag([word])[0][1][0].upper()
    tag_dict = {"J": wordnet.ADJ, "N": wordnet.NOUN, "V": wordnet.VERB, "R": wordnet.ADV}
    return tag_dict.get(tag, wordnet.NOUN)

print("\n--- NLTK WordNet Lemmatization ---")
lemmas_nltk = [lemmatizer.lemmatize(word, get_wordnet_pos(word)) for word in words_nltk]

for original, lemma in zip(words_nltk, lemmas_nltk):
    print(f"'{original}' -> '{lemma}'")

print("\n--- spaCy Lemmatization ---")
lemmas_spacy = [token.lemma_ for token in doc_spacy]

for original, lemma in zip(words_spacy, lemmas_spacy):
    print(f"'{original}' -> '{lemma}'")
```

```
'.' -> '.'
'Follow-up' -> 'Follow-up'
'is' -> 'be'
'recommended' -> 'recommend'
'for' -> 'for'
'cardiac' -> 'cardiac'
'rehabilitation' -> 'rehabilitation'
'.' -> '.'

--- spaCy Lemmatization ---
'The' -> 'the'
'patient' -> 'patient'
'presented' -> 'present'
```

```
'up' -> 'up'
'is' -> 'be'
'recommended' -> 'recommend'
'for' -> 'for'
'cardiac' -> 'cardiac'
'rehabilitation' -> 'rehabilitation'
'.' -> '.'
```

```
print("\n--- Comparison: Original vs. Stemmed vs. NLTK Lemmatized vs. spaCy Lemmatized ---")
print(f"{'Original':<15} {'Stemmed (NLTK)':<20} {'Lemmatized (NLTK)':<25} {'Lemmatized (spaCy)':<20}")
print(f"{'-'*15:<15} {'-'*20:<20} {'-'*25:<25} {'-'*20:<20}")

# Ensure all lists are of similar length for zipping, using NLTK words as base
min_len = min(len(words_nltk), len(stemmed_words_nltk), len(lemmas_nltk), len(lemmas_spacy))

for i in range(min_len):
    original_word = words_nltk[i]
    stemmed_word = stemmed_words_nltk[i]
    lemma_nltk = lemmas_nltk[i]
    lemma_spacy = lemmas_spacy[i]
    print(f"{original_word:<15} {stemmed_word:<20} {lemma_nltk:<25} {lemma_spacy:<20}")
```

```
--- Comparison: Original vs. Stemmed vs. NLTK Lemmatized vs. spaCy Lemmatized ---
Original        Stemmed (NLTK)       Lemmatized (NLTK)         Lemmatized (spaCy)
--------------- -------------------- ------------------------- --------------------
The             the                  The                       the
patient         patient              patient                   patient
presented       present              present                   present
with            with                 with                      with
acute           acut                 acute                     acute
myocardial      myocardi             myocardial                myocardial
infarction      infarct              infarction                infarction
,               ,                    ,                         ,
confirmed       confirm              confirm                   confirm
by              by                   by                        by
elevated        elev                 elevate                   elevated
troponin        troponin             troponin                  troponin
levels          level                level                     level
and             and                  and                       and
ECG             ecg                  ECG                       ECG
changes         chang                change                    change
.               .                    .                         .
A               a                    A                         a
history         histori              history                   history
of              of                   of                        of
hypertension    hypertens            hypertension              hypertension
and             and                  and                       and
hyperlipidemia  hyperlipidemia       hyperlipidemia            hyperlipidemia
was             wa                   be                        be
noted           note                 note                      note
.               .                    .                         .
Treatment       treatment            Treatment                 treatment
involved        involv               involve                   involve
immediate       immedi               immediate                 immediate
reperfusion     reperfus             reperfusion               reperfusion
therapy         therapi              therapy                   therapy
,               ,                    ,                         ,
antiplatelets   antiplatelet         antiplatelets             antiplatelet
,               ,                    ,                         ,
and             and                  and                       and
beta-blockers   beta-block           beta-blockers             beta
.               .                    .                         -
Follow-up       follow-up            Follow-up                 blocker
is              is                   be                        .
recommended     recommend            recommend                 follow
for             for                  for                       -
cardiac         cardiac              cardiac                   up
rehabilitation  rehabilit            rehabilitation            be
.               .                    .                         recommend
```

```
Start coding or generate with AI.
```