```
medical = """A 29-year-old female presents with "crushing fatigue" and muscle weakness that has persisted for six months. Sh

  She has a history of Iron Deficiency Anemia and reports Restless Leg Syndrome (RLS) that prevents her from reaching deep s
```

```
!pip install nltk
```

```
Requirement already satisfied: nltk in /usr/local/lib/python3.12/dist-packages (3.9.1)
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)
```

```
!pip install spacy
```

```
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-packages (from spacy) (3.0.12)
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in /usr/local/lib/python3.12/dist-packages (from spacy) (1.0.5)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/python3.12/dist-packages (from spacy) (1.0.15)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /usr/local/lib/python3.12/dist-packages (from spacy) (2.0.13)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python3.12/dist-packages (from spacy) (3.0.12)
Requirement already satisfied: thinc<8.4.0,>=8.3.4 in /usr/local/lib/python3.12/dist-packages (from spacy) (8.3.10)
Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in /usr/local/lib/python3.12/dist-packages (from spacy) (1.1.3)
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) (2.0.10)
Requirement already satisfied: weasel<0.5.0,>=0.4.2 in /usr/local/lib/python3.12/dist-packages (from spacy) (0.4.3)
Requirement already satisfied: typer-slim<1.0.0,>=0.3.0 in /usr/local/lib/python3.12/dist-packages (from spacy) (0.20.0)
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in /usr/local/lib/python3.12/dist-packages (from spacy) (4.67.1)
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) (2.32.4)
Requirement already satisfied: pydantic!=1.8,!=1.8.1,<3.0.0,>=1.7.4 in /usr/local/lib/python3.12/dist-packages (from spacy)
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!=1.8,!=1.8.1
Requirement already satisfied: pydantic-core==2.41.4 in /usr/local/lib/python3.12/dist-packages (from pydantic!=1.8,!=1.8.1,
Requirement already satisfied: typing-extensions>=4.14.1 in /usr/local/lib/python3.12/dist-packages (from pydantic!=1.8,!=1.
Requirement already satisfied: typing-inspection>=0.4.2 in /usr/local/lib/python3.12/dist-packages (from pydantic!=1.8,!=1.8
Requirement already satisfied: charset_normalizer<4,>=2 in /usr/local/lib/python3.12/dist-packages (from requests<3.0.0,>=2.
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.12/dist-packages (from requests<3.0.0,>=2.13.0->spacy)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.12/dist-packages (from requests<3.0.0,>=2.13.0->
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.12/dist-packages (from requests<3.0.0,>=2.13.0->
Requirement already satisfied: blis<1.4.0,>=1.3.0 in /usr/local/lib/python3.12/dist-packages (from thinc<8.4.0,>=8.3.4->spac
Requirement already satisfied: confection<1.0.0,>=0.0.1 in /usr/local/lib/python3.12/dist-packages (from thinc<8.4.0,>=8.3.4
Requirement already satisfied: click>=8.0.0 in /usr/local/lib/python3.12/dist-packages (from typer-slim<1.0.0,>=0.3.0->spacy
Requirement already satisfied: cloudpathlib<1.0.0,>=0.7.0 in /usr/local/lib/python3.12/dist-packages (from weasel<0.5.0,>=0.
Requirement already satisfied: smart-open<8.0.0,>=5.2.1 in /usr/local/lib/python3.12/dist-packages (from weasel<0.5.0,>=0.4.
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.12/dist-packages (from jinja2->spacy) (3.0.3)
Requirement already satisfied: wrapt in /usr/local/lib/python3.12/dist-packages (from smart-open<8.0.0,>=5.2.1->weasel<0.5.0
```

```
print(medical)
```

```
A 29-year-old female presents with "crushing fatigue" and muscle weakness that has persisted for six months. She recently tr

  She has a history of Iron Deficiency Anemia and reports Restless Leg Syndrome (RLS) that prevents her from reaching deep s
```

Tockenisation by words.

```
import nltk
nltk.download('punkt')
nltk.download('punkt_tab')
from nltk.tokenize import word_tokenize
word_tokenize(medical)
```

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt.zip.
[nltk_data] Downloading package punkt_tab to /root/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt_tab.zip.
['A',
 '29-year-old',
 'female',
 'presents',
 'with',
 '``',
 'crushing',
 'fatigue',
 "''",
 'and',
 'muscle',
 'weakness',
```

```
 'that',
 'has',
 'persisted',
 'for',
 'six',
 'months',
 '.',
 'She',
 'recently',
 'transitioned',
 'to',
 'a',
 'strict',
 'vegan',
 'diet',
 'and',
 'works',
 'in',
 'a',
 'windowless',
 'office',
 'environment',
 '.',
 'She',
 'has',
 'a',
 'history',
 'of',
 'Iron',
 'Deficiency',
 'Anemia',
 'and',
 'reports',
 'Restless',
 'Leg',
 'Syndrome',
 '(',
 'RLS',
 ')',
 'that',
 'prevents',
 'her'
```

Tockenisation by sentence.

```
from nltk.tokenize import sent_tokenize
sent_tokenize(medical)
```

```
['A 29-year-old female presents with "crushing fatigue" and muscle weakness that has persisted for six months.',
 'She recently transitioned to a strict vegan diet and works in a windowless office environment.',
 'She has a history of Iron Deficiency Anemia and reports Restless Leg Syndrome (RLS) that prevents her from reaching deep
sleep.',
 'She also mentions Raynaud’s Phenomenon (fingers turning white/blue in the cold).']
```

Stemming

```
from nltk.stem import PorterStemmer
from nltk.tokenize import word_tokenize
stemmer = PorterStemmer()
words = word_tokenize(medical)
stemmed_words = [stemmer.stem(word) for word in words]
stemmed_words
```

```
['a',
 '29-year-old',
 'femal',
 'present',
 'with',
 '``',
 'crush',
 'fatigu',
 "''",
 'and',
 'muscl',
 'weak',
 'that',
 'ha',
 'persist',
 'for',
 'six',
 'month',
 '.',
 'she',
 'recent',
 'transit',
```

```
 'to',
 'a',
 'strict',
 'vegan',
 'diet',
 'and',
 'work',
 'in',
 'a',
 'windowless',
 'offic',
 'environ',
 '.',
 'she',
 'ha',
 'a',
 'histori',
 'of',
 'iron',
 'defici',
 'anemia',
 'and',
 'report',
 'restless',
 'leg',
 'syndrom',
 '(',
 'rl',
 ')',
 'that',
 'prevent',
 'her',
 'from',
 'reach',
 'deep',
```

lemmatization

```
nltk.download('omw-1.4')
nltk.download('wordnet') # Added: Download the 'wordnet' corpus
from nltk.stem import WordNetLemmatizer
lemmatizer = WordNetLemmatizer()
words = word_tokenize(medical)
for word in words:
    print(word,"--->",lemmatizer.lemmatize(word))
```

```
[nltk_data] Downloading package omw-1.4 to /root/nltk_data...
[nltk_data]   Package omw-1.4 is already up-to-date!
[nltk_data] Downloading package wordnet to /root/nltk_data...
A ---> A
29-year-old ---> 29-year-old
female ---> female
presents ---> present
with ---> with
`` ---> ``
crushing ---> crushing
fatigue ---> fatigue
'' ---> ''
and ---> and
muscle ---> muscle
weakness ---> weakness
that ---> that
has ---> ha
persisted ---> persisted
for ---> for
six ---> six
months ---> month
. ---> .
She ---> She
recently ---> recently
transitioned ---> transitioned
to ---> to
a ---> a
strict ---> strict
vegan ---> vegan
diet ---> diet
and ---> and
works ---> work
in ---> in
a ---> a
windowless ---> windowless
office ---> office
environment ---> environment
. ---> .
She ---> She
has ---> ha
a ---> a
history ---> history
```

```
of ---> of
Iron ---> Iron
Deficiency ---> Deficiency
Anemia ---> Anemia
and ---> and
reports ---> report
Restless ---> Restless
Leg ---> Leg
Syndrome ---> Syndrome
( ---> (
RLS ---> RLS
) ---> )
that ---> that
prevents ---> prevents
her ---> her
from ---> from
```

Understemming

```
from nltk.stem import SnowballStemmer
snowball = SnowballStemmer(language='english')
words = word_tokenize(medical)
for word in words:
    print(word,"--->",snowball.stem(word))
```

```
A ---> a
29-year-old ---> 29-year-old
female ---> femal
presents ---> present
with ---> with
`` ---> ``
crushing ---> crush
fatigue ---> fatigu
'' ---> ''
and ---> and
muscle ---> muscl
weakness ---> weak
that ---> that
has ---> has
persisted ---> persist
for ---> for
six ---> six
months ---> month
. ---> .
She ---> she
recently ---> recent
transitioned ---> transit
to ---> to
a ---> a
strict ---> strict
vegan ---> vegan
diet ---> diet
and ---> and
works ---> work
in ---> in
a ---> a
windowless ---> windowless
office ---> offic
environment ---> environ
. ---> .
She ---> she
has ---> has
a ---> a
history ---> histori
of ---> of
Iron ---> iron
Deficiency ---> defici
Anemia ---> anemia
and ---> and
reports ---> report
Restless ---> restless
Leg ---> leg
Syndrome ---> syndrom
( ---> (
RLS ---> rls
) ---> )
that ---> that
prevents ---> prevent
her ---> her
from ---> from
reaching ---> reach
deep ---> deep
sleep ---> sleep
```

overstemming

```
from nltk.stem import RegexpStemmer
from nltk.tokenize import word_tokenize
regexp = RegexpStemmer('ing|e', min=4) # Fixed: combined regex pattern into a single line
words = word_tokenize(medical)
for word in words:
    print(word,"--->",regexp.stem(word)) # Fixed: changed 'Lanc' to 'regexp'
```

```
A ---> A
29-year-old ---> 29-year-old
female ---> fmal
presents ---> prsnts
with ---> with
`` ---> ``
crushing ---> crush
fatigue ---> fatigu
'' ---> ''
and ---> and
muscle ---> muscl
weakness ---> waknss
that ---> that
has ---> has
persisted ---> prsistd
for ---> for
six ---> six
months ---> months
. ---> .
She ---> She
recently ---> rcntly
transitioned ---> transitiond
to ---> to
a ---> a
strict ---> strict
vegan ---> vgan
diet ---> dit
and ---> and
works ---> works
in ---> in
a ---> a
windowless ---> windowlss
office ---> offic
environment ---> nvironmnt
. ---> .
She ---> She
has ---> has
a ---> a
history ---> history
of ---> of
Iron ---> Iron
Deficiency ---> Dficincy
Anemia ---> Anmia
and ---> and
reports ---> rports
Restless ---> Rstlss
Leg ---> Leg
Syndrome ---> Syndrom
( ---> (
RLS ---> RLS
) ---> )
that ---> that
prevents ---> prvnts
her ---> her
from ---> from
reaching ---> rach
deep ---> dp
sleep ---> slp
```

Filtering Stop Words

```
medical = """A 29-year-old female presents with "crushing fatigue" and muscle weakness that has persisted for six months. S

  She has a history of Iron Deficiency Anemia and reports Restless Leg Syndrome (RLS) that prevents her from reaching deep
import nltk
nltk.download("stopwords")
nltk.download('punkt_tab') # Added to ensure punkt_tab is available
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize

words_in_quote = word_tokenize(medical)
words_in_quote
```

```
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
[nltk_data] Downloading package punkt_tab to /root/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt_tab.zip.
['A',
 '29-year-old',
 'female',
```

```
    'presents',
    'with',
    '``',
    'crushing',
    'fatigue',
    "''",
    'and',
    'muscle',
    'weakness',
    'that',
    'has',
    'persisted',
    'for',
    'six',
    'months',
    '.',
    'She',
    'recently',
    'transitioned',
    'to',
    'a',
    'strict',
    'vegan',
    'diet',
    'and',
    'works',
    'in',
    'a',
    'windowless',
    'office',
    'environment',
    '.',
    'She',
    'has',
    'a',
    'history',
    'of',
    'Iron',
    'Deficiency',
    'Anemia',
    'and',
    'reports',
    'Restless',
    'Leg',
    'Syndrome',
    '(',
    'RLS',
    ')',
    'that',
    'prevents',
    'her'
```

Comparison

Discussion Section

1.NLP preprocessing techniques were applied to a clinical medical text using the NLTK library.

2.Word tokenization converted the raw text into individual words, making analysis easier.

3.Sentence tokenization separated the text into meaningful sentences representing different symptoms and observations.

4.Stemming reduced words to their root forms but sometimes produced inaccurate or incomplete words.

5.Lemmatization provided more meaningful base words compared to stemming, improving readability.

6.The experiment shows that basic NLP methods can structure unstructured medical data, but advanced models are needed for clinical accuracy.