

```
!pip install nltk spacy
!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) (1.5.3)
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) (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) (0.21.1)
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) (2.12.3)
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,<3.0.0,>=1.7.4->spacy) (0.7.0)
Requirement already satisfied: pydantic-core==2.41.4 in /usr/local/lib/python3.12/dist-packages (from pydantic!=1.8,!=1.8.1,<3.0.0,>=1.7.4->spacy) (2.41.4)
Requirement already satisfied: typing-extensions>=4.14.1 in /usr/local/lib/python3.12/dist-packages (from pydantic!=1.8,!=1.8.1,<3.0.0,>=1.7.4->spacy) (4.15.0)
Requirement already satisfied: typing-inspection>=0.4.2 in /usr/local/lib/python3.12/dist-packages (from pydantic!=1.8,!=1.8.1,<3.0.0,>=1.7.4->spacy) (0.4.2)
Requirement already satisfied: charset_normalizer<4,>=2 in /usr/local/lib/python3.12/dist-packages (from requests<3.0.0,>=2.13.0->spacy) (3.4.4)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.12/dist-packages (from requests<3.0.0,>=2.13.0->spacy) (3.11)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.12/dist-packages (from requests<3.0.0,>=2.13.0->spacy) (2.5.0)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.12/dist-packages (from requests<3.0.0,>=2.13.0->spacy) (2026.1.4)
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->spacy) (1.3.3)
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->spacy) (0.1.5)
Requirement already satisfied: cloudpathlib<1.0.0,>=0.7.0 in /usr/local/lib/python3.12/dist-packages (from weasel<0.5.0,>=0.4.2->spacy) (0.23.0)
Requirement already satisfied: smart-open<8.0.0,>=5.2.1 in /usr/local/lib/python3.12/dist-packages (from weasel<0.5.0,>=0.4.2->spacy) (7.5.0)
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,>=0.4.2->spacy) (2.0.1)
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_core_web_sm-3.8.0-py3-none-any.whl (12.8 MB)
    12.8/12.8 MB 94.0 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 libraries
import nltk
import spacy
import pandas as pd
```

```
import matplotlib.pyplot as plt
from collections import Counter
```

```
# Download NLTK resources
nltk.download('punkt')
nltk.download('averaged_perceptron_tagger')
```

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]  Unzipping tokenizers/punkt.zip.
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data]      /root/nltk_data...
[nltk_data]  Unzipping taggers/averaged_perceptron_tagger.zip.
True
```

```
essay_text = """
Academic writing plays a crucial role in higher education.
It allows researchers to present arguments, analyze data,
and communicate findings in a structured and formal manner.
Effective academic writing requires clarity, precision,
and the use of discipline-specific vocabulary.
"""

"""

```

```
print("Academic Essay Text:\n")
print(essay_text)
```

Academic Essay Text:

```
Academic writing plays a crucial role in higher education.
It allows researchers to present arguments, analyze data,
and communicate findings in a structured and formal manner.
Effective academic writing requires clarity, precision,
and the use of discipline-specific vocabulary.
```

```
nltk.download('punkt_tab')
tokens = nltk.word_tokenize(essay_text)
print(tokens)
```

```
[nltk_data] Downloading package punkt_tab to /root/nltk_data...
[nltk_data]  Unzipping tokenizers/punkt_tab.zip.
['Academic', 'writing', 'plays', 'a', 'crucial', 'role', 'in', 'higher', 'education', '.', 'It', 'allows', 'researchers', 'to', 'present', 'arguments', ',', 'analyze',
```

```
nltk.download('averaged_perceptron_tagger_eng')
nltk_pos_tags = nltk.pos_tag(tokens)
nltk_pos_tags
```

```
[nltk_data] Downloading package averaged_perceptron_tagger_eng to
[nltk_data]      /root/nltk_data...
[nltk_data]  Unzipping taggers/averaged_perceptron_tagger_eng.zip.
```

```
[('Academic', 'NNP'),
 ('writing', 'NN'),
 ('plays', 'VBZ'),
 ('a', 'DT'),
 ('crucial', 'JJ'),
 ('role', 'NN'),
 ('in', 'IN'),
 ('higher', 'JJR'),
 ('education', 'NN'),
 ('.', '.'),
 ('It', 'PRP'),
 ('allows', 'VBZ'),
 ('researchers', 'NNS'),
 ('to', 'TO'),
 ('present', 'JJ'),
 ('arguments', 'NNS'),
 ('', ''),
 ('analyze', 'NN'),
 ('data', 'NNS'),
 ('', ''),
 ('and', 'CC'),
 ('communicate', 'JJ'),
 ('findings', 'NNS'),
 ('in', 'IN'),
 ('a', 'DT'),
 ('structured', 'JJ'),
 ('and', 'CC'),
 ('formal', 'JJ'),
 ('manner', 'NN'),
 ('.', '.'),
 ('Effective', 'JJ'),
 ('academic', 'JJ'),
 ('writing', 'NN'),
 ('requires', 'VBZ'),
 ('clarity', 'NN'),
 ('', ''),
 ('precision', 'NN'),
 ('', ''),
 ('and', 'CC'),
 ('the', 'DT'),
 ('use', 'NN'),
 ('of', 'IN'),
 ('discipline-specific', 'JJ'),
 ('vocabulary', 'NN'),
 ('.', '.)]
```

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

spacy_pos_tags = [(token.text, token.pos_) for token in doc]
spacy_pos_tags
```

```
[('\n', 'SPACE'),
 ('Academic', 'ADJ'),
 ('writing', 'NOUN'),
 ('plays', 'VERB'),
 ('a', 'DET'),
 ('crucial', 'ADJ'),
 ('role', 'NOUN'),
```

```
('in', 'ADP'),
('higher', 'ADJ'),
('education', 'NOUN'),
('.', 'PUNCT'),
('\n', 'SPACE'),
('It', 'PRON'),
('allows', 'VERB'),
('researchers', 'NOUN'),
('to', 'ADP'),
('present', 'ADJ'),
('arguments', 'NOUN'),
('', 'PUNCT'),
('analyze', 'VERB'),
('data', 'NOUN'),
('', 'PUNCT'),
('\n', 'SPACE'),
('and', 'CCONJ'),
('communicate', 'VERB'),
('findings', 'NOUN'),
('in', 'ADP'),
('a', 'DET'),
('structured', 'ADJ'),
('and', 'CCONJ'),
('formal', 'ADJ'),
('manner', 'NOUN'),
('.', 'PUNCT'),
('\n', 'SPACE'),
('Effective', 'ADJ'),
('academic', 'ADJ'),
('writing', 'NOUN'),
('requires', 'VERB'),
('clarity', 'NOUN'),
('', 'PUNCT'),
('precision', 'NOUN'),
('', 'PUNCT'),
('\n', 'SPACE'),
('and', 'CCONJ'),
('the', 'DET'),
('use', 'NOUN'),
('of', 'ADP'),
('discipline', 'NOUN'),
('-', 'PUNCT'),
('specific', 'ADJ'),
('vocabulary', 'NOUN'),
('.', 'PUNCT'),
('\n', 'SPACE')]
```

```
nouns = [token.text.lower() for token in doc if token.pos_ == "NOUN"]
verbs = [token.text.lower() for token in doc if token.pos_ == "VERB"]

print("Nouns (Concepts):", nouns)
print("Verbs (Arguments):", verbs)
```

```
Nouns (Concepts): ['writing', 'role', 'education', 'researchers', 'arguments', 'data', 'findings', 'manner', 'writing', 'clarity', 'precision', 'use', 'discipline', 'vo
Verbs (Arguments): ['plays', 'allows', 'analyze', 'communicate', 'requires']
```

```
noun_freq = Counter(nouns)
verb_freq = Counter(verbs)

noun_freq, verb_freq
```

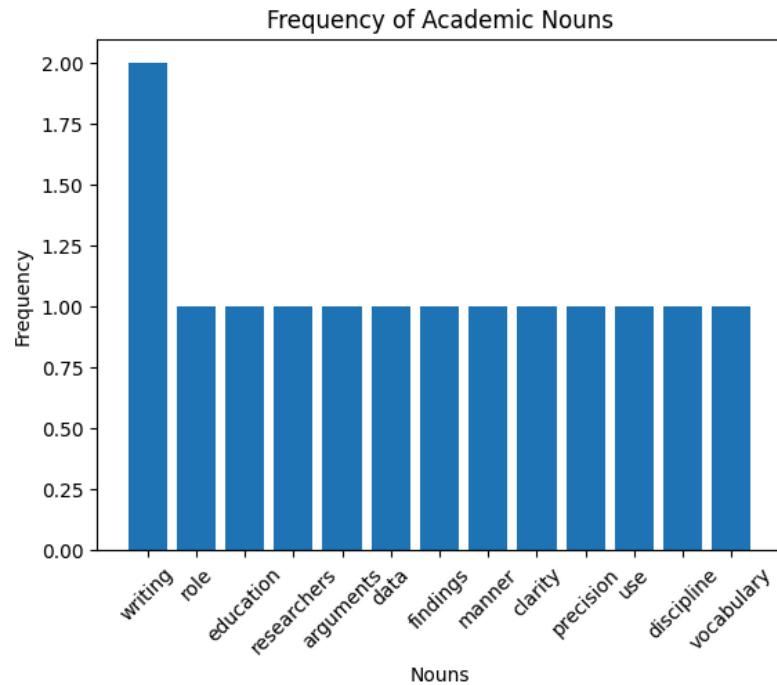
```
(Counter({'writing': 2,
          'role': 1,
          'education': 1,
          'researchers': 1,
          'arguments': 1,
          'data': 1,
          'findings': 1,
          'manner': 1,
          'clarity': 1,
          'precision': 1,
          'use': 1,
          'discipline': 1,
          'vocabulary': 1}),
Counter({'plays': 1,
          'allows': 1,
          'analyze': 1,
          'communicate': 1,
          'requires': 1}))
```

```
noun_df = pd.DataFrame(noun_freq.items(), columns=["Noun", "Frequency"])
verb_df = pd.DataFrame(verb_freq.items(), columns=["Verb", "Frequency"])
```

```
noun_df, verb_df
```

```
(   Noun  Frequency
0    writing      2
1      role      1
2   education      1
3  researchers      1
4   arguments      1
5     data      1
6   findings      1
7    manner      1
8    clarity      1
9   precision      1
10     use      1
11 discipline      1
12 vocabulary      1,
              Verb  Frequency
0      plays      1
1    allows      1
2   analyze      1
3  communicate      1
4   requires      1)
```

```
plt.figure()
plt.bar(noun_df["Noun"], noun_df["Frequency"])
plt.title("Frequency of Academic Nouns")
plt.xlabel("Nouns")
plt.ylabel("Frequency")
plt.xticks(rotation=45)
plt.show()
```



```
plt.figure()
plt.bar(verb_df["Verb"], verb_df["Frequency"])
plt.title("Frequency of Academic Verbs")
plt.xlabel("Verbs")
plt.ylabel("Frequency")
plt.xticks(rotation=45)
plt.show()
```

