

```
In [ ]: !pip install nltk
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: nltk in /usr/local/lib/python3.9/dist-packages (3.7)
Requirement already satisfied: tqdm in /usr/local/lib/python3.9/dist-packages (from nltk) (4.65.0)
Requirement already satisfied: regex>=2021.8.3 in /usr/local/lib/python3.9/dist-packages (from nltk) (2022.6.2)
Requirement already satisfied: click in /usr/local/lib/python3.9/dist-packages (from nltk) (8.1.3)
Requirement already satisfied: joblib in /usr/local/lib/python3.9/dist-packages (from nltk) (1.2.0)
```

```
In [ ]: nltk.download('punkt')
```

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

```
Out[ ]: True
```

```
In [ ]: import nltk
        from nltk.tokenize import sent_tokenize
```

```
In [ ]: text="India is a unique country with diversity. Unity is diversity is the main slogan
```

```
In [ ]: print(sent_tokenize(text))
```

```
['India is a unique country with diversity.', 'Unity is diversity is the main slogan of the country.']}
```

```
In [ ]: print(f'whitespace tokenization = {text.split()}')
```

```
whitespace tokenization = ['India', 'is', 'a', 'unique', 'country', 'with', 'diversity.', 'Unity', 'is', 'diversity', 'is', 'the', 'main', 'slogan', 'of', 'the', 'country.']}
```

```
In [ ]: from nltk.tokenize import wordpunct_tokenize
```

```
In [ ]: print(f'Punctuation-based tokenization = {wordpunct_tokenize(text)}')
```

```
Punctuation-based tokenization = ['India', 'is', 'a', 'unique', 'country', 'with', 'diversity', '.', 'Unity', 'is', 'diversity', 'is', 'the', 'main', 'slogan', 'of', 'the', 'country', '.']
```

```
In [ ]: from nltk.tokenize import TreebankWordTokenizer
```

```
In [ ]: sentence="What's your name?"
```

```
In [ ]: tokenizer = TreebankWordTokenizer()
print(f'Default/Treebank tokenization = {tokenizer.tokenize(sentence)}')
```

```
Default/Treebank tokenization = ['What', "'s", 'your', 'name', '?']
```

```
In [ ]: pip install emoji --upgrade
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Collecting emoji
  Downloading emoji-2.2.0.tar.gz (240 kB)
    Preparing metadata (setup.py) ... done
  Building wheels for collected packages: emoji
    Building wheel for emoji (setup.py) ... done
      Created wheel for emoji: filename=emoji-2.2.0-py3-none-any.whl size=234926 sha256=9242187f22f9ee5609d1d2d6b7ecff728350b4ca4c879e37f82b52fa7365bbfd
      Stored in directory: /root/.cache/pip/wheels/9a/b8/0f/f580817231cbf59f6ade9fd132ff60ada1de9f7dc85521f857
Successfully built emoji
Installing collected packages: emoji
Successfully installed emoji-2.2.0
```

```
In [ ]: import emoji
```

```
In [ ]: print(emoji.emojize('Hi Everyone! :grinning_face:'))
```

```
Hi Everyone! 😊
```

```
In [ ]: sentence1= emoji.emojize('Hi Everyone! :grinning face:')
```

```
In [ ]: from nltk.tokenize import TweetTokenizer
```

```
In [ ]: tokenizer = TweetTokenizer()
print(f'Tweet-rules based tokenization = {tokenizer.tokenize(sentence1)}')
```

```
Tweet-rules based tokenization = ['Hi', 'Everyone', '!', ':', 'grinning', 'face',
':']
```

```
In [ ]: sentence2="Hope, is the only thing stronger than fear! Hunger Games"
```

```
In [ ]: from nltk.corpus.reader.tagged import word_tokenize
print(word_tokenize(sentence2))
```

```
['Hope', ',', 'is', 'the', 'only', 'thing', 'stronger', 'than', 'fear', '!', 'Hunger',
', 'Games']
```

```
In [ ]: from nltk.tokenize import MWETokenizer
```

```
In [ ]: tokenizer = MWETokenizer()
tokenizer.add_mwe([('Hunger', 'Games')])
print(f'Multi-word expression (MWE) tokenization = {tokenizer.tokenize(word_tokenize(s))}')
```

```
Multi-word expression (MWE) tokenization = ['Hope', ',', 'is', 'the', 'only', 'thin
g', 'stronger', 'than', 'fear', '!', 'Hunger_Games']
```

```
In [ ]: # Import the toolkit and the full Porter Stemmer library
import nltk
from nltk.stem.porter import *
p_stemmer = PorterStemmer()
words = ['run', 'runner', 'running', 'ran', 'runs', 'easily', 'fairly']
for word in words:
  print(word+' --> '+p_stemmer.stem(word))
```

```
run --> run
runner --> runner
running --> run
ran --> ran
runs --> run
easily --> easili
fairly --> fairli
```

```
In [ ]: from nltk.stem.snowball import SnowballStemmer
# The Snowball Stemmer requires that you pass a Language parameter
s_stemmer = SnowballStemmer(language='english')
words = ['run', 'runner', 'running', 'ran', 'runs', 'easily', 'fairly']
for word in words:
    print(word+' --> '+s_stemmer.stem(word))
```

```
run --> run
runner --> runner
running --> run
ran --> ran
runs --> run
easily --> easili
fairly --> fair
```

```
In [ ]: !pip3 install spacy
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: spacy in /usr/local/lib/python3.9/dist-packages (3.4.4)
Requirement already satisfied: pydantic!=1.8,!_=1.8.1,<1.11.0,>=1.7.4 in /usr/local/lib/python3.9/dist-packages (from spacy) (1.10.5)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python3.9/dist-packages (from spacy) (3.0.8)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /usr/local/lib/python3.9/dist-packages (from spacy) (2.0.7)
Requirement already satisfied: jinja2 in /usr/local/lib/python3.9/dist-packages (from spacy) (3.1.2)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/python3.9/dist-packages (from spacy) (1.0.9)
Requirement already satisfied: numpy>=1.15.0 in /usr/local/lib/python3.9/dist-packages (from spacy) (1.22.4)
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.10 in /usr/local/lib/python3.9/dist-packages (from spacy) (3.0.12)
Requirement already satisfied: srsly<3.0.0,>=2.4.3 in /usr/local/lib/python3.9/dist-packages (from spacy) (2.4.6)
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in /usr/local/lib/python3.9/dist-packages (from spacy) (1.0.4)
Requirement already satisfied: setuptools in /usr/local/lib/python3.9/dist-packages (from spacy) (57.4.0)
Requirement already satisfied: smart-open<7.0.0,>=5.2.1 in /usr/local/lib/python3.9/dist-packages (from spacy) (6.3.0)
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in /usr/local/lib/python3.9/dist-packages (from spacy) (4.65.0)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in /usr/local/lib/python3.9/dist-packages (from spacy) (2.25.1)
Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in /usr/local/lib/python3.9/dist-packages (from spacy) (2.0.8)
Requirement already satisfied: wasabi<1.1.0,>=0.9.1 in /usr/local/lib/python3.9/dist-packages (from spacy) (0.10.1)
Requirement already satisfied: typer<0.8.0,>=0.3.0 in /usr/local/lib/python3.9/dist-packages (from spacy) (0.7.0)
Requirement already satisfied: thinc<8.2.0,>=8.1.0 in /usr/local/lib/python3.9/dist-packages (from spacy) (8.1.8)
Requirement already satisfied: pathy>=0.3.5 in /usr/local/lib/python3.9/dist-packages (from spacy) (0.10.1)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.9/dist-packages (from spacy) (23.0)
Requirement already satisfied: langcodes<4.0.0,>=3.2.0 in /usr/local/lib/python3.9/dist-packages (from spacy) (3.3.0)
Requirement already satisfied: typing-extensions>=4.2.0 in /usr/local/lib/python3.9/dist-packages (from pydantic!=1.8,!_=1.8.1,<1.11.0,>=1.7.4->spacy) (4.5.0)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in /usr/local/lib/python3.9/dist-packages (from requests<3.0.0,>=2.13.0->spacy) (1.26.14)
Requirement already satisfied: chardet<5,>=3.0.2 in /usr/local/lib/python3.9/dist-packages (from requests<3.0.0,>=2.13.0->spacy) (4.0.0)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.9/dist-packages (from requests<3.0.0,>=2.13.0->spacy) (2022.12.7)
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.9/dist-packages (from requests<3.0.0,>=2.13.0->spacy) (2.10)
Requirement already satisfied: confection<1.0.0,>=0.0.1 in /usr/local/lib/python3.9/dist-packages (from thinc<8.2.0,>=8.1.0->spacy) (0.0.4)
Requirement already satisfied: bliss<0.8.0,>=0.7.8 in /usr/local/lib/python3.9/dist-packages (from thinc<8.2.0,>=8.1.0->spacy) (0.7.9)
Requirement already satisfied: click<9.0.0,>=7.1.1 in /usr/local/lib/python3.9/dist-packages (from typer<0.8.0,>=0.3.0->spacy) (8.1.3)
```

Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.9/dist-packages (from jinja2->spacy) (2.1.2)

In [ ]: !python3 -m spacy download en\_core\_web\_sm

```

/usr/local/lib/python3.9/dist-packages/torch/cuda/__init__.py:497: UserWarning: Can't
initialize NVML
    warnings.warn("Can't initialize NVML")
2023-03-09 05:30:37.590741: W tensorflow/compiler/xla/stream_executor/platform/default/dso_loader.cc:64] Could not load dynamic library 'libnvinfer.so.7'; dlerror: libnvinfer.so.7: cannot open shared object file: No such file or directory; LD_LIBRARY_PATH: /usr/local/nvidia/lib:/usr/local/nvidia/lib64
2023-03-09 05:30:37.590869: W tensorflow/compiler/xla/stream_executor/platform/default/dso_loader.cc:64] Could not load dynamic library 'libnvinfer_plugin.so.7'; dlerror: libnvinfer_plugin.so.7: cannot open shared object file: No such file or directory; LD_LIBRARY_PATH: /usr/local/nvidia/lib:/usr/local/nvidia/lib64
2023-03-09 05:30:37.590889: W tensorflow/compiler/tf2tensorrt/utils/py_utils.cc:38] TF-TRT Warning: Cannot dlopen some TensorRT libraries. If you would like to use Nvidia GPU with TensorRT, please make sure the missing libraries mentioned above are installed properly.
2023-03-09 05:30:39.582751: E tensorflow/compiler/xla/stream_executor/cuda/cuda_driver.cc:267] failed call to cuInit: CUDA_ERROR_NO_DEVICE: no CUDA-capable device is detected
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Collecting en-core-web-sm==3.4.1
    Downloading https://github.com/explosion/spacy-models/releases/download/en_core_web_sm-3.4.1/en_core_web_sm-3.4.1-py3-none-any.whl (12.8 MB)
    ━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 12.8/12.8 MB 53.1 MB/s eta 0:00:00
Requirement already satisfied: spacy<3.5.0,>=3.4.0 in /usr/local/lib/python3.9/dist-packages (from en-core-web-sm==3.4.1) (3.4.4)
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.10 in /usr/local/lib/python3.9/dist-packages (from spacy<3.5.0,>=3.4.0->en-core-web-sm==3.4.1) (3.0.12)
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in /usr/local/lib/python3.9/dist-packages (from spacy<3.5.0,>=3.4.0->en-core-web-sm==3.4.1) (1.0.4)
Requirement already satisfied: srsly<3.0.0,>=2.4.3 in /usr/local/lib/python3.9/dist-packages (from spacy<3.5.0,>=3.4.0->en-core-web-sm==3.4.1) (2.4.6)
Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in /usr/local/lib/python3.9/dist-packages (from spacy<3.5.0,>=3.4.0->en-core-web-sm==3.4.1) (2.0.8)
Requirement already satisfied: numpy>=1.15.0 in /usr/local/lib/python3.9/dist-packages (from spacy<3.5.0,>=3.4.0->en-core-web-sm==3.4.1) (1.22.4)
Requirement already satisfied: thinc<8.2.0,>=8.1.0 in /usr/local/lib/python3.9/dist-packages (from spacy<3.5.0,>=3.4.0->en-core-web-sm==3.4.1) (8.1.8)
Requirement already satisfied: langcodes<4.0.0,>=3.2.0 in /usr/local/lib/python3.9/dist-packages (from spacy<3.5.0,>=3.4.0->en-core-web-sm==3.4.1) (3.3.0)
Requirement already satisfied: wasabi<1.1.0,>=0.9.1 in /usr/local/lib/python3.9/dist-packages (from spacy<3.5.0,>=3.4.0->en-core-web-sm==3.4.1) (0.10.1)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in /usr/local/lib/python3.9/dist-packages (from spacy<3.5.0,>=3.4.0->en-core-web-sm==3.4.1) (2.25.1)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python3.9/dist-packages (from spacy<3.5.0,>=3.4.0->en-core-web-sm==3.4.1) (3.0.8)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/python3.9/dist-packages (from spacy<3.5.0,>=3.4.0->en-core-web-sm==3.4.1) (1.0.9)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.9/dist-packages (from spacy<3.5.0,>=3.4.0->en-core-web-sm==3.4.1) (23.0)
Requirement already satisfied: pathy>=0.3.5 in /usr/local/lib/python3.9/dist-packages (from spacy<3.5.0,>=3.4.0->en-core-web-sm==3.4.1) (0.10.1)
Requirement already satisfied: setuptools in /usr/local/lib/python3.9/dist-packages (from spacy<3.5.0,>=3.4.0->en-core-web-sm==3.4.1) (57.4.0)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /usr/local/lib/python3.9/dist-packages (from spacy<3.5.0,>=3.4.0->en-core-web-sm==3.4.1) (2.0.7)
Requirement already satisfied: typer<0.8.0,>=0.3.0 in /usr/local/lib/python3.9/dist-packages (from spacy<3.5.0,>=3.4.0->en-core-web-sm==3.4.1) (0.7.0)
Requirement already satisfied: smart-open<7.0.0,>=5.2.1 in /usr/local/lib/python3.9/dist-packages (from spacy<3.5.0,>=3.4.0->en-core-web-sm==3.4.1) (6.3.0)

```

```
Requirement already satisfied: pydantic!=1.8,!=1.8.1,<1.11.0,>=1.7.4 in /usr/local/lib/python3.9/dist-packages (from spacy<3.5.0,>=3.4.0->en-core-web-sm==3.4.1) (1.10.5)
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in /usr/local/lib/python3.9/dist-packages (from spacy<3.5.0,>=3.4.0->en-core-web-sm==3.4.1) (4.65.0)
Requirement already satisfied: jinja2 in /usr/local/lib/python3.9/dist-packages (from spacy<3.5.0,>=3.4.0->en-core-web-sm==3.4.1) (3.1.2)
Requirement already satisfied: typing-extensions>=4.2.0 in /usr/local/lib/python3.9/dist-packages (from pydantic!=1.8,!=1.8.1,<1.11.0,>=1.7.4->spacy<3.5.0,>=3.4.0->en-core-web-sm==3.4.1) (4.5.0)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.9/dist-packages (from requests<3.0.0,>=2.13.0->spacy<3.5.0,>=3.4.0->en-core-web-sm==3.4.1) (2022.12.7)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in /usr/local/lib/python3.9/dist-packages (from requests<3.0.0,>=2.13.0->spacy<3.5.0,>=3.4.0->en-core-web-sm==3.4.1) (1.26.14)
Requirement already satisfied: chardet<5,>=3.0.2 in /usr/local/lib/python3.9/dist-packages (from requests<3.0.0,>=2.13.0->spacy<3.5.0,>=3.4.0->en-core-web-sm==3.4.1) (4.0.0)
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.9/dist-packages (from requests<3.0.0,>=2.13.0->spacy<3.5.0,>=3.4.0->en-core-web-sm==3.4.1) (2.10)
Requirement already satisfied: confection<1.0.0,>=0.0.1 in /usr/local/lib/python3.9/dist-packages (from thinc<8.2.0,>=8.1.0->spacy<3.5.0,>=3.4.0->en-core-web-sm==3.4.1) (0.0.4)
Requirement already satisfied: bliss<0.8.0,>=0.7.8 in /usr/local/lib/python3.9/dist-packages (from thinc<8.2.0,>=8.1.0->spacy<3.5.0,>=3.4.0->en-core-web-sm==3.4.1) (0.7.9)
Requirement already satisfied: click<9.0.0,>=7.1.1 in /usr/local/lib/python3.9/dist-packages (from typer<0.8.0,>=0.3.0->spacy<3.5.0,>=3.4.0->en-core-web-sm==3.4.1) (8.1.3)
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.9/dist-packages (from jinja2->spacy<3.5.0,>=3.4.0->en-core-web-sm==3.4.1) (2.1.2)
✓ Download and installation successful
You can now load the package via spacy.load('en_core_web_sm')
```

```
In [ ]: #Perform standard imports:
import spacy
# Load English tokenizer, tagger, parser and NER
nlp = spacy.load('en_core_web_sm')
def show_lemmas(text):
    for token in text:
        print(f'{token.text:{12}} {token.pos_:{6}} {token.lemma:{22}} {token.lemma_}')


In [ ]: doc = nlp(u"I am a runner running in a race because I love to run since I ran today.")
show_lemmas (doc)
```

I	PRON	4690420944186131903	I
am	AUX	10382539506755952630	be
a	DET	11901859001352538922	a
runner	NOUN	12640964157389618806	runner
running	VERB	12767647472892411841	run
in	ADP	3002984154512732771	in
a	DET	11901859001352538922	a
race	NOUN	8048469955494714898	race
because	SCONJ	16950148841647037698	because
I	PRON	4690420944186131903	I
love	VERB	3702023516439754181	love
to	PART	3791531372978436496	to
run	VERB	12767647472892411841	run
since	SCONJ	10066841407251338481	since
I	PRON	4690420944186131903	I
ran	VERB	12767647472892411841	run
today	NOUN	11042482332948150395	today
.	PUNCT	12646065887601541794	.