

Task 2: spaCy Hello World

```
In [1]: pip install spacy
```

Requirement already satisfied: spacy in /home/chattha/anaconda3/lib/python3.11/site-packages (3.7.2)
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in /home/chattha/anaconda3/lib/python3.11/site-packages (from spacy) (3.0.12)
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in /home/chattha/anaconda3/lib/python3.11/site-packages (from spacy) (1.0.5)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /home/chattha/anaconda3/lib/python3.11/site-packages (from spacy) (1.0.10)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /home/chattha/anaconda3/lib/python3.11/site-packages (from spacy) (2.0.8)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /home/chattha/anaconda3/lib/python3.11/site-packages (from spacy) (3.0.9)
Requirement already satisfied: thinc<8.3.0,>=8.1.8 in /home/chattha/anaconda3/lib/python3.11/site-packages (from spacy) (8.2.2)
Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in /home/chattha/anaconda3/lib/python3.11/site-packages (from spacy) (1.1.2)
Requirement already satisfied: srsly<3.0.0,>=2.4.3 in /home/chattha/anaconda3/lib/python3.11/site-packages (from spacy) (2.4.8)
Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in /home/chattha/anaconda3/lib/python3.11/site-packages (from spacy) (2.0.10)
Requirement already satisfied: weasel<0.4.0,>=0.1.0 in /home/chattha/anaconda3/lib/python3.11/site-packages (from spacy) (0.3.4)
Requirement already satisfied: typer<0.10.0,>=0.3.0 in /home/chattha/anaconda3/lib/python3.11/site-packages (from spacy) (0.9.0)
Requirement already satisfied: smart-open<7.0.0,>=5.2.1 in /home/chattha/anaconda3/lib/python3.11/site-packages (from spacy) (5.2.1)
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in /home/chattha/anaconda3/lib/python3.11/site-packages (from spacy) (4.65.0)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in /home/chattha/anaconda3/lib/python3.11/site-packages (from spacy) (2.31.0)
Requirement already satisfied: pydantic!=1.8,!1.8.1,<3.0.0,>=1.7.4 in /home/chattha/anaconda3/lib/python3.11/site-packages (from spacy) (1.10.8)
Requirement already satisfied: jinja2 in /home/chattha/anaconda3/lib/python3.11/site-packages (from spacy) (3.1.2)
Requirement already satisfied: setuptools in /home/chattha/anaconda3/lib/python3.11/site-packages (from spacy) (69.0.3)
Requirement already satisfied: packaging>=20.0 in /home/chattha/anaconda3/lib/python3.11/site-packages (from spacy) (23.1)
Requirement already satisfied: langcodes<4.0.0,>=3.2.0 in /home/chattha/anaconda3/lib/python3.11/site-packages (from spacy) (3.3.0)
Requirement already satisfied: numpy>=1.19.0 in /home/chattha/anaconda3/lib/python3.11/site-packages (from spacy) (1.24.3)

Requirement already satisfied: typing-extensions>=4.2.0 in /home/chattha/anaconda3/lib/python3.11/site-packages (from pydantic!=1.8,!<1.8.1,<3.0.0,>=1.7.4->spacy) (4.7.1)
Requirement already satisfied: charset-normalizer<4,>=2 in /home/chattha/anaconda3/lib/python3.11/site-packages (from requests<3.0.0,>=2.13.0->spacy) (2.0.4)
Requirement already satisfied: idna<4,>=2.5 in /home/chattha/anaconda3/lib/python3.11/site-packages (from requests<3.0.0,>=2.13.0->spacy) (3.4)
Requirement already satisfied: urllib3<3,>=1.21.1 in /home/chattha/anaconda3/lib/python3.11/site-packages (from requests<3.0.0,>=2.13.0->spacy) (1.26.16)
Requirement already satisfied: certifi>=2017.4.17 in /home/chattha/anaconda3/lib/python3.11/site-packages (from requests<3.0.0,>=2.13.0->spacy) (2023.7.22)
Requirement already satisfied: blis<0.8.0,>=0.7.8 in /home/chattha/anaconda3/lib/python3.11/site-packages (from thinc<8.3.0,>=8.1.8->spacy) (0.7.11)
Requirement already satisfied: confection<1.0.0,>=0.0.1 in /home/chattha/anaconda3/lib/python3.11/site-packages (from thinc<8.3.0,>=8.1.8->spacy) (0.1.4)
Requirement already satisfied: click<9.0.0,>=7.1.1 in /home/chattha/anaconda3/lib/python3.11/site-packages (from typer<0.10.0,>=0.3.0->spacy) (8.0.4)
Requirement already satisfied: cloudpathlib<0.17.0,>=0.7.0 in /home/chattha/anaconda3/lib/python3.11/site-packages (from weasel<0.4.0,>=0.1.0->spacy) (0.16.0)
Requirement already satisfied: MarkupSafe>=2.0 in /home/chattha/anaconda3/lib/python3.11/site-packages (from jinja2->spacy) (2.1.1)
Note: you may need to restart the kernel to use updated packages.

In [2]: `pip install -U pip setuptools wheel`

Requirement already satisfied: pip in /home/chattha/anaconda3/lib/python3.11/site-packages (23.3.2)
Requirement already satisfied: setuptools in /home/chattha/anaconda3/lib/python3.11/site-packages (69.0.3)
Requirement already satisfied: wheel in /home/chattha/anaconda3/lib/python3.11/site-packages (0.42.0)
Note: you may need to restart the kernel to use updated packages.

In [1]: `import spacy`
`import nltk`

```
In [4]: !python3 -m spacy download en_core_web_sm
```

Collecting en-core-web-sm==3.7.1

Downloading https://github.com/explosion/spacy-models/releases/download/en_core_web_sm-3.7.1/en_core_web_sm-3.7.1-py3-none-any.whl (https://github.com/explosion/spacy-models/releases/download/en_core_web_sm-3.7.1/en_core_web_sm-3.7.1-py3-none-any.whl) (12.8 MB)

12.8/12.8 MB 2.3 MB/s eta 0:00:00m eta 0:00:01[36m0:00:01
Requirement already satisfied: spacy<3.8.0,>=3.7.2 in /home/chattha/anaconda3/lib/python3.11/site-packages (from en-core-web-sm==3.7.1) (3.7.2)
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in /home/chattha/anaconda3/lib/python3.11/site-packages (from spacy<3.8.0,>=3.7.2->en-core-web-sm==3.7.1) (3.0.12)
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in /home/chattha/anaconda3/lib/python3.11/site-packages (from spacy<3.8.0,>=3.7.2->en-core-web-sm==3.7.1) (1.0.5)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /home/chattha/anaconda3/lib/python3.11/site-packages (from spacy<3.8.0,>=3.7.2->en-core-web-sm==3.7.1) (1.0.10)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /home/chattha/anaconda3/lib/python3.11/site-packages (from spacy<3.8.0,>=3.7.2->en-core-web-sm==3.7.1) (2.0.8)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /home/chattha/anaconda3/lib/python3.11/site-packages (from spacy<3.8.0,>=3.7.2->en-core-web-sm==3.7.1) (3.0.9)
Requirement already satisfied: thinc<8.3.0,>=8.1.8 in /home/chattha/anaconda3/lib/python3.11/site-packages (from spacy<3.8.0,>=3.7.2->en-core-web-sm==3.7.1) (8.2.2)
Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in /home/chattha/anaconda3/lib/python3.11/site-packages (from spacy<3.8.0,>=3.7.2->en-core-web-sm==3.7.1) (1.1.2)
Requirement already satisfied: srsly<3.0.0,>=2.4.3 in /home/chattha/anaconda3/lib/python3.11/site-packages (from spacy<3.8.0,>=3.7.2->en-core-web-sm==3.7.1) (2.4.8)
Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in /home/chattha/anaconda3/lib/python3.11/site-packages (from spacy<3.8.0,>=3.7.2->en-core-web-sm==3.7.1) (2.0.10)
Requirement already satisfied: weasel<0.4.0,>=0.1.0 in /home/chattha/anaconda3/lib/python3.11/site-packages (from spacy<3.8.0,>=3.7.2->en-core-web-sm==3.7.1) (0.3.4)
Requirement already satisfied: typer<0.10.0,>=0.3.0 in /home/chattha/anaconda3/lib/python3.11/site-packages (from spacy<3.8.0,>=3.7.2->en-core-web-sm==3.7.1) (0.9.0)
Requirement already satisfied: smart-open<7.0.0,>=5.2.1 in /home/chattha/anaconda3/lib/python3.11/site-packages (from spacy<3.8.0,>=3.7.2->en-core-web-sm==3.7.1) (5.2.1)
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in /home/chattha/anaconda3/lib/python3.11/site-packages (from spacy<3.8.0,>=3.7.2->en-core-web-sm==3.7.1) (4.65.0)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in /home/chattha/anaconda3/lib/python3.11/site-packages (from spacy<3.8.0,>=3.7.2->en-core-web-sm==3.7.1) (2.31.0)
Requirement already satisfied: pydantic!=1.8,!<1.8.1,<3.0.0,>=1.7.4 in /home/chattha/anaconda3/lib/python3.11/site-packages (from spacy<3.8.0,>=3.7.2->en-core-web-sm==3.7.1) (1.10.8)
Requirement already satisfied: jinja2 in /home/chattha/anaconda3/lib/python3.11/site-packages (from spacy<3.8.0,>=3.7.2->en-core-web-sm==3.7.1) (3.1.2)
Requirement already satisfied: setuptools in /home/chattha/anaconda3/lib/python3.11/site-packages (from spacy<3.8.0,>=3.7.2->en-core-web-sm==3.7.1) (69.0.3)

Requirement already satisfied: packaging>=20.0 in /home/chattha/anaconda3/lib/python3.11/site-packages (from spacy<3.8.0,>=3.7.2->en-core-web-sm==3.7.1) (23.1)

Requirement already satisfied: langcodes<4.0.0,>=3.2.0 in /home/chattha/anaconda3/lib/python3.11/site-packages (from spacy<3.8.0,>=3.7.2->en-core-web-sm==3.7.1) (3.3.0)

Requirement already satisfied: numpy>=1.19.0 in /home/chattha/anaconda3/lib/python3.11/site-packages (from spacy<3.8.0,>=3.7.2->en-core-web-sm==3.7.1) (1.24.3)

Requirement already satisfied: typing-extensions>=4.2.0 in /home/chattha/anaconda3/lib/python3.11/site-packages (from pydantic!=1.8,!1.8.1,<3.0.0,>=1.7.4->spacy<3.8.0,>=3.7.2->en-core-web-sm==3.7.1) (4.7.1)

Requirement already satisfied: charset-normalizer<4,>=2 in /home/chattha/anaconda3/lib/python3.11/site-packages (from requests<3.0.0,>=2.13.0->spacy<3.8.0,>=3.7.2->en-core-web-sm==3.7.1) (2.0.4)

Requirement already satisfied: idna<4,>=2.5 in /home/chattha/anaconda3/lib/python3.11/site-packages (from requests<3.0.0,>=2.13.0->spacy<3.8.0,>=3.7.2->en-core-web-sm==3.7.1) (3.4)

Requirement already satisfied: urllib3<3,>=1.21.1 in /home/chattha/anaconda3/lib/python3.11/site-packages (from requests<3.0.0,>=2.13.0->spacy<3.8.0,>=3.7.2->en-core-web-sm==3.7.1) (1.26.16)

Requirement already satisfied: certifi>=2017.4.17 in /home/chattha/anaconda3/lib/python3.11/site-packages (from requests<3.0.0,>=2.13.0->spacy<3.8.0,>=3.7.2->en-core-web-sm==3.7.1) (2023.7.22)

Requirement already satisfied: blis<0.8.0,>=0.7.8 in /home/chattha/anaconda3/lib/python3.11/site-packages (from thinc<8.3.0,>=8.1.8->spacy<3.8.0,>=3.7.2->en-core-web-sm==3.7.1) (0.7.11)

Requirement already satisfied: confection<1.0.0,>=0.0.1 in /home/chattha/anaconda3/lib/python3.11/site-packages (from thinc<8.3.0,>=8.1.8->spacy<3.8.0,>=3.7.2->en-core-web-sm==3.7.1) (0.1.4)

Requirement already satisfied: click<9.0.0,>=7.1.1 in /home/chattha/anaconda3/lib/python3.11/site-packages (from typer<0.10.0,>=0.3.0->spacy<3.8.0,>=3.7.2->en-core-web-sm==3.7.1) (8.0.4)

Requirement already satisfied: cloudpathlib<0.17.0,>=0.7.0 in /home/chattha/anaconda3/lib/python3.11/site-packages (from weasel<0.4.0,>=0.1.0->spacy<3.8.0,>=3.7.2->en-core-web-sm==3.7.1) (0.16.0)

Requirement already satisfied: MarkupSafe>=2.0 in /home/chattha/anaconda3/lib/python3.11/site-packages (from jinja2->spacy<3.8.0,>=3.7.2->en-core-web-sm==3.7.1) (2.1.1)

✓ Download and installation successful

You can now load the package via `spacy.load('en_core_web_sm')`

In []:

```
In [82]: from spacy.tokens.doc import Doc
from spacy.vocab import Vocab
doc = Doc(Vocab(),
words = [u'Hello', u'World!'])
print(doc)
```

Hello World!

```
In [83]: print(type(doc))
print(doc.vocab)
for token in doc:
    lexeme = doc.vocab[token.text]
    print(lexeme.text)

<class 'spacy.tokens.doc.Doc'>
<spacy.vocab.Vocab object at 0x7fac4d401120>
Hello
World!
```

```
In [84]: import spacy
nlp = spacy.load('en_core_web_sm')
doc = nlp(u'I want to learn spaCy.')
token_text1 = [token.text for token in doc]
token_text2 = [doc[i].text for i in range(len(doc))]
print(token_text1)
print(token_text2)

['I', 'want', 'to', 'learn', 'spaCy', '.']
['I', 'want', 'to', 'learn', 'spaCy', '.']
```

```
In [85]: doc = nlp(u'I want to learn spaCy.')
for i in range(len(doc)):
    print([t for t in doc[i].lefts])

[]
[I]
[]
[to]
[]
[]
```



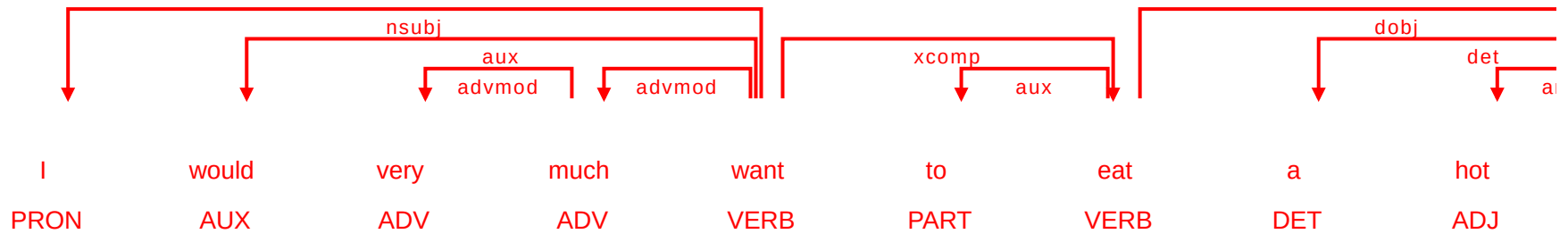
```
In [86]: doc = nlp(u'I want to learn spaCy.')
         for i in range(len(doc)):
             print([t for t in doc[i].rights])
```

```
[]
[learn, .]
[]
[spaCy]
[]
[]
```

```
In [87]: doc = nlp(u'I want to learn spaCy.')
         for i in range(len(doc)):
             print([t for t in doc[i].children])
```

```
[]
[I, learn, .]
[]
[to, spaCy]
[]
[]
```

```
In [95]: from spacy import displacy
displacy.render(doc, style='dep', options={'distance': 100, 'color': 'red', 'compact': True} , page=f'Left')
```



```
In [90]: doc = nlp(u'I would very much want to eat a hot dinner.')
         for i in range(len(doc)):
             print([t for t in doc[i].lefts])
```

```
[]
[]
[]
[very]
[I, would, much]
[]
[to]
[]
[]
[a, hot]
[]
```

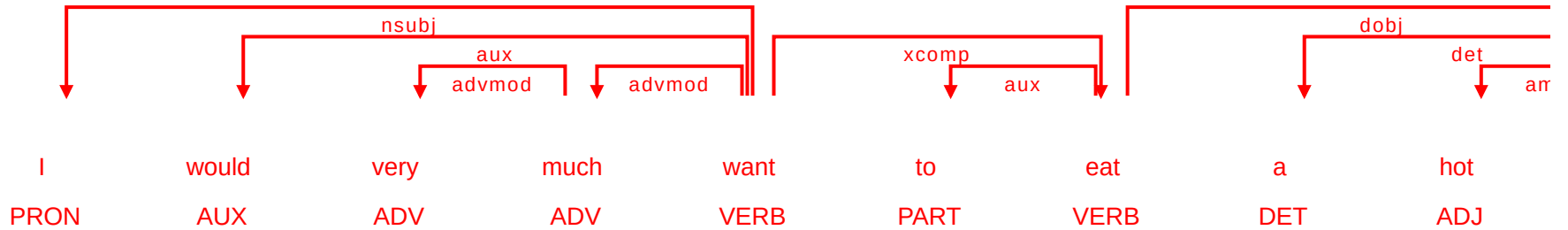
```
In [91]: doc = nlp(u'I would very much want to eat a hot dinner.')
         for i in range(len(doc)):
             print([t for t in doc[i].rights])
```

```
[]
[]
[]
[]
[eat, .]
[]
[dinner]
[]
[]
[]
[]
```

```
In [92]: doc = nlp(u'I would very much want to eat a hot dinner.')
         for i in range(len(doc)):
             print([t for t in doc1[i].children])
```

```
[]
[]
[]
[very]
[I, would, much, eat, .]
[]
[to, dinner]
[]
[]
[a, hot]
[]
```

```
In [94]: spacy.render(doc, style='dep', options={'distance': 100, 'color': 'red', 'compact': True}, page=f'Left and P
```



```
In [99]: sentences = ["I want to learn spaCy.", "I want to learn spaCy.", "I would very much want to eat a hot dinner"]
# Process each sentence
for idx, sentence in enumerate(sentences, start=1):
    print("\n Start of Sentence ", idx)
    doc = nlp(sentence)

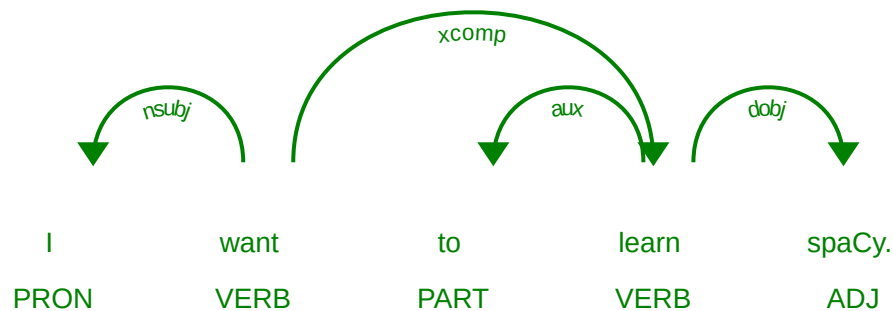
    # Draw left and right dependencies
    displacy.render(doc, style='dep', options={'distance': 100, 'color': 'GREEN', 'compact': False}, page=1)

    # Draw children dependencies
    displacy.render(doc, style='dep', options={'distance': 100, 'color': 'red', 'compact': True}, page=f'C{idx}')

    # List dependency grammars
    print(f"\nDependency Grammar of Sentence {idx}")
    for token in doc:
        print(f"{token.text} =====>({token.dep_}) =====> {token.head.text}")

    print("End of Sentence ", idx)
```

Start of Sentence 1



Task 3: NLTK vs spaCy Pipelines

In [59]: `sentence = [u"We are nearing the end of the semester at Peshawar. Final exams of the Fall 2023 semester wi`



```
In [63]: print("=== Input Sentence === \n\n", sentence)

for text in texts:
    sentences = nltk.sent_tokenize(text)          # sentence level segmentation
    print("\n === Input Sentence Segmentation === \n\n", sentences)
    for sentence in sentences:
        words = nltk.word_tokenize(sentence)      # word level segmentation
        print("\n === Sentence Tokenization === \n\n", words)
        tagged_words = nltk.pos_tag(words)        # POS
        print("\n === Sentence Tokenization === \n\n", tagged_words)
        ne_tagged_words = nltk.ne_chunk(tagged_words)
        print("\n === Name Entity Reconization === \n\n", ne_tagged_words)
```


=== Input Sentence ===

Final exams of the Fall 2023 semester will start soon.

=== Input Sentence Segmentation ===

```
['We are nearing the end of the semester at Peshawar.', 'Final exams of the Fall 2023 semester will start soon.']
```

=== Sentence Tokenization ===

```
['We', 'are', 'nearing', 'the', 'end', 'of', 'the', 'semester', 'at', 'Peshawar', '.']
```

=== Sentence Tokenization ===

```
[('We', 'PRP'), ('are', 'VBP'), ('nearing', 'VBG'), ('the', 'DT'), ('end', 'NN'), ('of', 'IN'), ('the', 'DT'), ('semester', 'NN'), ('at', 'IN'), ('Peshawar', 'NNP'), ('.', '.')]
```

=== Name Entity Reconization ===

```
(S
  We/PRP
  are/VBP
  nearing/VBG
  the/DT
  end/NN
  of/IN
  the/DT
  semester/NN
  at/IN
  (ORGANIZATION Peshawar/NNP)
  ./.)
```

=== Sentence Tokenization ===

```
['Final', 'exams', 'of', 'the', 'Fall', '2023', 'semester', 'will', 'start', 'soon', '.']
```

=== Sentence Tokenization ===

```
[('Final', 'JJ'), ('exams', 'NN'), ('of', 'IN'), ('the', 'DT'), ('Fall', 'NN'), ('2023', 'CD'), ('semester', 'NN'), ('will', 'MD'), ('start', 'VB'), ('soon', 'RB'), ('.', '.')]
```

=== Name Entity Reconization ===

```
(S
  Final/JJ
  exams/NN
  of/IN
  the/DT
  Fall/NN
  2023/CD
  semester/NN
  will/MD
  start/VB
  soon/RB
  ./.)
```

```
In [64]: nltk.download('punkt') # Sentence Tokenize
nltk.download('averaged_perceptron_tagger') # POS Tagging
nltk.download('maxent_ne_chunker') # Named Entity Chunking
nltk.download('words')
```

```
[nltk_data] Downloading package punkt to /home/chattha/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data]   /home/chattha/nltk_data...
[nltk_data]   Package averaged_perceptron_tagger is already up-to-
[nltk_data]   date!
[nltk_data] Downloading package maxent_ne_chunker to
[nltk_data]   /home/chattha/nltk_data...
[nltk_data]   Package maxent_ne_chunker is already up-to-date!
[nltk_data] Downloading package words to /home/chattha/nltk_data...
[nltk_data]   Package words is already up-to-date!
```

Out[64]: True

```
In [65]: from spacy import displacy
doc = nlp(u'We are nearing the end of the semester at Peshawar. Final exams of the Fall 2023 semester will
displacy.render(doc, style='ent')
```

We are nearing the end of the semester DATE at Peshawar GPE . Final exams of the Fall 2023 semester will start soon.

```
In [66]: for ent in doc.ents:
        print(ent.text, ent.label_)
```

the end of the semester DATE
Peshawar GPE

```
In [67]: import spacy
nlp = spacy.load('en_core_web_sm')
nlp.pipe_names
```

```
Out[67]: ['tok2vec', 'tagger', 'parser', 'attribute_ruler', 'lemmatizer', 'ner']
```

```
In [68]: nlp = spacy.load('en_core_web_sm', disable=['parser'])
```

```
In [69]: from spacy.language import Language
@Language.component("my_component")
def my_component(doc):
    # Do something to the doc here
    return doc
```

```
In [28]: nlp.add_pipe("my_component")
```

```
Out[28]: <function __main__.my_component(doc)>
```

```
In [29]: # nlp.add_pipe("my_component", first=True)
# nlp.add_pipe("my_component", last=True)
# nlp.add_pipe("my_component", before="parser")
# nlp.add_pipe("my_component", after="parser")
```

```
In [30]: nlp.pipe_names
```

```
Out[30]: ['tok2vec', 'tagger', 'attribute_ruler', 'lemmatizer', 'ner', 'my_component']
```

Task-4: Finding Patterns in Sentences

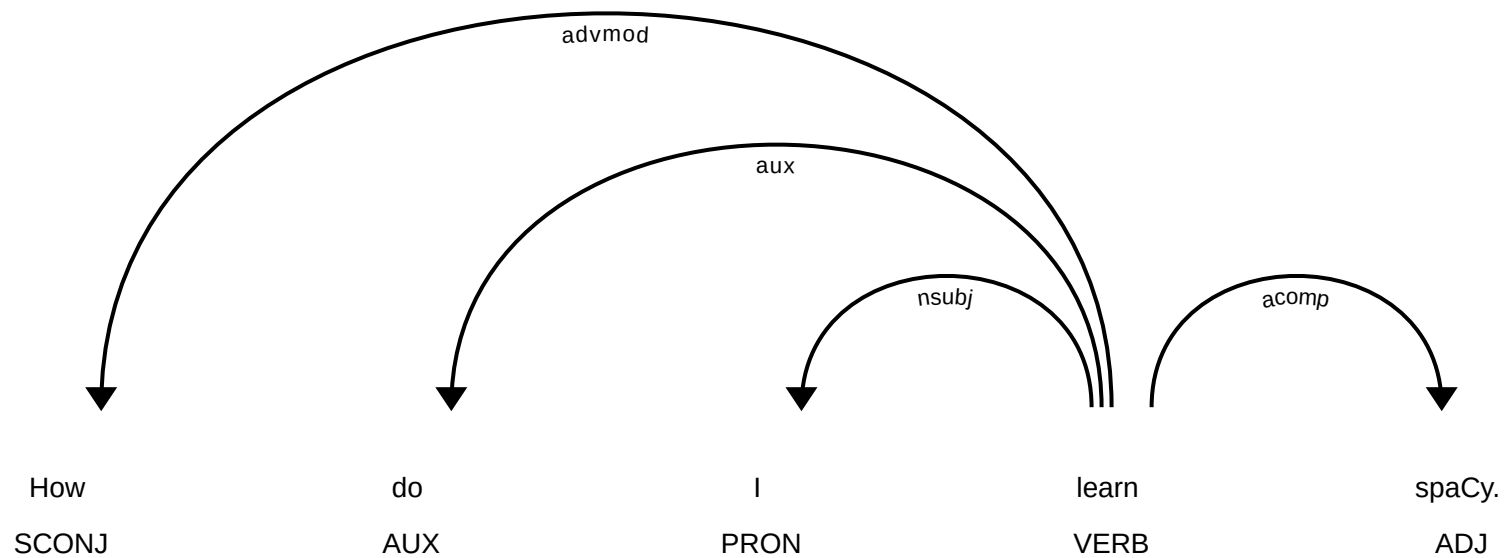
```
In [31]: doc = nlp(u'I want to learn spaCy.')  
displacy.render(doc, style='dep')
```

/home/chattha/anaconda3/lib/python3.11/site-packages/spacy/displacy/__init__.py:141: UserWarning: [W005]
Doc object not parsed. This means displaCy won't be able to generate a dependency visualization for it. Make sure the Doc was processed with a model that supports dependency parsing, and not just a language class like `English()`. For more info, see the docs:
<https://spacy.io/usage/models> (<https://spacy.io/usage/models>)
warnings.warn(Warnings.W005)

I	want	to	learn	spaCy.
PRON	VERB	PART	VERB	ADJ



```
In [34]: doc = nlp(u'How do I learn spaCy.')  
displacy.render(doc, style='dep', )
```



```
In [33]: import spacy
nlp = spacy.load('en_core_web_sm')
def dep_pattern(doc):
    for i in range(len(doc)-1):
        print(doc[i].dep_)

        if doc[i].dep_ == 'nsubj' and doc[i+1].dep_ == 'ROOT' and doc[i+2].dep_ == 'acomp':
            return True
    return False
doc = nlp(u'How do I learn spaCy.')
if dep_pattern(doc):

    print('Found')

else:

    print('Not found')
```

advmod
aux
nsubj
Found

In []:

In []: