```
!pip install spacy transformers torch pandas
!python -m spacy download en_core_web_sm
```

```
Requirement already satisfied: spacy in /usr/local/lib/python3.12/dist-packages (3.8.11)
Requirement already satisfied: transformers in /usr/local/lib/python3.12/dist-packages (5.0.0)
Requirement already satisfied: torch in /usr/local/lib/python3.12/dist-packages (2.10.0+cpu)
Requirement already satisfied: pandas in /usr/local/lib/python3.12/dist-packages (2.2.2)
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) (0.24.0)
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in /usr/local/lib/python3.12/dist-packages (from spacy) (4.67.3)
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) (26.0)
Requirement already satisfied: filelock in /usr/local/lib/python3.12/dist-packages (from transformers) (3.24.2)
Requirement already satisfied: huggingface-hub<2.0,>=1.3.0 in /usr/local/lib/python3.12/dist-packages (from transformers)
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.12/dist-packages (from transformers) (6.0.3)
Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.12/dist-packages (from transformers) (2025.11.3
Requirement already satisfied: tokenizers<=0.23.0,>=0.22.0 in /usr/local/lib/python3.12/dist-packages (from transformers)
Requirement already satisfied: safetensors>=0.4.3 in /usr/local/lib/python3.12/dist-packages (from transformers) (0.7.0)
Requirement already satisfied: typing-extensions>=4.10.0 in /usr/local/lib/python3.12/dist-packages (from torch) (4.15.0)
Requirement already satisfied: sympy>=1.13.3 in /usr/local/lib/python3.12/dist-packages (from torch) (1.14.0)
Requirement already satisfied: networkx>=2.5.1 in /usr/local/lib/python3.12/dist-packages (from torch) (3.6.1)
Requirement already satisfied: fsspec>=0.8.5 in /usr/local/lib/python3.12/dist-packages (from torch) (2025.3.0)
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.12/dist-packages (from pandas) (2.9.0.post
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.12/dist-packages (from pandas) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.12/dist-packages (from pandas) (2025.3)
Requirement already satisfied: hf-xet<2.0.0,>=1.2.0 in /usr/local/lib/python3.12/dist-packages (from huggingface-hub<2.0,>
Requirement already satisfied: httpx<1,>=0.23.0 in /usr/local/lib/python3.12/dist-packages (from huggingface-hub<2.0,>=1.3
Requirement already satisfied: shellingham in /usr/local/lib/python3.12/dist-packages (from huggingface-hub<2.0,>=1.3.0->t
Requirement already satisfied: annotated-types>=0.6.0 in /usr/local/lib/python3.12/dist-packages (from pydantic!=1.8,!=1.8
Requirement already satisfied: pydantic-core==2.41.4 in /usr/local/lib/python3.12/dist-packages (from pydantic!=1.8,!=1.8.
Requirement already satisfied: typing-inspection>=0.4.2 in /usr/local/lib/python3.12/dist-packages (from pydantic!=1.8,!=1
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.12/dist-packages (from python-dateutil>=2.8.2->pandas) (
Requirement already satisfied: charset_normalizer<4,>=2 in /usr/local/lib/python3.12/dist-packages (from requests<3.0.0,>=
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.12/dist-packages (from requests<3.0.0,>=2.13.0->spac
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: mpmath<1.4,>=1.1.0 in /usr/local/lib/python3.12/dist-packages (from sympy>=1.13.3->torch) (
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->sp
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
Requirement already satisfied: typer>=0.24.0 in /usr/local/lib/python3.12/dist-packages (from typer-slim<1.0.0,>=0.3.0->sp
Requirement already satisfied: cloudpathlib<1.0.0,>=0.7.0 in /usr/local/lib/python3.12/dist-packages (from weasel<0.5.0,>=
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.
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.12/dist-packages (from jinja2->spacy) (3.0.3)
Requirement already satisfied: anyio in /usr/local/lib/python3.12/dist-packages (from httpx<1,>=0.23.0->huggingface-hub<2.
Requirement already satisfied: httpcore==1.* in /usr/local/lib/python3.12/dist-packages (from httpx<1,>=0.23.0->huggingfac
Requirement already satisfied: h11>=0.16 in /usr/local/lib/python3.12/dist-packages (from httpcore==1.*->httpx<1,>=0.23.0-
Requirement already satisfied: wrapt in /usr/local/lib/python3.12/dist-packages (from smart-open<8.0.0,>=5.2.1->weasel<0.5
```

```
# spaCy - NLP library for NER and linguistic processing
import spacy

# displacy - Visualization tool for spaCy entities
from spacy import displacy

# Hugging Face transformer tools
from transformers import pipeline, AutoTokenizer, AutoModelForTokenClassification

# torch - backend framework for transformer models
import torch

# pandas - for storing results in table format
import pandas as pd
```

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

```
sentences = [
    "Prime Minister Narendra Modi met Joe Biden in Washington on Monday.",
    "Virat Kohli scored a century against Australia in Mumbai.",
    "Microsoft announced a new AI feature at the conference in California.",
    "The United Nations held a climate summit in New York.",
```

```
    "Elon Musk revealed the latest Tesla model in Texas."
]
```

```
for sentence in sentences:
    doc = nlp(sentence)
    print("\nSentence:", sentence)
    print("Entities:")
    for ent in doc.ents:
        print(f"{ent.text} --> {ent.label_}")
```

```
Sentence: Prime Minister Narendra Modi met Joe Biden in Washington on Monday.
Entities:
Narendra Modi --> PERSON
Joe Biden --> PERSON
Washington --> GPE
Monday --> DATE

Sentence: Virat Kohli scored a century against Australia in Mumbai.
Entities:
Virat Kohli --> PERSON
a century --> DATE
Australia --> GPE
Mumbai --> GPE

Sentence: Microsoft announced a new AI feature at the conference in California.
Entities:
Microsoft --> ORG
AI --> GPE
California --> GPE

Sentence: The United Nations held a climate summit in New York.
Entities:
The United Nations --> ORG
New York --> GPE

Sentence: Elon Musk revealed the latest Tesla model in Texas.
Entities:
Elon Musk --> PERSON
Tesla --> NORP
Texas --> GPE
```

```
spacy_results = []

for sentence in sentences:
    doc = nlp(sentence)
    for ent in doc.ents:
        spacy_results.append([sentence, ent.text, ent.label_])

spacy_df = pd.DataFrame(spacy_results, columns=["Sentence", "Entity", "Label"])
spacy_df
```

	Sentence	Entity	Label
0	Prime Minister Narendra Modi met Joe Biden in ...	Narendra Modi	PERSON
1	Prime Minister Narendra Modi met Joe Biden in ...	Joe Biden	PERSON
2	Prime Minister Narendra Modi met Joe Biden in ...	Washington	GPE
3	Prime Minister Narendra Modi met Joe Biden in ...	Monday	DATE
4	Virat Kohli scored a century against Australia...	Virat Kohli	PERSON
5	Virat Kohli scored a century against Australia...	a century	DATE
6	Virat Kohli scored a century against Australia...	Australia	GPE
7	Virat Kohli scored a century against Australia...	Mumbai	GPE
8	Microsoft announced a new AI feature at the co...	Microsoft	ORG
9	Microsoft announced a new AI feature at the co...	AI	GPE
10	Microsoft announced a new AI feature at the co...	California	GPE
11	The United Nations held a climate summit in Ne...	The United Nations	ORG
12	The United Nations held a climate summit in Ne...	New York	GPE
13	Elon Musk revealed the latest Tesla model in T...	Elon Musk	PERSON
14	Elon Musk revealed the latest Tesla model in T...	Tesla	NORP
15	Elon Musk revealed the latest Tesla model in T...	Texas	GPE

Next steps: Generate code with spacy_df | New interactive sheet

```
doc = nlp(sentences[0])
displacy.render(doc, style="ent", jupyter=True)
```

Prime Minister Narendra Modi **PERSON** met Joe Biden **PERSON** in Washington **GPE** on Monday **DATE** .

```
model_name = "dbmdz/bert-large-cased-finetuned-conll03-english"

tokenizer = AutoTokenizer.from_pretrained(model_name)
model = AutoModelForTokenClassification.from_pretrained(model_name)

ner_pipeline = pipeline("ner", model=model, tokenizer=tokenizer)
```

```
/usr/local/lib/python3.12/dist-packages/huggingface_hub/utils/_auth.py:94: UserWarning:
The secret `HF_TOKEN` does not exist in your Colab secrets.
To authenticate with the Hugging Face Hub, create a token in your settings tab (https://huggingface.co/settings/tokens), set
You will be able to reuse this secret in all of your notebooks.
Please note that authentication is recommended but still optional to access public models or datasets.
  warnings.warn(
config.json: 100% 998/998 [00:00<00:00, 50.9kB/s]
tokenizer_config.json: 100% 60.0/60.0 [00:00<00:00, 1.72kB/s]
vocab.txt: 213k/? [00:00<00:00, 4.43MB/s]
Warning: You are sending unauthenticated requests to the HF Hub. Please set a HF_TOKEN to enable higher rate limits and fast
WARNING:huggingface_hub.utils._http:Warning: You are sending unauthenticated requests to the HF Hub. Please set a HF_TOKEN t
model.safetensors: 100% 1.33G/1.33G [00:31<00:00, 71.4MB/s]
Loading weights: 100% 391/391 [00:03<00:00, 119.93it/s, Materializing param=classifier.weight]
BertForTokenClassification LOAD REPORT from: dbmdz/bert-large-cased-finetuned-conll03-english
Key                        | Status     |  |
---------------------------+------------+--+-
bert.pooler.dense.bias     | UNEXPECTED |  |
bert.pooler.dense.weight   | UNEXPECTED |  |

Notes:
- UNEXPECTED    :can be ignored when loading from different task/architecture; not ok if you expect identical arch.
```

```
for sentence in sentences:
    print("\nSentence:", sentence)
    results = ner_pipeline(sentence)

    for res in results:
        print(f"{res['word']} --> {res['entity']} (Score: {res['score']:.4f})")
```

```
Sentence: Prime Minister Narendra Modi met Joe Biden in Washington on Monday.
Na --> I-PER (Score: 0.9994)
##ren --> I-PER (Score: 0.9937)
##dra --> I-PER (Score: 0.9988)
Mo --> I-PER (Score: 0.9987)
##di --> I-PER (Score: 0.9920)
Joe --> I-PER (Score: 0.9985)
B --> I-PER (Score: 0.9965)
##iden --> I-PER (Score: 0.9941)
Washington --> I-LOC (Score: 0.9992)

Sentence: Virat Kohli scored a century against Australia in Mumbai.
V --> I-PER (Score: 0.9996)
##ira --> I-PER (Score: 0.9957)
##t --> I-PER (Score: 0.9990)
Ko --> I-PER (Score: 0.9995)
##hl --> I-PER (Score: 0.9903)
##i --> I-PER (Score: 0.9936)
Australia --> I-LOC (Score: 0.9997)
Mumbai --> I-LOC (Score: 0.9991)

Sentence: Microsoft announced a new AI feature at the conference in California.
Microsoft --> I-ORG (Score: 0.9994)
California --> I-LOC (Score: 0.9996)

Sentence: The United Nations held a climate summit in New York.
United --> I-ORG (Score: 0.9994)
Nations --> I-ORG (Score: 0.9994)
New --> I-LOC (Score: 0.9994)
York --> I-LOC (Score: 0.9994)

Sentence: Elon Musk revealed the latest Tesla model in Texas.
El --> I-PER (Score: 0.9989)
##on --> I-PER (Score: 0.9979)
Mu --> I-PER (Score: 0.9970)
##sk --> I-PER (Score: 0.9814)
Te --> I-MISC (Score: 0.9884)
```

```
##sla --> I-MISC (Score: 0.9759)
Texas --> I-LOC (Score: 0.9997)
```

```
hf_results = []

for sentence in sentences:
    results = ner_pipeline(sentence)

    merged_entities = []
    current_entity = ""
    current_label = ""
    current_score = 0

    for res in results:
        word = res['word'].replace("##", "")
        label = res['entity']
        score = res['score']

        if word.startswith("##"):
            current_entity += word
        else:
            if current_entity:
                merged_entities.append((current_entity, current_label, current_score))
            current_entity = word
            current_label = label
            current_score = score

    if current_entity:
        merged_entities.append((current_entity, current_label, current_score))

    for entity, label, score in merged_entities:
        hf_results.append([sentence, entity, label, score])

hf_df = pd.DataFrame(hf_results, columns=["Sentence", "Entity", "Label", "Confidence Score"])
hf_df
```

	Sentence	Entity	Label	Confidence Score
0	Prime Minister Narendra Modi met Joe Biden in ...	Na	I-PER	0.999380
1	Prime Minister Narendra Modi met Joe Biden in ...	ren	I-PER	0.993725
2	Prime Minister Narendra Modi met Joe Biden in ...	dra	I-PER	0.998827
3	Prime Minister Narendra Modi met Joe Biden in ...	Mo	I-PER	0.998749
4	Prime Minister Narendra Modi met Joe Biden in ...	di	I-PER	0.991973
5	Prime Minister Narendra Modi met Joe Biden in ...	Joe	I-PER	0.998531
6	Prime Minister Narendra Modi met Joe Biden in ...	B	I-PER	0.996507
7	Prime Minister Narendra Modi met Joe Biden in ...	iden	I-PER	0.994133
8	Prime Minister Narendra Modi met Joe Biden in ...	Washington	I-LOC	0.999161
9	Virat Kohli scored a century against Australia...	V	I-PER	0.999569
10	Virat Kohli scored a century against Australia...	ira	I-PER	0.995733
11	Virat Kohli scored a century against Australia...	t	I-PER	0.999028
12	Virat Kohli scored a century against Australia...	Ko	I-PER	0.999476
13	Virat Kohli scored a century against Australia...	hl	I-PER	0.990260
14	Virat Kohli scored a century against Australia...	i	I-PER	0.993635
15	Virat Kohli scored a century against Australia...	Australia	I-LOC	0.999722