

```
!python -m pip install --user spacy
!python -m spacy download en_core_web_sm
```

```
Requirement already satisfied: spacy in /usr/local/lib/python3.12/dist-packages (3.8.7)
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) (1.0.13)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /usr/local/lib/python3.12/dist-packages (from spacy) (2.0.11)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python3.12/dist-packages (from spacy) (3.0.10)
Requirement already satisfied: thinc<8.4.0,>=8.3.4 in /usr/local/lib/python3.12/dist-packages (from spacy) (8.3.6)
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) (2.5.1)
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,>=0.1.0 in /usr/local/lib/python3.12/dist-packages (from spacy) (0.4.1)
Requirement already satisfied: typer<1.0.0,>=0.3.0 in /usr/local/lib/python3.12/dist-packages (from spacy) (0.19.2)
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) (2.5.1)
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: langcodes<4.0.0,>=3.2.0 in /usr/local/lib/python3.12/dist-packages (from spacy) (3.5.0)
Requirement already satisfied: language-data>=1.2 in /usr/local/lib/python3.12/dist-packages (from langcodes<4.0.0,>=3.2.0-)
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==2.33.2 in /usr/local/lib/python3.12/dist-packages (from pydantic!=1.8,!<1.8.1)
Requirement already satisfied: typing-extensions>=4.12.2 in /usr/local/lib/python3.12/dist-packages (from pydantic!=1.8,!<1.8.1)
Requirement already satisfied: typing-inspection>=0.4.0 in /usr/local/lib/python3.12/dist-packages (from pydantic!=1.8,!<1.8.1)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.12/dist-packages (from requests<3.0.0,>=2.13.0)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.12/dist-packages (from requests<3.0.0,>=2.13.0)
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->spacy)
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<1.0.0,>=0.3.0->spacy) (8.1.8)
Requirement already satisfied: shellingham>=1.3.0 in /usr/local/lib/python3.12/dist-packages (from typer<1.0.0,>=0.3.0->spacy) (1.5.4)
Requirement already satisfied: rich>=10.11.0 in /usr/local/lib/python3.12/dist-packages (from typer<1.0.0,>=0.3.0->spacy) (13.7.1)
Requirement already satisfied: cloudpathlib<1.0.0,>=0.7.0 in /usr/local/lib/python3.12/dist-packages (from weasel<0.5.0,>=0.1.0)
Requirement already satisfied: smart-open<8.0.0,>=5.2.1 in /usr/local/lib/python3.12/dist-packages (from weasel<0.5.0,>=0.1.0)
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.12/dist-packages (from Jinja2->spacy) (3.0.3)
Requirement already satisfied: marisa-trie>=1.1.0 in /usr/local/lib/python3.12/dist-packages (from language-data>=1.2->lang
```

```
Requirement already satisfied: markdown-it-py>=2.2.0 in /usr/local/lib/python3.12/dist-packages (from rich>=10.11.0->typer<1.0.0,>=0.3.0)
Requirement already satisfied: pygments<3.0.0,>=2.13.0 in /usr/local/lib/python3.12/dist-packages (from rich>=10.11.0->typer<1.0.0,>=0.3.0)
Requirement already satisfied: wrapt in /usr/local/lib/python3.12/dist-packages (from smart-open<8.0.0,>=5.2.1->weasel<0.5.0,>=0.1.0)
Requirement already satisfied: mdurl~=0.1 in /usr/local/lib/python3.12/dist-packages (from markdown-it-py>=2.2.0->rich>=10.11.0)
Collecting en-core-web-sm==3.8.0
```

Using cached [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](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)

✓ 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 pandas as pd
```

```
df=pd.read_excel("mtsamples.xlsx",sheet_name="mtsamples")
df.to_csv("mtsamples.csv")
```

```
raw_df = pd.read_csv("mtsamples.csv", index_col=0)
#raw_df = pd.read_excel("C://dataset/mtsamples.xlsx", index_col=0)
raw_df.head()
```

Unnamed: 0		description	medical_specialty	sample_name	transcription	keywords
0	0	A 23-year-old white female presents with compl...	Allergy / Immunology	Allergic Rhinitis	SUBJECTIVE:, This 23-year-old white female pr...	allergy / immunology, allergic rhinitis, aller...
1	1	Consult for laparoscopic gastric bypass.	Bariatrics	Laparoscopic Gastric Bypass Consult - 2	PAST MEDICAL HISTORY:, He has difficulty climb...	bariatrics, laparoscopic gastric bypass, weigh...
2	2	Consult for laparoscopic gastric bypass.	Bariatrics	Laparoscopic Gastric Bypass Consult - 1	HISTORY OF PRESENT ILLNESS: , I have seen ABC ...	bariatrics, laparoscopic gastric bypass, heart...
3	3	2-D M-Mode. Doppler.	Cardiovascular / Pulmonary	2-D Echocardiogram - 1	2-D M-MODE: , ,1. Left atrial enlargement wit...	cardiovascular / pulmonary, 2-d m-mode, dopple...
4	4	2-D Echocardiogram	Cardiovascular / Pulmonary	2-D Echocardiogram - 2	1. The left ventricular cavity size and wall ...	cardiovascular / pulmonary, 2-d, doppler, echo...

Next steps: [Generate code with raw\\_df](#) [New interactive sheet](#)

```
import spacy
from spacy.lang.en import English
from spacy.pipeline import EntityRuler
```

```
import en_core_web_sm
nlp = en_core_web_sm.load()

# This function generate anotation for each entities and label
def generate_annotation(texts):
    annotations = []
    for text in texts:
        doc = nlp(text)
        entities = []
        for ent in doc.ents:
            entities.append((ent.start_char, ent.end_char, ent.label_))
        annotations.append((text, {'entities': entities}))
    return annotations
```

```
return annotations

# Extract text entities and labels from the dataset (transcription)
raw_df['transcription']=raw_df['transcription'].astype(str)
medical_doc = raw_df['transcription'].tolist()

# Let's generate annotations
annotations = generate_annotation(medical_doc)

# Let's print documents and annotations
print("Document:")
print(annotations[0][0]) # first document text
print("Annotations:")
print(annotations[0][1]) # annotation for the first document
```

```
Document:
SUBJECTIVE:, This 23-year-old white female presents with complaint of allergies. She used to have allergies when she live
Annotations:
{'entities': [(19, 30, 'DATE'), (128, 135, 'GPE'), (200, 208, 'PERSON'), (214, 220, 'LANGUAGE'), (303, 310, 'ORG'), (332, 3
```

```
import en_core_web_sm
nlp = en_core_web_sm.load()

# Create distict colours for labels

col_dict = {}
s_colours = ['#e6194B', '#3cb44b', '#ffe119', '#ffd8b1', '#f58231', '#f032e6', '#42d4f4']
for label, colour in zip(nlp.pipe_labels['ner'], s_colours):
    col_dict[label] = colour

options = {'ents': nlp.pipe_labels['ner'], 'colors':col_dict}

transcription = raw_df['transcription'][0]
doc = nlp(transcription)

spacy.displacy.render(doc, style = 'ent', jupyter = True, options = options)

[(ent.text, ent.label_) for ent in doc.ents]
```

SUBJECTIVE:; This 23-year-old DATE white female presents with complaint of allergies. She used to have allergies when she lived in Seattle GPE but she thinks they are worse here. In the past, she has tried Claritin PERSON , and Zyrtec LANGUAGE . Both worked for short time but then seemed to lose effectiveness. She has used Allegra ORG also. She used that last summer DATE and she began using it again two weeks ago DATE . It does not appear to be working very well. She has used over-the-counter sprays but no prescription nasal sprays. She does have asthma but does not require daily DATE medication for this and does not think it is flaring up.,MEDICATIONS: ; Her only medication currently is Ortho Tri-Cyclen PERSON and the Allegra ORG .,ALLERGIES: ; She has no known medicine allergies.,OBJECTIVE:; Vitals: Weight was 130 pounds QUANTITY and blood pressure 124/78.,HEENT: Her throat was mildly erythematous without exudate. Nasal ORG mucosa was erythematous and swollen. Only clear drainage was seen. TMs were clear.,Neck: Supple without adenopathy.,Lungs: Clear.,ASSESSMENT:; Allergic ORG rhinitis.,PLAN:; 1. She will try Zyrtec NORP instead of Allegra ORG again. Another option will be to use loratadine. She does not think she has prescription coverage so that might be cheaper.,2. Samples of Nasonex ORG two CARDINAL sprays in each nostril given for three weeks DATE . A prescription was written as well.

```
[('23-year-old', 'DATE'),
 ('Seattle', 'GPE'),
 ('Claritin', 'PERSON'),
 ('Zyrtec', 'LANGUAGE'),
 ('Allegra', 'ORG'),
 ('last summer', 'DATE'),
 ('two weeks ago', 'DATE'),
 ('daily', 'DATE'),
 ('Ortho Tri-Cyclen', 'PERSON'),
 ('Allegra', 'ORG'),
 ('130 pounds', 'QUANTITY'),
 ('Nasal', 'ORG'),
 ('Allergic', 'ORG'),
 ('Zyrtec', 'NORP'),
 ('Allegra', 'ORG'),
```

Start coding or generate with AI.