

▼ NLP - EXP - 6

Atharva Prashant Pawar (9427) - [Batch - D]

Note : Turn ON GPU if Available !!!

▼ Q1. Perform NER on given text using Spacy library.

```
import spacy
from spacy import displacy
import pandas as pd
#SpaCy 3.x brought significant speed and accuracy improvements
spacy.__version__
```

'3.6.1'

```
#Download spacy models
!python -m spacy download en_core_web_sm
```

```
2023-09-06 04:20:31.224157: W tensorflow/compiler/tf2tensorrt/utils/py_utils.cc:38] TF-TRT Warning: Could not find TensorRT
Collecting en-core-web-sm==3.6.0
  Downloading https://github.com/explosion/spacy-models/releases/download/en\_core\_web\_sm-3.6.0/en\_core\_web\_sm-3.6.0-py3-none-any.whl
12.8/12.8 MB 32.8 MB/s eta 0:00:00
Requirement already satisfied: spacy<3.7.0,>=3.6.0 in /usr/local/lib/python3.10/dist-packages (from en-core-web-sm==3.6.0) (3.6.1)
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (3.0.11)
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (1.0.0)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (1.0.0)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (2.0.2)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (3.0.2)
Requirement already satisfied: thinc<8.2.0,>=8.1.8 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (8.1.8)
Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (0.9.1)
Requirement already satisfied: srsly<3.0.0,>=2.4.3 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (2.4.3)
Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (2.0.6)
Requirement already satisfied: typer<0.10.0,>=0.3.0 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (0.10.0)
Requirement already satisfied: pathy<0.10.0 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (0.10.0)
Requirement already satisfied: smart-open<7.0.0,>=5.2.1 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (5.2.1)
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (4.38.0)
Requirement already satisfied: numpy<1.15.0 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (1.15.0)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (2.13.0)
Requirement already satisfied: pydantic!=1.8,!<1.8.1,<3.0.0,>=1.7.4 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (1.8.1)
Requirement already satisfied: Jinja2 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (2.11.2)
Requirement already satisfied: setuptools in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (59.0.1)
Requirement already satisfied: packaging<20.0 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (20.0)
Requirement already satisfied: langcodes<4.0.0,>=3.2.0 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (3.2.0)
Requirement already satisfied: annotated-types<0.4.0 in /usr/local/lib/python3.10/dist-packages (from pydantic!=1.8,!<1.8.1,<3.0.0,>=1.7.4->en-core-web-sm==3.6.0) (0.4.0)
Requirement already satisfied: pydantic-core==2.6.1 in /usr/local/lib/python3.10/dist-packages (from pydantic!=1.8,!<1.8.1,<3.0.0,>=1.7.4->en-core-web-sm==3.6.0) (2.6.1)
Requirement already satisfied: typing-extensions<4.6.1 in /usr/local/lib/python3.10/dist-packages (from pydantic!=1.8,!<1.8.1,<3.0.0,>=1.7.4->en-core-web-sm==3.6.0) (4.6.1)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.13.0->spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (2.0.12)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.13.0->spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (2.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.13.0->spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (1.26.15)
Requirement already satisfied: certifi<2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.13.0->spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (2017.4.17)
Requirement already satisfied: blis<0.8.0,>=0.7.8 in /usr/local/lib/python3.10/dist-packages (from thinc<8.2.0,>=8.1.8->spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (0.7.8)
Requirement already satisfied: confection<1.0.0,>=0.0.1 in /usr/local/lib/python3.10/dist-packages (from thinc<8.2.0,>=8.1.8->spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (0.0.1)
Requirement already satisfied: click<9.0.0,>=7.1.1 in /usr/local/lib/python3.10/dist-packages (from typer<0.10.0,>=0.3.0->spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (7.1.1)
Requirement already satisfied: MarkupSafe<2.0 in /usr/local/lib/python3.10/dist-packages (from Jinja2->spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (2.0.1)
✓ Download and installation successful
You can now load the package via spacy.load('en_core_web_sm')
```

```
# Load SpaCy model
nlp = spacy.load("en_core_web_sm")
#nlp = spacy.load("en_core_web_md")
#nlp = spacy.load("en_core_web_lg")

doc = nlp(text)

entities = []
labels = []
position_start = []
position_end = []

for ent in doc.ents:
    entities.append(ent)
    labels.append(ent.label_)
    position_start.append(ent.start_char)
    position_end.append(ent.end_char)
```

```
df = pd.DataFrame({'Entities':entities,'Labels':labels,'Position_Start':position_start, 'Position_End':position_end})
```

df

	Entities	Labels	Position_Start	Position_End	
0	(Apple)	ORG	0	5	
1	(Zoom)	GPE	15	19	
2	(China)	GPE	23	28	
3	(Wednesday, 6th, May)	DATE	32	49	
4	(Apple)	ORG	74	79	
5	(Google)	ORG	84	90	
6	(5, %)	PERCENT	105	107	
7	(Dow, Jones, Index)	ORG	111	126	
8	(the, United, States, of, America)	GPE	130	158	

```
spacy.explain("ORG")

'Companies, agencies, institutions, etc.'
```

Q2. Design and implement a customized NER model using Spacy.

```
import spacy

# !python -m spacy download en_core_web_sm
!python -m spacy download en_core_web_lg

2023-09-06 04:22:07.520256: W tensorflow/compiler/tf2tensorrt/utils/py_utils.cc:38] TF-TRT Warning: Could not find TensorRT
Collecting en-core-web-lg==3.6.0
  Downloading https://github.com/explosion/spacy-models/releases/download/en\_core\_web\_lg-3.6.0/en\_core\_web\_lg-3.6.0-py3-none-any.whl
    587.7/587.7 MB 1.7 MB/s eta 0:00:00
Requirement already satisfied: spacy<3.7.0,>=3.6.0 in /usr/local/lib/python3.10/dist-packages (from en-core-web-lg==3.6.0) (3.6.1)
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-lg==3.6.0) (3.0.11)
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-lg==3.6.0) (1.0.0)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-lg==3.6.0) (1.0.10)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-lg==3.6.0) (2.0.3)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-lg==3.6.0) (3.0.2)
Requirement already satisfied: thinc<8.2.0,>=8.1.8 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-lg==3.6.0) (8.1.8)
Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-lg==3.6.0) (1.1.2)
Requirement already satisfied: srsly<3.0.0,>=2.4.3 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-lg==3.6.0) (2.4.3)
Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-lg==3.6.0) (2.0.6)
Requirement already satisfied: typer<0.10.0,>=0.3.0 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-lg==3.6.0) (0.9.0)
Requirement already satisfied: pathy<0.10.0 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-lg==3.6.0) (0.10.0)
Requirement already satisfied: smart-open<7.0.0,>=5.2.1 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-lg==3.6.0) (7.0.0)
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-lg==3.6.0) (4.64.1)
Requirement already satisfied: numpy>=1.15.0 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-lg==3.6.0) (1.24.3)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-lg==3.6.0) (2.31.0)
Requirement already satisfied: pydantic!=1.8,!1.8.1,<3.0.0,>=1.7.4 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-lg==3.6.0) (2.6.1)
Requirement already satisfied: jinja2 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-lg==3.6.0) (3.1.2)
Requirement already satisfied: setuptools in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-lg==3.6.0) (68.0.0)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-lg==3.6.0) (23.1)
Requirement already satisfied: langcodes<4.0.0,>=3.2.0 in /usr/local/lib/python3.10/dist-packages (from spacy<3.7.0,>=3.6.0->en-core-web-lg==3.6.0) (3.3.0)
Requirement already satisfied: annotated-types>=0.4.0 in /usr/local/lib/python3.10/dist-packages (from pydantic!=1.8,!1.8.1,<3.0.0,>=1.7.4->pydantic-core==2.6.1) (0.5.0)
Requirement already satisfied: pydantic-core==2.6.1 in /usr/local/lib/python3.10/dist-packages (from pydantic!=1.8,!1.8.1,<3.0.0,>=1.7.4->pydantic-core==2.6.1) (2.6.1)
Requirement already satisfied: typing-extensions>=4.6.1 in /usr/local/lib/python3.10/dist-packages (from pydantic!=1.8,!1.8.1,<3.0.0,>=1.7.4->pydantic-core==2.6.1) (4.6.1)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.13.0->urllib3<3,>=1.21.1) (3.2.0)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.13.0->urllib3<3,>=1.21.1) (3.4)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.13.0->urllib3<3,>=1.21.1) (2.0.7)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.13.0->urllib3<3,>=1.21.1) (2023.7.22)
Requirement already satisfied: blis<0.8.0,>=0.7.8 in /usr/local/lib/python3.10/dist-packages (from thinc<8.2.0,>=8.1.8->spacy<3.7.0) (0.7.10)
Requirement already satisfied: confection<1.0.0,>=0.0.1 in /usr/local/lib/python3.10/dist-packages (from thinc<8.2.0,>=8.1.8->spacy<3.7.0) (0.0.4)
Requirement already satisfied: click<9.0.0,>=7.1.1 in /usr/local/lib/python3.10/dist-packages (from typer<0.10.0,>=0.3.0->spacy<3.7.0) (8.1.6)
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.10/dist-packages (from jinja2->spacy<3.7.0,>=3.6.0->en-core-web-lg==3.6.0) (2.1.2)
Installing collected packages: en-core-web-lg
Successfully installed en-core-web-lg-3.6.0
✓ Download and installation successful
You can now load the package via spacy.load('en_core_web_lg')
```

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

<spacy.lang.en.English at 0x7d4db2d9a650>
```

```
doc = nlp("Donad Trump was President of USA")
doc
```

```
Donad Trump was President of USA
```

```
type(doc)
```

```
spacy.tokens.doc.Doc
```

```
doc.ents
```

```
(Donad Trump, USA)
```

```
doc.ents[0], type(doc.ents[0])
```

```
(Donad Trump, spacy.tokens.span.Span)
```

```
from spacy import displacy
displacy.render(doc, style="ent", jupyter=True)
```

```
Donad Trump PERSON was President of USA GPE
```

```
import json
# https://www.kaggle.com/datasets/finaleepoch/medical-ner
with open('/content/Corona2.json', 'r') as f:
    data = json.load(f)
```

```
data['examples'][0]
```

```

        'annotator_id': 1,
        'tagged_token_id': 'f04a6b7e-8904-405c-9301-e4543238b7a5',
        'name': 'Ashpat123',
        'reason': 'exploration'}],
        'model_annotations': []],
        'classifications': {}

data['examples'][0].keys()

dict_keys(['id', 'content', 'metadata', 'annotations', 'classifications'])

data['examples'][0]['content']

'While bismuth compounds (Pepto-Bismol) decreased the number of bowel movements in those with traveler
s' diarrhea, they do not decrease the length of illness.[91] Anti-motility agents like loperamide are
also effective at reducing the number of stools but not the duration of disease.[8] These agents shoul
d be used only if bloody diarrhea is not present.[92]\n\nDiosmectite, a natural aluminomagnesium silic
ate clay, is effective in alleviating symptoms of acute diarrhea in children,[93] and also has some ef
fects in chronic functional diarrhea. radiation-induced diarrhea. and chemotherapy-induced diarrhea.[4

data['examples'][0]['annotations'][0]

{'id': '0825a1bf-6a6e-4fa2-be77-8d104701eaed',
 'tag_id': 'c06bd022-6ded-44a5-8d90-f17685bb85a1',
 'end': 371,
 'start': 360,
 'example_id': '18c2f619-f102-452f-ab81-d26f7e283ffe',
 'tag_name': 'Medicine',
 'value': 'Diosmectite',
 'correct': None,
 'human_annotations': [{'timestamp': '2020-03-21T00:24:32.098000Z',
 'annotator_id': 1,
 'tagged_token_id': '0825a1bf-6a6e-4fa2-be77-8d104701eaed',
 'name': 'Ashpat123',
 'reason': 'exploration'}],
 'model_annotations': []}

training_data = []
for example in data['examples']:
    temp_dict = {}
    temp_dict['text'] = example['content']
    temp_dict['entities'] = []
    for annotation in example['annotations']:
        start = annotation['start']
        end = annotation['end']
        label = annotation['tag_name'].upper()
        temp_dict['entities'].append((start, end, label))
    training_data.append(temp_dict)

print(training_data[0])

{'text': "While bismuth compounds (Pepto-Bismol) decreased the number of bowel movements in those with travelers' diarrhea, they do

training_data[0]['text']

'While bismuth compounds (Pepto-Bismol) decreased the number of bowel movements in those with traveler
s' diarrhea, they do not decrease the length of illness.[91] Anti-motility agents like loperamide are
also effective at reducing the number of stools but not the duration of disease.[8] These agents shoul
d be used only if bloody diarrhea is not present.[92]\n\nDiosmectite, a natural aluminomagnesium silic
ate clay, is effective in alleviating symptoms of acute diarrhea in children,[93] and also has some ef
fects in chronic functional diarrhea. radiation-induced diarrhea. and chemotherapy-induced diarrhea.[4

training_data[0]['entities']

[(360, 371, 'MEDICINE'),
 (383, 408, 'MEDICINE'),
 (104, 112, 'MEDICALCONDITION'),
 (679, 689, 'MEDICINE'),
 (6, 23, 'MEDICINE'),
 (25, 37, 'MEDICINE'),
 (461, 470, 'MEDICALCONDITION'),
 (577, 589, 'MEDICINE'),
 (853, 865, 'MEDICALCONDITION'),
 (188, 198, 'MEDICINE'),
 (754, 762, 'MEDICALCONDITION'),
 (870, 880, 'MEDICALCONDITION'),
 (823, 833, 'MEDICINE'),
 (852, 853, 'MEDICALCONDITION'),
 (461, 469, 'MEDICALCONDITION'),
 (535, 543, 'MEDICALCONDITION'),
 (692, 704, 'MEDICINE'),
 (563, 571, 'MEDICALCONDITION')]

```

```

training_data[0]['text'][360:371]

'Diosmectite'

from spacy.tokens import DocBin
from tqdm import tqdm

nlp = spacy.blank("en") # load a new spacy model
doc_bin = DocBin()

from spacy.util import filter_spans

for training_example in tqdm(training_data):
    text = training_example['text']
    labels = training_example['entities']
    doc = nlp.make_doc(text)
    ents = []
    for start, end, label in labels:
        span = doc.char_span(start, end, label=label, alignment_mode="contract")
        if span is None:
            print("Skipping entity")
        else:
            ents.append(span)
    filtered_ents = filter_spans(ents)
    doc.ents = filtered_ents
    doc_bin.add(doc)

doc_bin.to_disk("train.spacy")

```

```

100%|██████████| 31/31 [00:00<00:00, 352.17it/s]Skipping entity
Skipping entity
Skipping entity
Skipping entity
Skipping entity
Skipping entity
Skipping entity
Skipping entity
Skipping entity
Skipping entity
Skipping entity
Skipping entity
Skipping entity
Skipping entity
Skipping entity
Skipping entity

```

<https://spacy.io/usage/training#quickstart>

...

create a base_config.cfg file and paste the below content in it

--- File Name: base_config.cfg

--- File Content:

```

# This is an auto-generated partial config. To use it with 'spacy train'
# you can run spacy init fill-config to auto-fill all default settings:
# python -m spacy init fill-config ./base_config.cfg ./config.cfg

```

```

[paths]
train = null
dev = null
vectors = "en_core_web_lg"
[system]
gpu_allocator = null

```

```

[nlp]
lang = "en"
pipeline = ["tok2vec", "ner"]
batch_size = 1000

```

```

[components]

```

```

[components.tok2vec]
factory = "tok2vec"

```

```

[components.tok2vec.model]
@architectures = "spacy.Tok2Vec.v2"

```

```

[components.tok2vec.model.embed]
@architectures = "spacy.MultiHashEmbed.v2"

```

```
width = ${components.tok2vec.model.encode.width}
attrs = ["NORM", "PREFIX", "SUFFIX", "SHAPE"]
rows = [5000, 1000, 2500, 2500]
include_static_vectors = true
```

```
[components.tok2vec.model.encode]
@architectures = "spacy.MaxoutWindowEncoder.v2"
width = 256
depth = 8
window_size = 1
maxout_pieces = 3
```

```
[components.ner]
factory = "ner"
```

```
[components.ner.model]
@architectures = "spacy.TransitionBasedParser.v2"
state_type = "ner"
extra_state_tokens = false
hidden_width = 64
maxout_pieces = 2
use_upper = true
n0 = null
```

```
[components.ner.model.tok2vec]
@architectures = "spacy.Tok2VecListener.v1"
width = ${components.tok2vec.model.encode.width}
```

```
[corpora]
```

```
[corpora.train]
@readers = "spacy.Corpus.v1"
path = ${paths.train}
max_length = 0
```

```
[corpora.dev]
@readers = "spacy.Corpus.v1"
path = ${paths.dev}
max_length = 0
```

```
[training]
dev_corpus = "corpora.dev"
train_corpus = "corpora.train"
```

```
[training.optimizer]
@optimizers = "Adam.v1"
```

```
[training.batcher]
@batchers = "spacy.batch_by_words.v1"
discard_oversize = false
tolerance = 0.2
```

```
[training.batcher.size]
@schedules = "compounding.v1"
start = 100
stop = 1000
compound = 1.001
```

```
[initialize]
vectors = ${paths.vectors}
'''
```

```
!python -m spacy init fill-config base_config.cfg config.cfg
```

```
2023-09-06 04:34:38.926520: W tensorflow/compiler/tf2tensorrt/utils/py_utils.cc:38] TF-TRT Warning: Could not find TensorRT
✓ Auto-filled config with all values
✓ Saved config
config.cfg
You can now add your data and train your pipeline:
python -m spacy train config.cfg --paths.train ./train.spacy --paths.dev ./dev.spacy
```

```
!python -m spacy train config.cfg --output ./ --paths.train ./train.spacy --paths.dev ./train.spacy
```

```
2023-09-06 04:34:56.684124: W tensorflow/compiler/tf2tensorrt/utils/py_utils.cc:38] TF-TRT Warning: Could not find TensorRT
i Saving to output directory: .
i Using CPU

===== Initializing pipeline =====
✓ Initialized pipeline

===== Training pipeline =====
i Pipeline: ['tok2vec', 'ner']
```

i Initial learn rate: 0.001

E	#	LOSS	TOK2VEC	LOSS NER	ENTS_F	ENTS_P	ENTS_R	SCORE
0	0	0.00		153.29	0.55	0.91	0.39	0.01
7	200	499.25		3640.10	70.00	82.80	60.63	0.70
14	400	192.16		1153.36	87.20	97.10	79.13	0.87
22	600	72.20		471.57	93.66	97.45	90.16	0.94
30	800	117.10		435.09	95.67	95.67	95.67	0.96
40	1000	97.26		314.46	96.43	97.20	95.67	0.96
51	1200	96.63		232.85	97.84	97.65	98.03	0.98
65	1400	311.37		219.57	97.46	96.89	98.03	0.97
81	1600	129.49		227.94	98.43	98.43	98.43	0.98
101	1800	125.51		232.96	98.81	99.21	98.43	0.99
127	2000	440.32		253.24	98.82	98.82	98.82	0.99
159	2200	92.65		240.60	98.81	99.21	98.43	0.99
198	2400	48.37		250.91	98.82	98.44	99.21	0.99
246	2600	85.11		303.28	98.82	98.44	99.21	0.99
295	2800	68.32		287.54	98.82	98.82	98.82	0.99

Aborted.

```
nlp_ner = spacy.load("model-best")
```

Custom Spacy Model Inference

```
doc = nlp_ner("While bismuth compounds (Pepto-Bismol) decreased the number of bowel movements in those with travelers' diarrhea, they do
```

```
colors = {"PATHOGEN": "#F67DE3", "MEDICINE": "#7DF6D9", "MEDICALCONDITION": "#a6e22d"}
options = {"colors": colors}
```

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

While bismuth compounds **MEDICINE** (Pepto-Bismol **MEDICINE**) decreased the number of bowel movements in those with travelers' diarrhea **MEDICALCONDITION** , they do not decrease the length of illness.[91] Anti-motility agents like

Postlab - Question

Q1. Extract all the Geographical (cities, Countries, states) names from a given text

```
Q1text = """Kiran want to know the famous foods in each state of India.
So, he opened Google and search for this question. Google showed that
in Delhi it is Chaat, in Gujarat it is Dal Dhokli, in Tamilnadu it is
Pongal, in Andhrapradesh it is Biryani, in Assam it is Papaya Khar,
in Bihar it is Litti Chowkha and so on for all other states"""
```

```
import spacy
nlp = spacy.load("en_core_web_sm")
doc = nlp(Q1text)
geographical_names = []

for entity in doc.ents:
    # Checking: if the entity is a geographical location (GPE: Geo-Political Entity)
    if entity.label_ == "GPE":
        geographical_names.append(entity.text)

# Removing duplicates from the list
geographical_names = list(set(geographical_names))

print("Geographical location Names:", geographical_names)

Geographical location Names: ['Bihar', 'Pongal', 'Andhrapradesh', 'Assam', 'Gujarat', 'India', 'Delhi', 'Tamilnadu']
```

Q2. Extract all the birth dates of cricketers in the given Text

```
Q2text = """Sachin Tendulkar was born on 24 April 1973, Virat Kohli was
born on 5 November 1988, Dhoni was born on 7 July 1981
and finally Ricky ponting was born on 19 December 1974."""
```

```
import spacy
```

```
nlp = spacy.load("en_core_web_sm")
doc = nlp(Q2text)
birth_dates = []

for entity in doc.ents:
    # Checking: if the entity is recognized as a DATE entity (DATE: Date)
    if entity.label_ == "DATE":
        birth_dates.append(entity.text)

print("All Birth Dates:", birth_dates)

All Birth Dates: ['24 April 1973', '5 November 1988', '7 July 1981', '19 December 1974']
```

✓ 0s completed at 10:42 AM

