

# Ingénierie cognitive

## NER\_covid\_article

### RÉALISÉ PAR :

- EL-HASSANI Mohamed
- EL JAMAAI Mohamed

### ENCADRÉ PAR :

- Mme. Maryem RHANOU
- Mme. Mounia MIKRAM

## table de matière

|   |           |
|---|-----------|
| <b>I.INTRODUCTION :</b> .....             | <b>3</b>  |
| <b>II. DESCRIPTION DU SUJET :</b> .....   | <b>4</b>  |
| <b>III. SOURCES DE DONNEES :</b> .....    | <b>4</b>  |
| <b>IV.ANNOTATION :</b> .....              | <b>6</b>  |
| 1.Annotation par NER Annotation : .....   | 6         |
| 2.Annotation avec le modèle Scapy : ..... | 7         |
| <b>V.CONCLUSION :</b> .....               | <b>11</b> |

## I.Introduction :

La prise en charge des patients atteints de Covid-19 s'est améliorée au fil du temps, en particulier grâce à l'acquisition de connaissances sur la maladie et à l'élaboration de protocoles de traitement plus efficaces. En France, les hôpitaux ont mis en place des unités de soins dédiées aux patients atteints de Covid-19, qui leur permettent de recevoir des soins adaptés à leur état de santé. De nombreux professionnels de santé ont également été formés pour gérer les cas graves de Covid-19 et pour utiliser les traitements disponibles de manière appropriée.

Il est important de noter que la prise en charge des patients atteints de Covid-19 reste un défi pour les systèmes de santé de nombreux pays, en particulier lorsque le nombre de cas est élevé. Les hôpitaux peuvent être surchargés et il peut être difficile de fournir des soins adéquats à tous les patients. C'est pourquoi il est important de continuer à respecter les mesures de prévention recommandées, telles que le lavage fréquent des mains et le port d'un masque, pour aider à freiner la propagation de la maladie.

## II. Description du sujet :

Le contexte de notre projet est d'annoter un fichier texte qui a été obtenu en utilisant un outil de "scraping" de données. Dans notre cas, on utilise un article du journal "le monde" sous le nom 'La prise en charge du Covid-19 se normalise de plus en plus dans les hôpitaux français', comme source de données pour étudier la prise en charge du Covid-19, ainsi que nous spécifions les informations que nous voulons extraire dans le texte. Par exemple, nous pouvons vouloir définir des termes techniques liés au COVID-19, des dates, des lieux, des personnes mentionnées, etc. Pour cela, on annotera notre fichier scrapper par le logiciel ou bien site web 'NER ANNOTATOR', après on fait l'extraction de notre fichier JSON qui nous appliquons sur elle le modèle spacy.

## III. Sources de données :



Nos données seront scrappées URL :

' [https://www.lemonde.fr/planete/article/2022/12/20/la-prise-en-charge-du-covid-19-se-normalise-de-plus-en-plus-dans-les-hopitaux-francais\\_6155158\\_3244.html](https://www.lemonde.fr/planete/article/2022/12/20/la-prise-en-charge-du-covid-19-se-normalise-de-plus-en-plus-dans-les-hopitaux-francais_6155158_3244.html) '

- Installation des bibliothèques utilisées :

```
[ ] 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/python3.8/dist-packages (3.7)
Requirement already satisfied: joblib in /usr/local/lib/python3.8/dist-packages (from nltk) (1.2.0)
Requirement already satisfied: click in /usr/local/lib/python3.8/dist-packages (from nltk) (7.1.2)
Requirement already satisfied: regex<=2021.8.3 in /usr/local/lib/python3.8/dist-packages (from nltk) (2022.6.2)
Requirement already satisfied: tqdm in /usr/local/lib/python3.8/dist-packages (from nltk) (4.64.1)
```

```
[ ] pip install newspaper3k
```

```
Downloading cssselect-1.2.0-py2.py3-none-any.whl (18 kB)
Requirement already satisfied: PyYAML<=3.11 in /usr/local/lib/python3.8/dist-packages (from newspaper3k) (6.0)
Collecting feedfinder2<=0.0.4
  Downloading feedfinder2-0.0.4.tar.gz (3.3 kB)
Collecting tldextract<=2.0.1
  Downloading tldextract-3.4.0-py3-none-any.whl (93 kB)
    |████████████████████████████████████████| 93 kB 2.3 MB/s
Requirement already satisfied: BeautifulSoup4<=4.4.1 in /usr/local/lib/python3.8/dist-packages (from newspaper3k) (4.6.3)
Requirement already satisfied: Pillow<=3.3.0 in /usr/local/lib/python3.8/dist-packages (from newspaper3k) (7.1.2)
Requirement already satisfied: nltk<=3.2.1 in /usr/local/lib/python3.8/dist-packages (from newspaper3k) (3.7)
Collecting tinysegmenter==0.3
  Downloading tinysegmenter-0.3.tar.gz (16 kB)
```

- TEXT Scraper :

```
[ ] print(article.text)
```

```
Dans la zone d'accueil des patients atteints du Covid-19, au service d'urgences de l'hôpital Pasteur 2 de Nice, le 26 février 2021. LAURENT CARRÉ

Après presque trois ans d'épidémie, quel poids le Covid-19 représente-t-il pour l'hôpital ? Alors que les virus respiratoires de la saison hivernale mettent les établissements sous t

Lire aussi : Article réservé à nos abonnés Covid-19 : en France, l'épidémie progresse moins vite

« On n'est plus du tout dans la même dimension émotionnelle et de risque infectieux qu'en 2020, souligne Marc Lambert, chef de service de médecine polyvalente de post-urgence au cent

Un mois et demi après le début de la neuvième vague de Covid-19, début novembre, le nombre de contaminations semble avoir atteint un plateau haut, avec 62 000 cas recensés le 12 déc

Le profil des patients a changé

Dans la plupart des établissements interrogés, les patients traités pour Covid-19 représentent environ 10 % des hospitalisés, à peu près au même niveau, voire en deçà, que la grippe

Il vous reste 60.03% de cet article à lire. La suite est réservée aux abonnés.
```

```
#get the top image
article.top_image
```

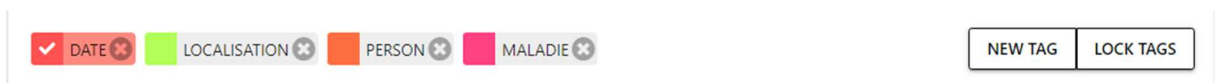
```
'https://img.lemde.fr/2022/12/19/0/0/4000/2666/1440/960/60/0/68ba269_1671491838794-hopital-pasteur-laurent-carre-66.jpg'
```

```
[ ] print(article.summary)
```

## IV.Annotation :

### 1.Annotation par NER Annotation :


- Crée nos entités manuellement :



- Appliquer nos entités à notre fichier text importer :



- Exporter notre fichier JSON :

 annotations

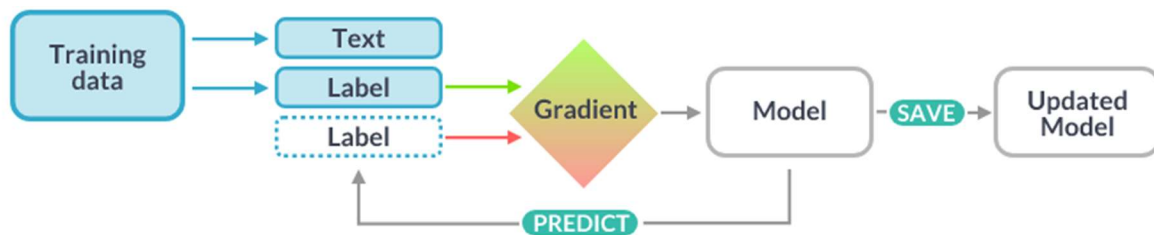
21/12/2022 01:29

Fichier source JSON

3 Ko



## Entraînez notre pipeline NER personnalisé avec Spacy en 5 étapes simples.



- Instalation de notre environnement :

```
[ ] import spacy

/usr/local/lib/python3.8/dist-packages/torch/cuda/_init__.py:497: UserWarning: Can't initialize NVML
warnings.warn("Can't initialize NVML")

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

/usr/local/lib/python3.8/dist-packages/torch/cuda/_init__.py:497: UserWarning: Can't initialize NVML
warnings.warn("Can't initialize NVML")
2022-12-21 00:39:57.551839: E tensorflow/stream_executor/cuda/cuda_driver.cc:271] failed call to cuInit: CUDA_ERROR_NO_DEVICE: no CUDA-capable de
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Collecting en-core-web-lg==3.4.1
  Downloading https://github.com/explosion/spacy-models/releases/download/en\_core\_web\_lg-3.4.1/en\_core\_web\_lg-3.4.1-py3-none-any.whl (587.7 MB)
  [REDACTED] 587.7 MB 16 kB/s
Requirement already satisfied: spacy<3.5.0,>=3.4.0 in /usr/local/lib/python3.8/dist-packages (from en-core-web-lg==3.4.1) (3.4.4)
Requirement already satisfied: typer<0.8.0,>=0.3.0 in /usr/local/lib/python3.8/dist-packages (from spacy<3.5.0,>=3.4.0->en-core-web-lg==3.4.1) (0
Requirement already satisfied: pydantic<1.8.1,>=1.8.1,<1.11.0,>=1.7.4 in /usr/local/lib/python3.8/dist-packages (from spacy<3.5.0,>=3.4.0->en-core
Requirement already satisfied: smart-open<7.0.0,>=5.2.1 in /usr/local/lib/python3.8/dist-packages (from spacy<3.5.0,>=3.4.0->en-core-web-lg==3.4.
Requirement already satisfied: Jinja2 in /usr/local/lib/python3.8/dist-packages (from spacy<3.5.0,>=3.4.0->en-core-web-lg==3.4.1) (2.11.3)
Requirement already satisfied: setuptools in /usr/local/lib/python3.8/dist-packages (from spacy<3.5.0,>=3.4.0->en-core-web-lg==3.4.1) (57.4.0)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python3.8/dist-packages (from spacy<3.5.0,>=3.4.0->en-core-web-lg==3.4.1) (4
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in /usr/local/lib/python3.8/dist-packages (from spacy<3.5.0,>=3.4.0->en-core-web-lg==3.4.1) (4
Requirement already satisfied: packaging<20.0 in /usr/local/lib/python3.8/dist-packages (from spacy<3.5.0,>=3.4.0->en-core-web-lg==3.4.1) (21.3)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in /usr/local/lib/python3.8/dist-packages (from spacy<3.5.0,>=3.4.0->en-core-web-lg==3.4.1) (8
Requirement already satisfied: thinc<8.2.0,>=8.1.0 in /usr/local/lib/python3.8/dist-packages (from spacy<3.5.0,>=3.4.0->en-core-web-lg==3.4.1) (8
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/python3.8/dist-packages (from spacy<3.5.0,>=3.4.0->en-core-web-lg==3.4
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in /usr/local/lib/python3.8/dist-packages (from spacy<3.5.0,>=3.4.0->en-core-web-lg==3.4
```

- Importer notre data file :

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

<spacy.lang.en.English at 0x7f5b5ae61a60>

[ ] import json
data=json.load(open('/content/annotations.json'))
print(annotations)

': [{"La prise en charge du Covid-19 se normalise de plus en plus dans les hôpitaux français\r\n\r\nDans la zone d'accueil des patients atteints d

[ ] data['classes'][0]

'DATE'

[ ] data['annotations']

[["La prise en charge du Covid-19 se normalise de plus en plus dans les hôpitaux français\r\n\r\n\r\nDans la zone d'accueil des patients atteints du
Covid-19, au service d'urgences de l'hôpital Pasteur 2 de Nice, le 26 février 2021. LAURENT CARRÉ\r\n\r\n\r\nAprès presque trois ans d'épidémie, quel
poids le Covid-19 représente-t-il pour l'hôpital ? Alors que les virus respiratoires de la saison hivernale mettent les établissements sous
tension, le Covid complique encore la difficile comptabilité des lits hospitaliers, s'ajoutant aux épidémies de grippe et de bronchiolite qui
submergent les services. Pour autant, une forme de normalisation de la maladie s'observe depuis plusieurs mois déjà. Les « unités Covid »
regroupant les patients touchés par le virus, isolés du reste de l'hôpital, semblent un lointain souvenir dans de nombreux établissements. De
plus en plus, le Covid-19 s'installe dans le quotidien des soignants.\r\n\r\n\r\nLire aussi : Article réservé à nos abonnés Covid-19 : en France,
l'épidémie progresse moins vite\r\n\r\n\r\n« On n'est plus du tout dans la même dimension émotionnelle et de risque infectieux qu'en 2020, souligne
Marc Lambert, chef de service de médecine polyvalente de post-urgence au centre hospitalier universitaire (CHU) de Lille. Notre vie a changé ; on
```

- Train notre data :

```
[ ] entity_name = "LOCALISATION"
train_data = data['annotations']
train_data = [tuple(i) for i in train_data]

[ ] train_data

[["La prise en charge du Covid-19 se normalise de plus en plus dans les hôpitaux français\r\n\r\n\r\nDans la zone d'accueil
des patients atteints du Covid-19, au service d'urgences de l'hôpital Pasteur 2 de Nice, le 26 février 2021. LAURENT
CARRÉ\r\n\r\n\r\nAprès presque trois ans d'épidémie, quel poids le Covid-19 représente-t-il pour l'hôpital ? Alors que les
virus respiratoires de la saison hivernale mettent les établissements sous tension, le Covid complique encore la difficile
comptabilité des lits hospitaliers, s'ajoutant aux épidémies de grippe et de bronchiolite qui submergent les services.
Pour autant, une forme de normalisation de la maladie s'observe depuis plusieurs mois déjà. Les « unités Covid »
regroupant les patients touchés par le virus, isolés du reste de l'hôpital, semblent un lointain souvenir dans de nombreux
établissements. De plus en plus, le Covid-19 s'installe dans le quotidien des soignants.\r\n\r\n\r\nLire aussi : Article
réservé à nos abonnés Covid-19 : en France, l'épidémie progresse moins vite\r\n\r\n\r\n« On n'est plus du tout dans la même
dimension émotionnelle et de risque infectieux qu'en 2020, souligne Marc Lambert, chef de service de médecine polyvalente
de post-urgence au centre hospitalier universitaire (CHU) de Lille. Notre vie a changé ; on peut dire "merci" à la
vaccination. » Pour l'ancien responsable de l'unité Covid mise en place lors de la première vague et répartie dans les
services depuis 2021 - comme dans beaucoup d'établissements -, « le Covid doit être intégré comme un élément quotidien
dans la prise en charge des patients ». « Ce n'est plus logique de traiter ces patients à part, les soignants s'adaptent
et apprennent à gérer ce risque infectieux », ajoute le médecin.\r\n\r\n\r\nUn mois et demi après le début de la neuvième
vague de Covid-19, début novembre, le nombre de contaminations semble avoir atteint un plateau haut, avec 62 000 cas
recensés le 12 décembre, ce que confirmera peut-être un pic dans les prochains jours. Le nombre d'admissions à l'hôpital
continue, lui, d'augmenter, et dépasse déjà les niveaux de la septième vague, en juillet, avec 1 500 personnes
hospitalisées en moyenne par jour. Des indicateurs élevés, donc, mais qui n'inquiètent plus les soignants de la même
manière \r\n\r\n\r\nLe profil des patients a changé"]
```



- Convertir notre data to spacy format :

Conversion of Data to .spacy format

Copy the training data and below script and run from command line

```
[ ] import pandas as pd
import os
from tqdm import tqdm
import spacy
from spacy.tokens import DocBin

#nlp = spacy.blank("en") # load a new spacy model
nlp = spacy.load("en_core_web_sm") # load other spacy model

db = DocBin() # create a DocBin object

for text, annot in tqdm(train_data): # data in previous format
    doc = nlp.make_doc(text) # create doc object from text
    ents = []
    for start, end, label in annot["entities"]: # add character indexes
        span = doc.char_span(start, end, label=label, alignment_mode="contract")
        if span is None:
            print("Skipping entity")
        else:
            ents.append(span)
    doc.ents = ents # label the text with the ents
    db.add(doc)

os.chdir(r'./content')
db.to_disk("./train.spacy") # save the docbin object

100%|██████████| 1/1 [00:00<00:00, 53.97it/s]
```

- Tester notre model :

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

/usr/local/lib/python3.8/dist-packages/torch/cuda/__init__.py:497: UserWarning: Can't initialize NVML
warnings.warn("Can't initialize NVML")
2022-12-21 03:43:46.753621: E tensorflow/stream_executor/cuda/cuda_driver.cc:271] failed call to cuInit: CUDA_ERROR_NO_DEVICE: no CUDA-capable device is detected
✓ Created output directory: output
i Saving to output directory: output
i Using CPU

===== Initializing pipeline =====
[2022-12-21 03:43:47,823] [INFO] Set up nlp object from config
INFO:spacy:Set up nlp object from config
[2022-12-21 03:43:47,841] [INFO] Pipeline: ['tok2vec', 'ner']
INFO:spacy:Pipeline: ['tok2vec', 'ner']
[2022-12-21 03:43:47,845] [INFO] Created vocabulary
INFO:spacy:Created vocabulary
[2022-12-21 03:43:47,846] [INFO] Finished initializing nlp object
INFO:spacy:Finished initializing nlp object
[2022-12-21 03:43:48,128] [INFO] Initialized pipeline components: ['tok2vec', 'ner']
INFO:spacy:Initialized pipeline components: ['tok2vec', 'ner']
✓ 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 0.00 201.50 0.00 0.00 0.00 0.00
200 200 2726.99 5403.23 100.00 100.00 100.00 1.00
400 400 0.00 0.00 100.00 100.00 100.00 1.00
600 600 0.00 0.00 100.00 100.00 100.00 1.00
800 800 0.00 0.00 100.00 100.00 100.00 1.00
1000 1000 0.00 0.00 100.00 100.00 100.00 1.00
1200 1200 0.00 0.00 100.00 100.00 100.00 1.00
1400 1400 0.00 0.00 100.00 100.00 100.00 1.00
1600 1600 0.00 0.00 100.00 100.00 100.00 1.00
1800 1800 0.00 0.00 100.00 100.00 100.00 1.00
✓ Saved pipeline to output directory
output/model-last
```

- Execution final :

```
[ ] nlp=spacy.load("en_core_web_sm")
doc = nlp("La prise en charge du Covid-19 se normalise de plus en plus dans les hôpitaux français Dans la zone d'accueil c
spacy.displacy.render(doc, style="ent", jupyter=True) # display in Jupyter
```

La GPE prise en charge du Covid-19 se normalise de plus en plus dans NORP les hôpitaux français Dans NORP la zone d'accueil des patients atteints du Covid-19, au service d'urgences de l'hôpital Pasteur 2 de Nice, le 26 CARDINAL février 2021. LAURENT CARRÉ Après presque trois ans d'épidémie, quel poids PERSON le Covid-19 représente-t-il pour l'hôpital ? Alors que les virus respiratoires de la saison PERSON hivernale mettent les établissements sous tension, le Covid PERSON complique encore la difficile comptabilité des lits hospitaliers, s'ajoutant aux épidémies de grippe et de bronchiolite qui submergent les services. Pour autant, une forme de normalisation de la maladie s'observe depuis plusieurs mois déjà. Les « unités Covid PERSON » regroupant les patients touchés par le virus, isolés du reste de l'hôpital ORG, semblent un ORG lointain souvenir dans de nombreux établissements ORG. De plus en plus, le Covid-19 s'installe dans NORP le quotidien des soignants. Lire aussi : Article réservé à nos abonnés Covid-19 : en France GPE, l'épidémie progresse moins vite On n'est plus du tout dans NORP la même dimension émotionnelle PERSON et de risque infectieux qu'en 2020 DATE, souligne Marc Lambert PERSON, chef de service de médecine polyvalente de post-urgence au centre ORG hospitalier universitaire (CHU) de Lille. Notre GPE vie a changé ; on peut dire "merci" à la vaccination. » Pour l'ancien responsable de l'unité Covid PERSON mise en place lors de la première vague et répartie dans NORP les services depuis 2021

➤ Pour plus de détaille :

[https://github.com/EL-HASSANI/NER\\_article\\_covid](https://github.com/EL-HASSANI/NER_article_covid)

## V.Conclusion :

Il y a plusieurs raisons possibles pour lesquelles le modèle d'entraînement ne pas être tout à fait efficace, comme des biais dans les données ou une mauvaise préparation des données. Pour améliorer les performances du modèle, il peut être utile de diagnostiquer le problème et de tester différentes approches pour le résoudre.

Ce projet nous a permis de mettre en pratique les connaissances acquises lors de cours d'Ingénierie Cognitive, telles que l'utilisation de Python dans divers contextes, allant de l'extraction de données web à l'établissement d'un modèle de reconnaissance d'entités à partir de textes. Le projet nous a également donné une idée de la répartition des différentes compétences sur le marché de l'emploi au Maroc, ce qui nous sera utile pour approfondir nos connaissances en science des données ou acquérir les compétences les plus demandées dans notre carrière .