

Exploration des fonctions de SpaCy

```
In [ ]: # Importe spacy
import spacy

# Crée un objet nlp français vide
nlp = spacy.blank("fr")

# Charge le moyen pipeline français
nlp = spacy.load("fr_core_news_md")
```

Nous allons travailler avec le texte : "Depuis environs 3 ans j'ai une énorme douleurs au sternum qui m'irradie dans le dos mais qui se calme voir disparaît le plus souvent en buvant de l'eau. Depuis 1 semaines la douleurs revient plus souvent et me réveil même la nuit je suis donc obligé de boire pour me soulager. Et aujourd'hui en buvant de l'eau car je sentais que ça arrivé j'avais comme une sensation de boule qui se débloque entre l'oesophage et l'estomac. Cela pourrait être seulement un symptôme d'un reflux ? Pour info il y a 1 an j'ai consulté un gastro qui m'a fait une gastroscopie. Très mauvais souvenir sans anesthésie car je me débattais dans tout les sens. Le gastro à apparemment seulement vu une béance cardiaque et m'a dit de perdre du poids. J'ai 35ans 1m94 105kg"

```
In [ ]: # Crée en traitant une chaîne de caractères avec l'objet nlp
doc = nlp("Depuis environs 3 ans j'ai une énorme douleurs au sternum qui m'irradie
```

Nous allons, dans un premier temps, vérifier si on peut récupérer un groupe de mot avec la méthode json.

Groupes de mots en json

On crée donc un fichier json avec les groupes de mots que l'on veut récupérer. On récupère ce fichier.

```
In [ ]: import json

with open("vocabulaire.json", encoding="utf8") as f:
    VOCABULAIRE = json.loads(f.read())

print(VOCABULAIRE)
```

```
['douleurs au sternum', 'mal', 'mal au ventre', 'irradie dans le dos', 'gorge', 'd
ouleurs', 'boire pour me soulager', 'boule', 'oesophage', 'estomac']
```

On compare le texte avec le fichier json.

```
In [ ]: # Importe le PhraseMatcher et initialise-le
from spacy.matcher import PhraseMatcher

matcher = PhraseMatcher(nlp.vocab)

# Crée des motifs objets Doc et ajoute-les au matcher
# C'est la version rapide de : [nlp(country) for country in COUNTRIES]
patterns = list(nlp.pipe(VOCABULAIRE))
matcher.add("VOCABULAIRE", patterns)
```

```
# Appelle Le matcher sur Le document de test et affiche le résultat
matches = matcher(doc)
symptomes_patient = [doc[start:end] for match_id, start, end in matches]
print(symptomes_patient)
```

[douleurs, douleurs au sternum, irradie dans le dos, douleurs, boire pour me soulager, boule, oesophage, estomac]

On remarque qu'avec cette méthode on peut en effet récupérer les groupes de mots.

Associer symptômes et synonymes en json

Maintenant on va associer ces mots à leur synonyme pour ne récupérer que les synonymes.

On récupère la bibliothèque.

```
In [ ]: with open("bibliotheque.json", encoding="utf8") as f:
        BIBLIOTHEQUE = json.loads(f.read())

        print(BIBLIOTHEQUE)
```

```
{'douleur au sternum': ['douleurs au sternum', 'mal austernum'], 'douleur au ventre': ['mal au ventre', 'mal au bide', 'douleur au ventre'], 'douleur au dos': ['irradier dans le dos', 'douleur au dos']}
```

```
In [ ]: # On créer une bibliothèque pour associer les symptômes avec leurs synonymes
        dictionnaire_symptomes = {}

        for symptome_patient in symptomes_patient:
            for symptome, synonymes in BIBLIOTHEQUE.items():
                if(str(symptome_patient) in synonymes):
                    dictionnaire_symptomes[symptome_patient]=symptome

        # Affiche notre bibliothèque
        for symptome_recherche, symptome_correspondant in dictionnaire_symptomes.items():
            print(f"'{symptome_recherche}' correspond à '{symptome_correspondant}'")
```

'douleurs au sternum' correspond à 'douleur au sternum'

Cette méthode fonctionne et permet de récupérer les symptômes à partir des synonymes que l'on a récupéré dans le texte.

La fonctionnalité marche mais malheureusement on perd énormément d'information car il faut que les symptômes soit écrit d'une manière précise. Par exemple on peut écrire "boule dans l'oesophage" comme synonyme mais comme dans ce cas où le patient écrit "sensation de boule qui se débloque entre l'oesophage et l'estomac" le code ne le reconnaît pas.

Ce qu'on peut faire pour y remédier serait de faire un traitement sur le texte.

Lemmatisation et PoS Tagging

On ne récupère que les noms.

```
In [ ]: text_lemma = ''
        for token in doc:
            if(token.pos_ == 'NOUN'):
                text_lemma = text_lemma+' '+token.lemma_

        print(text_lemma)
```

environ an douleur sternum dos eau semaine douleur nuit eau sensation boule oesop
hage estomac symptôme reflux info an gastro gastroscopie souvenir anesthésie sens
gastro béance poids 1m94 kilogramme

On fait le traitement sur notre nouveau texte avec notre fichier json modifié.

```
In [ ]: # Import de notre nouveau fichier json
with open("bibliotheque_traitee.json", encoding="utf8") as f:
    bibliotheque_traitee = json.loads(f.read())

vocabulaire=[]
for k,v in bibliotheque_traitee.items():
    for valeur in v:
        vocabulaire.append(valeur)
print(vocabulaire)

['douleur sternum', 'mal sternum', 'mal ventre', 'mal bide', 'douleur ventre', 'ir
radier dos', 'douleur dos', 'boule oesophage', 'grosseur oesophage']
```

```
In [ ]: doc = nlp(text_lemma)

matcher = PhraseMatcher(nlp.vocab)

patterns = list(nlp.pipe(vocabulaire))
matcher.add("VOCABULAIRE", patterns)

# Appelle le matcher sur le document de test et affiche le résultat
matches = matcher(doc)
symptomes_patient = [doc[start:end] for match_id, start, end in matches]

# On crée une bibliothèque pour associer les symptômes avec leurs synonymes
dictionnaire_symptomes = {}

for symptome_patient in symptomes_patient:
    for symptome, synonymes in bibliotheque_traitee.items():
        if(str(symptome_patient) in synonymes):
            dictionnaire_symptomes[symptome_patient]=symptome

# Affiche notre bibliothèque
for k,v in dictionnaire_symptomes.items():
    print(v)
```

douleur sternum
boule oesophage

On remarque que cette fois-ci on ne récupère pas le symptôme "mal dos". Pour régler ce problème, on pourrait alors faire un traitement par étapes. Un premier traitement plus large et ensuite on réduit le texte pour récupérer des symptômes qui nous auraient échappés.

Recherches symptômes par couche

```
In [ ]: # Crée en traitant une chaîne de caractères avec l'objet nlp
doc = nlp("Depuis environs 3 ans j'ai une énorme douleurs au sternum qui m'irradie")

text_lemma = ''
dictionnaire_symptomes = []

# On récupère les noms et les verbes
for token in doc:
    if token.pos_ == 'NOUN' or token.pos_ == 'VERB':
        text_lemma = text_lemma+' '+token.lemma_
```

```

doc = nlp(text_lemma)

# On compare Le doc avec Le fichier de vocabulaire
matcher = PhraseMatcher(nlp.vocab)
patterns = list(nlp.pipe(vocabulaire))
matcher.add("VOCABULAIRE", patterns)

matches = matcher(doc)
symptomes_patient = [doc[start:end] for match_id, start, end in matches]

# On récupère Les symptomes a partir de Leur synonymes
for symptome_patient in symptomes_patient:
    for symptome, synonymes in bibliotheque_traitee.items():
        if(str(symptome_patient) in synonymes):
            dictionnaire_symptomes.append(symptome)

print("Première couche : ")
print(dictionnaire_symptomes)

# Deuxième couche
text_lemma = ''
# On recupère seulement Les noms
for token in doc:
    if token.pos_ == 'NOUN':
        text_lemma = text_lemma+' '+token.lemma_

doc = nlp(text_lemma)

# On compare Le doc avec Le fichier de vocabulaire
matcher = PhraseMatcher(nlp.vocab)
patterns = list(nlp.pipe(vocabulaire))
matcher.add("VOCABULAIRE", patterns)

matches = matcher(doc)
symptomes_patient = [doc[start:end] for match_id, start, end in matches]

# On récupère Les symptomes a partir de Leur synonymes
for symptome_patient in symptomes_patient:
    for symptome, synonymes in bibliotheque_traitee.items():
        if(str(symptome_patient) in synonymes):
            if symptome not in dictionnaire_symptomes:
                dictionnaire_symptomes.append(symptome)

print("Deuxième couche : ")
print(dictionnaire_symptomes)

```

```

Première couche :
['douleur sternum', 'douleur dos']
Deuxième couche :
['douleur sternum', 'douleur dos', 'boule oesophage']

```

Avec cette méthode on récupère bien le mal de dos ainsi que la boule a l'oesophage. C'est un code pour l'instant lourd mais qui peut être condensé. En additionnant les couches on pourra alors récupérer plusieurs symptomes.