

SPACY (avec des messages d'erreur)

Spacy est une bibliothèque qui permet d'effectuer les opérations d'analyse sur des textes dans plus de 50 langues3 : Tokenization (analyse lexicale) ; Reconnaissance d'entités nommées (La reconnaissance d'entités nommées est une sous-tâche de l'activité d'extraction d'information dans des corpus documentaires. Elle consiste à rechercher des objets textuels (c'est-à-dire un mot, ou un groupe de mots) catégorisables dans des classes telles que noms de personnes, noms d'organisations ou d'entreprises, noms de lieux, quantités, distances, valeurs, dates, etc.).

```
In [9]: ! pip install --user --upgrade pip
```

```
Requirement already satisfied: pip in c:\users\tomdu\anaconda3\lib\site-packages (21.1)
WARNING: Value for scheme.headers does not match. Please report this to <https://github.com/pypa/pip/issues/9617>
distutils: C:\Users\tomdu\AppData\Roaming\Python\Python38\Include\UNKNOWN
sysconfig: C:\Users\tomdu\AppData\Roaming\Python\Python38\Include
WARNING: Additional context:
user = True
home = None
root = None
prefix = None
```

```
In [11]: !python -m spacy download en_core_web_sm --user
```

```
Collecting en-core-web-sm==3.0.0
  Downloading https://github.com/explosion/spacy-models/releases/download/en_core_web_sm-3.0.0/en_core_web_sm-3.0.0-py3-none-any.whl (1
3.7 MB)
Requirement already satisfied: spacy<3.1.0,>=3.0.0 in c:\users\tomdu\anaconda3\lib\site-packages (from en-core-web-sm==3.0.0) (3.0.6)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in c:\users\tomdu\anaconda3\lib\site-packages (from spacy<3.1.0,>=3.0.0->en-cor
e-web-sm==3.0.0) (1.0.5)
Requirement already satisfied: thinc<8.1.0,>=8.0.3 in c:\users\tomdu\anaconda3\lib\site-packages (from spacy<3.1.0,>=3.0.0->en-core-web-
sm==3.0.0) (8.0.3)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in c:\users\tomdu\anaconda3\lib\site-packages (from spacy<3.1.0,>=3.0.0->en-core-web-
sm==3.0.0) (2.0.5)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in c:\users\tomdu\anaconda3\lib\site-packages (from spacy<3.1.0,>=3.0.0->en-core-
web-sm==3.0.0) (2.24.0)
Requirement already satisfied: blis<0.8.0,>=0.4.0 in c:\users\tomdu\anaconda3\lib\site-packages (from spacy<3.1.0,>=3.0.0->en-core-web-s
m==3.0.0) (0.7.4)
Requirement already satisfied: wasabi<1.1.0,>=0.8.1 in c:\users\tomdu\anaconda3\lib\site-packages (from spacy<3.1.0,>=3.0.0->en-core-web
-sm==3.0.0) (0.8.2)
Requirement already satisfied: jinja2 in c:\users\tomdu\anaconda3\lib\site-packages (from spacy<3.1.0,>=3.0.0->en-core-web-sm==3.0.0)
(2.11.2)
Requirement already satisfied: setuptools in c:\users\tomdu\anaconda3\lib\site-packages (from spacy<3.1.0,>=3.0.0->en-core-web-sm==3.0.
0) (50.3.1.post20201107)
Requirement already satisfied: packaging>=20.0 in c:\users\tomdu\anaconda3\lib\site-packages (from spacy<3.1.0,>=3.0.0->en-core-web-sm==
3.0.0) (20.4)
Requirement already satisfied: numpy>=1.15.0 in c:\users\tomdu\anaconda3\lib\site-packages (from spacy<3.1.0,>=3.0.0->en-core-web-sm==3.
0.0) (1.19.2)
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in c:\users\tomdu\anaconda3\lib\site-packages (from spacy<3.1.0,>=3.0.0->en-core-web-
sm==3.0.0) (4.50.2)
Requirement already satisfied: catalogue<2.1.0,>=2.0.3 in c:\users\tomdu\anaconda3\lib\site-packages (from spacy<3.1.0,>=3.0.0->en-core-
web-sm==3.0.0) (2.0.3)
Requirement already satisfied: pathy>=0.3.5 in c:\users\tomdu\anaconda3\lib\site-packages (from spacy<3.1.0,>=3.0.0->en-core-web-sm==3.
0.0) (0.5.2)
Requirement already satisfied: pydantic<1.8.0,>=1.7.1 in c:\users\tomdu\anaconda3\lib\site-packages (from spacy<3.1.0,>=3.0.0->en-core-w
eb-sm==3.0.0) (1.7.3)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in c:\users\tomdu\anaconda3\lib\site-packages (from spacy<3.1.0,>=3.0.0->en-core-we
b-sm==3.0.0) (3.0.5)
Requirement already satisfied: typer<0.4.0,>=0.3.0 in c:\users\tomdu\anaconda3\lib\site-packages (from spacy<3.1.0,>=3.0.0->en-core-web-
sm==3.0.0) (0.3.2)
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.4 in c:\users\tomdu\anaconda3\lib\site-packages (from spacy<3.1.0,>=3.0.0->en-co
re-web-sm==3.0.0) (3.0.5)
Requirement already satisfied: srsly<3.0.0,>=2.4.1 in c:\users\tomdu\anaconda3\lib\site-packages (from spacy<3.1.0,>=3.0.0->en-core-web-
sm==3.0.0) (2.4.1)
Requirement already satisfied: six in c:\users\tomdu\anaconda3\lib\site-packages (from packaging>=20.0->spacy<3.1.0,>=3.0.0->en-core-web
-sm==3.0.0) (1.15.0)
Requirement already satisfied: pyparsing>=2.0.2 in c:\users\tomdu\anaconda3\lib\site-packages (from packaging>=20.0->spacy<3.1.0,>=3.0.0
->en-core-web-sm==3.0.0) (2.4.7)
Requirement already satisfied: smart-open<4.0.0,>=2.2.0 in c:\users\tomdu\anaconda3\lib\site-packages (from pathy>=0.3.5->spacy<3.1.0,>=
3.0.0->en-core-web-sm==3.0.0) (3.0.0)
Requirement already satisfied: idna<3,>=2.5 in c:\users\tomdu\anaconda3\lib\site-packages (from requests<3.0.0,>=2.13.0->spacy<3.1.0,>=
3.0.0->en-core-web-sm==3.0.0) (2.10)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\tomdu\anaconda3\lib\site-packages (from requests<3.0.0,>=2.13.0->spacy<3.
1.0,>=3.0.0->en-core-web-sm==3.0.0) (2020.6.20)
Requirement already satisfied: chardet<4,>=3.0.2 in c:\users\tomdu\anaconda3\lib\site-packages (from requests<3.0.0,>=2.13.0->spacy<3.1.
0,>=3.0.0->en-core-web-sm==3.0.0) (3.0.4)
Requirement already satisfied: urllib3!=1.25.0,!1.25.1,<1.26,>=1.21.1 in c:\users\tomdu\anaconda3\lib\site-packages (from requests<3.0.
0,>=2.13.0->spacy<3.1.0,>=3.0.0->en-core-web-sm==3.0.0) (1.25.11)
Requirement already satisfied: click<7.2.0,>=7.1.1 in c:\users\tomdu\anaconda3\lib\site-packages (from typer<0.4.0,>=0.3.0->spacy<3.1.0,
>=3.0.0->en-core-web-sm==3.0.0) (7.1.2)
Requirement already satisfied: MarkupSafe>=0.23 in c:\users\tomdu\anaconda3\lib\site-packages (from jinja2->spacy<3.1.0,>=3.0.0->en-core
-web-sm==3.0.0) (1.1.1)
2021-04-29 14:56:45.684941: W tensorflow/stream_executor/platform/default/dso_loader.cc:60] Could not load dynamic library 'cudart64_11
0.dll'; dlderror: cudart64_110.dll not found
2021-04-29 14:56:45.684975: I tensorflow/stream_executor/cuda/cudart_stub.cc:29] Ignore above cudart dlerror if you do not have a GPU se
t up on your machine.
WARNING: Value for scheme.headers does not match. Please report this to <https://github.com/pypa/pip/issues/9617>
distutils: C:\Users\tomdu\AppData\Roaming\Python\Python38\Include\UNKNOWN
sysconfig: C:\Users\tomdu\AppData\Roaming\Python\Python38\Include
WARNING: Additional context:
user = True
home = None
root = None
prefix = None
Installing collected packages: en-core-web-sm
Successfully installed en-core-web-sm-3.0.0
[+] Download and installation successful
You can now load the package via spacy.load('en_core_web_sm')
```

```
In [12]: import spacy
```

```
In [13]: nlp = spacy.load("en_core_web_sm")
```

```
In [14]: doc = nlp("Lewis (first name unknown) was briefly a professional baseball player; he played one game for the Buffalo Bisons of the Major
```

```
In [4]: for token in doc:  
        print(token.text)
```

```
Lewis  
(  
first  
name  
unknown  
)  
was  
briefly  
a  
professional  
baseball  
player  
;  
he  
played  
one  
game  
for  
the  
Buffalo  
Bisons  
of  
the  
Major  
League  
Baseball  
(  
MLB  
)  
Players  
,  
League  
(  
PL  
)  
on  
July  
12  
,  
1890  
,  
Having  
approached  
the  
team  
for  
a  
tryout  
,  
Lewis  
was  
the  
starting  
pitcher  
for  
the  
visiting  
Bisons  
against  
the  
Brooklyn  
Ward  
's  
Wonders  
,  
During  
the  
three  
innings  
he  
pitched  
,  
Lewis  
allowed  
20  
runs  
,  
all  
earned  
,  
before  
being  
moved  
to  
left  
field  
,  
where  
he  
played  
the  
remainder  
of  
the  
game  
.
```

```
In [15]: for token in doc:
        print(token.text, token.lemma_, token.pos_, token.tag_, token.dep_,
              token.shape_, token.is_alpha, token.is_stop)
```

```
Lewis Lewis PROPN NNP nsubj Xxxxx True False
( ( PUNCT -LRB- punct ( False False
first first ADJ JJ amod xxxx True True
name name NOUN NN compound xxxx True True
unknown unknown ADJ JJ amod xxxx True False
) ) PUNCT -RRB- punct ) False False
was be AUX VBD ccomp xxx True True
briefly briefly ADV RB advmod xxxx True False
a a DET DT det x True True
professional professional ADJ JJ amod xxxx True False
baseball baseball NOUN NN compound xxxx True False
player player NOUN NN attr xxxx True False
; ; PUNCT : punct ; False False
he he PRON PRP nsubj xx True True
played play VERB VBD ROOT xxxx True False
one one NUM CD nummod xxx True True
game game NOUN NN dobj xxxx True False
for for ADP IN prep xxx True True
the the DET DT det xxx True True
Buffalo Buffalo PROPN NNP compound Xxxxx True False
Bisons Bisons PROPN NNPS pobj Xxxxx True False
of of ADP IN prep xx True True
the the DET DT det xxx True True
Major Major PROPN NNP compound Xxxxx True False
League League PROPN NNP compound Xxxxx True False
Baseball Baseball PROPN NNP pobj Xxxxx True False
( ( PUNCT -LRB- punct ( False False
MLB MLB PROPN NNP appos XXX True False
) ) PUNCT -RRB- punct ) False False
Players Players PROPN NNPS poss Xxxxx True False
' ' PART POS case ' False False
League League PROPN NNP pobj Xxxxx True False
( ( PUNCT -LRB- punct ( False False
PL PL PROPN NNP appos XX True False
) ) PUNCT -RRB- punct ) False False
on on ADP IN prep xx True True
July July PROPN NNP pobj Xxxx True False
12 12 NUM CD nummod dd False False
, , PUNCT , punct , False False
1890 1890 NUM CD nummod dddd False False
. . PUNCT . punct . False False
Having having AUX VBG aux Xxxxx True False
approached approach VERB VBN advcl xxxx True False
the the DET DT det xxx True True
team team NOUN NN dobj xxxx True False
for for ADP IN prep xxx True True
a a DET DT det x True True
tryout tryout NOUN NN pobj xxxx True False
, , PUNCT , punct , False False
Lewis Lewis PROPN NNP nsubj Xxxxx True False
was be AUX VBD ROOT xxx True True
the the DET DT det xxx True True
starting start VERB VBG amod xxxx True False
pitcher pitcher NOUN NN attr xxxx True False
for for ADP IN prep xxx True True
the the DET DT det xxx True True
visiting visit VERB VBG amod xxxx True False
Bisons Bisons PROPN NNPS pobj Xxxxx True False
against against ADP IN prep xxxx True True
the the DET DT det xxx True True
Brooklyn Brooklyn PROPN NNP compound Xxxxx True False
Ward Ward PROPN NNP poss Xxxx True False
's 's PART POS case 'x False True
Wonders wonder NOUN NNS pobj Xxxxx True False
. . PUNCT . punct . False False
During during ADP IN prep Xxxxx True True
the the DET DT det xxx True True
three three NUM CD nummod xxxx True True
innings inning NOUN NNS pobj xxxx True False
he he PRON PRP nsubj xx True True
pitched pitch VERB VBD relcl xxxx True False
, , PUNCT , punct , False False
Lewis Lewis PROPN NNP nsubj Xxxxx True False
allowed allow VERB VBD ROOT xxxx True False
20 20 NUM CD nummod dd False False
runs run NOUN NNS dobj xxxx True False
, , PUNCT , punct , False False
all all DET DT cc xxx True True
earned earn VERB VBN acl xxxx True False
, , PUNCT , punct , False False
before before ADP IN prep xxxx True True
being be AUX VBG auxpass xxxx True True
moved move VERB VBN pcomp xxxx True False
to to ADP IN prep xx True True
left left ADJ JJ amod xxxx True False
field field NOUN NN pobj xxxx True False
, , PUNCT , punct , False False
where where ADV WRB advmod xxxx True True
he he PRON PRP nsubj xx True True
played play VERB VBD relcl xxxx True False
the the DET DT det xxx True True
remainder remainder NOUN NN dobj xxxx True False
of of ADP IN prep xx True True
the the DET DT det xxx True True
game game NOUN NN pobj xxxx True False
. . PUNCT . punct . False False
```

```
In [16]: for ent in doc.ents:
        print(ent.text, ent.start_char, ent.end_char, ent.label_)
```

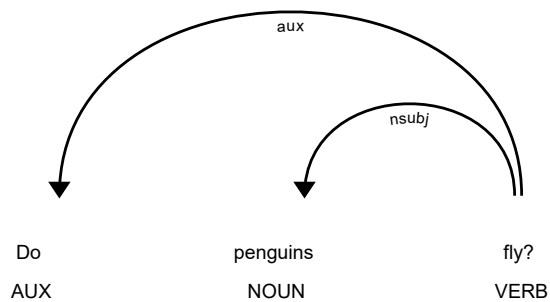
```
first 7 12 ORDINAL
one 81 84 CARDINAL
```

the Buffalo Bisons 94 112 ORG
the Major League Baseball 116 141 ORG
MLB 143 146 ORG
Players' League 148 163 ORG
July 12, 1890 172 185 DATE
Lewis 228 233 PERSON
Bisons 276 282 PERSON
the Brooklyn Ward's 291 310 ORG
three 331 336 CARDINAL
Lewis 357 362 PERSON
20 371 373 CARDINAL

In [17]: `from spacy import displacy`

```
nlp = spacy.load("en_core_web_sm")
doc = nlp("Do penguins fly?")
displacy.serve(doc, style="dep")
```

C:\Users\tomdu\anaconda3\lib\site-packages\spacy\displacy_init_.py:97: UserWarning: [W011] It looks like you're calling displacy.serve from within a Jupyter notebook or a similar environment. This likely means you're already running a local web server, so there's no need to make displacy start another one. Instead, you should be able to replace displacy.serve with displacy.render to show the visualization.
warnings.warn(Warnings.W011)

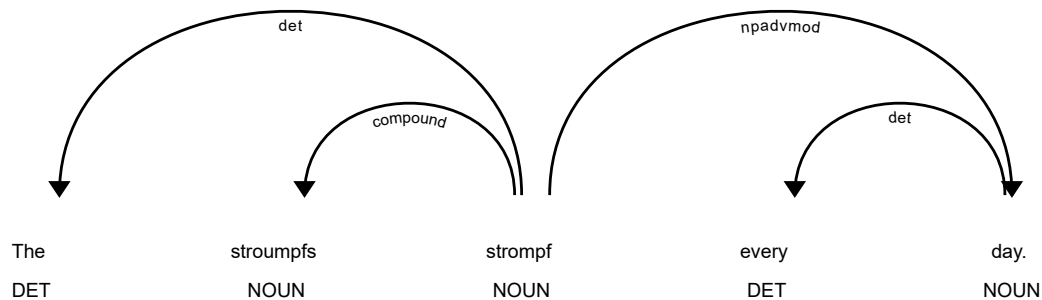


Using the 'dep' visualizer
Serving on http://0.0.0.0:5000 ...

Shutting down server on port 5000.

In [20]: `import spacy`
`from spacy import displacy`

```
nlp = spacy.load("en_core_web_sm")
doc = nlp("The stroumpfs strompf every day.")
displacy.serve(doc, style="dep")
```



Using the 'dep' visualizer
Serving on http://0.0.0.0:5000 ...

Shutting down server on port 5000.