



Missão Prática | Nível 4 | Mundo 5

João Gabriel Cesconetto - 202208324053

Dando inteligência ao software

Objetivos da prática:

- 1) Configurar o ambiente Google Colab;**
- 2) Descrever tarefas diversas relacionadas ao Processamento de Linguagem Natural;**
- 3) Descrever o processo de identificação de entidades a partir de textos;**
- 4) Descrever o processo de extração de frases-chave a partir de textos;**
- 5) Descrever o processo de identificação de linguagem predominante a partir de textos.**

1) Descrever tarefas diversas relacionadas ao Processamento de Linguagem Natural

Passo 1: Instalando a biblioteca e recarregando o ambiente.

```
!pip install spacy

import pkg_resources,imp

imp.reload(pkg_resources)
```

Show hidden output

Passo 2: Baixando o modelo a ser utilizado e recarregando o ambiente.

```
[2] import spacy.cli

spacy.cli.download("en_core_web_trf")

import pkg_resources,imp

imp.reload(pkg_resources)
```

✓ Download and installation successful
You can now load the package via `spacy.load('en_core_web_trf')`
⚠ 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.
<module 'pkg_resources' from '/usr/local/lib/python3.10/dist-packages/pkg_resources/__init__.py'>

Passo 3: Importando a biblioteca e definindo o modelo a ser utilizado.

```
✓ 2s [3] import spacy

nlp = spacy.load('en_core_web_sm')
```

Passo 4: Definindo o texto a ser utilizado nas tarefas de NLP

```
✓ 0s text = """National Park Week starts on Saturday, and it also starts off with a bang-
for-your-buck.

That's because every US National Park Service site will have free entry on
Saturday. NPS manages almost 430 sites, and the majority of them already offer
free entry every day.

But this is your chance to get into the coveted, big-name national parks and other
sites without paying a fee.

That includes legendary parks such as Yosemite, which normally has an entry fee
of $20 per person or $35 per vehicle.

(Note: You'll still need a reservation to drive into Yosemite on weekends and
holidays from April 13 to June 30.)

"""

doc = nlp(text)
```

Passo 5: Realizando a tokenização do texto

✓
Os



```
for token in doc:  
    print(token)
```



```
National  
Park  
Week  
starts  
on  
Saturday  
,  
and  
it  
also  
starts  
off  
with  
a  
bang-  
  
for  
-  
your  
-  
buck  
.
```

Passo 6: Obtendo as tags de classe gramatical para os tokens individuais.

```
for token in doc:|
    print(token, token.pos_, token.tag_)
```

```
legendary ADJ JJ
parks NOUN NNS
such ADJ JJ
as ADP IN
Yosemite PROPN NNP
, PUNCT ,
which PRON WDT
normally ADV RB
has VERB VBZ
an DET DT
entry NOUN NN
fee NOUN NN
```

```
SPACE _SP
of ADP IN
$ SYM $
20 NUM CD
per ADP IN
person NOUN NN
or CCONJ CC
$ SYM $
35 NUM CD
per ADP IN
vehicle NOUN NN
. PUNCT .
```

```
SPACE _SP
( PUNCT -LRB-
Note NOUN NN
: PUNCT :
You PRON PRP
'll AUX MD
still ADV RB
need VERB VB
a DET DT
```

Passo 7: Imprimindo os tokens e sua análise morfológica

0s



```
for token in doc:
    print(token, token.morph)
```



```
National Number=Sing
Park Number=Sing
Week Number=Sing
starts Number=Sing|Person=3|Tense=Pres|VerbForm=Fin
on
Saturday Number=Sing
, PunctType=Comm
and ConjType=Cmp
it Case=Nom|Gender=Neut|Number=Sing|Person=3|PronType=Prs
also
starts Number=Sing|Person=3|Tense=Pres|VerbForm=Fin
off
with
a Definite=Ind|PronType=Art
bang- Degree=Pos

for
- PunctType=Dash
your Person=2|Poss=Yes|PronType=Prs
- PunctType=Dash
buck Number=Sing
. PunctType=Peri

That Number=Sing|PronType=Dem
's Number=Sing|Person=3|Tense=Pres|VerbForm=Fin
because
every
```

0s



```
from spacy import displacy
displacy.render(doc, style='dep', options={'compact': True})
```

