



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

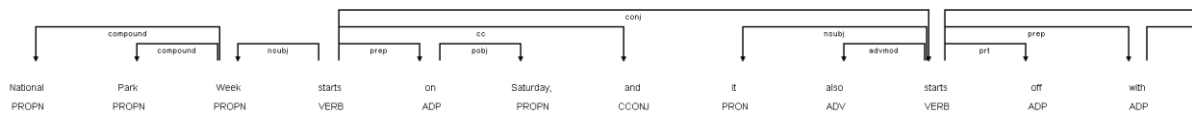
for
- PunctType=Dash
your Person=2|Poss=Yes|PronType=Prs
- PunctType=Dash
buck Number=Sing
. PunctType=Peri

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})
```



2) Descrever o processo de identificação de entidades a partir de textos

Importando as bibliotecas python e configurando o modelo a ser utilizado pelo spacy

```
[1] import pandas as pd

import spacy

import requests

from bs4 import BeautifulSoup

nlp = spacy.load("en_core_web_sm")

pd.set_option("display.max_rows", 200)
```

```
doc = nlp("Apple is looking at buying U.K. startup for $1 billion")

for ent in doc.ents:

    print(ent.text, ent.start_char, ent.end_char, ent.label_)
```

```
Apple 0 5 ORG
U.K. 27 31 GPE
$1 billion 44 54 MONEY
```

Importando a lib displacy e exibindo o resultado da análise NER de forma visual

```
from spacy import displacy

displacy.render(doc, style="ent")
```

```
Apple ORG is looking at buying U.K. GPE startup for $1 billion MONEY
```

Analisando um bloco de texto



```
sample_txt = """  
  
Hello Zhang Wei. Your AnyCompany Financial Services, LLC credit card account  
1111-0000-1111-0000 has a minimum payment of $24.53 that is due by July 31st.  
  
Based on your autopay settings, we will withdraw your payment on the due date  
from  
  
your bank account XXXXXX1111 with the routing number XXXXX0000.  
  
  
Your latest statement was mailed to 100 Main Street, Anytown, WA 98121.  
  
After your payment is received, you will receive a confirmation text message  
at 206-555-0100.  
  
If you have questions about your bill, AnyCompany Customer Service is available  
by  
  
phone at 206-555-0199 or email at support@anycompany.com.  
  
"""  
  
newdoc = nlp(sample_txt)  
  
displacy.render(newdoc, style="ent")
```

Hello **Zhang Wei** **PERSON** . Your **AnyCompany Financial Services** **ORG** , **LLC** **ORG** credit card account

1111-0000-1111-0000 **DATE** has a minimum payment of \$ **24.53** **MONEY** that is due by **July 31st** **DATE** .

Based on your autopay settings, we will withdraw your payment on the due date
from

your bank account **XXXXXX1111** **ORG** with the routing number **XXXXX0000**.

Your latest statement was mailed to **100** **CARDINAL** **Main Street** **FAC** , **Anytown** **GPE** , WA98121.

After your payment is received, you will receive a confirmation text message

at **206-555-0100** **CARDINAL** .

If you have questions about your bill, **AnyCompany Customer Service** **ORG** is available
by

phone at **206-555-0199** **CARDINAL** or email at support@anycompany.com.

Visualizando o resultado da análise de forma tabular

✓
Ds



```
entities = [(ent.text, ent.label_, ent.lemma_) for ent in newdoc.ents]  
  
df = pd.DataFrame(entities, columns=['text', 'type', 'lemma'])  
  
df.head()
```



	text	type	lemma
0	Zhang Wei	PERSON	Zhang Wei
1	AnyCompany Financial Services	ORG	AnyCompany Financial Services
2	LLC	ORG	LLC
3	1111-0000-1111-0000	DATE	1111-0000-1111-0000
4	24.53	MONEY	24.53





```
import spacy

import requests

from bs4 import BeautifulSoup

from spacy import displacy

nlp = spacy.load("en_core_web_sm")

txt_br="""

ONU aprova missão internacional para restabelecer segurança no Haiti

O Conselho de Segurança das Nações Unidas (ONU) aprovou na noite de segunda-
feira (2) a criação e o envio de uma força internacional para a manutenção de paz
no Haiti, devido aos conflitos entre as gangues que dominam o país.

O Haiti é o país mais pobre do hemisfério ocidental. A história do país é marcada
por golpes, deposições e massacres que geraram grande instabilidade política,
turbulência econômica e crise social. Em 30 de abril de 2004 o Conselho de
Segurança da Organização das Nações Unidas aprovou, por unanimidade, a
criação da Missão de Estabilização do Haiti, a MINUSTAH.

A missão, planejada para ter uma duração inicial de seis meses, vem sendo
prorrogada. O objetivo é combater a insegurança no país após a crise que forçou a
saída do ex-presidente Jean Bertrand Aristide, em fevereiro de 2004.

Coube ao Brasil a chefia da missão, instalada no mês de junho de 2004, bem como
constituir o maior contingente nacional de "boinas azuis" : 1.470 militares, de um
total de 6.700, militares de 21 países. Apesar das dificuldades estruturais do Haiti,
as atividades da MINUSTAH trouxeram avanços, como a realização das eleições
ocorridas em fevereiro deste ano.

""

docbr = nlp(txt_br)

displacy.render(docbr, style="ent")
```

ONU aprova missão internacional para restabelecer segurança no Haiti

O Conselho de Segurança PERSON das Nações Unidas (ONU) aprovou na noite de segunda-feira PERSON (2 CARDINAL) a criação e o envio de uma força internacional para a manutenção de paz no Haiti GPE , devido aos conflitos PERSON entre as gangues que dominam PERSON o país.

O Haiti GPE é o país mais pobre do hemisfério ocidental. A história GPE do país é marcada por golpes, deposições e massacres que geraram grande instabilidade política, turbulência econômica e crise social. Em 30 de abril de QUANTITY 2004 DATE o Conselho de GPE Segurança da Organização PERSON das Nações Unidas ORG aprovou, por unanimidade, a criação da Missão de Estabilização do Haiti GPE , a MINUSTAH ORG .

A missão, planejada para PERSON ter uma duração inicial de seis meses, vem sendo prorrogada. O objetivo é combater a insegurança no país após a crise que forçou a saída do ex-presidente Jean Bertrand Aristide PRODUCT , em fevereiro de 2004.

Coube ORG ao Brasil a chefia da missão, instalada GPE no mês de ORG junho de 2004 DATE , bem como constituir o maior contingente nacional de "boinas azuis" : 1.470 militares QUANTITY , de um total de 6.700 CARDINAL , militares CARDINAL de 21 CARDINAL países. Apesar das dificuldades estruturais PERSON do Haiti GPE , as atividades da MINUSTAH ORG trouxeram avanços ORG , como a realização das eleições ocorridas em fevereiro GPE deste ano.

Missão Prática

Passo 1: Instalando as bibliotecas e recarregando o ambiente

```
[1] !pip install -U pip setuptools wheel
!pip install -U spacy
!python -m spacy download en_core_web_sm
!pip install spacytextblob --upgrade

import pkg_resources,imp

imp.reload(pkg_resources)
```



Show hidden output

Passo 2: Importando as bibliotecas para análise de sentimento

```
[14] import spacy
from spacytextblob.spacytextblob import SpacyTextBlob
```

Passo 3: Definindo o modelo e a pipeline a serem utilizadas na análise

✓
Ds

```
[15] nlp = spacy.load('en_core_web_sm')
nlp.add_pipe('spacytextblob')
```



<spacytextblob.spacytextblob.SpacyTextBlob at 0x7edd701da170>

Passo 4: Definindo o texto inicial a ser analisado para verificação/validação da biblioteca

✓
Ds

```
[16] user_input = 'This is a wonderful campsite. I loved the serenity and the birds chirping in the morning.'

doc = nlp(user_input)
```

Passo 5: Exibindo o resultado da primeira análise (um range entre -1 [avaliação negativa] e 1 [avaliação positiva])

✓
Ds

```
input_polarity = doc._.blob.polarity

sentiment = {
    'score': input_polarity
}

print(sentiment)
```



{'score': 0.85}

Passo 6: Definindo a lista de tweets a serem analisadas

```
[18] tweets = ["Bayer Leverkusen goalkeeper Bernd Leno will not be going to Napoli. His agent Uli Ferber to Bild: I can confirm that there were negotiations with Napoli, which we have broken off. Napoli is not an option. Atleti",
"@ChelseaFC Don't make him regret it and start him over Hoofiz",
"@LiverpoolFF @AnfieldEdition He's a liar, made up. I've unfollowed him as loads of others have. Pure blagger. #LFC",
"@theesk @Everton Didn't realise Kenwright is due to leave at the end of the month. In all seriousness could you see him being interested in us?",
"@hasanshahbaz19 @LFC My knowledge has decreased somewhat in the past few seasons",
"Report: Linked with Everton and Wolves, Italians set to sign 24-year-old winger",
"Are seeing tweets that there's been a fall out @Everton between the money man... I'm presuming it's just a quiet news day or some kopite with nothing better to do! @ALANWYERSMEDIA",
"@LFC @OfficialALDO @intChampionsCup @ManUtd Expect loads of excuses after tonight's game",
"@MartinOmond17 @asrymad @aren_D @MathewLewis1997 @amhaithu @DJMarzio @AllisonBecker @LFC @skySportsNews @skySport @OfficialASRoma I'm just fine I have your fanbase angry over stating facts should ask them hun. Xo",
"Was a weekend of football results! @ManUtd @Glentoran @RangersFC & Hearts ????",
"@ChelseaFC For the first time in a long while, my heart was relaxed while watching Chelsea. Really enjoyed it today. Come on, CHELSEA!!!",
"@ChelseaFC @CesarAzpi What a fantastic signing worth every single penny ??",
"Pogba scores, Pogba assists. But tomorrow papers won't be telling you this, instead they will tell you how he'll end up at Juve because he's unhappy, frustrated, have grudges with Mourinho and so on and so forth #mufc",
"@westhumbd we need to keep @Utd and get @IrvingLozano20 to compliment",
"@kevdeo @Everton Shouldn't be happening! Needs to stay away with his venomous attitude until he is sold",
"@thefootball @aguerocergioakun @ManCity What a genius. Pep taking winning mentality with him, conquering league after league. Baller",
"@H020799 Can we get a RT for our #Lfc Mo Salah Liverpool Enamel Pin Badge"]
```

Passo 7: Analisando os tweets

```
✓ [19] for item in tweets:
Ds
    doc = nlp(item)
    input_polarity = doc._.blob.polarity
    sentiment = {
        'tweet': item,
        'score': input_polarity
    }
    print(sentiment)
```

Output:

{'tweet': 'Bayer Leverkusen goalkeeper Bernd Leno will not be going to Napoli. His agent Uli Ferber to Bild: I can confirm that there were negotiations with Napoli, which we have broken off. Napoli is not an option. Atletico Madrid and Arsenal are the other strong rumours. #B04

#AFC', 'score': -0.030555555555555575}

{'tweet': 'Gary Speed v Blackburn at St James in 2001/02 anyone? #NUFC #BEL #JAP #WorldCup', 'score': 0.0}

{'tweet': '@ChelseaFC Don't make him regret it and start him over Hoofiz', 'score': 0.0}

{'tweet': '@LiverpoolFF @AnfieldEdition He's a liar, made up. I've unfollowed him as loads of others have. Pure blagger. #LFC', 'score': 0.21428571428571427}

{'tweet': '@theesk @Everton Didn't realise Kenwright is due to leave at the end of the month. In all seriousness could you see him being interested in us?', 'score': 0.0625}

{'tweet': '@hasanshahbaz19 @LFC My knowledge has decreased somewhat in the past few seasons', 'score': -0.28333333333333334}

{'tweet': 'Report: Linked with #Everton and #Wolves, Italians set to sign £4.5m-rated winger',
'score': 0.0}

{'tweet': 'Am seeing tweets that there's been a fall out @Everton between the money men... I'm presuming it's just a quiet news day or some kopite with nothing better to do!
@ALANMYERSMEDIA', 'score': 0.3125}

{'tweet': '@LFC @officialAL20 @IntChampionsCup @ManUtd Expect loads of excuses after tonight's game', 'score': -0.4}

{'tweet': '@MartinDiamond17 @azryahmad @Baren_D @Mathewlewis1997 @iamheinthu @DiMarzio @Alissonbecker @LFC @SkySportsNews @SkySport @OfficialASRoma I'm just fine I have your fanbase angry over stating facts should ask them hun. Xo', 'score': -0.04166666666666666}

{'tweet': 'What a weekend of football results! @ManUtd @Glentoran @RangersFC & Hearts ???', 'score': 0.0}

{'tweet': '@ChelseaFC For the first time in a long while, my heart was relaxed while watching Chelsea. Really enjoyed it today. Come on, CHELSEA!!!', 'score': 0.39218749999999997}

{'tweet': '@ChelseaFC @CesarAzpi What a fantastic signing worth every single penny ??', 'score': 0.20952380952380953}

{'tweet': '"Pogba scores, Pogba assists. But tomorrow papers won't be telling you this, instead they will tell you how he'll end up at Juve because he's unhappy, frustrated, have grudges with Mourinho and so on and so forth #mufc", 'score': -0.6499999999999999}

{'tweet': '@WestHamUtd we need to keep @CH14_ and get @HirvingLozano70 to compliment@kevdev9 @Everton Shouldn't be happening! Needs to stay away with his venomous attitude until he is sold!', 'score': 0.0}

{'tweet': '@brfootball @aguerosergiokun @ManCity What a genius. Pep taking winning mentality with him, conquering league after league. Baller', 'score': 0.5}

{'tweet': '@HMZ0709 Can we get a RT for our #lfc Mo Salah Liverpool Enamel Pin Badge', 'score': 0.0}