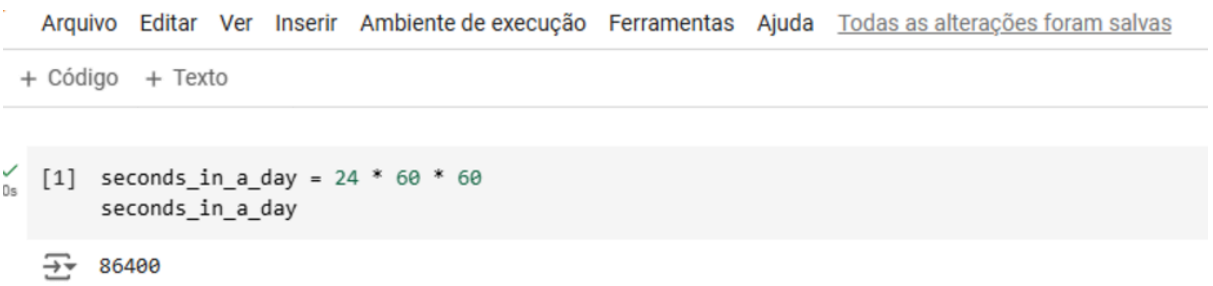
Documentação Microatividades

Estácio
Italo Gabriel de Almeida Costa

Microatividade 1



Microatividade 2

!pip install spacy

import pkg_resources, imp

imp.reload(pkg_resources)

import spacy.cli

spacy.cli.download("en_core_web_trf")

import pkg_resources, imp

imp.reload(pkg_resources)

import spacy

nlp = spacy.load('en_core_web_sm')

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)


 for token in doc:

print(token)


 for token in doc:

print(token, token.pos_, token.tag_)


 for token in doc:

print(token, token.morph)


 from spacy import displacy

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

Microatividade 3

```
 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_)

from spacy import displacy
displacy.render(doc, style="ent")

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.
...
"""
newdoc = nlp(sample_txt)
displacy.render(newdoc, style="ent")

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

!python -m spacy download pt

txt_br = """
ONU aprova missão internacional para restabelecer segurança no Haiti
...
"""
docbr = nlp(txt_br)
displacy.render(docbr, style="ent")
```

Microatividade 4

```
!pip install git+https://github.com/boudinfl/pke.git
!python -m spacy download en_core_web_sm
import pkg_resources, imp
imp.reload(pkg_resources)

import pke
# initialize a TopicRank keyphrase extraction model
extractor = pke.unsupervised.TopicRank()

sample = """Tesla has been ordered to recall nearly 4,000 of its Cybertrucks due to
an accelerator pedal that can stick in place when pressed down.
The cause, according to the regulator: soap.
“An unapproved change introduced lubricant (soap) to aid in the component
assembly of the pad onto the accelerator pedal.
Residual lubricant reduced the retention of the pad to the pedal,” the NHTSA
wrote in the recall document.
Tesla has yet to detail how many of the futuristic looking Cybertrucks it has
produced. But it has said that it would be slow ramping up production of the vehicle,
which had its first deliveries in late November.
The NHTSA said the recall affects “all Model Year (‘MY’) 2024 Cybertruck vehicles
manufactured from November 13, 2023, to April 4, 2024.”""".replace("\n", " ")

extractor.load_document(input=sample, language='en')
```

```
for i, sentence in enumerate(extractor.sentences):

print("sentence {}:".format(i))

print(" - words: {} ...".format(' '.join(sentence.words[:5])))

print(" - stems: {} ...".format(' '.join(sentence.stems[:5])))

print(" - PoS: {} ...".format(' '.join(sentence.pos[:5])))


extractor.candidate_selection()

for i, candidate in enumerate(extractor.candidates):

print("candidate {}: {} (stemmed form)".format(i, candidate))

print(" - surface forms:", [" ".join(u) for u in extractor.candidates[candidate].surface_forms])

print(" - offsets:", extractor.candidates[candidate].offsets)

print(" - sentence_ids:", extractor.candidates[candidate].sentence_ids)

print(" - pos_patterns:", extractor.candidates[candidate].pos_patterns)


extractor.candidate_weighting()

for i, topic in enumerate(extractor.topics):

print("topic {}: {} ".format(i, ';'.join(topic)))
```

Microatividade 5

```
!pip install langdetect


text = [

"Wear masks, keep distance, wash hands, be safe in these difficult days.",

"Viseljen maszkot, tartson távolságot, mosson kezet, legyen biztonságban ezekben a nehéz napokban",

"Deaths are increasing, be vigilant.",

"Носите маски, соблюдайте дистанцию, мойте руки, будьте осторожны в эти тяжелые дни.",
```

"Covid-19: Indians flock to vaccination centers as vaccines are now available for 60+ in India since the 1st of March",

"Indossa maschere, mantieni le distanze, lavati le mani, sii al sicuro in questi giorni difficili.",

"Portez des masques, gardez vos distances, lavez-vous les mains, soyez en sécurité en ces jours difficiles.",

"Brug masker, hold afstand, vask hænder, vær sikker i disse vanskelige dage.",

"We are facing a global education crisis. No effort should be spared to safely bring every child back into the classroom.",

"Bruk masker, hold avstand, vask hendene, vær trygg i disse vanskelige dagene.",

"Portu maskojn, tenu distancon, lavu manojn, estu sekuraj en ĉi tiuj malfacilaj tagoj.",

"Tragen Sie Masken, halten Sie Abstand, waschen Sie Ihre Hände, seien Sie in diesen schwierigen Tagen sicher.",

"Носіть маски, тримайтеся на відстані, мийте руки, будьте в безпеці в ці важкі дні.",

"Lock down, working from home are the keys words for these days.",

"Lives have changed drastically across the planet and this period will forever be remembered as the beginning of something we have yet to witness.",

"Este é um exemplo de texto escrito em português."

]

```
import pkg_resources, imp

imp.reload(pkg_resources)
```

```
from langdetect import detect
```

```
for x in text:

print('Frase: ', x)

print('Idioma: ', detect(x), '\n\n')
```