

 Estácio	<p align="center"> UNIVERSIDADE ESTÁCIO DE SÁ POLO DALPLAZA CENTER – SÃO LUÍS/MA DESENVOLVIMENTO FULL STACK - 22.3 Relatório da Missão Prática Nível 3 Mundo 5 </p>
Aluno:	Lucas Silva Costa
Professor:	Jhonatan Alves
Repositório:	https://github.com/LutchasDev/M5-nivel4-main

Título da Prática: Dando inteligência ao software

Objetivos da Prática:

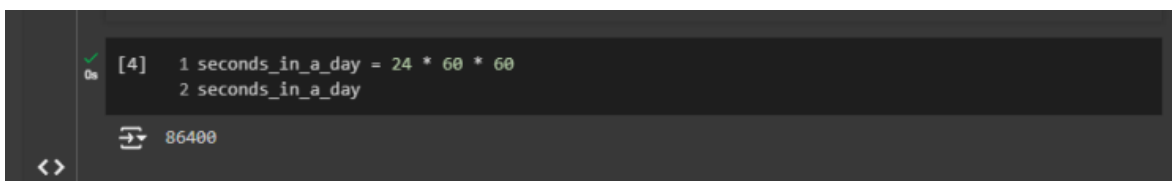
- Configurar o ambiente Google Colab;
- Descrever tarefas diversas relacionadas ao Processamento de Linguagem Natural;
- Descrever o processo de identificação de entidades a partir de textos;
- Descrever o processo de extração de frases-chave a partir de textos;
- Descrever o processo de identificação de linguagem predominante a partir de textos;

Microatividade 1

Configurar o ambiente Google Colab

Procedimentos:

1. Acesse a página do Google Colab: <https://colab.google>
2. No menu superior direito, selecione a opção "Open Colab";
3. Na nova aba aberta no navegador, faça login (botão no canto superior direito) com uma conta Google;
4. Após o login, feche a janela modal exibida, para visualizar o notebook previamente existente "Olá, este é o Colaboratory";
5. Navegue pela página e repare que existem blocos de texto seguidos de blocos de código. Esses últimos são caracterizados por possuírem uma cor de fundo diferente, em cinza, e por terem uma setinha que pode ser clicada, o que fará com o código contido no bloco seja executado e seu resultado apresentado a seguir. Veja o print abaixo (onde a setinha que permite a execução foi destacada com um quadrado vermelho):



The screenshot shows a code cell in Google Colab with a dark background. On the left, there is a green checkmark and the text '0s'. The code inside the cell is: `[4] 1 seconds_in_a_day = 24 * 60 * 60` followed by `2 seconds_in_a_day`. Below the code, there is a red square icon with a white right-pointing arrow, which is the 'Run' button. To the right of this icon, the output '86400' is displayed in red text. At the bottom left of the cell, there is a '<>' icon.

6. Leia os blocos de texto de cada seção e, a seguir, clique para executar os exemplos de código. Isso lhe permitirá um bom entendimento do funcionamento do Google Colab.

Microatividade 2

Descrever tarefas diversas relacionadas ao Processamento de Linguagem Natural

Procedimientos:

1. Estando logado no Google Colab, clique no menu “Arquivo” e selecione a opção “Novo notebook”;
2. Na nova aba aberta no navegador, dê um nome ao seu notebook, clicando e alterando o nome automaticamente gerado – Untitled0.ipynb – para nlp.ipynb;
3. Na janela de código, clique na opção “+Texto” (destacada no print abaixo) para inserir um bloco de texto;

Resultado:

Passo 1:

```
colab.research.google.com/drive/1zRPeUMPW...SCH1t5sD8SdpXZZauthuser=3#scrollTo=xwXt1mWda
```

nlipyb

Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda

Código + Texto

RAM Disco Gemini

MICROATIVIDADES

Tarefas diversas de Processamento de Linguagem Natural

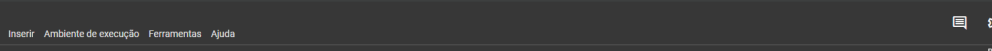
Passo 1: Instalando a biblioteca e recarregando o ambiente.*# Nova seção

```
!pip install spacy
import pkg_resources,imp
imp.reload(pkg_resources)
```

```
Requirement already satisfied: spacy in /usr/local/lib/python3.10/dist-packages (3.7.5)
Requirement already satisfied: spacy-loggers<3.1.0,>=3.1.1 in /usr/local/lib/python3.10/dist-packages (from spacy) (3.0.12)
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (1.0.5)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (1.0.10)
Requirement already satisfied: cymem<1.0,>=2.0.2 in /usr/local/lib/python3.10/dist-packages (from spacy) (2.0.8)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python3.10/dist-packages (from spacy) (3.0.9)
Requirement already satisfied: thinc<8.3.0,>=8.2.2 in /usr/local/lib/python3.10/dist-packages (from spacy) (8.2.5)
Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in /usr/local/lib/python3.10/dist-packages (from spacy) (1.1.3)
Requirement already satisfied: srsly<3.0,>=2.4.3 in /usr/local/lib/python3.10/dist-packages (from spacy) (2.4.8)
Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in /usr/local/lib/python3.10/dist-packages (from spacy) (2.0.10)
Requirement already satisfied: weasel<0.5.0,>=0.1.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (0.4.1)
Requirement already satisfied: typer<1.0.0,>=0.3.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (0.12.5)
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (4.66.0)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (2.32.3)
Requirement already satisfied: pydantic<1.8.1,>=1.8.1,<3.0.0,>=1.7.4 in /usr/local/lib/python3.10/dist-packages (from spacy) (2.9.2)
Requirement already satisfied: Jinja2 in /usr/local/lib/python3.10/dist-packages (from spacy) (3.1.4)
Requirement already satisfied: setuptools in /usr/local/lib/python3.10/dist-packages (from spacy) (75.1.0)
Requirement already satisfied: packaging<20.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (24.1)
Requirement already satisfied: langcodes<4.0.0,>=3.2.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (3.4.1)
Requirement already satisfied: numpy<=1.19.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (1.26.4)
Requirement already satisfied: language-data<=1.2 in /usr/local/lib/python3.10/dist-packages (from langcodes<4.0.0,>=3.2.0->spacy) (1.2.0)
Requirement already satisfied: annotated-types<0.6.0 in /usr/local/lib/python3.10/dist-packages (from pydantic<1.8.1,>=1.8.1,<3.0.0,>=1.7.4->spacy) (0.7.0)
Requirement already satisfied: pydantic-core<=2.23.4 in /usr/local/lib/python3.10/dist-packages (from pydantic<1.8.1,>=1.8.1,<3.0.0,>=1.7.4->spacy) (2.23.4)
Requirement already satisfied: typing-extensions<=4.6.1 in /usr/local/lib/python3.10/dist-packages (from pydantic<1.8.1,>=1.8.1,<3.0.0,>=1.7.4->spacy) (4.12.2)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.13.0->spacy) (3.4.0)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.13.0->spacy) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.13.0->spacy) (2.2.3)
Requirement already satisfied: certifi<2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.13.0->spacy) (2024.8.30)
Requirement already satisfied: h11<0.8.0,>=0.7.2 in /usr/local/lib/python3.10/dist-packages (from thinc<8.3.0,>=8.2.2->spacy) (0.7.11)
Requirement already satisfied: confection<1.0.0,>=0.0.1 in /usr/local/lib/python3.10/dist-packages (from thinc<8.3.0,>=8.2.2->spacy) (0.1.5)
```

98 conclusão: 11/32

Passo 2:



```
import spacy.cli
spacy.cli.download("en_core_web_trf")

import pkg_resources, imp
imp.reload(pkg_resources)
```

✓ Download and installation successful

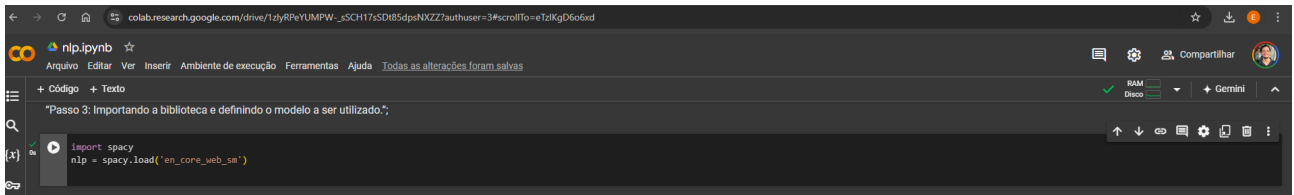
You can now load the package via `spacy.load('en_core_web_trf')`

⚠ 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' button in the 'Runtime' option.

```
!pip install spacy
code pkg_resources from '/usr/local/lib/python3.10/dist-packages/pkg_resources/' init .py'
```

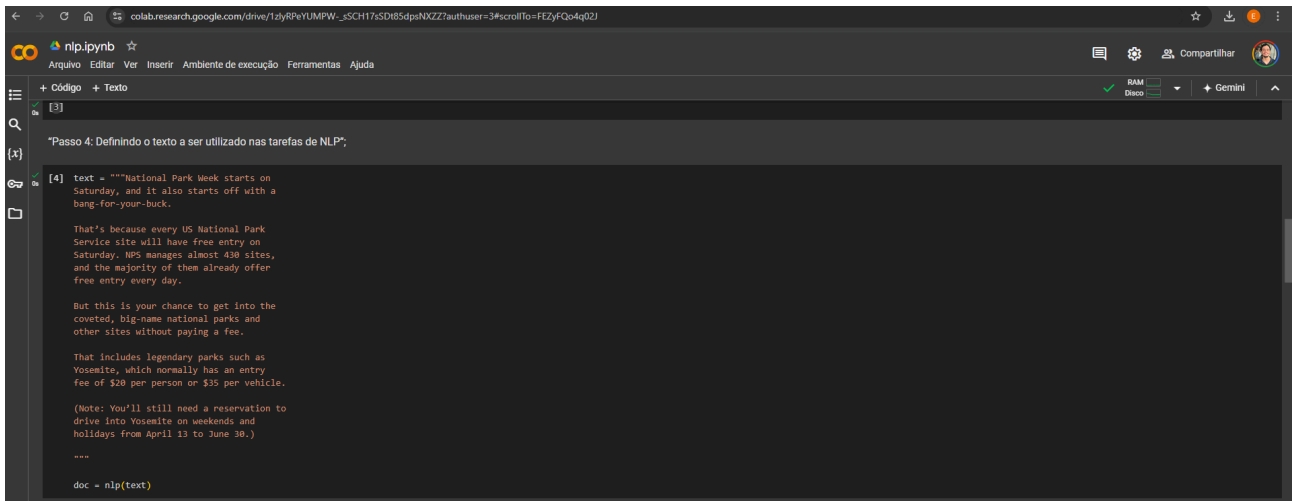
Passo 3:



The screenshot shows a Google Colab notebook interface. The title bar indicates the notebook is named 'nlp.ipynb'. The menu bar includes 'Arquivo', 'Editar', 'Ver', 'Inserir', 'Ambiente de execução', 'Ferramentas', and 'Ajuda'. The toolbar shows 'RAM' and 'Disco' usage. The main code cell contains the following Python code:

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

Passo 4:



The screenshot shows a Google Colab notebook interface. The title bar indicates the notebook is named 'nlp.ipynb'. The menu bar includes 'Arquivo', 'Editar', 'Ver', 'Inserir', 'Ambiente de execução', 'Ferramentas', and 'Ajuda'. The toolbar shows 'RAM' and 'Disco' usage. The main code cell contains the following Python code:

```
[4] 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.

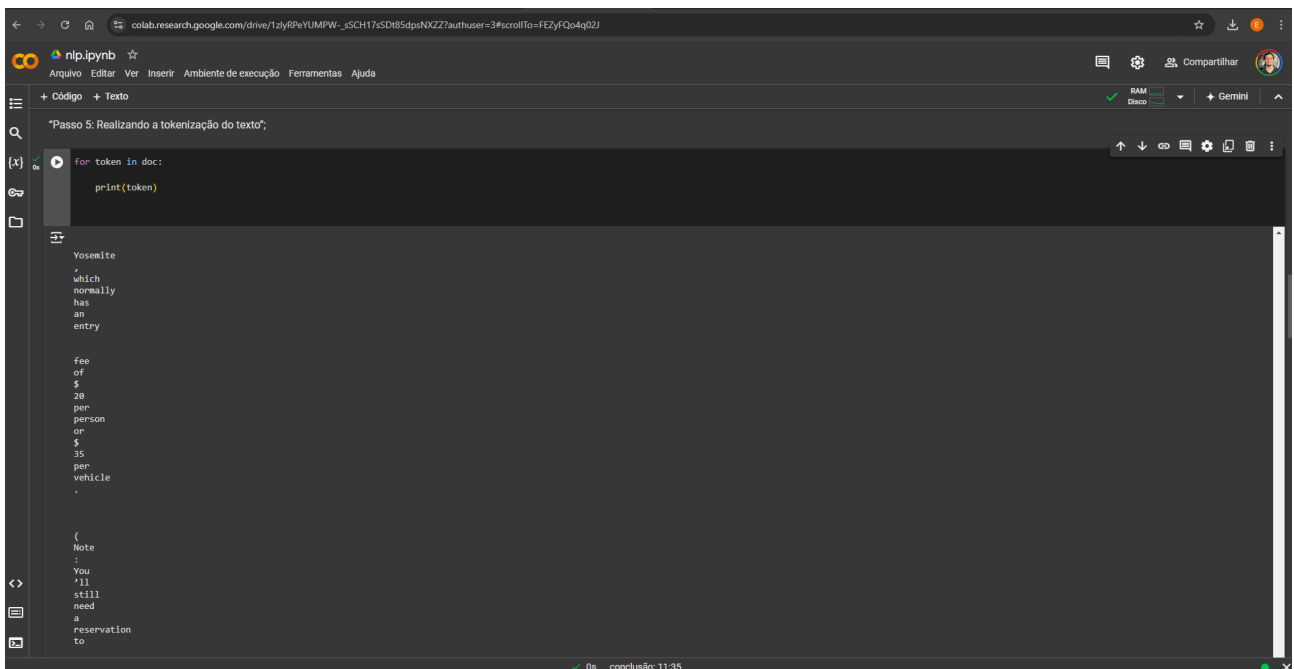
That includes legendary parks such as
Yosemite, which normally has an entry
fee of $20 per person or $35 per vehicle.

(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:



The screenshot shows a Google Colab notebook interface. The title bar indicates the notebook is named 'nlp.ipynb'. The menu bar includes 'Arquivo', 'Editar', 'Ver', 'Inserir', 'Ambiente de execução', 'Ferramentas', and 'Ajuda'. The toolbar shows 'RAM' and 'Disco' usage. The main code cell contains the following Python code:

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

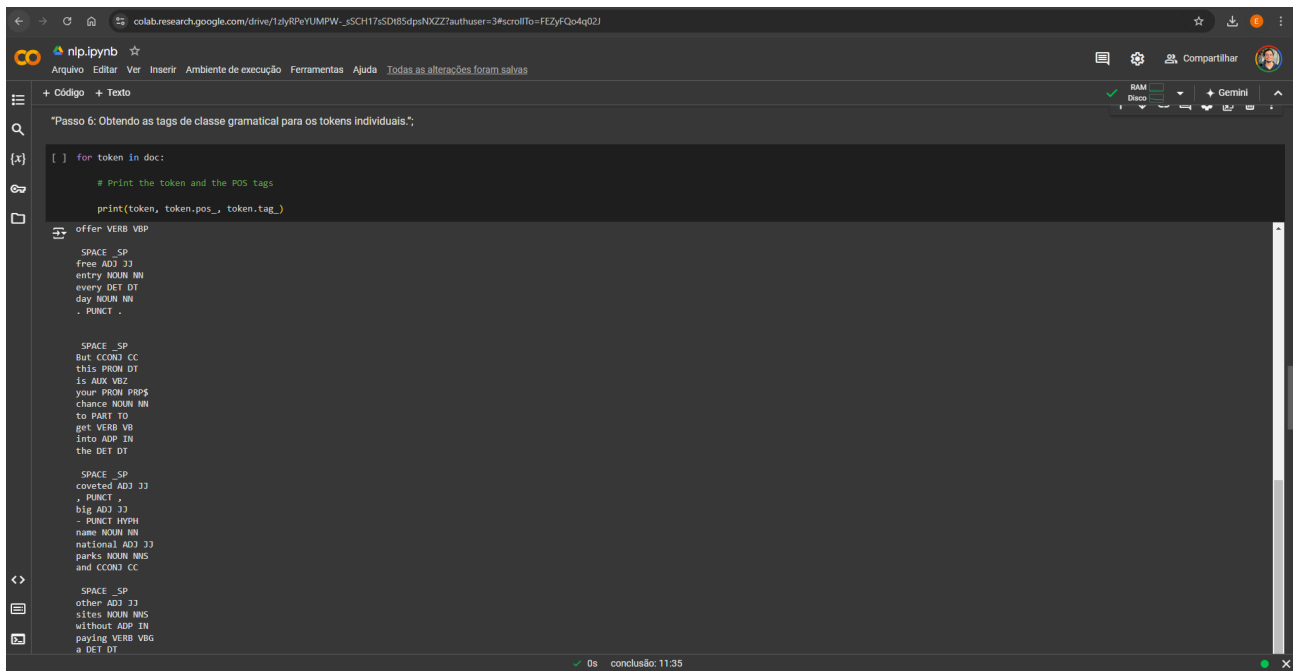
The output of the code is displayed below the code cell, showing the tokens of the text defined in Step 4:

```
Yosemite
,
which
normally
has
an
entry

fee
of
$
20
per
person
on
$
35
per
vehicle
.

(
Note
:
You
'll
still
need
a
reservation
to
```

Passo 6:



```
colab.research.google.com/drive/1zyRPeYUMPW...SCH17sSDt85dpsHXZZ?authuser=3#scrollTo=FEZyFQo4q02
```

nlipyb

Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda Todas as alterações foram salvas

+ Código + Texto

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

```
[ ] for token in doc:
    # Print the token and the POS tags
    print(token, token.pos_, token.tag_)
```

```
offer VERB VBP
SPACE _SP
free ADJ JJ
entry NOUN NN
every DET DT
day NOUN NN
. PUNCT .

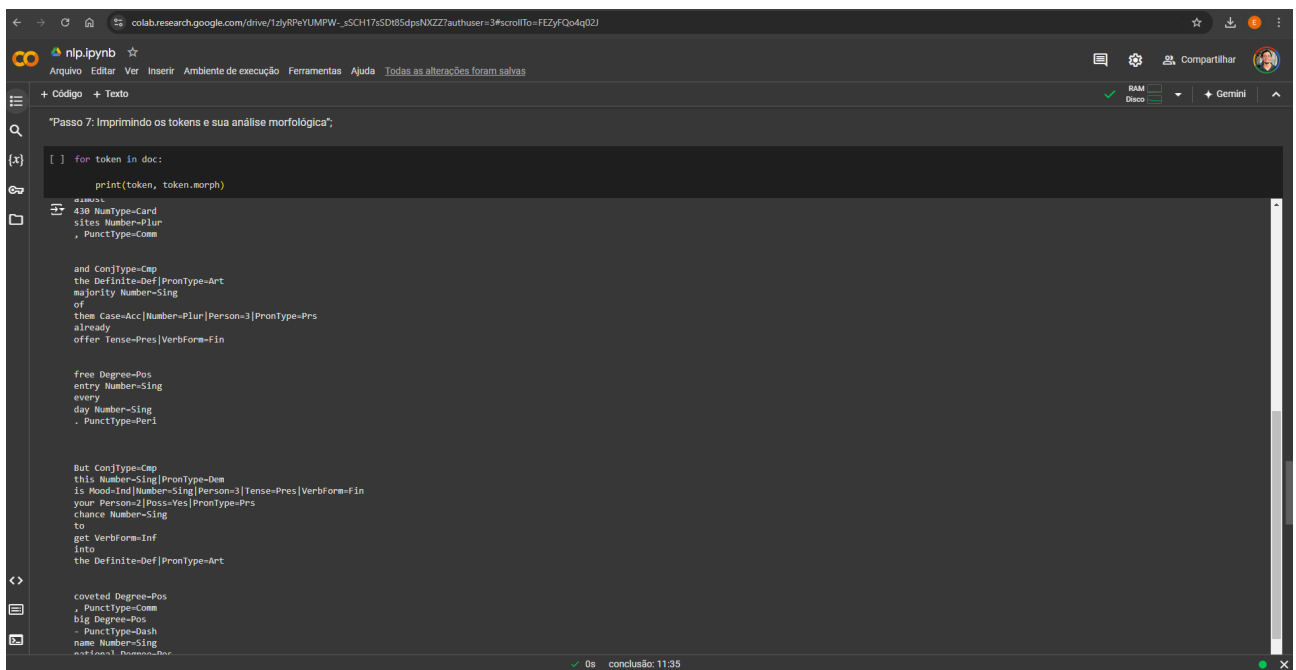
SPACE _SP
But CONJ CC
this PRON DT
is AUX VBZ
your PRON PRP$
chance NOUN NN
to PART TO
get VERB VB
into ADP IN
the DET DT

SPACE _SP
coveted ADJ JJ
, PUNCT ,
big ADJ JJ
- PUNCT HYPH
name NOUN NN
national ADJ JJ
parks NOUN NNS
and CONJ CC

SPACE _SP
other ADJ JJ
sites NOUN NNS
without ADP IN
paying VERB VBG
a DET DT
```

0s conclusão: 11:35

Passo 7:



```
colab.research.google.com/drive/1zyRPeYUMPW...SCH17sSDt85dpsHXZZ?authuser=3#scrollTo=FEZyFQo4q02
```

nlipyb

Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda Todas as alterações foram salvas

+ Código + Texto

"Passo 7: Imprimindo os tokens e sua análise morfológica;"

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

```
offer
438 NumType=Card
sites Number=Plur
, PunctType=Comm

and ConjType=Cmp
the Definite=Def|PronType=Art
majority Number=Sing
of
them Case=Acc|Number=Plur|Person=3|PronType=Prs
already
offer Tense=Pres|VerbForm=Fin

free Degree=Pos
entry Number=Sing
every
day Number=Sing
. PunctType=Peri

But ConjType=Cmp
this Number=Sing|PronType=Dem
is Mood=Ind|Number=Sing|Person=3|Tense=Pres|VerbForm=Fin
your Person=2|Poss=Yes|PronType=Prs
chance Number=Sing
to
get VerbForm=Inf
into
the Definite=Def|PronType=Art

coveted Degree=Pos
, PunctType=Comm
big Degree=Pos
- PunctType=Dash
name Number=Sing
national Number=Plur
```

0s conclusão: 11:35

Passo 8:

colab.research.google.com/drive/1zy8PeYUMPW...sSCH17sSDt85dp10QZZ?authuser=3#scrollTo=F2ZyFQo4q02J

nlipyb

Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda Todas as alterações foram salvas

+ Código + Texto

RAM Disco Gemini

"Passo 8: Visualizando a árvore de análise sintática por frase e a relação entre as palavras,"

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

National PROPN Park PROPN Week PROPN starts VERB on ADP SPACE dep Saturday, PROPN and CONJ it PRON also ADV starts VERB off ADP

0s conclusão: 11:35

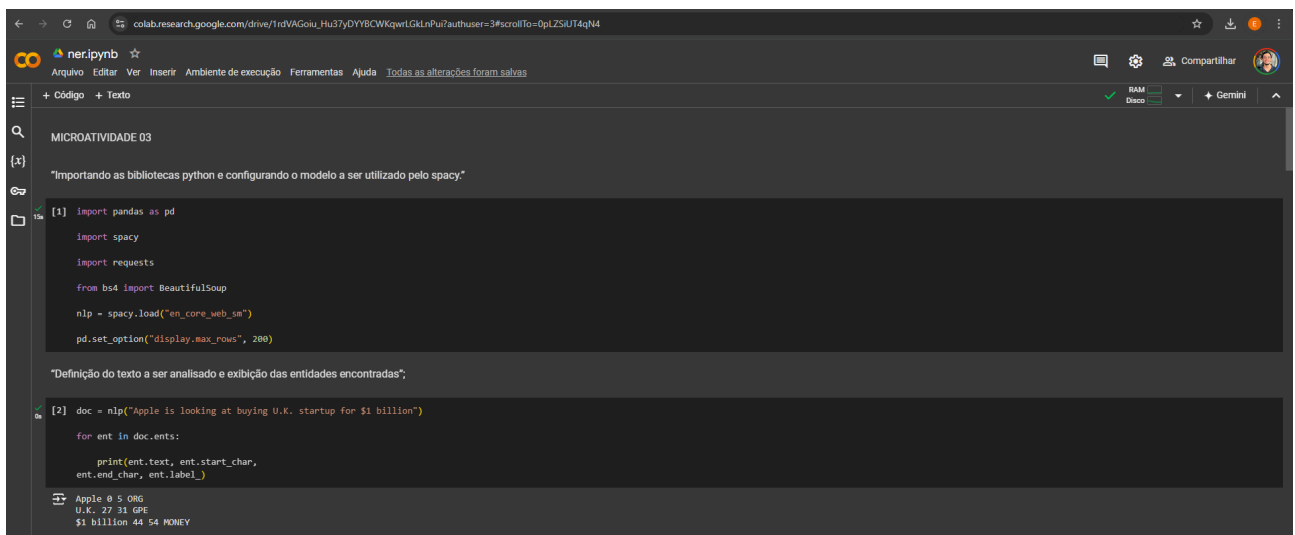
Microatividade 3

Descrever o processo de identificação de entidades a partir de textos.

Procedimentos:

1. Estando logado no Google Colab, clique no menu "Arquivo" e selecione a opção "Novo notebook";
2. Na nova aba aberta no navegador, dê um nome ao seu notebook, clicando e alterando o nome automaticamente gerado – Untitled0.ipynb – para ner.ipynb;
3. Na janela de código, clique na opção "+Texto" (destacada no print abaixo) para inserir um bloco de texto;

Resultados:



```
colab.research.google.com/drive/1rdVAGou_Hu37yDYY8CWKqwrLGkLnPui?authuser=3#scrollTo=0pLZSIU4qN4
```

ner.ipynb

Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda Todas as alterações foram salvas

+ Código + Texto

MICROATIVIDADE 03

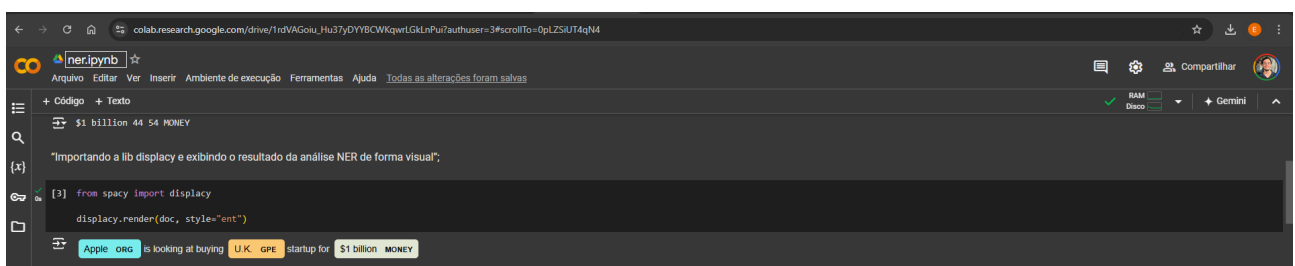
"Importando as bibliotecas python e configurando o modelo a ser utilizado pelo spacy."

```
[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)
```

"Definição do texto a ser analisado e exibição das entidades encontradas;"

```
[2] 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



```
colab.research.google.com/drive/1rdVAGou_Hu37yDYY8CWKqwrLGkLnPui?authuser=3#scrollTo=0pLZSIU4qN4
```

ner.ipynb

Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda Todas as alterações foram salvas

+ Código + Texto

\$1 billion 44 54 MONEY

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

```
[3] from spacy import displacy
displacy.render(doc, style="ent")
```

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

```
colab.research.google.com/drive/1ndVAGou_Hu37yDYB8CWKqwtLGkLnPuf?authuser=3#scrollTo=0pLZ5U14qN4
ner.ipynb
Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda Todas as alterações foram salvas
+ Código + Texto
Apple Card is coming at a dizzy rate. OFC LIFE Setup for 2 FOUND 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)
```

```
colab.research.google.com/drive/1ndVAGou_Hu37yDYB8CWKqwtLGkLnPuf?authuser=3#scrollTo=0pLZ5U14qN4
ner.ipynb
Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda Todas as alterações foram salvas
+ Código + Texto
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]
```


colab.research.google.com/drive/1rdVAGou_Hu37jDY8CWKqwtGkLnPu?authuser=3#scrollTo=0plZ5U14qN4

ner.ipynb

Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda Todas as alterações foram salvas

+ Código + Texto

at 206-555-0100 CARDINAL

If you have questions about your bill, [AnyCompany Customer Service](#) is available by

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

"Visualizando o resultado da análise de forma tabular":

```
[5] 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	Financial Services	ORG	Financial Services
2	LLC	ORG	LLC
3	1111-0000-1111-0000	DATE	1111-0000-1111-0000
4	24.53	MONEY	24.53

Próximas etapas: Gerar código com df Ver gráficos recomendados New Interactive sheet

3a conclusão: 11:42

colab.research.google.com/drive/1rdVAGou_Hu37jDY8CWKqwtGkLnPu?authuser=3#scrollTo=0plZ5U14qN4

ner.ipynb

Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda Todas as alterações foram salvas

+ Código + Texto

"Utilizando modelos no idioma português":

```
!python -m spacy download pt

import pkg_resources, imp
imp.reload(pkg_resources)
```

As of spacy v3.0, shortcuts like 'pt' are deprecated. Please use the full pipeline package name 'pt_core_news_sm' instead.

Collecting pt-core-news-sm==3.7.0
Downloading https://github.com/explosion/spacy-models/releases/download/pt_core_news_sm-3.7.0/pt_core_news_sm-3.7.0-py3-none-any.whl (13.0 MB)
13.0/13.0 MB 64.0 MB/s eta 0:00:00

Requirement already satisfied: spacy<3.8.0,>=3.7.0 in /usr/local/lib/python3.10/dist-packages (from pt-core-news-sm==3.7.0) (3.7.5)
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in /usr/local/lib/python3.10/dist-packages (from spacy<3.8.0,>=3.7.0>pt-core-news-sm==3.7.0) (3.0.12)
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in /usr/local/lib/python3.10/dist-packages (from spacy<3.8.0,>=3.7.0>pt-core-news-sm==3.7.0) (1.0.5)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/python3.10/dist-packages (from spacy<3.8.0,>=3.7.0>pt-core-news-sm==3.7.0) (1.0.10)
Requirement already satisfied: cymem<2.0.0,>=2.0.2 in /usr/local/lib/python3.10/dist-packages (from spacy<3.8.0,>=3.7.0>pt-core-news-sm==3.7.0) (2.0.6)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python3.10/dist-packages (from spacy<3.8.0,>=3.7.0>pt-core-news-sm==3.7.0) (3.0.9)
Requirement already satisfied: thinc<8.3.0,>=8.2.2 in /usr/local/lib/python3.10/dist-packages (from spacy<3.8.0,>=3.7.0>pt-core-news-sm==3.7.0) (8.2.5)
Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in /usr/local/lib/python3.10/dist-packages (from spacy<3.8.0,>=3.7.0>pt-core-news-sm==3.7.0) (1.1.3)
Requirement already satisfied: srsly<3.0.0,>=2.4.3 in /usr/local/lib/python3.10/dist-packages (from spacy<3.8.0,>=3.7.0>pt-core-news-sm==3.7.0) (2.4.6)
Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in /usr/local/lib/python3.10/dist-packages (from spacy<3.8.0,>=3.7.0>pt-core-news-sm==3.7.0) (2.0.10)
Requirement already satisfied: weasel<0.5.0,>=0.1.0 in /usr/local/lib/python3.10/dist-packages (from spacy<3.8.0,>=3.7.0>pt-core-news-sm==3.7.0) (0.4.1)
Requirement already satisfied: typer<1.0.0,>=0.3.0 in /usr/local/lib/python3.10/dist-packages (from spacy<3.8.0,>=3.7.0>pt-core-news-sm==3.7.0) (0.12.5)
Requirement already satisfied: tqdm<4.0.0,>=4.38.0 in /usr/local/lib/python3.10/dist-packages (from spacy<3.8.0,>=3.7.0>pt-core-news-sm==3.7.0) (4.66.6)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in /usr/local/lib/python3.10/dist-packages (from spacy<3.8.0,>=3.7.0>pt-core-news-sm==3.7.0) (2.32.3)
Requirement already satisfied: pydantic<1.8.1,>=1.8.1,<3.0.0,>=1.7.4 in /usr/local/lib/python3.10/dist-packages (from spacy<3.8.0,>=3.7.0>pt-core-news-sm==3.7.0) (2.9.2)
Requirement already satisfied: Jinja2 in /usr/local/lib/python3.10/dist-packages (from spacy<3.8.0,>=3.7.0>pt-core-news-sm==3.7.0) (3.1.4)
Requirement already satisfied: setuptools in /usr/local/lib/python3.10/dist-packages (from spacy<3.8.0,>=3.7.0>pt-core-news-sm==3.7.0) (75.1.0)
Requirement already satisfied: packaging<=20.0 in /usr/local/lib/python3.10/dist-packages (from spacy<3.8.0,>=3.7.0>pt-core-news-sm==3.7.0) (24.1)
Requirement already satisfied: langcodes<4.0.0,>=3.2.0 in /usr/local/lib/python3.10/dist-packages (from spacy<3.8.0,>=3.7.0>pt-core-news-sm==3.7.0) (3.4.1)
Requirement already satisfied: numpy<1.19.0 in /usr/local/lib/python3.10/dist-packages (from spacy<3.8.0,>=3.7.0>pt-core-news-sm==3.7.0) (1.26.4)
Requirement already satisfied: language-data<=2.1.2 in /usr/local/lib/python3.10/dist-packages (from langcodes<4.0.0,>=3.2.0>spacy<3.8.0,>=3.7.0>pt-core-news-sm==3.7.0) (1.2.0)
Requirement already satisfied: annotated-types<=0.6.0 in /usr/local/lib/python3.10/dist-packages (from pydantic<1.8.1,>=1.8.1,<3.0.0,>=1.7.4>spacy<3.8.0,>=3.7.0>pt-core-news-sm==3.7.0) (0.7.0)
Requirement already satisfied: pydantic-core==2.23.4 in /usr/local/lib/python3.10/dist-packages (from pydantic<1.8.1,>=1.8.1,<3.0.0,>=1.7.4>spacy<3.8.0,>=3.7.0>pt-core-news-sm==3.7.0) (2.23.4)
Requirement already satisfied: typing-extensions<=4.6.1 in /usr/local/lib/python3.10/dist-packages (from pydantic<1.8.1,>=1.8.1,<3.0.0,>=1.7.4>spacy<3.8.0,>=3.7.0>pt-core-news-sm==3.7.0) (4.12.2)
Requirement already satisfied: charset-normalizer<4.0,>=2 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.13.0>spacy<3.8.0,>=3.7.0>pt-core-news-sm==3.7.0) (3.4.0)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.13.0>spacy<3.8.0,>=3.7.0>pt-core-news-sm==3.7.0) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.13.0>spacy<3.8.0,>=3.7.0>pt-core-news-sm==3.7.0) (2.2.3)
Requirement already satisfied: certifi<=2021.4.17 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.13.0>spacy<3.8.0,>=3.7.0>pt-core-news-sm==3.7.0) (2024.8.30)
Requirement already satisfied: blis<0.8.0,>=0.2.2 in /usr/local/lib/python3.10/dist-packages (from thinc<8.3.0,>=8.2.2>spacy<3.8.0,>=3.7.0>pt-core-news-sm==3.7.0) (0.7.11)
Requirement already satisfied: confection<1.0.0,>=0.0.1 in /usr/local/lib/python3.10/dist-packages (from thinc<8.3.0,>=8.2.2>spacy<3.8.0,>=3.7.0>pt-core-news-sm==3.7.0) (0.1.5)
Requirement already satisfied: click==8.0.0 in /usr/local/lib/python3.10/dist-packages (from thinc<8.3.0,>=8.2.2>spacy<3.8.0,>=3.7.0>pt-core-news-sm==3.7.0) (8.1.7)

3a conclusão: 11:42

```
colab.research.google.com/drive/1rdVAGou_Hu37yDYBCKWkqWLGkLnPui?authuser=3#scrollTo=0pLZSIU4qN4
ner.ipynb
Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda Todas as alterações foram salvas
+ Código + Texto
Adicione um novo bloco de código com as seguintes linhas:
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.
3s conclusão: 11:42
```

```
colab.research.google.com/drive/1rdVAGou_Hu37yDYBCKWkqWLGkLnPui?authuser=3#scrollTo=0pLZSIU4qN4
ner.ipynb
Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda Todas as alterações foram salvas
+ Código + Texto
docbr = nlp(txt_br)
displacy.render(docbr, style="ent")
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 PERSON, em fevereiro de 2004.
Coube ORG ao Brasil a chefia da missão,
instalada no mês de ORG junho de 2004 DATE bem
3s conclusão: 11:42
```

Descrever o processo de extração de frases-chave a partir de textos

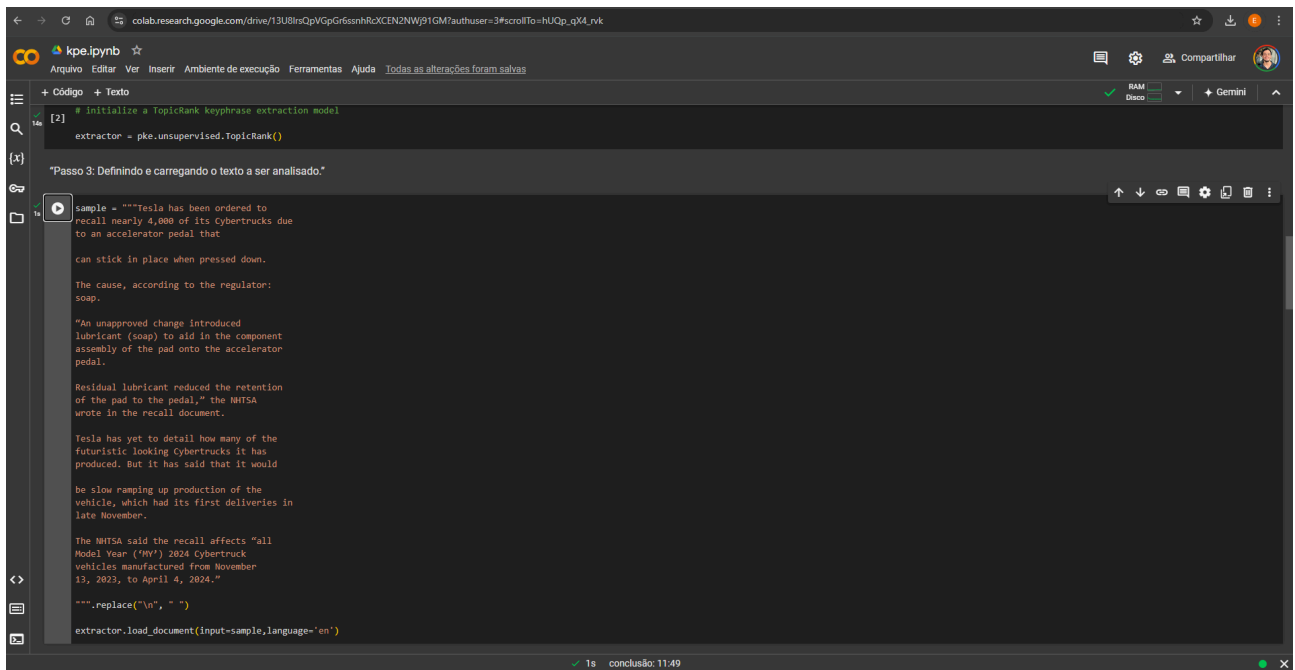
1. Estando logado no Google Colab, clique no menu “Arquivo” e selecione a opção “Novo notebook”;
2. Na nova aba aberta no navegador, dê um nome ao seu notebook, clicando e alterando o nome automaticamente gerado – Untitled0.ipynb – para kpe.ipynb;
3. Na janela de código, clique na opção “+Texto” (destacada no print abaixo) para inserir um bloco de texto;

Passo 1:

Passo 2:

11

Passo 3:



The screenshot shows a Jupyter Notebook interface with a dark theme. The top bar includes the Google Colab logo and a URL. The notebook has two tabs: 'Código' and 'Texto'. The 'Código' tab is active, showing a Python cell with the following code:

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

Below the code cell, there is a text area with the following text:

"Passo 3: Definindo e carregando o texto a ser analisado."

The text area contains a sample text about Tesla's recall of Cybertrucks. The text is as follows:

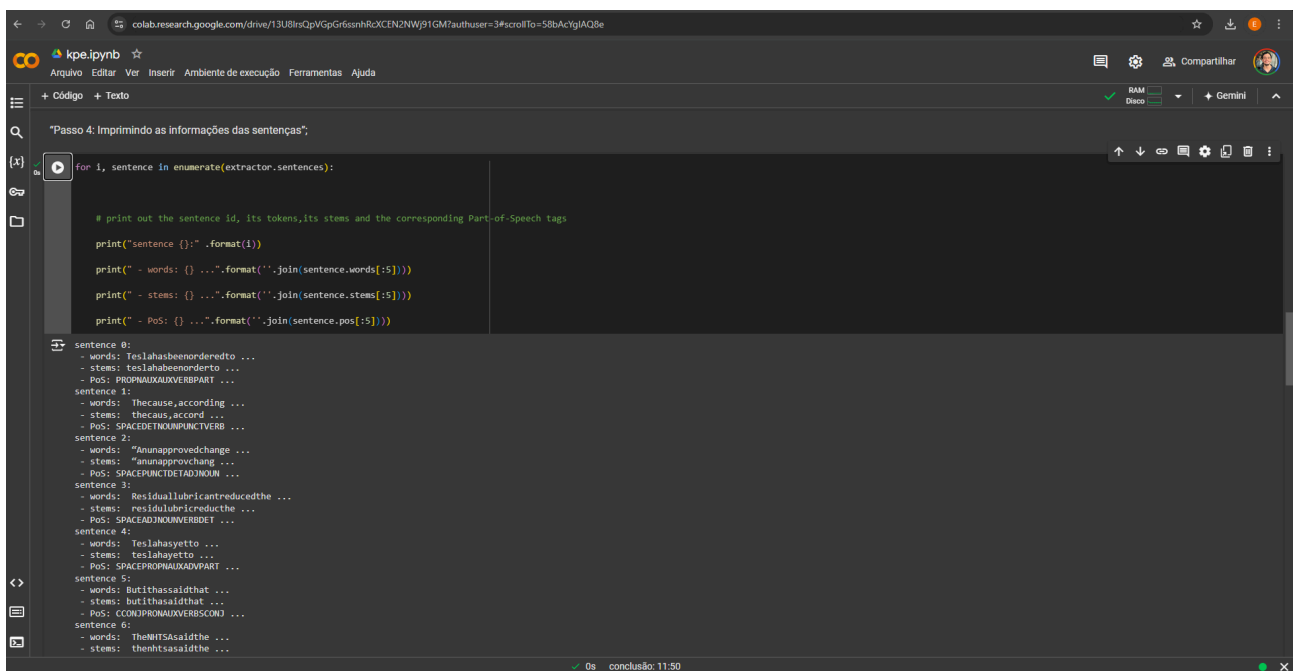
sample = """Tesla has been ordered to recall nearly 4,000 of its Cybertrucks due to an accelerator pedal that can stick in place when pressed down. The cause, according to the regulator: soap. "An unapproved change introduced lubricant (soap) to aid in the component assembly of the pad onto the accelerator 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." """

The code cell is executed, and the output shows the text being loaded into the extractor.

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

The bottom status bar shows '1s' and 'conclusão: 11:49'.

Passo 4:



The screenshot shows a Jupyter Notebook interface with a dark theme. The top bar includes the Google Colab logo and a URL. The notebook has two tabs: 'Código' and 'Texto'. The 'Código' tab is active, showing a Python cell with the following code:

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

    # print out the sentence id, its tokens, its stems and the corresponding Part-of-Speech tags
    print("sentence {}: ".format(i))

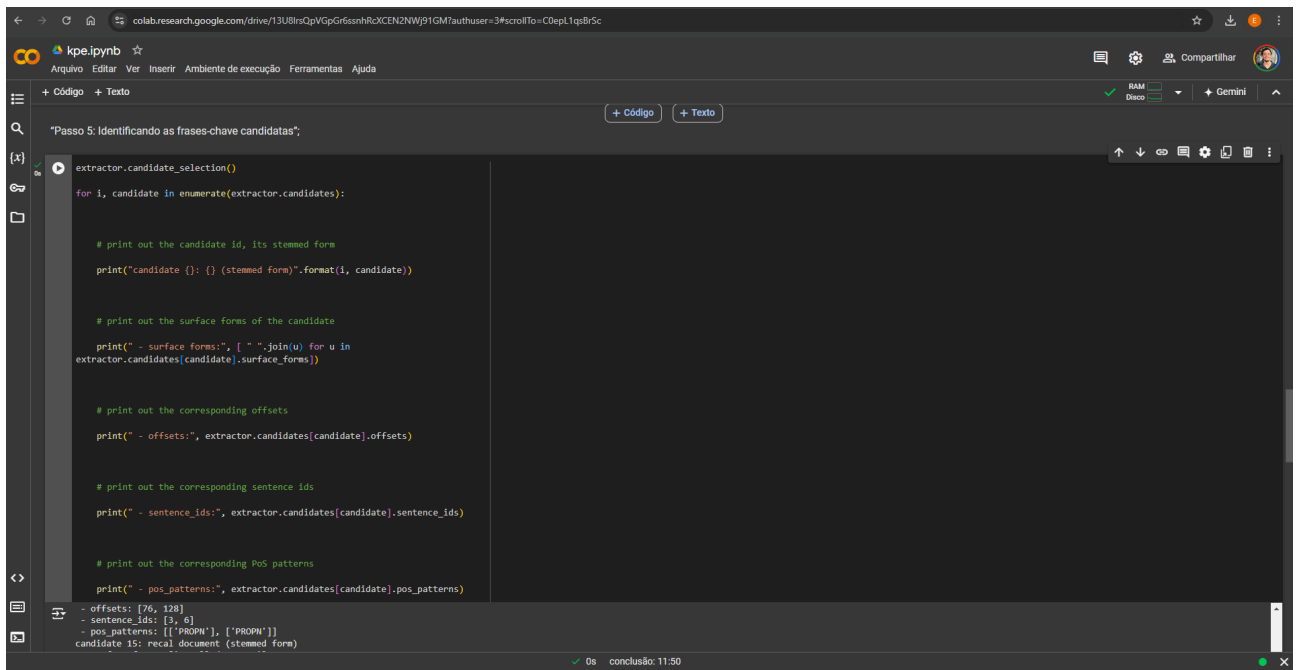
    print(" - words: {} ...".format(' '.join(sentence.words[:5])))
    print(" - stems: {} ...".format(' '.join(sentence.stems[:5])))
    print(" - PoS: {} ...".format(' '.join(sentence.pos[:5])))
```

Below the code cell, there is a text area showing the output of the code. The output is as follows:

```
sentence 0:
- words: teslahasbeenorderedto ...
- stems: teslahasbeenorderto ...
- PoS: PROPNAUXAUXVERBPART ...
sentence 1:
- words: Thecause,according ...
- stems: thecaus,accord ...
- PoS: SPACEDETINOUNPUNCTVERB ...
sentence 2:
- words: "Anunapprovedchange ...
- stems: "anunapprovchang ...
- PoS: SPACEPUNCTIDETADJINOUN ...
sentence 3:
- words: Residuallubricantreducedthe ...
- stems: residulubricreducedthe ...
- PoS: SPACEADJINOUNVERBDEI ...
sentence 4:
- words: Teslahasyetto ...
- stems: teslahayetto ...
- PoS: SPACEPROPNAUXADVPART ...
sentence 5:
- words: Butithassaidthat ...
- stems: butithassaidthat ...
- PoS: CCONDPROMAUXVERASCOM ...
sentence 6:
- words: TheNHTSAsaidthe ...
- stems: thenhtsasaidthe ...
```

The bottom status bar shows '0s' and 'conclusão: 11:50'.

Passo 5:



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

        # print out the candidate id, its stemmed form
        print("candidate {}: {} (stemmed form)".format(i, candidate))

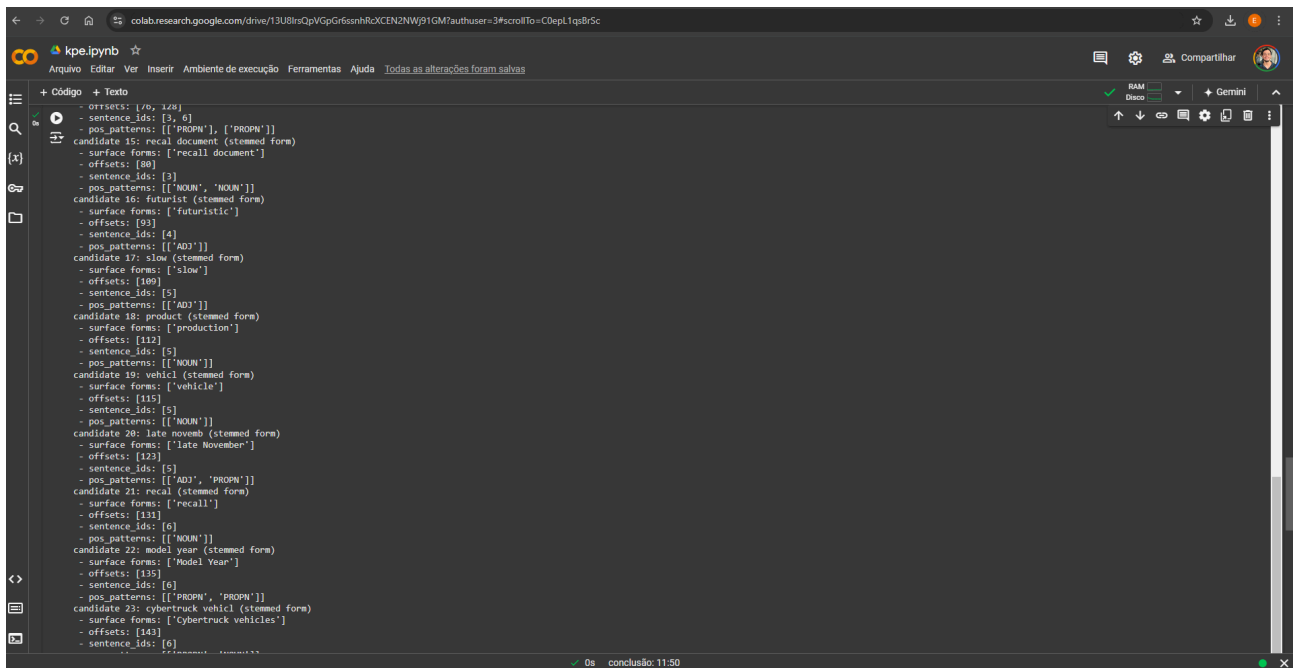
        # print out the surface forms of the candidate
        print("- surface forms:", [ " ".join(u) for u in
            extractor.candidates[candidate].surface_forms])

        # print out the corresponding offsets
        print("- offsets:", extractor.candidates[candidate].offsets)

        # print out the corresponding sentence ids
        print("- sentence_ids:", extractor.candidates[candidate].sentence_ids)

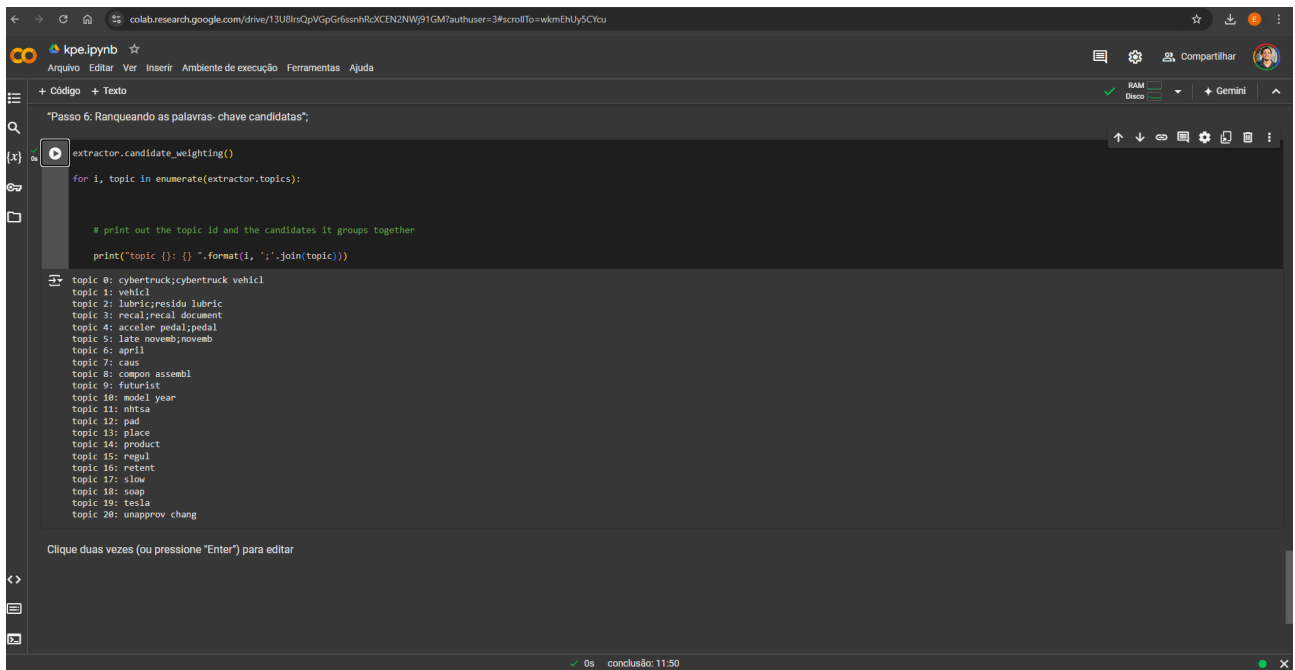
        # print out the corresponding PoS patterns
        print("- pos_patterns:", extractor.candidates[candidate].pos_patterns)

    - offsets: [76, 128]
    - sentence_ids: [3, 6]
    - pos_patterns: [['PROPN'], ['PROPN']]
    candidate 15: recal document (stemmed form)
```



```
- offsets: [76, 128]
- sentence_ids: [3, 6]
- pos_patterns: [['PROPN'], ['PROPN']]
candidate 15: recal document (stemmed form)
- surface forms: ['recall document']
- offsets: [68]
- sentence_ids: [3]
- pos_patterns: [['PROPN'], ['PROPN']]
candidate 16: futurist (stemmed form)
- surface forms: ['futuristic']
- offsets: [93]
- sentence_ids: [4]
- pos_patterns: [['ADJ']]
candidate 17: slow (stemmed form)
- surface forms: ['slow']
- offsets: [109]
- sentence_ids: [5]
- pos_patterns: [['ADJ']]
candidate 18: product (stemmed form)
- surface forms: ['production']
- offsets: [112]
- sentence_ids: [5]
- pos_patterns: [['NOUN']]
candidate 19: vehicl (stemmed form)
- surface forms: ['vehicle']
- offsets: [115]
- sentence_ids: [5]
- pos_patterns: [['NOUN']]
candidate 20: late novebm (stemmed form)
- surface forms: ['late November']
- offsets: [123]
- sentence_ids: [5]
- pos_patterns: [['ADJ', 'PROPN']]
candidate 21: recal (stemmed form)
- surface forms: ['recall']
- offsets: [131]
- sentence_ids: [6]
- pos_patterns: [['NOUN']]
candidate 22: model year (stemmed form)
- surface forms: ['Model Year']
- offsets: [135]
- sentence_ids: [6]
- pos_patterns: [['PROPN', 'PROPN']]
candidate 23: cybertruck vehicl (stemmed form)
- surface forms: ['Cybertruck vehicles']
- offsets: [143]
- sentence_ids: [6]
```

Passo 6:



colab.research.google.com/drive/13U8lrsQpVGpGr6ssnhRxCEN2NWj91GM7authuser=3#scrollTo=wkmEHUy5CYcu

kpe.ipynb

Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda

+ Código + Texto

Passo 6: Ranqueando as palavras-chave candidatas;

```
extractor.candidate_weighting()

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

    # print out the topic id and the candidates it groups together

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

topic 0: cybertruck;cybertruck vehicl
topic 1: vehicl
topic 2: lubric;residu lubric
topic 3: recal;recal document
topic 4: acceler pedal;pedal
topic 5: late novemb;novemb
topic 6: april
topic 7: caus
topic 8: compon assembl
topic 9: futurist
topic 10: model year
topic 11: nhisa
topic 12: pad
topic 13: place
topic 14: product
topic 15: regul
topic 16: retent
topic 17: slow
topic 18: soap
topic 19: tesia
topic 20: unapprov chang

Clique duas vezes (ou pressione "Enter") para editar

0s conclusão: 11:50

Microatividade 5

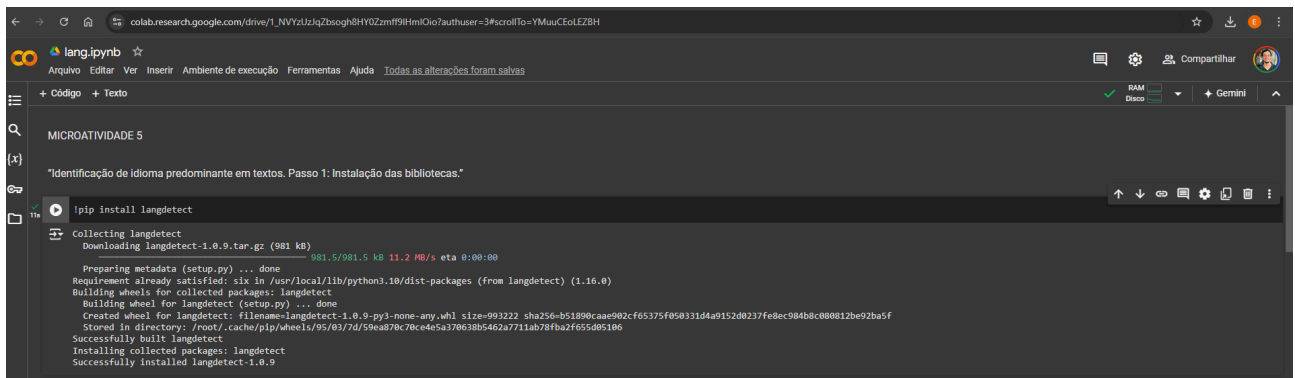
Descrever o processo de identificação de linguagem predominante a partir de textos

Procedimentos:

1. Estando logado no Google Colab, clique no menu "Arquivo" e selecione a opção "Novo notebook";
2. Na nova aba aberta no navegador, dê um nome ao seu notebook, clicando e alterando o nome automaticamente gerado – Untitled0.ipynb – para lang.ipynb;
3. Na janela de código, clique na opção "+Texto" (destacada no print abaixo) para inserir um bloco de texto;

Resultado:

Passo 1:

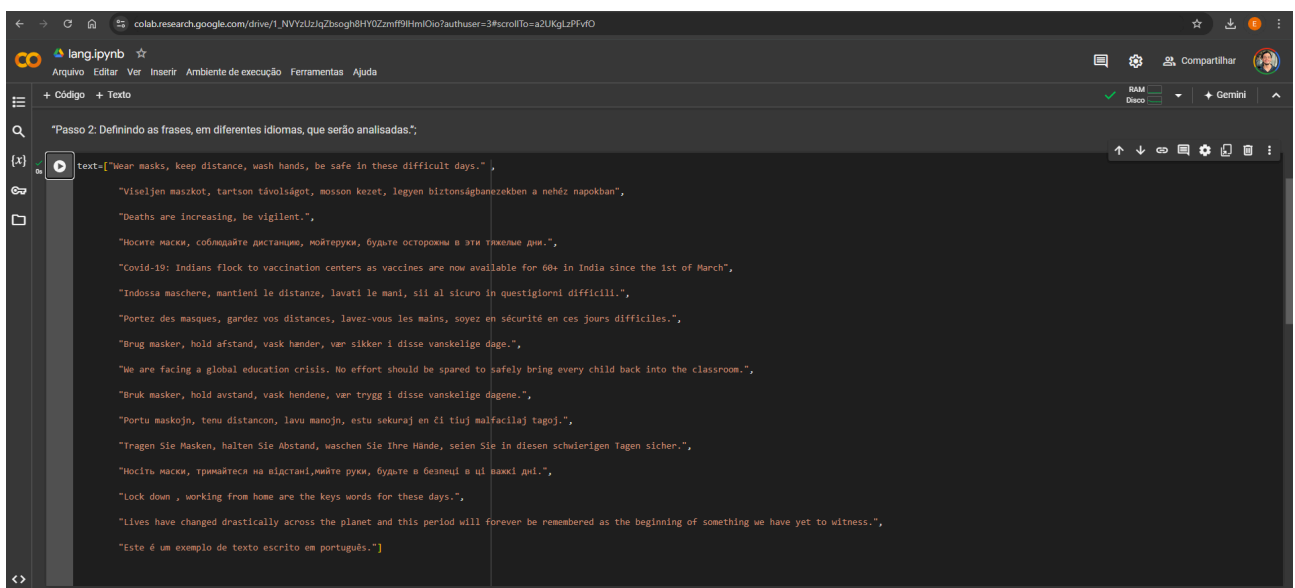


The screenshot shows the Google Colab interface with a notebook named 'lang.ipynb'. The code cell contains the command `!pip install langdetect`. The output shows the successful installation of langdetect-1.0.9. The interface includes a menu bar with 'Arquivo', 'Editar', 'Ver', 'Inserir', 'Ambiente de execução', 'Ferramentas', and 'Ajuda'. The left sidebar shows the notebook structure with a code cell selected.

```
!pip install langdetect

Collecting langdetect
  Downloading langdetect-1.0.9.tar.gz (981 kB)
    981.5/981.5 kB 11.2 MB/s eta 0:00:00
  Preparing metadata (setup.py) ... done
Requirement already satisfied: six in /usr/local/lib/python3.10/dist-packages (from langdetect) (1.16.0)
Building wheels for collected packages: langdetect
  Building wheel for langdetect (setup.py) ... done
Created wheel for langdetect: filename=langdetect-1.0.9-py3-none-any.whl size=993222 sha256=b61890ca902cf65375f050331d4e9152d0237fe8ec984b0c080812be92ba5f
Stored in directory: /root/.cache/pip/wheels/95/03/7d/59ca870c78ce4e5a370638b5462a7711ab78fb2f655085106
Successfully built langdetect
Installing collected packages: langdetect
Successfully installed langdetect-1.0.9
```

Passo 2:



The screenshot shows the Google Colab interface with a notebook named 'lang.ipynb'. The code cell contains a text block with multilingual content. The interface includes a menu bar with 'Arquivo', 'Editar', 'Ver', 'Inserir', 'Ambiente de execução', 'Ferramentas', and 'Ajuda'. The left sidebar shows the notebook structure with a text cell selected.

```
text=["Wear masks, keep distance, wash hands, be safe in these difficult days.",
      "Viseljén 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."]


```

Passo 3:

The screenshot shows the lang.ipynb Jupyter Notebook interface. The top bar includes the lang.ipynb logo, a star icon, and a menu with 'Arquivo', 'Editar', 'Ver', 'Inserir', 'Ambiente de execução', 'Ferramentas', and 'Ajuda'. On the right, there are icons for chat, settings, sharing, and a user profile. The left sidebar has icons for file explorer, search, and other functions. The main area displays the notebook content with the title 'Passo 3: Recarregando o ambiente python após a instalação da lib.'. Below the title, there is a code cell with the following Python code:

```
import pkg_resources, imp
imp.reload(pkg_resources)
<module 'pkg_resources' from '/usr/local/lib/python3.10/dist-packages/pkg_resources/__init__.py'>
```

Passo 4:

The screenshot shows the lang.ipynb Jupyter Notebook interface. The top bar includes the lang.ipynb logo, a star icon, and a menu with 'Arquivo', 'Editar', 'Ver', 'Inserir', 'Ambiente de execução', 'Ferramentas', and 'Ajuda'. On the right, there are icons for chat, settings, sharing, and a user profile. The left sidebar has icons for file explorer, search, and other functions. The main area displays the notebook content with the title 'Passo 4: Detectando a linguagem predominante na lista de frases;'. Below the title, there is a code cell with the following Python code:

```
from langdetect import detect
for x in text:
    print ('Frase: ', x)
    print ('Idioma: ', detect(x), '\n\n')
```

Below the code cell, the output of the code is displayed, showing the detected language for each phrase in the list:

```
Frase: Deaths are increasing, be vigilant.
Idioma: en

Frase: Носите маски, соблюдайте дистанцию, мойте руки, будьте осторожны в эти тяжелые дни.
Idioma: ru

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

Frase: Indossa maschere, mantieni le distanze, lavati le mani, sii al sicuro in questi giorni difficili.
Idioma: it

Frase: Portez des masques, gardez vos distances, lavez-vous les mains, soyez en sécurité en ces jours difficiles.
Idioma: fr

Frase: Brug masker, hold afstand, vask hænder, vær sikker i disse vanskelige dage.
Idioma: da

Frase: We are facing a global education crisis. No effort should be spared to safely bring every child back into the classroom.
Idioma: en

Frase: Bruk masker, hold avstand, vask hendene, vær trygg i disse vanskelige dagene.
Idioma: da
```

The bottom status bar shows '1s' and 'conclusão: 11:54'.

The screenshot shows the lang.ipynb Jupyter Notebook interface. The top bar includes the lang.ipynb logo, a star icon, and a menu with 'Arquivo', 'Editar', 'Ver', 'Inserir', 'Ambiente de execução', 'Ferramentas', and 'Ajuda'. On the right, there are icons for chat, settings, sharing, and a user profile. The left sidebar has icons for file explorer, search, and other functions. The main area displays the notebook content with the title 'Passo 4: Detectando a linguagem predominante na lista de frases;'. Below the title, there is a code cell with the following Python code:

```
from langdetect import detect
for x in text:
    print ('Frase: ', x)
    print ('Idioma: ', detect(x), '\n\n')
```

Below the code cell, the output of the code is displayed, showing the detected language for each phrase in the list:

```
Frase: Носите маски, соблюдайте дистанцию, мойте руки, будьте осторожны в эти тяжелые дни.
Idioma: ru

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

Frase: Indossa maschere, mantieni le distanze, lavati le mani, sii al sicuro in questi giorni difficili.
Idioma: it

Frase: Portez des masques, gardez vos distances, lavez-vous les mains, soyez en sécurité en ces jours difficiles.
Idioma: fr

Frase: Brug masker, hold afstand, vask hænder, vær sikker i disse vanskelige dage.
Idioma: da

Frase: We are facing a global education crisis. No effort should be spared to safely bring every child back into the classroom.
Idioma: en

Frase: Bruk masker, hold avstand, vask hendene, vær trygg i disse vanskelige dagene.
Idioma: da

Frase: Portu maskojn, tenu distancon, lavu manojn, estu sekuraj en ĉi tiuj malfacilaj tagoj.
Idioma: hr

Frase: Tragen Sie Masken, halten Sie Abstand, waschen Sie Ihre Hände, seien Sie in diesen schwierigen Tagen sicher.
Idioma: de

Frase: Носіть маски, тримайтеся на відстані, мийте руки, будьте в безпеці в ці важкі дні.
Idioma: uk

Frase: Lock down , working from home are the keys words for these days.
Idioma: en

Frase: Lives have changed drastically across the planet and this period will forever be remembered as the beginning of something we have yet to witness.
Idioma: en
```

The bottom status bar shows '1s' and 'conclusão: 11:54'.

MISSÃO PRÁTICA

Dando Inteligência ao Software

Contextualização

Recentemente a empresa em que você trabalha, como Analista de Data Science, foi contratada por uma grande empresa interessada em abrir, no Brasil, centros de treinamento esportivos vinculados a grandes clubes de futebol da Inglaterra. Nesse contexto, a empresa contratante deseja saber a percepção das pessoas em relação aos clubes citados, i.e., de uma forma geral, qual o sentimento delas, expressos através de textos publicados em redes sociais, sobre os mesmos.

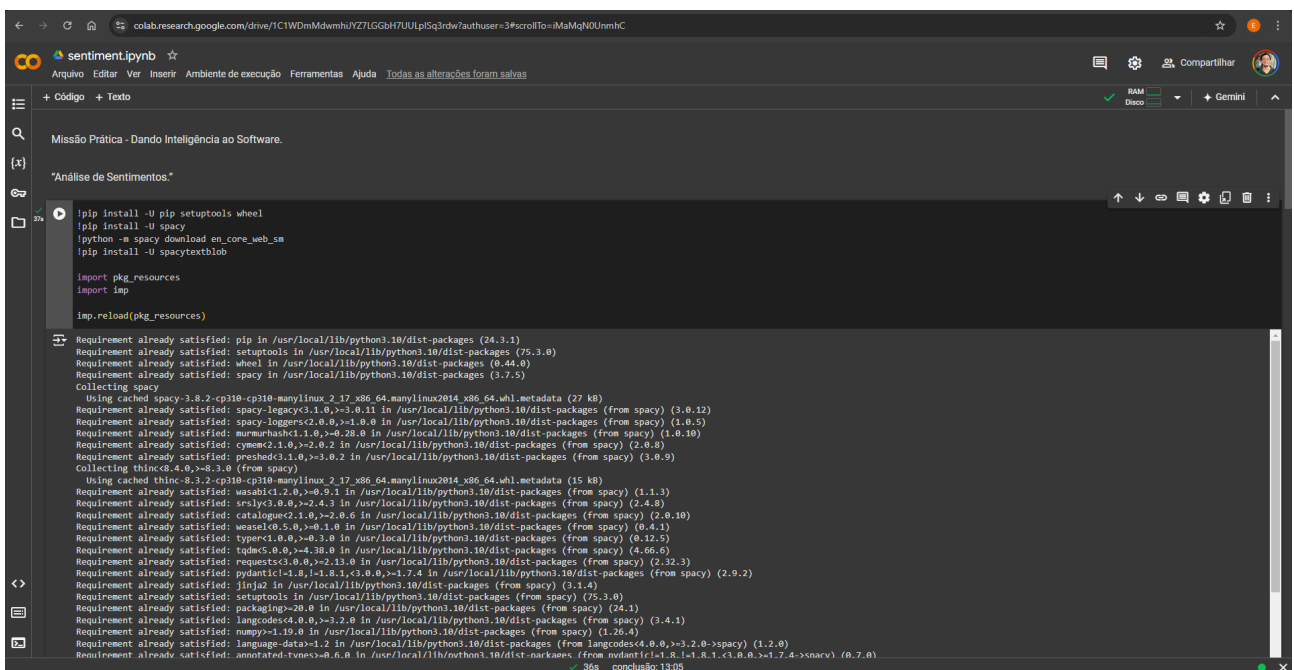
Para essa atividade você deverá aplicar a Análise de Sentimentos, tarefa de Processamento de Linguagem Natural com uso de Machine Learning. Todo o passo-a-passo necessário para a atividade é descrito a seguir.

Roteiro:

1. Estando logado no Google Colab, clique no menu "Arquivo" e selecione a opção "Novo notebook";
2. Na nova aba aberta no navegador, dê um nome ao seu notebook, clicando e alterando o nome automaticamente gerado – Untitled0.ipynb – para sentiment.ipynb;
3. Na janela de código, clique na opção "+Texto" (destacada no print abaixo) para inserir um bloco de texto;
4. No bloco de texto, insira um texto que explique o que será executado, a seguir, no bloco de código a ser inserido. Segue uma sugestão, que pode ser complementada posteriormente por você:
5. Insira um novo bloco de texto com o conteúdo: "Passo 1: Instalando as bibliotecas e recarregando o ambiente";
6. Insira um bloco de código com o conteúdo abaixo:
7. Execute o código acima. Durante o processo, caso receba, na tela, a mensagem dizendo que a sessão precisa ser reiniciada, clique no respectivo botão;
8. Insira um novo bloco de texto com o conteúdo: "Passo 2: Importando as bibliotecas para análise de sentimento";
9. Insira um bloco de código com as linhas abaixo e execute:
10. Insira um novo bloco de texto com o conteúdo: "Passo 3: Definindo o modelo e a pipeline a serem utilizadas na análise";
11. Crie um bloco de código com as linhas abaixo e execute:
12. Insira um novo bloco de texto com o conteúdo: "Passo 4: Definindo o texto inicial a ser analisado para verificação/validação da biblioteca";

13. Crie um bloco de código com o conteúdo abaixo e o execute:
14. Insira um novo bloco de texto com o conteúdo: "Passo 5: Exibindo o resultado da primeira análise (um range entre -1 [avaliação negativa] e 1 [avaliação positiva]);"
15. Crie um bloco de código com as linhas abaixo e o execute:
16. Insira um novo bloco de texto com o conteúdo: "Passo 6: Definindo a lista de tweets a serem analisadas";
17. Insira um bloco de código com as linhas abaixo e execute:
18. Insira um novo bloco de texto com o conteúdo: "Passo 7: Analisando os tweets";
19. Insira e executa o bloco de código abaixo:
20. Por fim, caso queira, você poderá salvar uma cópia do código no Google Drive ou no Github. Tais opções encontram-se disponíveis a partir do menu Arquivo.

Resultado:



```
!pip install -U pip setuptools wheel
!pip install -U spacy
!python -m spacy download en_core_web_sm
!pip install -U spacytextblob

import pkg_resources
import ip

ip.reload(pkg_resources)

Requirement already satisfied: pip in /usr/local/lib/python3.10/dist-packages (24.3.1)
Requirement already satisfied: setuptools in /usr/local/lib/python3.10/dist-packages (75.3.0)
Requirement already satisfied: wheel in /usr/local/lib/python3.10/dist-packages (0.44.0)
Requirement already satisfied: spacy in /usr/local/lib/python3.10/dist-packages (3.7.5)
Collecting spacy
  Using cached spacy-3.8.2-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (27 kB)
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in /usr/local/lib/python3.10/dist-packages (from spacy) (3.0.12)
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (1.0.5)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (1.0.10)
Requirement already satisfied: cyemes<2.1.0,>=2.0.2 in /usr/local/lib/python3.10/dist-packages (from spacy) (2.0.8)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python3.10/dist-packages (from spacy) (3.0.9)
Collecting thinc<8.4.0,>=8.3.0 (from spacy)
  Using cached thinc-8.3.2-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (15 kB)
Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in /usr/local/lib/python3.10/dist-packages (from spacy) (1.1.3)
Requirement already satisfied: srsly<3.0.0,>=2.4.3 in /usr/local/lib/python3.10/dist-packages (from spacy) (2.4.8)
Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in /usr/local/lib/python3.10/dist-packages (from spacy) (2.0.10)
Requirement already satisfied: weasel<0.5.0,>=0.1.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (0.4.1)
Requirement already satisfied: typer<4.0.0,>=3.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (0.12.5)
Requirement already satisfied: tqdm<4.60.0,>=4.38.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (4.66.6)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (2.32.3)
Requirement already satisfied: pydantic<1.8.1,>=1.8.1,<3.0.0,>=1.7.4 in /usr/local/lib/python3.10/dist-packages (from spacy) (2.9.2)
Requirement already satisfied: Jinja2 in /usr/local/lib/python3.10/dist-packages (from spacy) (2.11.4)
Requirement already satisfied: setuptools in /usr/local/lib/python3.10/dist-packages (from spacy) (75.3.0)
Requirement already satisfied: packaging>20.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (24.1)
Requirement already satisfied: langcodes<4.0.0,>=3.2.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (3.4.1)
Requirement already satisfied: numpy>=1.19.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (1.26.4)
Requirement already satisfied: language-data<1.2 in /usr/local/lib/python3.10/dist-packages (from langcodes<4.0.0,>=3.2.0->spacy) (1.2.0)
Requirement already satisfied: annotated-types<0.6.0 in /usr/local/lib/python3.10/dist-packages (from pydantic<1.8.1,>=1.8.1,<3.0.0,>=1.7.4->spacy) (0.7.0)
36s conclusão: 13:05
```



```
colab.research.google.com/drive/1C1WDmMdwmlhUZY7LGGbH7UULpISq3rdw?authuser=3#scrollTo=BzACUo5Yz2Wq

sentiment.ipynb
Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda

+ Código + Texto

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]);"

input_polarity=doc._.polarity
sentiment = {'score': input_polarity}
print(sentiment)

"Passo 6: Definindo a lista de tweets a serem analisadas;"

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. Atletico Madrid and Ar",
    "Gary Speed v Blackburn at St James in 2001/02 anyone? #NUFC #BEL #JAP #WorldCup",
    "ChelseaFC Don't make his regret it and start his over Moorli!",
    "LiverpoolFC @AnfieldEdition He's a liar, made up. I've unfollowed him as loads of others have. Pure blagger. #LFC",
    "Gthesk @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?",
    "@hasanshabbaz19 @LFC My knowledge has decreased somewhat in the past few seasons",
    "Report: Linked with @Everton and @holves. Italians set to sign £4.5m-rated winger",
    "Am seeing tweets that there's been a fall out @Everton between the money man... I'm presuming it's just a quiet news day on some kopite with nothing better to do! @BALANWERSMEDIA",
    "@LFC @officialAL20 @IntChampionsCup @ManUtd Expect loads of excuses after tonight's game",
    "Martindiamond17 @azryahmad @Aren_D @Matthewlewis1997 @lamheinthu @DMario @Alissonbecker @LFC @skySportsNews @skySport @OfficialASRoma I'm just fine I have your fanbase angry over stating facts should ask them hun. Xo",
    "What a weekend of football results! @ManUtd @clintonan @angersFC & 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 @cesarapi 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",
    "@westhamUtd we need to keep @CH1d and get @irvinglozano70 to compliment",
    "@kevdev9 @Everton Shouldn't be happening! Needs to stay away with his venomous attitude until he is sold!",
    "Football @gaguerosgiokun @ankcity What a genius. Pep taking winning mentality with him, conquering league after league. Baller",
    "@w120709 Can we get a RT for our #Ifc Mo Salah Liverpool Enamel Pin Badge"
```

```
colab.research.google.com/drive/1C1WDmMdwmlhUZY7LGGbH7UULpISq3rdw?authuser=3#scrollTo=P-BUY5INle6j

sentiment.ipynb
Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda Todas as alterações foram salvas

+ Código + Texto

"Passo 7: Analisando os tweets;"

for item in tweets:
    doc = nlp
    p(item)
    input_polarity = doc._.polarity
    sentiment = {
        'tweet': item,
        'score': input_polarity
    }
    print(sentiment)
```