Missão Prática | Nível 4 | Mundo 5

Dando inteligência ao software.

Aluna: Simone Ramos de Jesus.

Matricula: 202208290965.

Curso EAD: Desenvolvimento Full Stack.

Campus Virtual EAD: Polo Prado – Belo Horizonte – MG.

Ano: 2024.

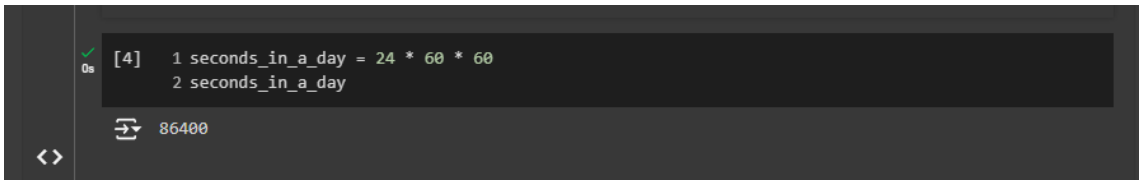
Microatividade 1

Configurar o ambiente Google Colab

- Procedimentos:

1. Acesse a página do Google Colab: https://colab.google
2. No menu superior direito, seleciona 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):



Microatividade 2

Descrever tarefas diversas relacionadas ao Processamento de Linguagem Natural

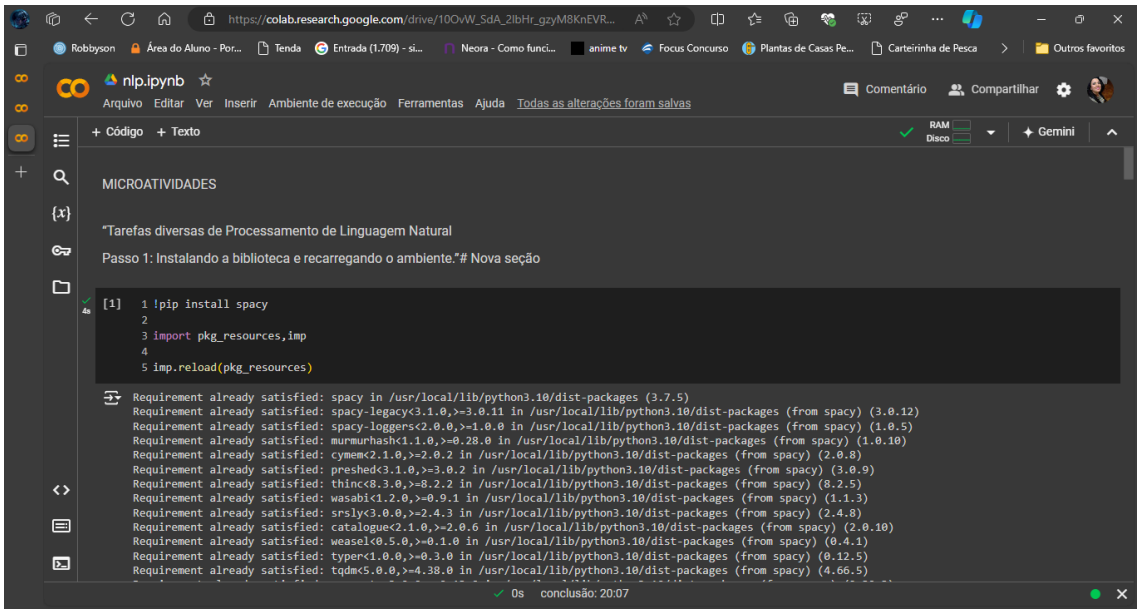
- Procedimentos:

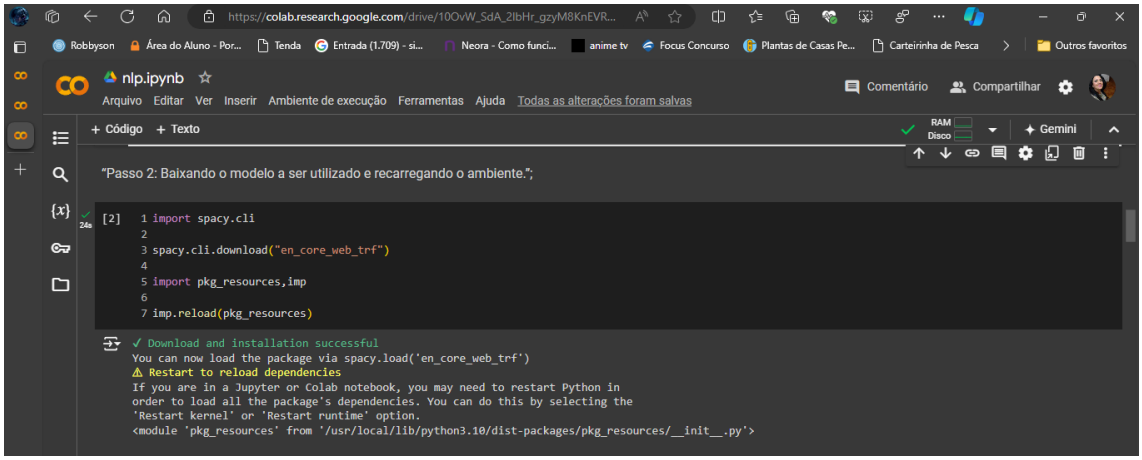
1. Estando logado no Google Colab, clique no menu “Arquivo” e selecione a opção “Novo notebook”;

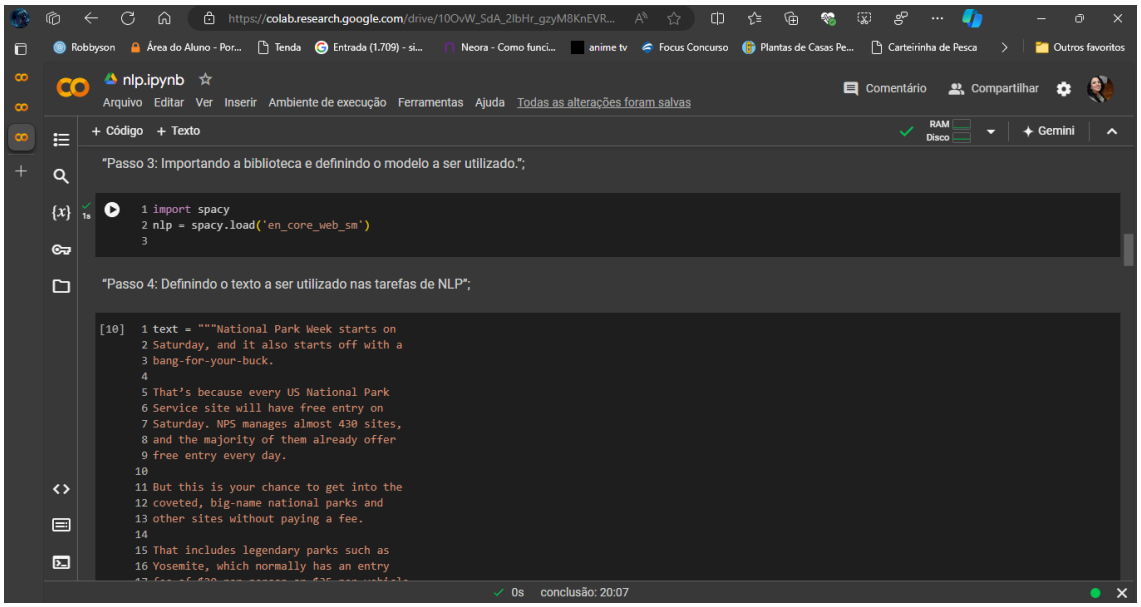
2. Na nova aba aberta no navegador, dê um nome ao seu notebook, clicando e alterando o nome automaticamente gerado

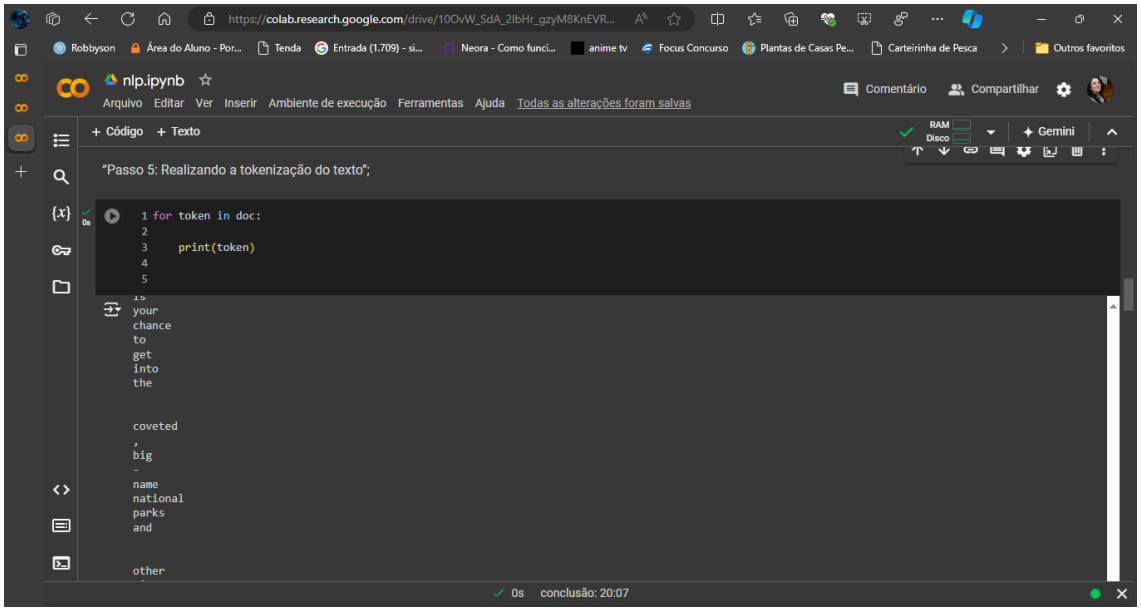
– Untitled0.ipynb – para nlp.ipynb;

3. Na janela de código, clique na opção “+Texto” (destacada no print abaixo) para inserir um bloco de texto;






nlp.ipynb
Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda Todas as alterações foram salvas
Comentário
Compartilhar
+ Código + Texto
RAM
Disco
Gemini
"Passo 3: Importando a biblioteca e definindo o modelo a ser utilizado.";
1 import spacy
2 nlp = spacy.load('en_core_web_sm')
3
"Passo 4: Definindo o texto a ser utilizado nas tarefas de NLP";
[10] 1 text = """National Park Week starts on
2 Saturday, and it also starts off with a
3 bang-for-your-buck.
4
5 That's because every US National Park
6 Service site will have free entry on
7 Saturday. NPS manages almost 430 sites,
8 and the majority of them already offer
9 free entry every day.
10
11 But this is your chance to get into the
12 coveted, big-name national parks and
13 other sites without paying a fee.
14
15 That includes legendary parks such as
16 Yosemite, which normally has an entry
0s conclusão: 20:07


nlp.ipynb
Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda Todas as alterações foram salvas
Comentário
Compartilhar
+ Código + Texto
RAM
Disco
Gemini
"Passo 5: Realizando a tokenização do texto";
1 for token in doc:
2
3 print(token)
4
5
your
chance
to
get
into
the
coveted
,
big
-
name
national
parks
and
other
0s conclusão: 20:07

nlp.ipynb

"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_)
```

```
 SPACE _SP
Yosemite PROPN NNP
, PUNCT ,
which PRON WDT
normally ADV RB
has VERB VBZ
an DET DT
entry NOUN NN

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

0s conclusão: 20:07

nlp.ipynb

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

```
for token in doc:

    print(token, token.morph)
```

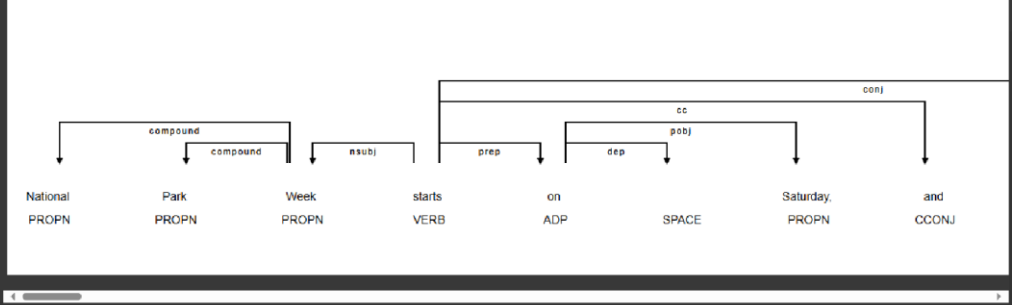
```
Saturday Number=Sing
, PunctType=Comm
and ConjType=Cmp
it Case=Nom|Gender=Neut|Number=Sing|Person=3|PronType=Prs
also
starts Number=Sing|Person=3|Tense=Pres|VerbForm=Fin
off
with
a Definite=Ind|PronType=Art


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

0s conclusão: 20:07

nlp.ipynb

"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})
```



0s conclusão: 20:07

Microatividade 3

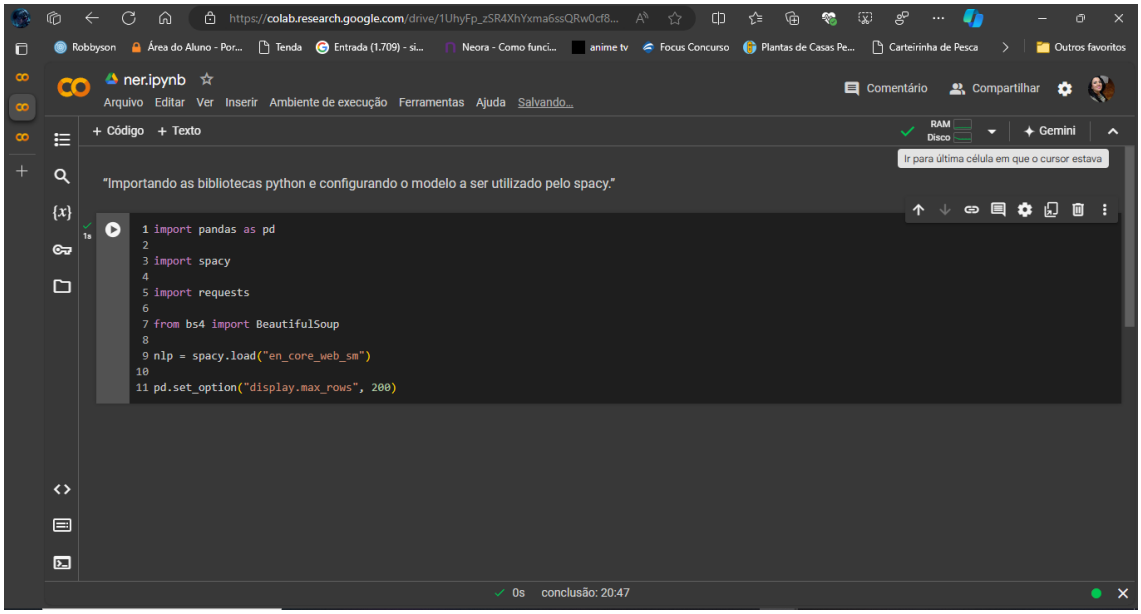
Descrever o processo de identificação de entidades a partir de textos.

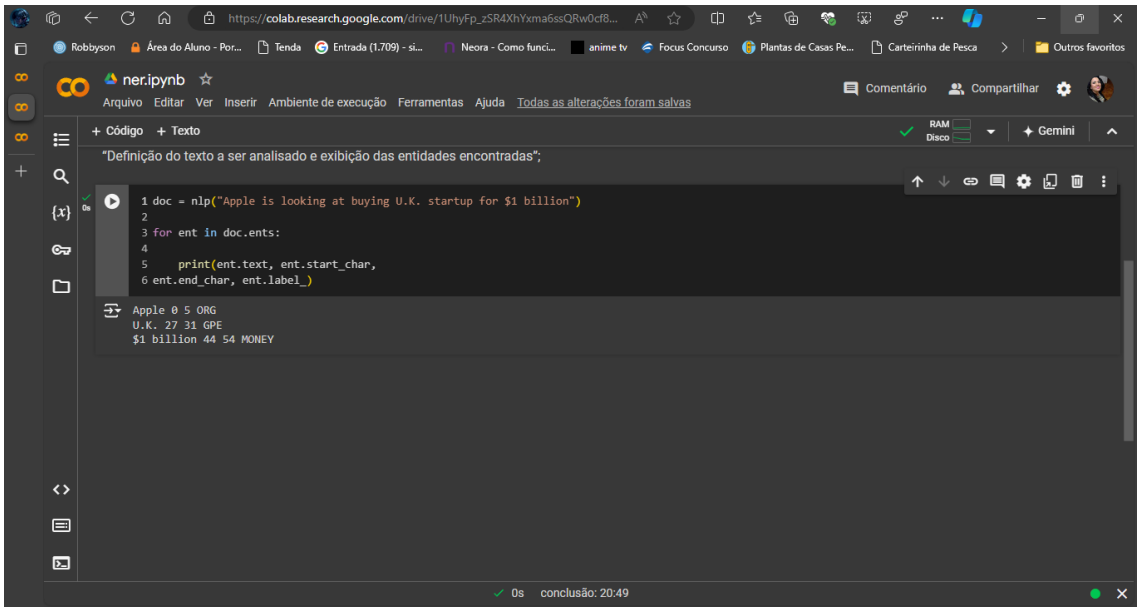
- Procedimentos:

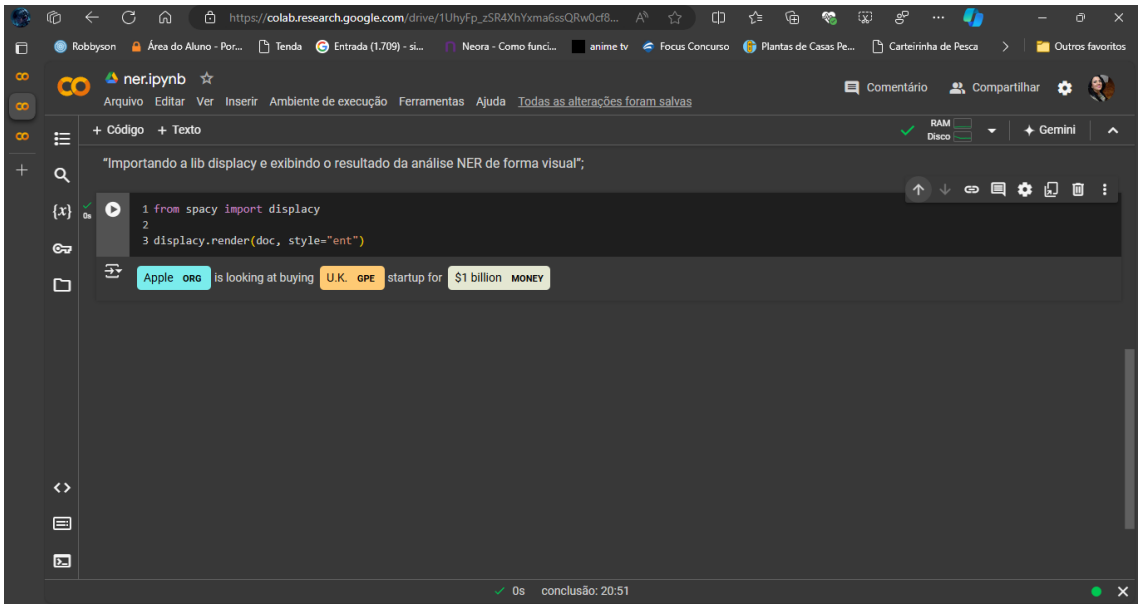
1. Estando logado no Google Colab, clique no menu “Arquivo” e selecione a opção “Novo notebook”;

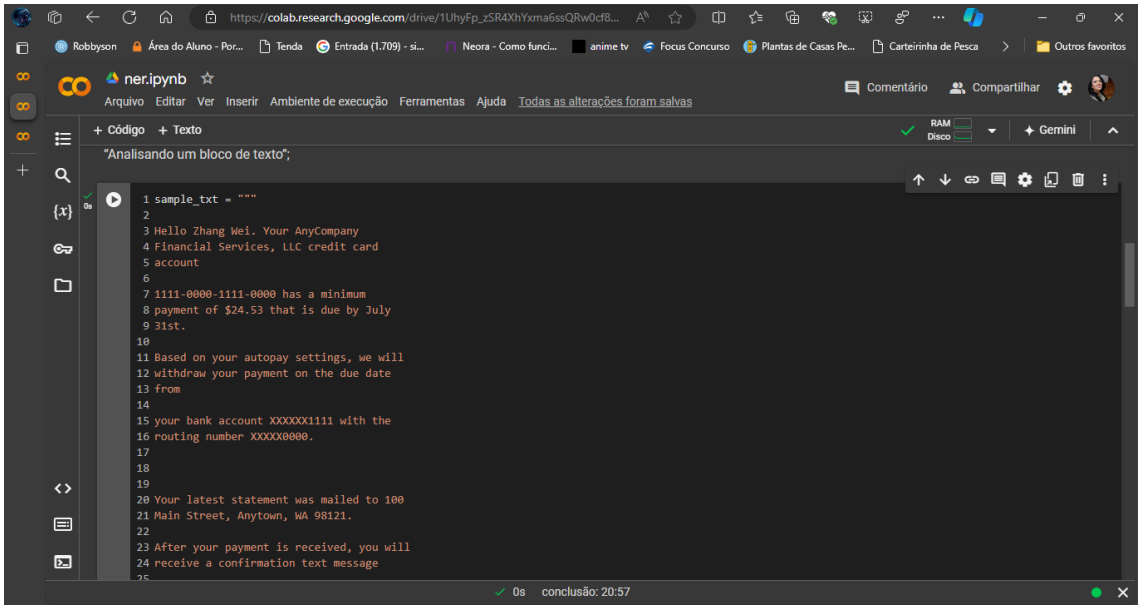
2. Na nova aba aberta no navegador, dê um nome ao seu notebook, clicando e alterando o nome automaticamente gerado – Untitled0.ipynb – para ner.ipynb;

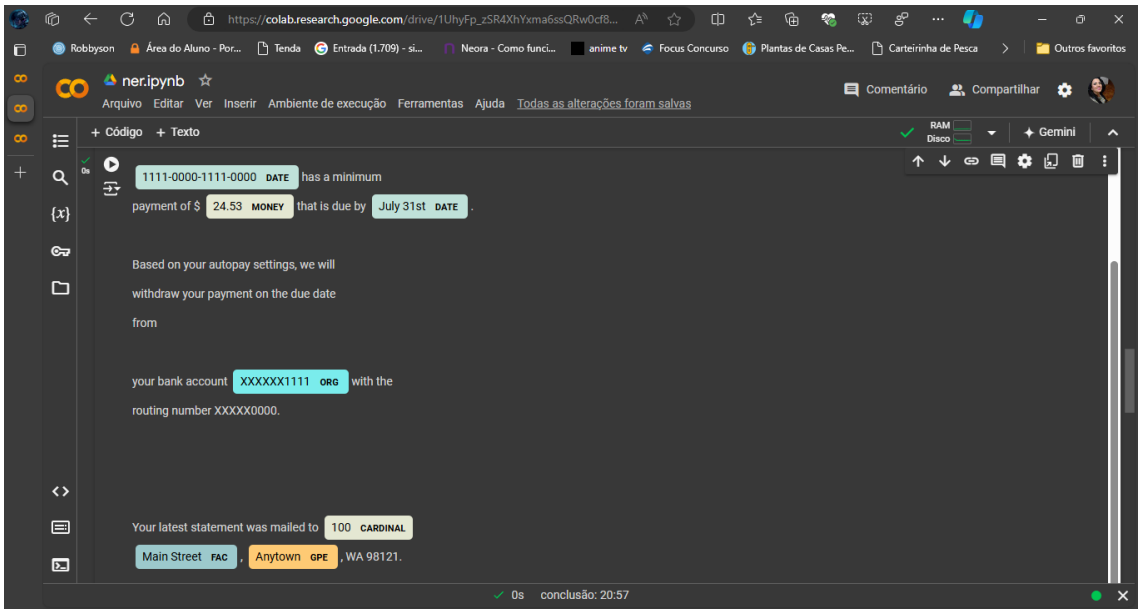
3. Na janela de código, clique na opção “+Texto” (destacada no print abaixo) para inserir um bloco de texto;

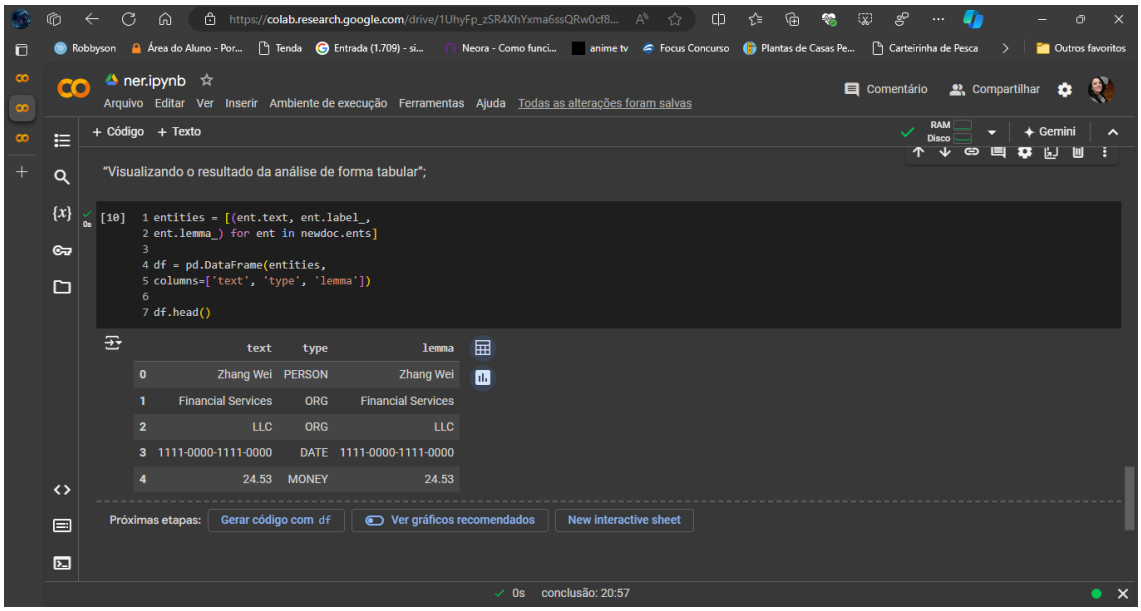


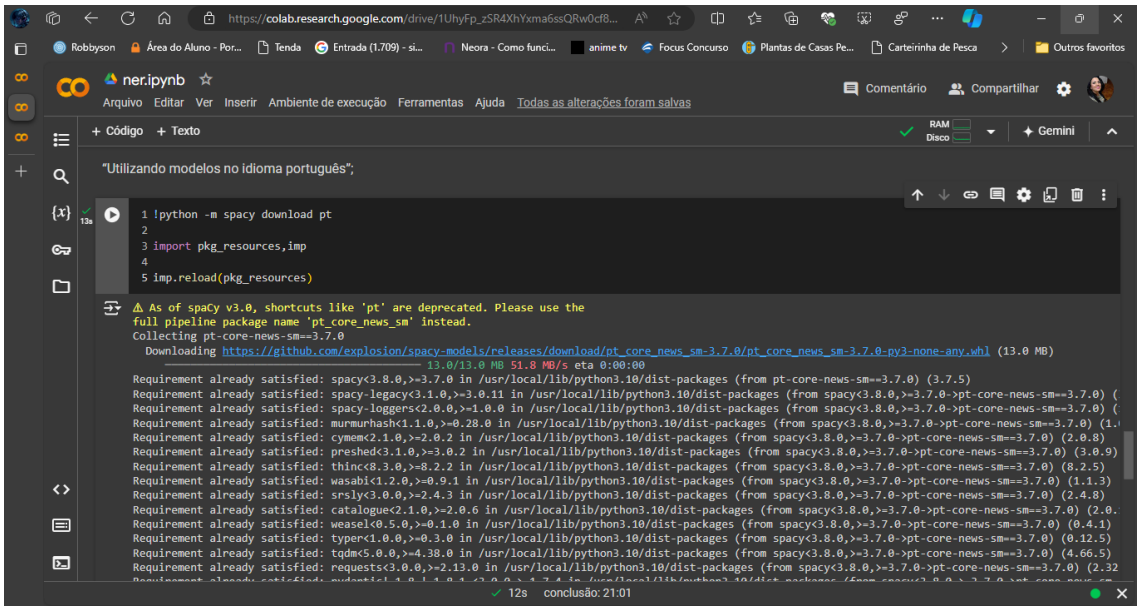


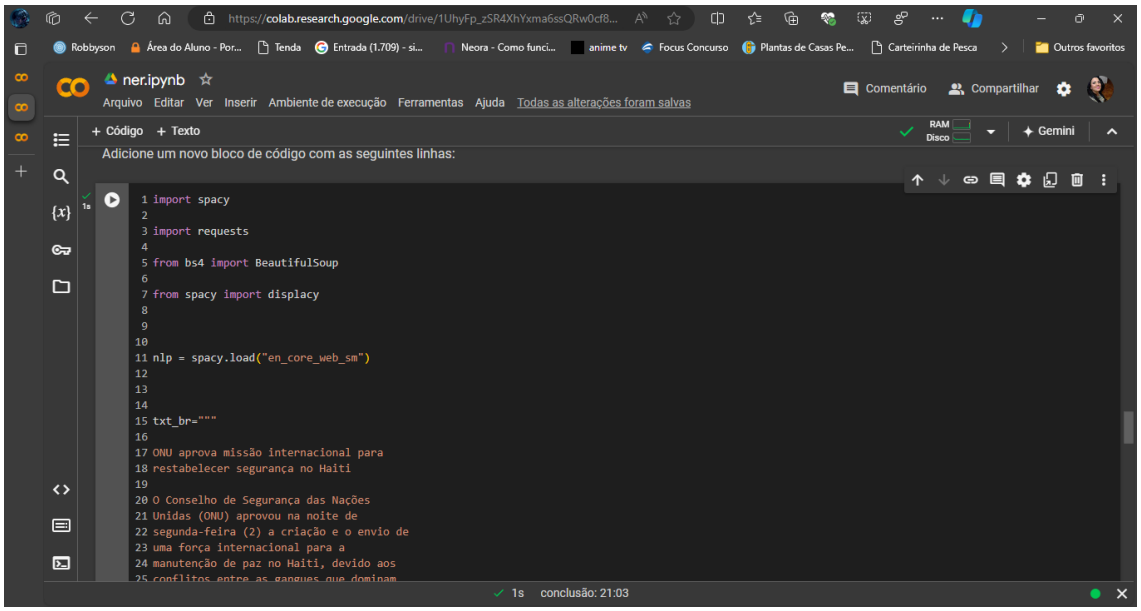

ner.ipynb
Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda Todas as alterações foram salvas
Comentário Compartilhar
+ Código + Texto
Gemini
"Importando a lib displacy e exibindo o resultado da análise NER de forma visual";
1 from spacy import displacy
2
3 displacy.render(doc, style="ent")
Apple ORG is looking at buying U.K. GPE startup for $1 billion MONEY
0s conclusão: 20:51

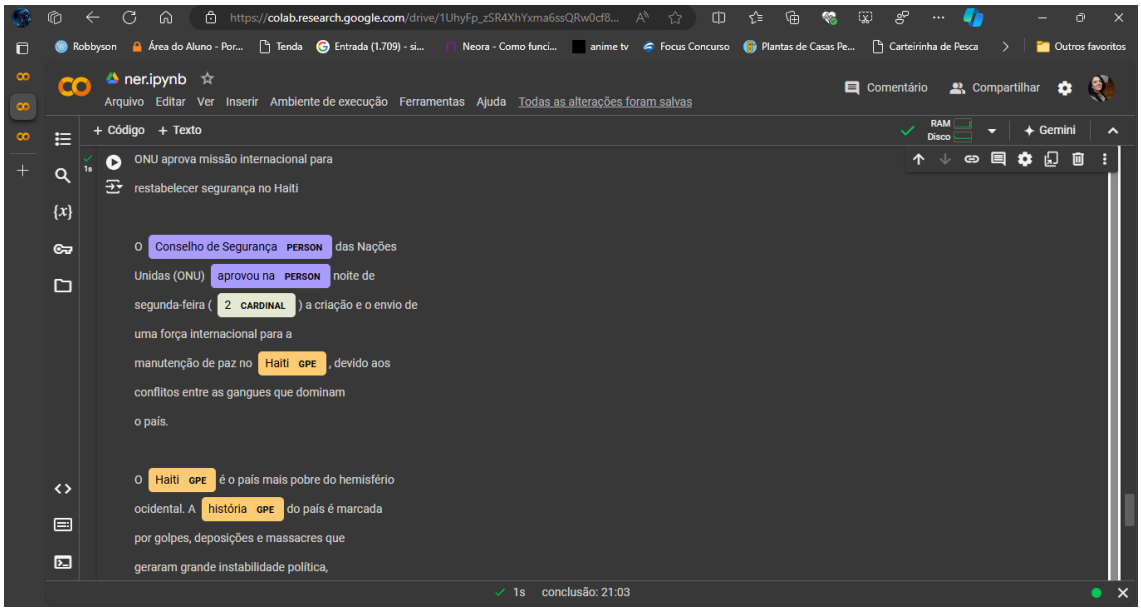

ner.ipynb
Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda Todas as alterações foram salvas
Comentário Compartilhar
+ Código + Texto
Gemini
"Analisando um bloco de texto";
1 sample_txt = """
2
3 Hello Zhang Wei. Your AnyCompany
4 Financial Services, LLC credit card
5 account
6
7 1111-0000-1111-0000 has a minimum
8 payment of $24.53 that is due by July
9 31st.
10
11 Based on your autopay settings, we will
12 withdraw your payment on the due date
13 from
14
15 your bank account XXXXXX1111 with the
16 routing number XXXXX0000.
17
18
19
20 Your latest statement was mailed to 100
21 Main Street, Anytown, WA 98121.
22
23 After your payment is received, you will
24 receive a confirmation text message
0s conclusão: 20:57


ner.ipynb
Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda Todas as alterações foram salvas
Comentário Compartilhar
+ Código + Texto
Gemini
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 , WA 98121.
0s conclusão: 20:57


ner.ipynb
Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda Todas as alterações foram salvas
Comentário
Compartilhar
+ Código + Texto
Gemini
"Visualizando o resultado da análise de forma tabular";
[10] 1 entities = [(ent.text, ent.label_,
2 ent.lemma_) for ent in newdoc.ents]
3
4 df = pd.DataFrame(entities,
5 columns=['text', 'type', 'lemma'])
6
7 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
0s conclusão: 20:57


ner.ipynb
Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda Todas as alterações foram salvas
Comentário
Compartilhar
+ Código + Texto
Gemini
"Utilizando modelos no idioma português";
1 !python -m spacy download pt
2
3 import pkg_resources,imp
4
5 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 51.8 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) (
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) (
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.
Requirement already satisfied: cymem<2.1.0,>=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) (2.0.8)
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) (2.4.8)
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.
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<5.0.0,>=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) (4.66.5)
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
12s conclusão: 21:01


ner.ipynb
Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda Todas as alterações foram salvas
Comentário
Compartilhar
+ Código + Texto
RAM
Disco
Gemini
Adicione um novo bloco de código com as seguintes linhas:
1 import spacy
2
3 import requests
4
5 from bs4 import BeautifulSoup
6
7 from spacy import displacy
8
9
10
11 nlp = spacy.load("en_core_web_sm")
12
13
14
15 txt_br="""
16
17 ONU aprova missão internacional para
18 restabelecer segurança no Haiti
19
20 O Conselho de Segurança das Nações
21 Unidas (ONU) aprovou na noite de
22 segunda-feira (2) a criação e o envio de
23 uma força internacional para a
24 manutenção de paz no Haiti, devido aos
25 conflitos entre as gangues que dominam
1s conclusão: 21:03


ner.ipynb
Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda Todas as alterações foram salvas
Comentário
Compartilhar
+ Código + Texto
RAM
Disco
Gemini
ONU aprova missão internacional para
restabelecer segurança no Haiti
O Conselho de Segurança PERSON das Nações
Unidas (ONU) aprovou na PERSON noite de
segunda-feira (2 CARDINAL) a criação e o envio de
uma força internacional para a
manutenção de paz no Haiti GPE , devido aos
conflitos entre as gangues que dominam
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,
1s conclusão: 21:03

Microatividade 4

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

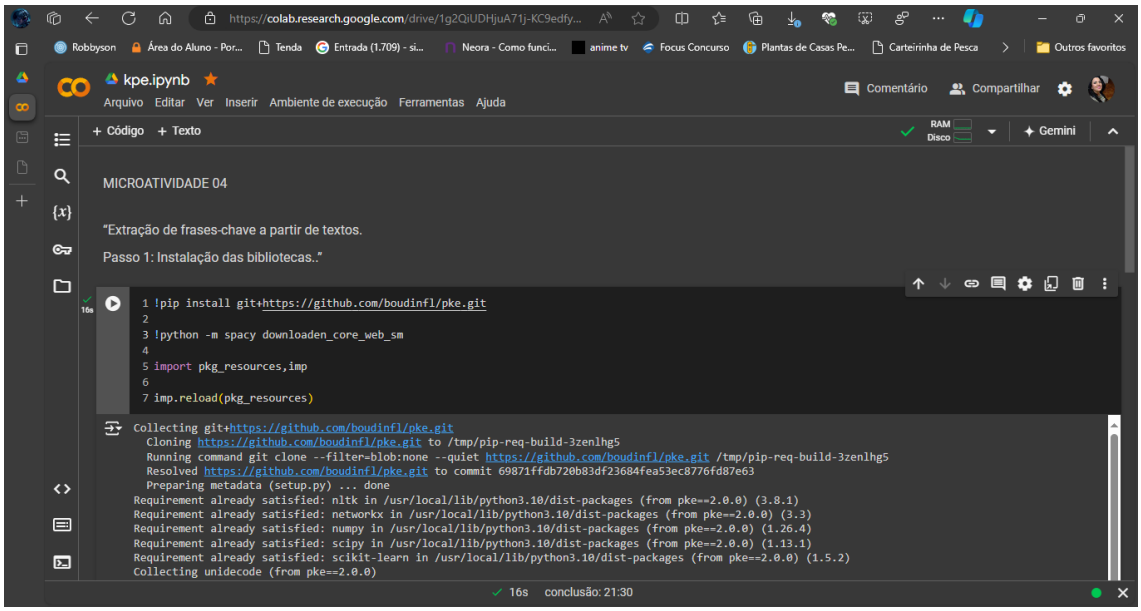
- Procedimentos:

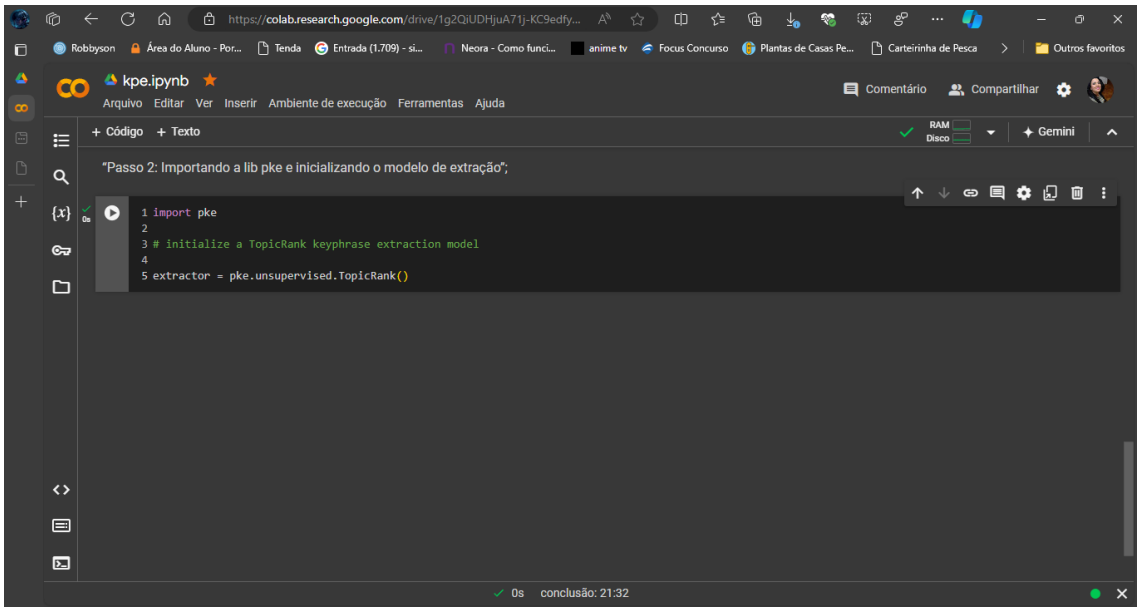
1. Estando logado no Google Colab, clique no menu “Arquivo” e selecione a opção “Novo notebook”;

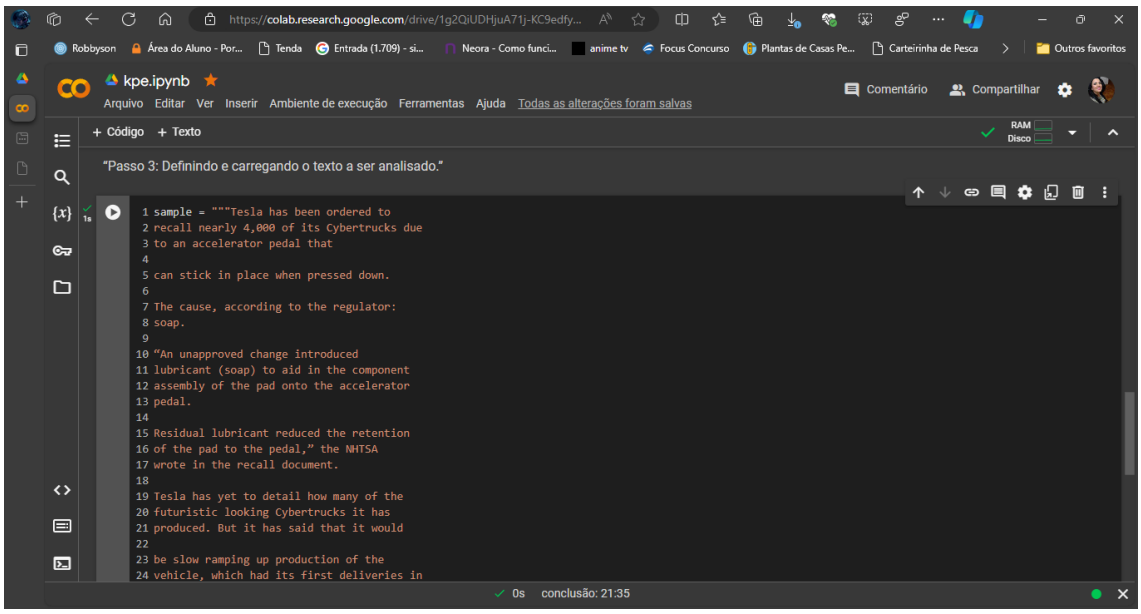
2. Na nova aba aberta no navegador, dê um nome ao seu notebook, clicando e alterando o nome automaticamente gerado

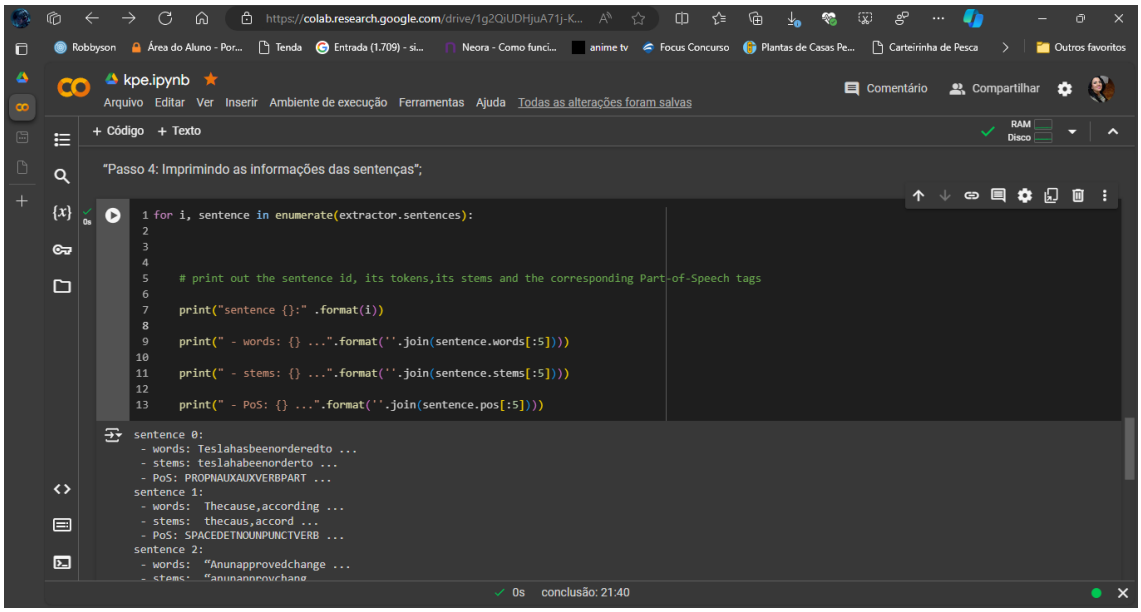
– Untitled0.ipynb – para kpe.ipynb;

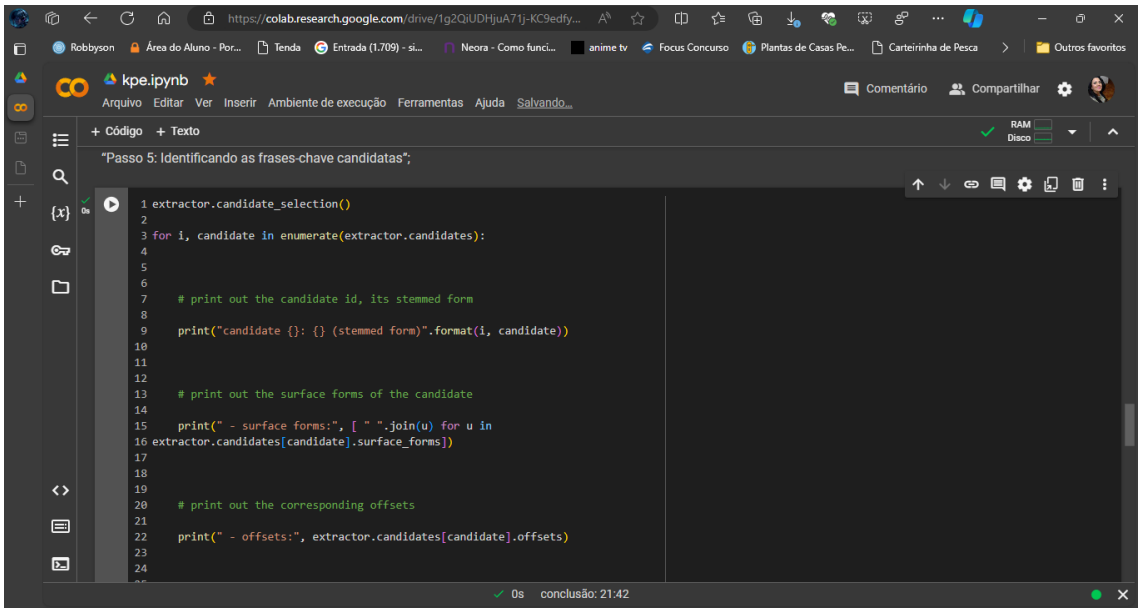
3. Na janela de código, clique na opção “+Texto” (destacada no print abaixo) para inserir um bloco de texto;

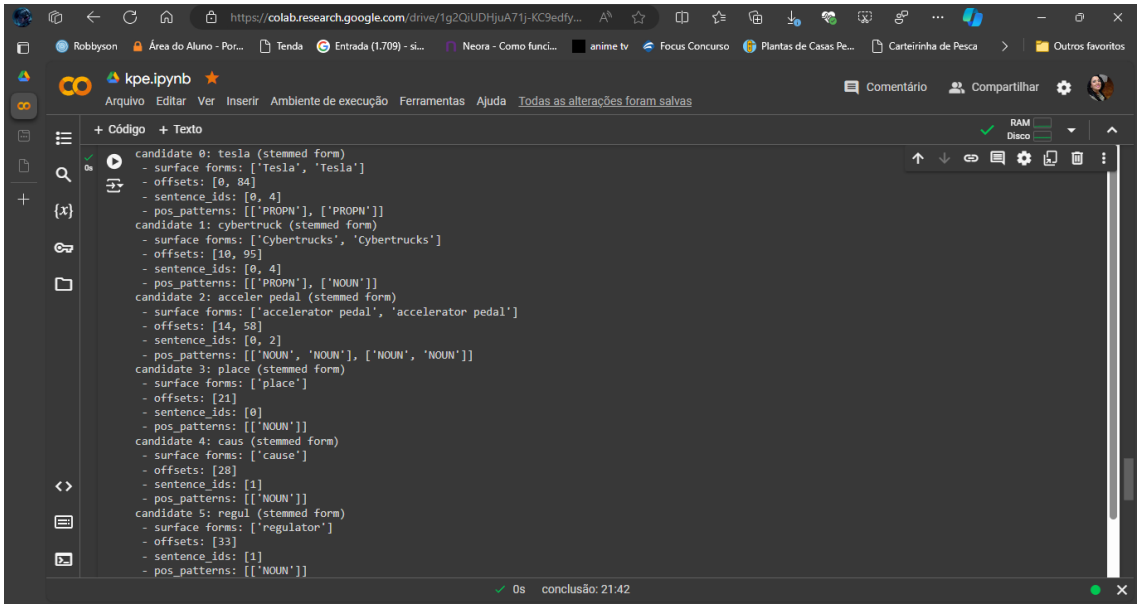


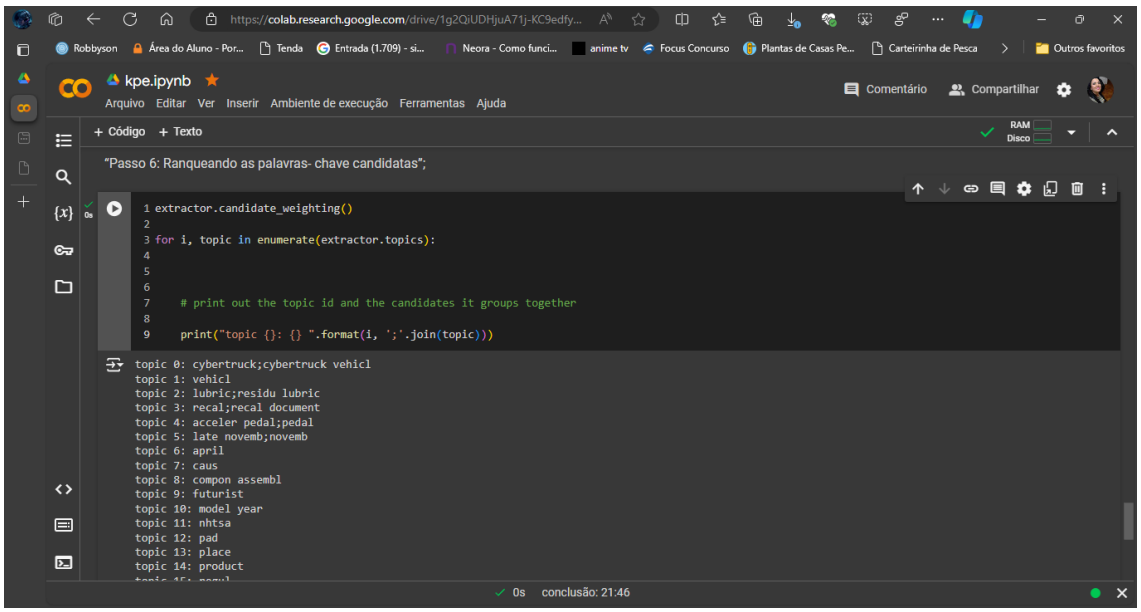

kpe.ipynb
Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda
Comentário
Compartilhar
+ Código + Texto
Gemini
"Passo 2: Importando a lib pke e inicializando o modelo de extração";
1 import pke
2
3 # initialize a TopicRank keyphrase extraction model
4
5 extractor = pke.unsupervised.TopicRank()
0s conclusão: 21:32


kpe.ipynb
Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda Todas as alterações foram salvas
Comentário
Compartilhar
+ Código + Texto
"Passo 3: Definindo e carregando o texto a ser analisado."
1 sample = """Tesla has been ordered to
2 recall nearly 4,000 of its Cybertrucks due
3 to an accelerator pedal that
4
5 can stick in place when pressed down.
6
7 The cause, according to the regulator:
8 soap.
9
10 “An unapproved change introduced
11 lubricant (soap) to aid in the component
12 assembly of the pad onto the accelerator
13 pedal.
14
15 Residual lubricant reduced the retention
16 of the pad to the pedal,” the NHTSA
17 wrote in the recall document.
18
19 Tesla has yet to detail how many of the
20 futuristic looking Cybertrucks it has
21 produced. But it has said that it would
22
23 be slow ramping up production of the
24 vehicle, which had its first deliveries in
0s conclusão: 21:35


kpe.ipynb
Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda Todas as alterações foram salvas
Comentário
Compartilhar
+ Código + Texto
"Passo 4: Imprimindo as informações das sentenças";
1 for i, sentence in enumerate(extractor.sentences):
2
3
4
5 # print out the sentence id, its tokens,its stems and the corresponding Part-of-Speech tags
6
7 print("sentence {}:" .format(i))
8
9 print(" - words: {} ...".format(''.join(sentence.words[:5])))
10
11 print(" - stems: {} ...".format(''.join(sentence.stems[:5])))
12
13 print(" - PoS: {} ...".format(''.join(sentence.pos[:5])))
sentence 0:
- words: Teslahasbeenorderedto ...
- stems: teslahabeenorderto ...
- PoS: PROPNAUXAUXVERBPART ...
sentence 1:
- words: Thecause,according ...
- stems: thecaus,accord ...
- PoS: SPACEDETNOUNPUNCTVERB ...
sentence 2:
- words: “Anunapprovedchange ...
0s conclusão: 21:40


kpe.ipynb
Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda Salvando...
Comentário
Compartilhar
+ Código + Texto
"Passo 5: Identificando as frases-chave candidatas";
1 extractor.candidate_selection()
2
3 for i, candidate in enumerate(extractor.candidates):
4
5
6
7 # print out the candidate id, its stemmed form
8
9 print("candidate {}: {} (stemmed form)".format(i, candidate))
10
11
12
13 # print out the surface forms of the candidate
14
15 print(" - surface forms:", [" ".join(u) for u in
16 extractor.candidates[candidate].surface_forms])
17
18
19
20 # print out the corresponding offsets
21
22 print(" - offsets:", extractor.candidates[candidate].offsets)
23
24
0s conclusão: 21:42


kpe.ipynb
Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda Todas as alterações foram salvas
Comentário
Compartilhar
+ Código + Texto
candidate 0: tesla (stemmed form)
- surface forms: ['Tesla', 'Tesla']
- offsets: [0, 84]
- sentence_ids: [0, 4]
- pos_patterns: [['PROPN'], ['PROPN']]
candidate 1: cybertruck (stemmed form)
- surface forms: ['Cybertrucks', 'Cybertrucks']
- offsets: [10, 95]
- sentence_ids: [0, 4]
- pos_patterns: [['PROPN'], ['NOUN']]
candidate 2: acceler pedal (stemmed form)
- surface forms: ['accelerator pedal', 'accelerator pedal']
- offsets: [14, 58]
- sentence_ids: [0, 2]
- pos_patterns: [['NOUN', 'NOUN'], ['NOUN', 'NOUN']]
candidate 3: place (stemmed form)
- surface forms: ['place']
- offsets: [21]
- sentence_ids: [0]
- pos_patterns: [['NOUN']]
candidate 4: caus (stemmed form)
- surface forms: ['cause']
- offsets: [28]
- sentence_ids: [1]
- pos_patterns: [['NOUN']]
candidate 5: regul (stemmed form)
- surface forms: ['regulator']
- offsets: [33]
- sentence_ids: [1]
- pos_patterns: [['NOUN']]
0s conclusão: 21:42


kpe.ipynb
Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda
Comentário
Compartilhar
+ Código + Texto
"Passo 6: Ranqueando as palavras- chave candidatas";
1 extractor.candidate_weighting()
2
3 for i, topic in enumerate(extractor.topics):
4
5
6
7 # print out the topic id and the candidates it groups together
8
9 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: nhtsa
topic 12: pad
topic 13: place
topic 14: product
0s conclusão: 21:46

Microatividade 5

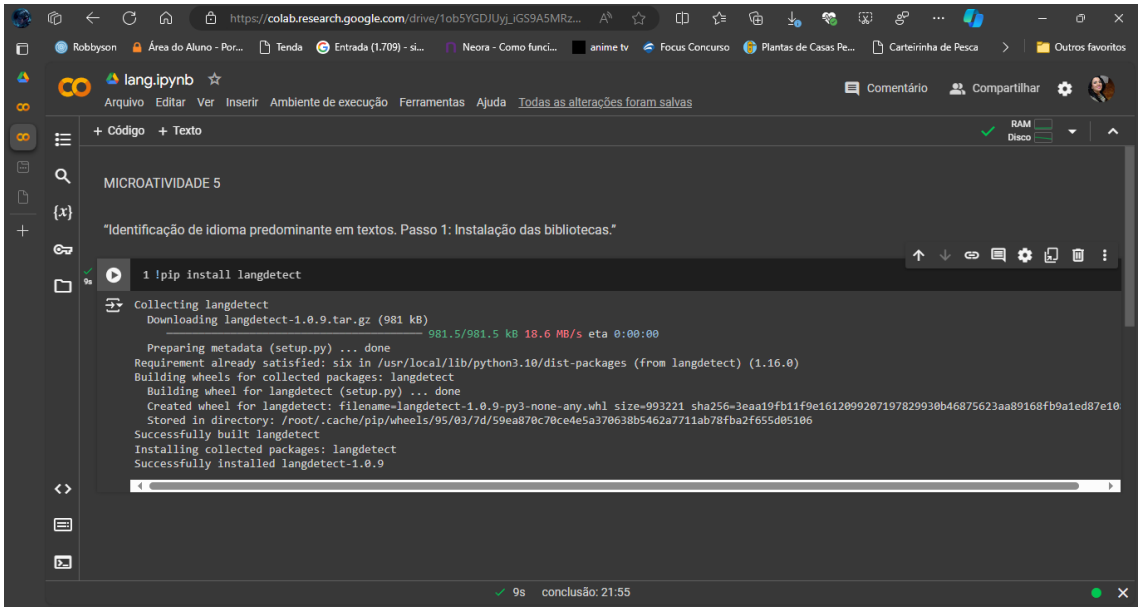
Descrever o processo de identificação de linguagem predominante a partir de textos

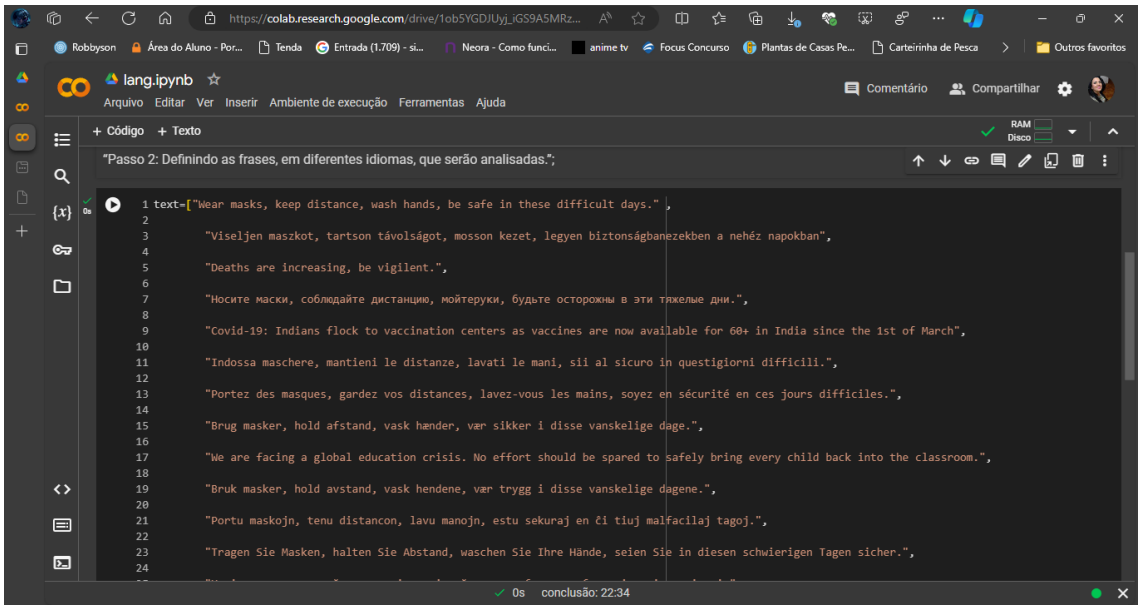
- Procedimentos

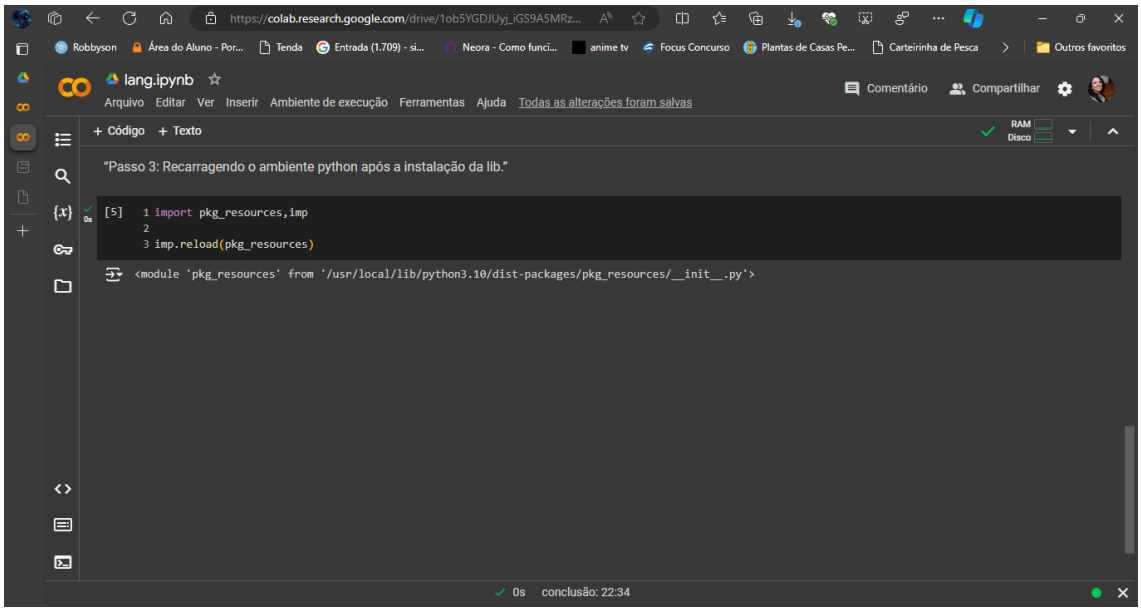
1. Estando logado no Google Colab, clique no menu “Arquivo” e selecione a opção “Novo notebook”;

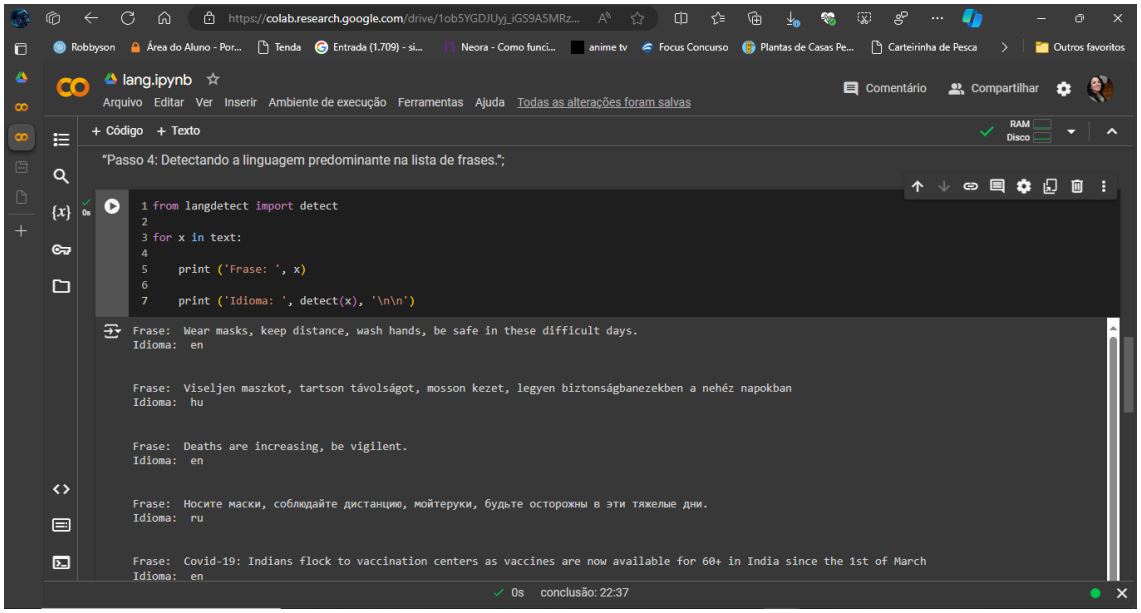
2. Na nova aba aberta no navegador, dê um nome ao seu notebook, clicando e alterando o nome automaticamente gerado – Untitled0.ipynb – para lang.ipynb;

3. Na janela de código, clique na opção “+Texto” (destacada no print abaixo) para inserir um bloco de texto;






lang.ipynb
Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda Todas as alterações foram salvas
Comentário
Compartilhar
+ Código + Texto
RAM
Disco
"Passo 3: Recarregando o ambiente python após a instalação da lib."
[5] 1 import pkg_resources,imp
2
3 imp.reload(pkg_resources)
<module 'pkg_resources' from '/usr/local/lib/python3.10/dist-packages/pkg_resources/__init__.py'>
0s conclusão: 22:34


lang.ipynb
Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda Todas as alterações foram salvas
Comentário
Compartilhar
+ Código + Texto
RAM
Disco
"Passo 4: Detectando a linguagem predominante na lista de frases.";
1 from langdetect import detect
2
3 for x in text:
4
5 print ('Frase: ', x)
6
7 print ('Idioma: ', detect(x), '\n\n')
Frase: Wear masks, keep distance, wash hands, be safe in these difficult days.
Idioma: en
Frase: Viseljen maszkot, tartson távolságot, mosson kezet, legyen biztonságbanezekben a nehéz napokban
Idioma: hu
Frase: Deaths are increasing, be vigilent.
Idioma: en
Frase: Носите маски, соблюдайте дистанцию, мойтеруки, 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
0s conclusão: 22:37