

Missão Prática – Mundo 05 – Nível 04

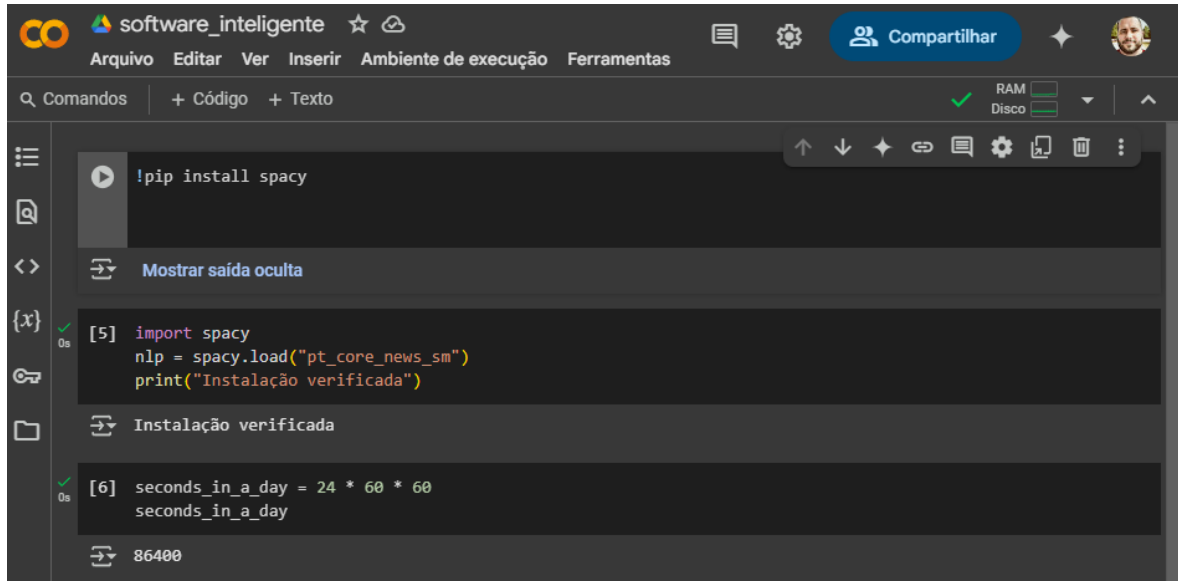
Gilvan Pereira de Oliveira – 2023.01.53256-61197

Polo Centro – São Lourenço Da Mata – PE

**RPG0033 – DANDO INTELIGÊNCIA AO SOFTWARE – 9001 – 2025.1**

<https://github.com/GilvanPOliveira/FullStack/tree/main/Mundo05/softwareInteligente>

### Micro atividade 1: Configurar o ambiente Google Colab



The screenshot shows a Google Colab notebook interface. The top bar includes the Google Colab logo, the notebook title 'software\_inteligente', and various icons for file management, editing, and sharing. Below the top bar, there are tabs for 'Comandos', '+ Código', and '+ Texto'. The main area of the notebook displays a code cell with the following content:

```
!pip install spacy
```

Below the code cell, there is a section for 'Mostrar saída oculta' (Show hidden output). This section contains two output blocks:

```
[5] import spacy
    nlp = spacy.load("pt_core_news_sm")
    print("Instalação verificada")
```

The output of the first cell is 'Instalação verificada' (Installation verified).

```
[6] seconds_in_a_day = 24 * 60 * 60
    seconds_in_a_day
```

The output of the second cell is '86400'.

- Criado um novo notebook no Google Colab, denominado: software\_inteligente;
- Configurando o Google Colab para trabalhar com Python;
- Instalado a biblioteca Spacy, que é uma biblioteca de processamento de linguagem natural (NLP) em Python;
- Verificando se a instalação da biblioteca foi feita corretamente;
- Executando o exemplo fornecido na micro atividade para complementar a execução da mesma.

## Micro atividade 2: Descrever tarefas diversas relacionadas ao Processamento de Linguagem Natural

**Tarefas diversas de Processamento de Linguagem Natural**

- Passo 1: Instalando a biblioteca e recarregando o ambiente;

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

Mostrar saída oculta

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

```
[3] import spacy.cli
spacy.cli.download("en_core_web_trf")
import pkg_resources,imp
imp.reload(pkg_resources)
```

Mostrar saída oculta

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

```
[4] import spacy
nlp = spacy.load('en_core_web_sm')
```

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

```
[5] 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;

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

National  
Park  
Week  
starts  
on  
Saturday  
,  
and  
it  
also

starts  
off  
with  
a  
bang  
-  
for

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

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

```
National PROPN NNP
Park PROPN NNP
Week PROPN NNP
starts VERB VBZ
on ADP IN
Saturday PROPN NNP
, PUNCT ,
and CONJ CC
it PRON PRP
also ADV RB
```

```
SPACE _SP
starts VERB VBZ
off ADP RP
with ADP IN
a DET DT
bang NOUN NN
- PUNCT HYPH
for ADP IN
- PUNCT HYPH
your PRON PRP$
- PUNCT HYPH
buck NOUN NN
. PUNCT .
```

```
SPACE _SP
That PRON DT
's VERB VBZ
because SCONJ IN
every DET DT
US PROPN NNP
National PROPN NNP
Park PROPN NNP
Service PROPN NNP
site NOUN NN
will AUX MD
have VERB VB
```

```
SPACE _SP
free ADJ JJ
```

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

```
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 Number=Sing
- PunctType=Dash
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
US Number=Sing
National Number=Sing
Park Number=Sing
Service Number=Sing
site Number=Sing
will VerbForm=Fin
have VerbForm=Inf
```

- 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})
```

Diagram illustrating the dependency parse tree for the sentence "National Park Week starts on Saturday,". The tree shows the following structure:

- starts** (VERB) is the root node.
- starts** is connected to **National** (PROPN) and **Park** (PROPN) via a **compound** relation.
- National** and **Park** are connected via a **compound** relation.
- starts** is connected to **Week** (PROPN) via a **nsubj** relation.
- starts** is connected to **on** (ADP) via a **prep** relation.
- on** is connected to **Saturday,** (PROPN) via a **pobj** relation.

- Passo 5: Realizando a tokenização do texto;

```
[6] for token in doc:
    print(token)
```

Mostrar saída oculta

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

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

Mostrar saída oculta

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

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

Mostrar saída oculta

- 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})
```

Mostrar saída oculta

### Micro atividade 3: Descrever o processo de identificação de entidades a partir de textos

The screenshot shows a Jupyter Notebook environment with the following content:

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

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

The screenshot shows a Jupyter Notebook environment with the following content:

Analisando um bloco de texto

```
[6] 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 , WA 98121 DATE .  
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

```

[7] 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

Próximas etapas: [Gerar código com df](#) [Ver gráficos recomendados](#) [New interactive sheet](#)

## Utilizando modelos no idioma português

```

[8] !python -m spacy download pt

import pkg_resources, imp

imp.reload(pkg_resources)

```

[Mostrar saída oculta](#)

```

[17] 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")

```

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.

#### Micro atividade 4: Descrever o processo de extração de frases-chave a partir de textos

Extração de frases-chave a partir de textos.

- Passo 1: Instalação das bibliotecas;

```
[1] !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)
```

Mostrar saída oculta

- Passo 2: Importando a lib pke e inicializando o modelo de extração;

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

- Passo 3: Definindo e carregando o texto a ser analisado;

```

1s 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')

```

- Passo 4: Imprimindo as informações das sentenças;

```

0s 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])))

```

```

🔍 sentence 0:
- words: Tesla has been ordered to ...
- stems: tesla ha been order to ...
- PoS: PROPON AUX AUX VERB PART ...
sentence 1:
- words: The cause , according ...
- stems: the caus , accord ...
- PoS: SPACE DET NOUN PUNCT VERB ...
sentence 2:
- words: " An unapproved change ...
- stems: " an unapprov chang ...
- PoS: SPACE PUNCT DET ADJ NOUN ...
sentence 3:
- words: Residual lubricant reduced the ...
- stems: residu lubric reduc the ...
- PoS: SPACE ADJ NOUN VERB DET ...
sentence 4:
- words: Tesla has yet to ...
- stems: tesla ha yet to ...
- PoS: SPACE PROPON AUX ADV PART ...
sentence 5:
- words: But it has said that ...
- stems: but it ha said that ...
- PoS: CCONJ PRON AUX VERB SCONJ ...
sentence 6:
- words: The NHTSA said the ...
- stems: the nhtsa said the ...
- PoS: SPACE DET PROPON VERB DET ...
sentence 7:
- words: ...
- stems: ...
- PoS: SPACE ...

```



- Passo 5: Identificando as frases-chave candidatas;

```

0s 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)
- surface forms: ['recall document']
- offsets: [80]
- sentence_ids: [3]
- pos_patterns: [['NOUN', 'NOUN']]
candidate 16: futurist (stemmed form)
- surface forms: ['futuristic']
- offsets: [93]
- sentence_ids: [4]
- pos_patterns: [['ADJ']]
candidate 17: slow (stemmed form)
- surface forms: ['slow']

```

- Passo 6: Ranqueando as palavras-chave candidatas;

```

0s extractor.candidate_weighting()

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

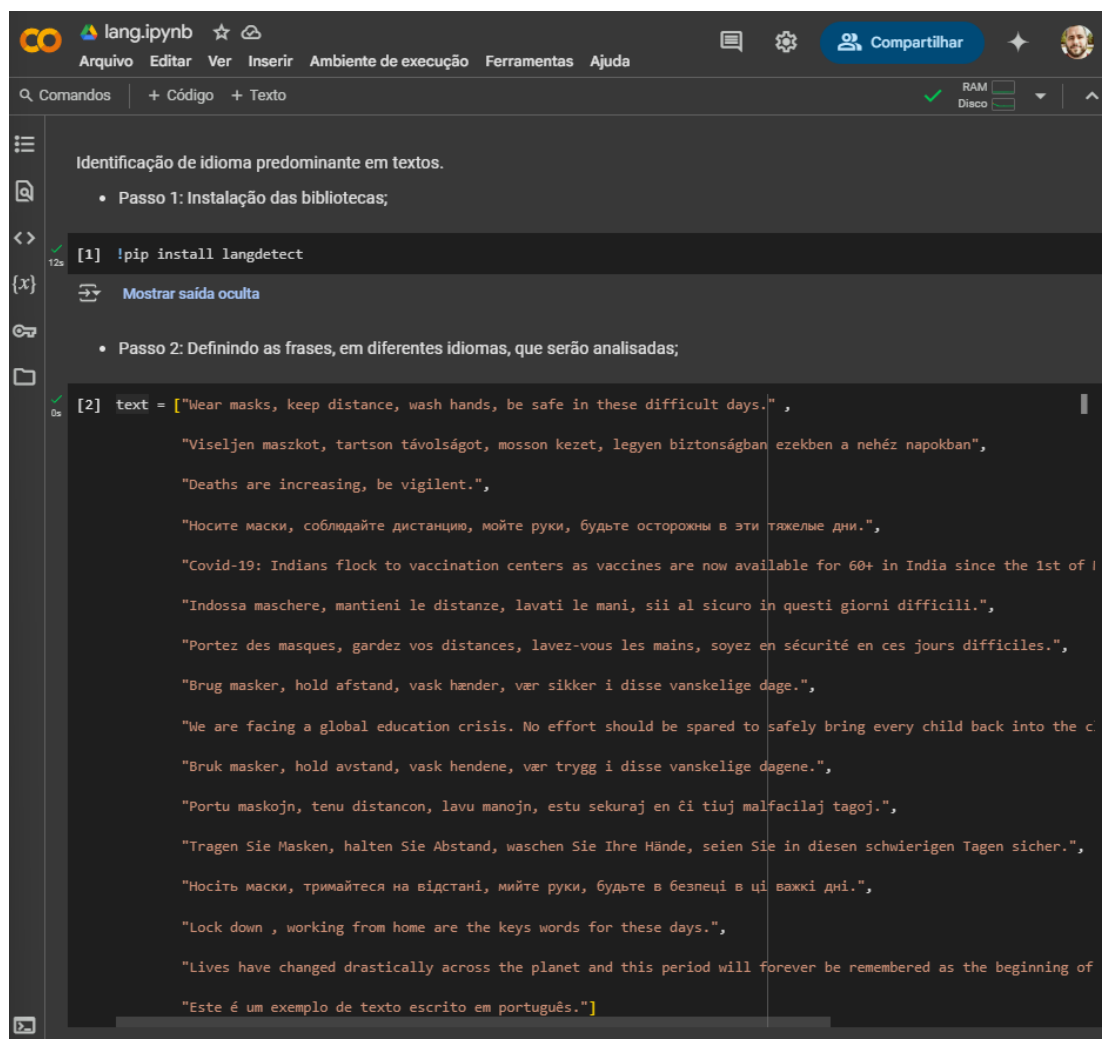
    # print out the topic id and the candidates it groups together

    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: nhntsa
topic 12: pad
topic 13: place
topic 14: product
topic 15: regul
topic 16: retent
topic 17: slow
topic 18: soap
topic 19: tesla
topic 20: unapprov chang

```

## Micro atividade 5: Descrever o processo de identificação de linguagem predominante a partir de textos



The screenshot shows a Jupyter Notebook titled "lang.ipynb" with a dark theme. The interface includes a top bar with navigation icons and a "Compartilhar" button. Below the bar, there are tabs for "Comandos", "Código", and "Texto". The notebook content is divided into two cells. The first cell, labeled "[1]", contains the command `!pip install langdetect` and a "Mostrar saída oculta" button. The second cell, labeled "[2]", contains a list of text samples in various languages, including English, Hungarian, German, Russian, Italian, French, Danish, Chinese, Portuguese, and a final line in Portuguese. The text samples are as follows:

```
[2] 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 I",  
            "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 c",  
            "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",  
            "Este é um exemplo de texto escrito em português."]
```

- Passo 3: Recarregando o ambiente python após a instalação da lib;

```
[3] import pkg_resources, imp
    imp.reload(pkg_resources)
```

Mostrar saída oculta

- Passo 4: Detectando a linguagem predominante na lista de frases;

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

Frase: Wear masks, keep distance, wash hands, be safe in these difficult days.  
Idioma: en

Frase: Viseljén maszkot, tartson távolságot, mosson kezet, legyen biztonságban ezekben a nehéz napokban  
Idioma: hu

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  
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

- Passo 5: Detectando a linguagem predominante na lista de frases, com mais idiomas.

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

Frase: Wear masks, keep distance, wash hands, be safe in these difficult days.  
Idioma: en

Frase: Viseljén maszkot, tartson távolságot, mosson kezet, legyen biztonságban ezekben a nehéz napokban.  
Idioma: hu

Frase: Les décès augmentent, soyez vigilants.  
Idioma: fr

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

Frase: Covid-19: Indios acuden en masa a los centros de vacunación ya que las vacunas están disponibles para mayores de  
Idioma: es

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

Frase: 戴口罩, 保持距离, 勤洗手, 在这些困难时期保持安全.  
Idioma: zh-cn

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

Frase: Estamos enfrentando uma crise global na educação. Nenhum esforço deve ser poupado para trazer todas as crianças  
Idioma: pt