



# Estácio

## Estácio - Mundo 5 - Missão Nível 4

Faculdade Estácio - Polo Itaipava - Petrópolis/RJ.

Curso: Desenvolvimento Full Stack.

Disciplina: Nível 4: Dando Inteligência ao Software

Semestre Letivo: 5.

Integrante: Jeferson Jones Smith da Rocha.

Repositório: <https://github.com/JefersonSmith/estacio-mundo5-nivel4>

# Dando Inteligência ao Software com Processamento de Linguagem Natural

## Introdução

O Processamento de Linguagem Natural (PLN) emerge como um campo fundamental da Inteligência Artificial (IA), capacitando as máquinas a compreender, interpretar e gerar linguagem humana de maneira significativa. No cenário tecnológico atual, onde a interação homem-máquina se torna cada vez mais prevalente e a quantidade de dados textuais cresce exponencialmente, as técnicas de PLN são indispensáveis. Ferramentas como assistentes virtuais, tradutores automáticos, sistemas de análise de sentimento e chatbots são apenas alguns exemplos de aplicações que transformaram a forma como interagimos com a tecnologia e acessamos informações.

Este trabalho acadêmico, referente à disciplina RPG0034 - Dando inteligência ao software, explora diversas facetas do PLN através de uma série de microatividades práticas e culmina em uma missão prática focada na análise de sentimentos. Utilizando o ambiente colaborativo Google Colab e bibliotecas Python proeminentes como Spacy, pke, langdetect e spacytextblob, o objetivo é fornecer uma compreensão aprofundada e aplicada das principais tarefas de PLN. Abordaremos desde a configuração do ambiente de desenvolvimento até a execução de tarefas complexas como identificação de entidades nomeadas (NER), extração de frases-chave, detecção de idioma e, finalmente, a análise de polaridade em textos de redes sociais.

A contextualização fornecida no material de orientação destaca a relevância de ferramentas de IA na nuvem, como AWS Comprehend e Google Cloud Natural Language API, mas opta por uma abordagem híbrida, combinando o ambiente Google Colab com bibliotecas Python. Essa escolha permite não apenas a familiarização com tecnologias de ponta acessíveis na nuvem, mas também a construção de um conhecimento sólido que pode ser replicado em ambientes de desenvolvimento locais. Ao longo das atividades, exploraremos conceitos como tokenização, etiquetagem de classes gramaticais (POS tagging), análise morfológica e sintática, preparando o terreno para a aplicação prática final.

A missão prática proposta simula um cenário real de análise de dados, onde uma empresa busca compreender a percepção pública sobre clubes de futebol ingleses através da análise de sentimentos expressos em tweets. Esta aplicação demonstra o poder do PLN em extrair insights valiosos de grandes volumes de texto não estruturado, uma habilidade crucial para profissionais de Data Science e áreas correlatas. Ao final deste trabalho, espera-se que o leitor tenha adquirido não apenas o conhecimento teórico, mas também a experiência prática necessária para aplicar técnicas de PLN em diversos contextos, utilizando ferramentas modernas e eficazes.

## Microatividade 1: Configurar o ambiente Google Colab

Conheça o Colab

Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda

Comandos + Código + Texto Copiar para o Drive

Conectar

### Vamos começar

O documento que você está lendo não é uma página da Web estática, mas sim um ambiente interativo chamado **notebook Colab** que permite escrever e executar código.

Por exemplo, aqui está uma **célula de código** com um breve script Python que calcula um valor, armazena-o em uma variável e imprime o resultado:

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

86400

Para executar o código na célula acima, clique nela e depois pressione o botão Play à esquerda do código ou use o atalho do teclado "Command/Ctrl+Enter". Para editar o código, basta clicar na célula e começar a editar.

As variáveis definidas em uma célula podem ser usadas mais tarde em outras células:

```
[ ] seconds_in_a_week = 7 * seconds_in_a_day
seconds_in_a_week
```

604800

Os notebooks do Colab permitem combinar **código executável** e **rich text** em um só documento, além de **imagens**, **HTML**, **LaTeX** e muito mais. Quando você cria seus próprios notebooks do Colab, eles são armazenados na sua conta do Google Drive. É possível compartilhar os notebooks do Colab facilmente com colegas de trabalho ou amigos e permitir que eles façam comentários ou até editem o documento. Para saber mais, consulte a [Visão Geral do Colab](#). Para criar um novo bloco do Colab, use o menu Arquivo acima ou acesse o seguinte: [criar um novo bloco do Colab](#).

Os notebooks do Colab são notebooks do Jupyter hospedados no Colab. Para saber mais sobre o projeto Jupyter, acesse [jupyter.org](#).

### Ciência de dados

Com o Colab, você pode aproveitar todo o potencial das conhecidas bibliotecas Python para analisar e ver dados. A célula de códigos abaixo

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

Conheça o Colab

Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda

Comandos + Código + Texto

Conectar

### Tarefas diversas de Processamento de Linguagem Natural

Passo 1: Instalando a biblioteca e recarregando o ambiente.

```
[ ] pip install spacy
import pkg_resources; import sys; reload(pkg_resources)
```

Requirement already satisfied: spacy in /usr/local/lib/python3.11/dist-packages (3.8.5)

Requirement already satisfied: spacy-legacy<1.0.0,>=1.0.11 in /usr/local/lib/python3.11/dist-packages (from spacy) (3.0.12)

Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in /usr/local/lib/python3.11/dist-packages (from spacy) (1.0.5)

Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/python3.11/dist-packages (from spacy) (1.0.12)

Requirement already satisfied: cytoolz<2.10.0,>=2.0.2 in /usr/local/lib/python3.11/dist-packages (from spacy) (2.0.11)

Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python3.11/dist-packages (from spacy) (3.0.9)

Requirement already satisfied: thinc<8.4.0,>=8.3.4 in /usr/local/lib/python3.11/dist-packages (from spacy) (8.3.0)

Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in /usr/local/lib/python3.11/dist-packages (from spacy) (1.1.3)

Requirement already satisfied: srsly<3.0.0,>=2.4.3 in /usr/local/lib/python3.11/dist-packages (from spacy) (2.5.1)

Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in /usr/local/lib/python3.11/dist-packages (from spacy) (2.0.10)

Requirement already satisfied: weasel<0.5.0,>=0.1.0 in /usr/local/lib/python3.11/dist-packages (from spacy) (0.4.1)

Requirement already satisfied: typer<1.0.0,>=0.3.0 in /usr/local/lib/python3.11/dist-packages (from spacy) (0.15.3)

Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in /usr/local/lib/python3.11/dist-packages (from spacy) (4.67.1)

Requirement already satisfied: numpy<1.19.0 in /usr/local/lib/python3.11/dist-packages (from spacy) (2.0.2)

Requirement already satisfied: requests<3.0.0,>=2.13.0 in /usr/local/lib/python3.11/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/python3.11/dist-packages (from spacy) (2.11.4)

Requirement already satisfied: Jinja2 in /usr/local/lib/python3.11/dist-packages (from spacy) (3.1.0)

Requirement already satisfied: setuptools in /usr/local/lib/python3.11/dist-packages (from spacy) (65.2.0)

Requirement already satisfied: packaging<20.0 in /usr/local/lib/python3.11/dist-packages (from spacy) (24.2)

Requirement already satisfied: langcodes<4.0.0,>=3.2.0 in /usr/local/lib/python3.11/dist-packages (from spacy) (3.5.0)

Requirement already satisfied: language-data<1.2 in /usr/local/lib/python3.11/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/python3.11/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<2.33.2 in /usr/local/lib/python3.11/dist-packages (from pydantic<1.8.1,>=1.8.1, <3.0.0,>=1.7.4->spacy) (2.33.2)

Requirement already satisfied: typing-extensions<4.12.2 in /usr/local/lib/python3.11/dist-packages (from pydantic<1.8.1,>=1.8.1, <3.0.0,>=1.7.4->spacy) (4.13.2)

Requirement already satisfied: typing-inspection<0.4.0 in /usr/local/lib/python3.11/dist-packages (from pydantic<1.8.1,>=1.8.1, <3.0.0,>=1.7.4->spacy) (0.4.0)

Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.11/dist-packages (from requests<3.0.0,>=2.13.0->spacy) (3.4.2)

Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-packages (from requests<3.0.0,>=2.13.0->spacy) (3.10)

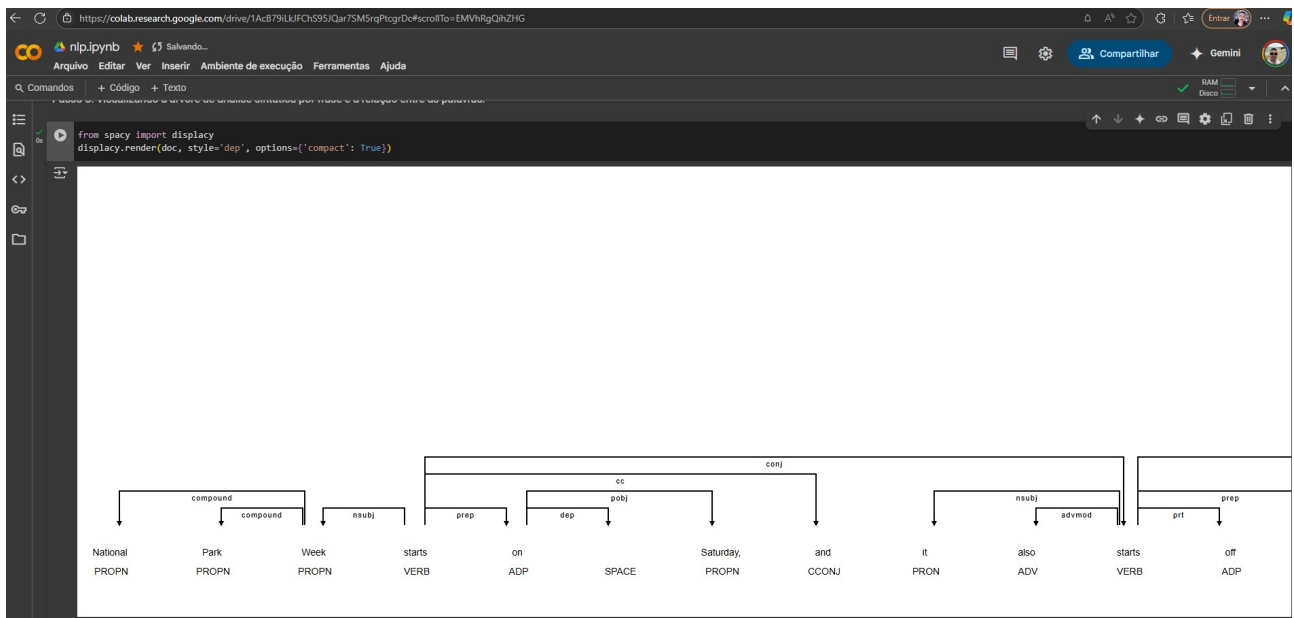
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11/dist-packages (from requests<3.0.0,>=2.13.0->spacy) (2.4.0)

Requirement already satisfied: certifi<2017.4.17 in /usr/local/lib/python3.11/dist-packages (from requests<3.0.0,>=2.13.0->spacy) (2025.4.26)

Requirement already satisfied: blis<1.4.0,>=1.3.0 in /usr/local/lib/python3.11/dist-packages (from thinc<8.4.0,>=8.3.4->spacy) (1.3.0)

Requirement already satisfied: confection<1.0.0,>=0.0.1 in /usr/local/lib/python3.11/dist-packages (from thinc<8.4.0,>=8.3.4->spacy) (0.1.5)

Requirement already satisfied: click<8.0.0 in /usr/local/lib/python3.11/dist-packages (from typer<1.0.0,>=0.3.0->spacy) (8.2.0)



**Microatividade 3: Descrever o processo de identificação de entidades a partir de textos**

The screenshot shows a Jupyter Notebook interface with the following code in the cell:

```
[1] Importando as bibliotecas python e configurando o modelo a ser utilizado pelo spacy.
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)

[2] Definição do texto a ser analisado e exibição das entidades encontradas
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_)

[3] Importando a lib displacy e exibindo o resultado da análise NER de forma visual
from spacy import displacy
displacy.render(doc, style="ent")
```

The output shows the text "Apple is looking at buying U.K. startup for \$1 billion" with the following entities identified: Apple (ORG), U.K. (GPE), \$1 billion (MONEY), and MONEY (MONEY).

The screenshot shows a Jupyter Notebook interface with the following code in the cell:

```
sample_text = """
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
"""

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

The output shows the text "ONU aprova missão internacional para restabelecer segurança no Haiti. O Conselho de Segurança das Nações Unidas aprovou, por unanimidade, a criação da Missão de Estabilização do Haiti 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". The entities identified are: ONU (ORG), Haiti (LOC), Conselho de Segurança das Nações Unidas (ORG), MINUSTAH (ORG), Jean Bertrand Aristide (PER), and Brasil (LOC).

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

```
https://colab.research.google.com/drive/1g_77-derVIEF9p8IEeGlvPKVN7pBaXC#scrollTo=77KHpRckWRP

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

Comandos + Código + Texto

- sentence_ids: [6]
- pos_patterns: [['PROPN']]
candidate 25: april (stemmed form)
- surface forms: ['April']
- offsets: [146]
- sentence_ids: [6]
- pos_patterns: [['PROPN']]

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: cybertuck;cybertuck vehicl
topic 1: vehicl
topic 2: lubric;residu lubric
topic 3: recal;recal document
topic 4: accelar 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
topic 15: regul
topic 16: retent
topic 17: slow
topic 18: soap
topic 19: tesla
topic 20: unapprov chang
```

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

```
https://colab.research.google.com/drive/18Q480G49mby7nDOvq5fA8GluwKnLpG#scrollTo=C4u-D4lioXJ

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

Comandos + Código + Texto

<module 'pkg_resources' from '/usr/local/lib/python3.11/dist-packages/pkg_resources/__init__.py'>

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: 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

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

# Missão Prática Final: Dando Inteligência ao Software

 missao\_pratica.ipynb

Arquivo Editar Ver Inserir Ambiente de execução Ferramentas Ajuda

Comandos + Código + Texto Copiar para o Drive Conectar

Passo 7: Analisando os tweets

+ Código + Texto

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

```
{
  "tweet": "Bayer Leverkusen goalkeeper Bernd Leno will not be going to Napoli. His agent Uli Ferber to Bild: I can confirm that there were negotiations with Napoli, which we have broken off. Napoli is not an option. Atletico Madrid",
  "score": 0.0
},
{
  "tweet": "Gary Speed v Blackburn at St James in 2001/02 anyone? #MUFC #BEL #3AP #WorldCup", "score": 0.0
},
{
  "tweet": "@ChelseaFC Don't make him regret it and start him over Hoofli", "score": 0.0
},
{
  "tweet": "@LiverpoolFF @AnfieldEdition He's a liar, made up. I've unfollowed him as loads of others have. Pure blagger. #LFC", "score": 0.21428571428571427
},
{
  "tweet": "@heshk @everton Didn't realise Kewright is due to leave at the end of the month. In all seriousness could you see him being interested in us?", "score": 0.0625
},
{
  "tweet": "@hasanshabbaz19 @LFC My knowledge has decreased somewhat in the past few seasons", "score": -0.28333333333333334
},
{
  "tweet": "Report: Linked with #Everton and #Wolves, Italians set to sign €4.5m-rated winger", "score": 0.0
},
{
  "tweet": "Am seeing tweets that there's been a fall out @everton between the money men... I'm presuming it's just a quiet news day or some kopite with nothing better to do! @ALANWYERSMEDIA", "score": 0.3125
},
{
  "tweet": "@LFC @OfficialAL20 @intchampionsCup @ManUtd Expect loads of excuses after tonight's game", "score": -0.4
},
{
  "tweet": "@MartinDiamond17 @azryahmad @Aren_D @matthewlewis1997 @iamheinthu @DlMarzio @Alissonbecker @LFC @skySportsNews @skySport @OfficialASRoma I'm just fine I have your fanbase angry over stating facts should ask them hun. Xo",
  "score": 0.0
},
{
  "tweet": "What a weekend of football results! @ManUtd @Glentoran @RangersFC & Hearts ???", "score": 0.0
},
{
  "tweet": "@ChelseaFC for the first time in a long while, my heart was relaxed while watching Chelsea. Really enjoyed it today. Come on, CHELSEA!!!", "score": 0.39218749999999997
},
{
  "tweet": "@ChelseaFC @cesarazpi what a fantastic signing worth every single penny ??", "score": 0.28952388952388953
},
{
  "tweet": "Pogba scores, Pogba assists. But tomorrow papers won't be telling you this, instead they will tell you how he'll end up at Juve because he's unhappy, frustrated, have grudges with Mourinho and so on and so forth #muFC",
  "score": 0.0
},
{
  "tweet": "@MestliamUtd we need to keep @CH14 and get @IrvingLozano70 to compliment", "score": 0.0
},
{
  "tweet": "@kevde99 @everton Shouldn't be happening! Needs to stay away with his venomous attitude until he is sold!", "score": 0.0
},
{
  "tweet": "@FootballBagueosergiokun @Kancity what a genius. Pro taking winning mentality with him, conquering league after league. Baller", "score": 0.5
},
{
  "tweet": "@#20709 Can we get a RT for our #lfc Mo Salah Liverpool Enamel Pin Badge", "score": 0.0
}
```

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