

Turma: 9001

Matrícula: 202211381232

Aluno: Marco Sergio Albino Vittorio Barozzi Disciplina : Dando Inteligência ao Software

Código:RPG0034

Missão Prática | Dando Inteligência ao Software

- Objetivos

Este projeto tem como objetivo os seguintes pontos :

- 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



Turma: 9001

Matrícula: 202211381232

Aluno: Marco Sergio Albino Vittorio Barozzi Disciplina : Dando Inteligência ao Software

Código:RPG0034

Missão Prática | Dando Inteligência ao Software

- Procedimentos

Os procedimentos das micro atividades assim como a missão pratica podem ser encontrados no link abaixo.

Link da missão https://sway.cloud.microsoft/s/4I0FaBWMTOk9e6Dg/embed

O foco aqui será apenas nos itens principais do procedimento da missão pratica



Turma: 9001

Matrícula: 202211381232

Aluno: Marco Sergio Albino Vittorio Barozzi Disciplina: Dando Inteligência ao Software

Código:RPG0034

Missão Prática | Dando Inteligência ao Software

- Pontos de veificação

6. Insira um bloco de código com o conteúdo abaixo:

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

import pkg_resources,imp
imp.reload(pkg_resources)

Durante a instalação das bibliotecas no Google Colab, surgiram problemas relacionados a conflitos de dependências, especialmente com pacotes como 'numpy', 'spacy', e o modelo 'en_core_web_sm'. Além disso, foi emitido um aviso sobre a ausência do pacote 'jedi', que é uma dependência necessária para o 'ipython'.



Turma: 9001

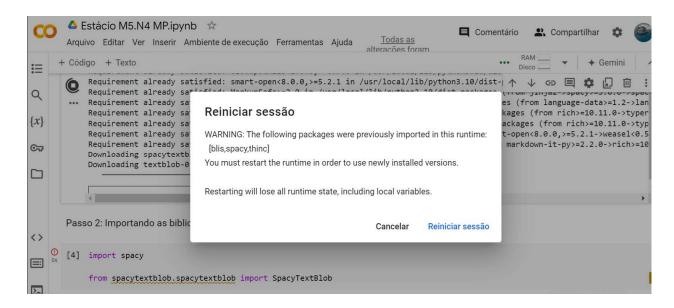
Matrícula: 202211381232

Aluno: Marco Sergio Albino Vittorio Barozzi Disciplina: Dando Inteligência ao Software

Código: RPG0034

Missão Prática | Dando Inteligência ao Software

Esses problemas geravam quebras constantes com solicitações de reinicialização da sessão



Descrições dos problemas

:

- 1. Conflitos de dependências: Algumas versões mais recentes de pacotes, como 'numpy 2.0.2' e 'spacy 3.8.2', eram incompatíveis com outros pacotes instalados, como 'tensorflow', 'gensim', e o próprio modelo 'en_core_web_sm', resultando em mensagens de erro e avisos de incompatibilidade.
- 2. Ausência do 'jedi': O ambiente reclamou a falta do pacote 'jedi', que é necessário para o correto funcionamento do 'ipython'.



Turma: 9001

Matrícula: 202211381232

Aluno: Marco Sergio Albino Vittorio Barozzi Disciplina: Dando Inteligência ao Software

Código: RPG0034

Missão Prática | Dando Inteligência ao Software

- Soluções:
- 1. Instalação de versões compatíveis : Optamos por reverter o `spaCy` para a versão 3.7.5 e o modelo `en_core_web_sm` 3.7.1 , que são compatíveis entre si e com outras dependências.
- 2. Instalação do pacote 'jedi: O pacote 'jedi' foi instalado para resolver o erro relacionado ao 'ipython'.

```
Passo 1: Instalando as bibliotecas e recarregando o ambiente
# Atualizando pip, setuptools e wheel
     !pip install -U pip setuptools wheel
     !pip uninstall numpy -y
     !pip install numpy==1.26.4
    # Instalando a versão mais recente compatível do spaCy (3.7.5)
    !pip install spacy==3.7.5
    # Baixar e instalar o modelo de linguagem en_core_web_sm versão 3.7.1, que é compatível com spacy 3.7.5
    !python -m spacy download en_core_web_sm
    # Instalando spacytextblob para análise de sentimentos
    !pip install spacytextblob
     !pip install jedi
     # Recarregar pkg_resources para garantir que todas as dependências sejam reconhecidas corretamente
     import pkg_resources, imp
     imp.reload(pkg_resources)
→ Requirement already satisfied: pip in /usr/local/lib/python3.10/dist-packages (24.2)
```



Turma: 9001

Matrícula: 202211381232

Aluno: Marco Sergio Albino Vittorio Barozzi Disciplina: Dando Inteligência ao Software

Código:RPG0034

Missão Prática | Dando Inteligência ao Software

Objetivo de cada importação

1. Atualização do pip, setuptools e wheel

O primeiro passo consiste em atualizar as ferramentas `pip`, `setuptools` e `wheel`. Essas ferramentas são essenciais para a instalação e gerenciamento de pacotes no ambiente Python. Manter essas ferramentas atualizadas garante a compatibilidade com as bibliotecas que serão instaladas posteriormente.

- 2. Desinstalação do numpy incompatível e instalação da versão correta Neste passo, a versão incompatível do `numpy` (2.0.2) é removida e, em seguida, a versão correta e compatível (1.26.4) é instalada. O `numpy` é uma biblioteca fundamental para cálculos numéricos, e a versão correta é necessária para evitar conflitos com outras bibliotecas como o `spaCy`.
- 3. Instalação do spaCy na versão compatível Aqui, instalamos a versão 3.7.5 do `spaCy`, uma biblioteca poderosa para processamento de linguagem natural (NLP). A escolha dessa versão específica é feita para garantir a compatibilidade com os outros pacotes e modelos que serão utilizados no projeto.
- 4. Download e instalação do modelo de linguagem en_core_web_sm Após a instalação do `spaCy`, é necessário fazer o download do modelo de linguagem `en_core_web_sm`, que contém as regras e dados linguísticos necessários para analisar textos em inglês. Esse modelo permite ao `spaCy` entender e processar textos de forma eficiente.



Turma: 9001

Matrícula: 202211381232

Aluno: Marco Sergio Albino Vittorio Barozzi Disciplina: Dando Inteligência ao Software

Código: RPG0034

Missão Prática | Dando Inteligência ao Software

5. Instalação do spacytextblob para análise de sentimentos
Neste passo, instalamos o pacote `spacytextblob`, que adiciona ao `spaCy` a
capacidade de realizar análise de sentimentos. Essa funcionalidade será utilizada para
determinar a polaridade (positiva ou negativa) e a subjetividade (fato ou opinião) dos
textos analisados.

6. Instalação do pacote jedi

O 'jedi' é instalado para melhorar a experiência de uso no ambiente de desenvolvimento interativo, como o IPython ou Jupyter. Essa biblioteca fornece funcionalidades de autocompletar, sugerindo métodos e funções conforme você digita, facilitando a programação.

7. Recarregamento do módulo pkg resources

Finalmente, o módulo `pkg_resources` é recarregado. Esse passo é importante para garantir que todas as novas dependências instaladas sejam reconhecidas pelo ambiente Python sem a necessidade de reiniciar o kernel ou a sessão, assegurando que tudo funcione corretamente.

9. Insira um bloco de código com as linhas abaixo e execute:

import spacy

from spacytextblob.spacytextblob import SpacyTextBlob



Turma: 9001

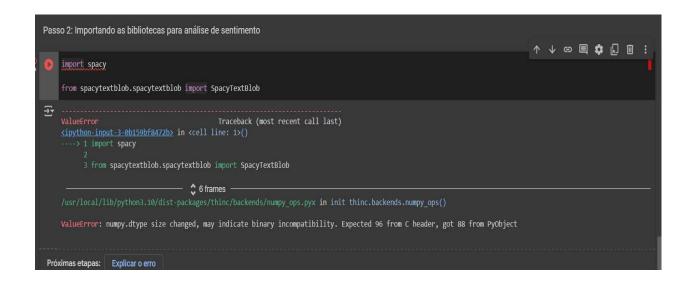
Matrícula: 202211381232

Aluno: Marco Sergio Albino Vittorio Barozzi Disciplina: Dando Inteligência ao Software

Código:RPG0034

Missão Prática | Dando Inteligência ao Software

Nessa etapa ocorreu o seguinte erro



Esse problema estava relacionado a uma incompatibilidade entre o numpy e o spaCy. O erro ocorria porque a versão do numpy instalada inicialmente estáva causando um conflito de compatibilidade binária. O spaCy espera que o tipo numpy.dtype tenha um tamanho diferente do que foi encontrado. Esse tipo de erro geralmente acontece quando há uma discrepância entre as versões das bibliotecas instaladas.



Turma: 9001

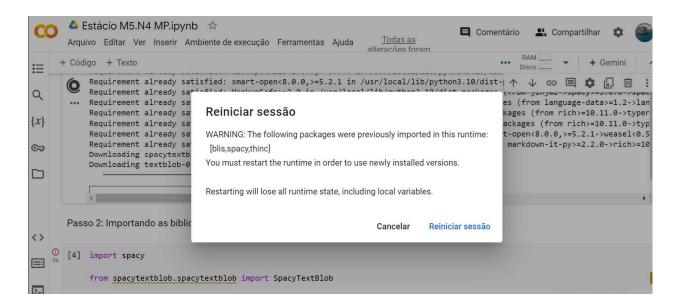
Matrícula: 202211381232

Aluno: Marco Sergio Albino Vittorio Barozzi Disciplina: Dando Inteligência ao Software

Código: RPG0034

Missão Prática | Dando Inteligência ao Software

Esse problema também causava quebras e corrigido posteriormente na instalação inicial



11. Crie um bloco de código com as linhas abaixo e execute:

nlp = spacy.load('en core web sm')

nlp.add pipe('spacytextblob')



Turma: 9001

Matrícula: 202211381232

Aluno: Marco Sergio Albino Vittorio Barozzi Disciplina: Dando Inteligência ao Software

Código: RPG0034

Missão Prática | Dando Inteligência ao Software

Os comandos acima configuram o spaCy e a extensão spacytextblob para realizar análises de sentimento em um texto. Aqui está o que cada linha faz:

- 1. 'nlp = spacy.load('en core web sm')':
- Carrega o modelo de linguagem em inglês en_core_web_sm do spaCy. Esse modelo contém regras de análise linguística, como tokenização, POS tagging (marcações de partes do discurso), e outras funcionalidades que permitem processar textos em inglês.
- 2. 'nlp.add_pipe('spacytextblob')':
- Adiciona o pipeline spacytextblob ao modelo. Esse pipeline será responsável por analisar o sentimento do texto processado. O spacytextblob usa a biblioteca TextBlob para realizar análise de sentimentos, retornando pontuações de polaridade (positiva, negativa, neutra).

O resultado indica que o spacytextblob foi adicionado com sucesso à pipeline do spaCy, o que permite que você processe textos e obtenha análises de sentimento diretamente.



Turma: 9001

Matrícula: 202211381232

Aluno: Marco Sergio Albino Vittorio Barozzi Disciplina: Dando Inteligência ao Software

Código:RPG0034

Missão Prática | Dando Inteligência ao Software

13. Crie um bloco de código com o conteúdo abaixo e o execute:

user_input = 'This is a wonderful campsite. I loved the serenity and the birds chirping in the morning.'

doc = nlp(user_input)

Passo 4: Definindo o texto inicial a ser analisado para verificação/validação da biblioteca

[5] user_input = 'This is a wonderful campsite. I loved the serenity and the birds chirping in the morning.'

doc = nlp(user_input)

objetivo de cada item

1. Variável 'user_input': Armazena o texto que será analisado. Nesse caso, é uma frase positiva sobre um acampamento: *"This is a wonderful campsite. I loved the serenity and the birds chirping in the morning."



Turma: 9001

Matrícula: 202211381232

Aluno: Marco Sergio Albino Vittorio Barozzi Disciplina : Dando Inteligência ao Software

Código:RPG0034

Missão Prática | Dando Inteligência ao Software

2. Função 'nlp(user_input)': Utiliza a pipeline 'nlp' criada anteriormente com 'spaCy' para processar o texto armazenado em 'user_input'. O texto é transformado em um documento processado ('doc'), que pode então ser analisado com base em aspectos de linguagem, como polaridade de sentimento, tokens, entre outros.

Essencialmente, esse bloco de código é responsável por definir o texto a ser analisado e preparar o processamento do mesmo usando as funcionalidades do 'spaCy' junto com a extensão 'spacytextblob'.



Turma: 9001

Matrícula: 202211381232

Aluno: Marco Sergio Albino Vittorio Barozzi Disciplina: Dando Inteligência ao Software

Código: RPG0034

Missão Prática | Dando Inteligência ao Software

15. Crie um bloco de código com as linhas abaixo e o execute:

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



Turma: 9001

Matrícula: 202211381232

Aluno: Marco Sergio Albino Vittorio Barozzi Disciplina: Dando Inteligência ao Software

Código: RPG0034

Missão Prática | Dando Inteligência ao Software

Aqui o erro ocorre porque o atributo `polarity` não está registrado corretamente para o objeto `doc` do spaCy. Para corrigir isso, você precisa usar a maneira adequada de acessar os atributos de polaridade e subjetividade da extensão <u>spacytextblob</u>.

```
Passo 5: Exibindo o resultado da primeira análise (um range entre-1 [avaliação negativa] e 1 [avaliação positiva]

# Acessando a polaridade e subjetividade usando o spacytextblob input_polarity = doc._.blob.polarity # Usar '_.blob.polarity' para acessar a polaridade input_subjectivity = doc._.blob.subjectivity # Usar '_.blob.subjectivity' para acessar a subjetividade

# Criando um dicionário para armazenar os resultados sentiment = {
    'score': input_polarity,
    'subjectivity': input_subjectivity
}

# Exibindo os resultados
print(sentiment)

# ('score': 0.85, 'subjectivity': 0.9)
```

O resultado exibido na imagem mostra a análise de sentimentos realizada no texto fornecido. O dicionário 'sentiment' exibe dois valores:

- 1. Polarity (ou score): O valor de 0.85 indica que o texto tem uma polaridade positiva (próxima de 1). Isso significa que o sentimento predominante no texto é positivo.
- 2. Subjectivity: O valor de 0.9 indica que o texto é altamente subjetivo. A subjectividade está relacionada a opiniões pessoais e emoções; valores próximos de 1 indicam uma forte presença de sentimentos ou opiniões, enquanto valores próximos de 0 indicam que o texto é mais objetivo e baseado em fatos.



Turma : 9001

Matrícula: 202211381232

Aluno: Marco Sergio Albino Vittorio Barozzi Disciplina: Dando Inteligência ao Software

Código: RPG0034

Missão Prática | Dando Inteligência ao Software

Aplicação:

- Polaridade: Esse valor pode ser usado para classificar automaticamente comentários ou textos como positivos, negativos ou neutros. Um exemplo de aplicação é analisar feedbacks de clientes, avaliações de produtos ou postagens em redes sociais para identificar o tom emocional.
- Subjectividade: Pode ser útil para distinguir textos que são baseados em opiniões ou fatos, o que tem diversas aplicações em processamento de linguagem natural, como em análise de críticas, notícias ou resenhas.

Esses indicadores são amplamente utilizados em áreas como marketing, atendimento ao cliente, e análise de redes sociais para avaliar o sentimento geral dos usuários ou clientes.

17. Insira um bloco de código com as linhas abaixo e execute:

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 Arsenal are the other strong rumours. #B04 #AFC", ...etc etc ...



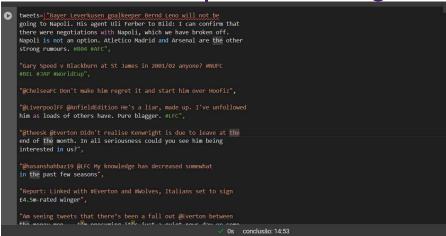
Turma: 9001

Matrícula: 202211381232

Aluno: Marco Sergio Albino Vittorio Barozzi Disciplina: Dando Inteligência ao Software

Código: RPG0034

Missão Prática | Dando Inteligência ao Software



Aqui o problema apresentado na imagem está relacionado à formatação do código, onde as quebras de linha dentro de uma string de texto no Python estão causando problemas de leitura e execução. Em Python, strings precisam estar em uma linha contínua ou, caso sejam divididas em múltiplas linhas, é necessário usar aspas triplas (""") ou a continuação explícita de linha (\) para evitar erros.



Turma: 9001

Matrícula: 202211381232

Aluno: Marco Sergio Albino Vittorio Barozzi Disciplina: Dando Inteligência ao Software

Código: RPG0034

Missão Prática | Dando Inteligência ao Software

Nesse caso foi adotada a complementação com "\n"





Turma: 9001

Matrícula: 202211381232

Aluno: Marco Sergio Albino Vittorio Barozzi Disciplina: Dando Inteligência ao Software

Código:RPG0034

Missão Prática | Dando Inteligência ao Software

19. Insira e executa o bloco de código abaixo:

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

Aqui, o erro apresentado, AttributeError: [E046] Can't retrieve unregistered extension attribute 'polarity', indica que o atributo polarity não foi corretamente registrado ou adicionado à pipeline do spaCy.

Esse problema ocorre porque o componente SpacyTextBlob (responsável por adicionar os atributos de análise de sentimento como polarity e subjectivity) provavelmente não foi carregado ou não está presente na pipeline do spaCy.



Turma: 9001

Matrícula: 202211381232

Aluno: Marco Sergio Albino Vittorio Barozzi Disciplina: Dando Inteligência ao Software

Código: RPG0034

Missão Prática | Dando Inteligência ao Software

O erro anterior indicava que o atributo 'polarity' não estava registrado porque o 'SpacyTextBlob' não havia sido corretamente adicionado à pipeline do spaCy. A linha que corrige isso:

nlp.add_pipe('spacytextblob')



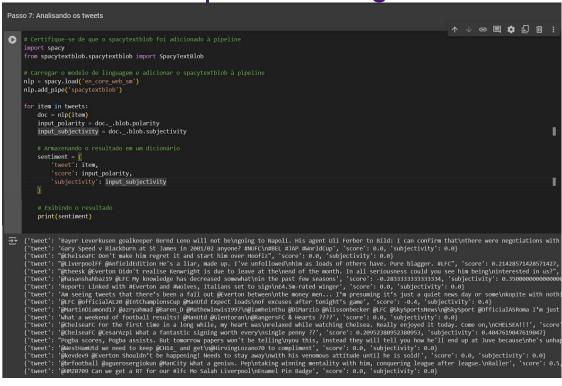
Turma: 9001

Matrícula: 202211381232

Aluno: Marco Sergio Albino Vittorio Barozzi Disciplina: Dando Inteligência ao Software

Código: RPG0034

Missão Prática | Dando Inteligência ao Software



Aqui foi feita a correção utilizando o componente 'SpacyTextBlob' para acessar a polaridade ('polarity') e subjetividade ('subjectivity') dos tweets.

Agora está presente no código, garantindo que o 'SpacyTextBlob' seja carregado na pipeline e os atributos necessários sejam acessíveis, como demonstrado nos resultados exibidos no terminal.

Os valores de polaridade e subjetividade estão sendo calculados corretamente para cada tweet.