

1. PLN

É uma área da inteligência artificial que estuda as interações entre computadores e a linguagem humana. Seu objetivo principal é permitir que máquinas compreendam, interpretem, gerem e respondam a textos ou falas humanas de maneira útil e significativa.

Em outras palavras, o PLN busca “ensinar” os computadores a entenderem o idioma humano da mesma forma como nós o usamos — com todas as suas ambiguidades, variações, contextos e nuances.

Principais tarefas do PLN incluem:

- **Análise sintática:** entender a estrutura gramatical de frases (ex.: identificar sujeito, verbo, objeto).
- **Reconhecimento de entidades nomeadas (NER):** localizar nomes de pessoas, lugares, organizações, etc.
- **Análise de sentimentos:** determinar a emoção ou intenção expressa em um texto.
- **Lematização e stemming:** reduzir palavras à sua forma base ou raiz.
- **Geração de linguagem natural (NLG):** produzir texto coerente e natural a partir de dados.
- **Tradução automática:** converter texto de um idioma para outro.
- **Resumo automático:** extrair as partes mais importantes de um texto.

2. Biblioteca spaCy

SpaCy é uma biblioteca de Processamento de Linguagem Natural (**PLN**) escrita em Python projetada para ser rápida, robusta e produtiva. É amplamente utilizada em aplicações que envolvem análise de texto, como extração de informações, classificação de textos, sistemas de perguntas e respostas, entre outros.

Principais funcionalidades:

- Tokenização
- Lematização
- Análise sintática (Part-of-Speech tagging)
- Reconhecimento de Entidades Nomeadas (NER)
- Detecção de sentenças
- Cálculo de similaridade
- Identificação de stop words

3. Exemplos - Funcionalidades

a) Tokenização

Identificação das menores unidades semânticas em um texto.

Arquivos: 01-Exemplo-Tokenização.py

```
import spacy

print('\n\n----- [ text ] ----- \n')
# Texto em portugues
nlp = spacy.load('pt_core_news_sm')
doc = nlp('Olá, meu nome é Marcos')

for token in doc:
    print(token.text)

print('\n\n FIM \n\n')
```

b) Part-of-Speech Tagging (POS)

Associação de categorias gramaticais aos tokens.

Arquivos: 02-Exemplo-POS.py

```
import spacy

print('\n\n----- [ text ] ----- \n')
# Texto em portugues
nlp = spacy.load('pt_core_news_sm')
doc = nlp('Olá, meu nome é Marcos')
```

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

print("\n\n FIM \n\n')
```

c) Lematização

Redução das palavras para sua forma canônica.

Arquivo: 03-Exemplo-Lematização.py

```
import spacy

# Carregar o modelo de idioma
nlp = spacy.load("pt_core_news_sm")

# Processamento do texto
texto = "Eles estão estudando pouco, mas ontem estudaram muito."
doc = nlp(texto)

# Lematização
for token in doc:
    # print(token.text, token.lemma_)
    print(f'{token.text:15} → {token.lemma_}')
```

d) Stop Words

Filtragem de palavras comuns que não agregam significado relevante.

Arquivo: 04-Exemplo-Stop Words .py

```
import spacy

# Carregar o modelo de idioma
```

```
nlp = spacy.load("pt_core_news_sm")

# Processamento do texto

texto = "O spaCy é uma biblioteca de processamento de linguagem natural em Python,
ela bem é bem utilizada, mas existem outras."

doc = nlp(texto)

print("\n\n ----- \n\n")

# Remoção de Stop Words

tokens_sem_stop_words = [token.text for token in doc if not token.is_stop]

print("Texto sem stop words\n",tokens_sem_stop_words)
```

e) Reconhecimento de Entidades Nomeadas (NER)

Identificação de nomes próprios, datas, organizações, etc.

Arquivos: 05-Exemplo-Entidades Nomeadas.py

```
import spacy

nlp = spacy.load("pt_core_news_sm")

doc = nlp("Paulo, A Fatec é uma faculdade localizada em Tatuí, estado de São Paulo e
California. Ela tem parceria com a Microsoft e Google e Bradesco.")

for ent in doc.ents:
    print(ent.text, '-->', ent.label_)
    print(str(spacy.explain(ent.label_)))
```

f) Segmentação de Sentenças

Divisão do texto em sentenças.

Arquivos: 06-Exemplo-Sentencas.py

```
import spacy
```

```

# Carregar o modelo de idioma
nlp = spacy.load("en_core_web_sm")

# Processamento do texto
texto = "O spaCy é uma biblioteca de processamento de linguagem natural em Python.
É útil para análise de texto."
doc = nlp(texto)

# Extraíndo frases
for sent in doc.sents:
    print(sent.text)

```

g) Morfologia

Extração de atributos morfológicos como gênero e tipo pronominal.

Arquivos: 07-Exemplo-Morfologia.py

```

import spacy

# Carregando o modelo da língua portuguesa
nlp = spacy.load("pt_core_news_sm")

# Processando o texto
doc = nlp("Elas compraram laranjas. Eu vou comprar banana. Colocarei no carro.")

# Extraíndo frases
print("\n classificação morfológica \n")
for token in doc:
    print(token.text, " --> ", token.morph)
    print("-----")

```

```
if token.morph.get("Gender") == ['Masc']:
    print(" Masculino \n")
```

h) Similaridade entre Tokens

Cálculo de similaridade semântica entre palavras.

Arquivo: 08-Exemplo - Similaridade entre Tokens.py

```
import spacy

# Carregar o modelo de idioma
nlp = spacy.load("pt_core_news_sm")

# -----
# 1- Processamento dos textos
# -----
texto1 = "gato"
texto2 = "gata"
token1 = nlp(texto1)[0]
token2 = nlp(texto2)[0]

# Similaridade
similaridade = token1.similarity(token2)
print("\n\n Similaridade entre '{}' e '{}': {:.2f}".format(texto1, texto2, similaridade))

# -----
# 2- Processamento dos textos
# -----
texto3 = "gato"
texto4 = "rato"
token3 = nlp(texto3)[0]
```

```
token4 = nlp(texto4)[0]

# Similaridade
similaridade2 = token3.similarity(token4)
print("\nSimilaridade entre '{}' e '{}': {:.2f}".format(texto3, texto4, similaridade2))
print("\n\n\n\n FIM \n\n\n\n")
```