

PROCESSAMENTO DE LINGUAGEM NATURAL COM PYTHON

O processamento da linguagem natural (PLN) é um ramo da inteligência artificial (IA) dedicado a desenvolver algoritmos capazes de entender e manipular a linguagem humana. No centro da criação desses algoritmos está uma ferramenta poderosa: o Python. Com sua sintaxe clara, vasto ecossistema de bibliotecas e comunidade ativa, o Python oferece uma plataforma perfeita para aplicações com processamento de linguagem. Neste roteiro, nós iremos abordar duas bibliotecas utilizadas em tarefas de PLN, sendo estas: NLTK (Natural Language Toolkit) e Spacy.

CONFIGURANDO O AMBIENTE

Antes de começar, é necessário configurar o ambiente de desenvolvimento com as bibliotecas necessárias. Você pode instalar as bibliotecas usando o comando pip.

```
pip install nltk  
pip install spacy
```

Ambas as bibliotecas necessitam do download de recursos adicionais. Para o NLTK, digite no prompt Python o comando abaixo e execute:

```
import nltk  
nltk.download('all')
```

O Spacy requer modelos de linguagem específicos. Você pode escolher um modelo e instalá-lo. Por exemplo, para o modelo em inglês, digite no prompt Python o comando abaixo e execute:

```
python -m spacy download en_core_web_sm
```

Para outros idiomas, você pode instalar modelos específicos de acordo com a documentação oficial do Spacy disponível no [link](#).

TESTANDO O AMBIENTE

Para garantir que tudo esteja funcionando corretamente, você pode testar as bibliotecas com alguns comandos simples apresentados a seguir:

1 – Testando o NLTK:

Digite o comando no Prompt Python e execute:

```
import nltk
print(nltk.corpus.gutenberg.fileids())
```

A saída esperada deve ser similar a apresentada abaixo:

```
['austen-emma.txt', 'austen-persuasion.txt', 'austen-sense.txt', 'bible-kjv.txt',
'blake-poems.txt', 'bryant-stories.txt', 'burgess-busterbrown.txt', 'carroll-
alice.txt', 'chesterton-ball.txt', 'chesterton-brown.txt', 'chesterton-thursday.txt',
'edgeworth-parents.txt', 'melville-moby_dick.txt', 'milton-paradise.txt',
'shakespeare-caesar.txt', 'shakespeare-hamlet.txt', 'shakespeare-macbeth.txt',
'whitman-leaves.txt']
```

2 – Digite o comando no Prompt Python e execute:

```
import spacy
nlp = spacy.load("en_core_web_sm")
doc = nlp("Hello, world!")
print([(token.text, token.pos_) for token in doc])
```

A saída esperada deve ser similar a apresentada abaixo:

```
[('Hello', 'PROPN'), (',', 'PUNCT'), ('world', 'NOUN'), ('!', 'PUNCT')]
```