

NLP

NATURAL
LANGUAGE
PROCESSING

NLP COM SPACY

Tópicos especiais em gestão de TI.

SPACY

- Vamos conhecer algumas métricas usadas para analisar textos e explorá-las através de exemplos, utilizando a biblioteca spaCy.

IMPORTANDO A BIBLIOTECA

- Para fazer o uso da biblioteca spacy, devemos importá-la.

```
import spacy
```

BAIXANDO O MODELO EM PORTUGUÊS

- Para trabalhar com o texto utilizando a língua portuguesa, é necessário fazer o download do modelo.

```
!python -m spacy download pt_core_news_sm
```

- O modelo deve ser carregado para utilizá-lo posteriormente.

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

PASSANDO O TEXTO PARA O MODELO

- Para realizar as tarefas de NLP, é necessário passar o texto para o modelo.

```
doc = nlp(u'Você já ouviu falar em processamento de linguagem natural?')
```

TOKENIZATION

- O processo de separar o texto em unidades menores é conhecido como Tokenization.
- Apesar do processo ser bem simples, podemos utilizar essa tarefa da biblioteca Spacy.

`doc.text.split()`

PART-OF-SPEECH TAGGING (POS)

- O modelo também é capaz de fazer a análise gramatical do nosso texto.
- Cada palavra é rotulada com uma classe gramatical correspondente.

`[(token.orth_, token.pos_) for token in doc]`

LEMMATIZING (LEMMA)

- No lemmatizing os tokens não sofrem flexão verbal, temporal ou de gênero.

`[token.lemma_ for token in doc]x'`

RECONHECIMENTO DE ENTIDADE NOMEADA

- As entidades nomeadas são os nomes próprios (Pessoas, Lugares, Marcas, Etc).

```
for entity in doc.ents:  
    print(entity, entity.label_)
```

DEPENDÊNCIA SINTÁTICA

- É possível descobrir a dependência sintática entre as palavras, comumente utilizado para analisar sintaxe de códigos-fonte.

```
for word in doc:
```

```
    print(word.text, word.dep_)
```

```
spacy.displacy.render(doc, style='dep', jupyter=True, options={'distance': 90})
```

SIMILARIDADE ENTRE PALAVRAS

- A biblioteca Spacy possibilita calcular a similaridade entre dois termos distintos.

```
my,dog,and_,cat,and__,horse = nlp(u'my dog and cat and horse')  
print(cat.similarity(dog))  
print(cat.similarity(horse))  
print(dog.similarity(horse))
```