

Prática: Processamento de Linguagem Natural (NLP) (III)

Eduardo Prasniewski

1 TensorFlow Playlist

Explica os conceitos básicos práticos de Processamento de linguagem natural com keras. Desde o momento de tokenização até LSTM NLP (prever a próxima palavra ainda com influência de várias palavras anteriores e não somente a última).

2 Spacy e Python

Primeiramente é discutido o que de fato é Processamento de Linguagem Natural (NLP) e suas subdivisões, como: categorização de texto, tradução de linguas etc. Após isso introduz como instalar o spacy e seus módulos. Com Spacy instalado corretamente, foi possível visualizar as nuances que a biblioteca faz com os textos (tokenização), no qual não é um simples *split* entre espaços, além de inúmeras outras ferramentas, como *lemma* e *ents*.

Após isso comenta superficialmente como funciona por baixo dos panos o funcionamento de uma NLP, com os vetores de palavras e suas similaridades.

Posteriormente, entra em um assunto vasto, as *pipelines*, que basicamente é uma extensão personalizada em um modelo já existente. Por exemplo, adicionar um contexto fictício de um filme (assim como foi utilizado no vídeo). O conceito de pipeline foi discutido de várias formas ao decorrer do resto do curso, cada vez introduzindo uma nova forma, como: Entity Ruler, Matcher, trocar tipos de entidades e por fim com o uso de expressões regulares (regex).