

Módulo 7 | Aula #3

Indo de caracteres para números

Tempo estimado de leitura:
5 min

O que é *embedding*?

Uma das técnicas para transformar os dados textuais para um formato em que as máquinas têm capacidade de compreender é o *embedding*.

O *embedding* é uma técnica utilizada para representar as palavras em futuras análises textuais. Normalmente, as palavras ficam armazenadas em um vetor de modo que aquelas que possuam significados similares fiquem próximas umas das outras. Dessa forma, é mais simples identificar os sinônimos.

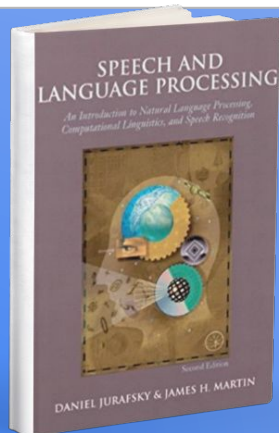
Talvez, neste momento, você esteja se perguntando quais seriam as possíveis aplicações de NLP, então trouxemos aqui alguns exemplos para você:

- Identificar o quanto duas frases ou textos são semelhantes - **análise de semelhança textual**;
- Classificar se **uma sentença contendo uma premissa pode ser respondida por uma outra sentença contendo uma hipótese**;
- Identificar **qual é a resposta para uma pergunta**;
- Identificar em um texto **o tipo de informação compartilhada**;
- De acordo com um texto escrito e publicado por algum usuário, **identificar qual sentimento foi colocado na publicação**;
- Identificar em qual **idioma o texto foi escrito**;
- Predizer e sugerir qual será **a próxima palavra na frase**;
- **Identificar o assunto** do texto/áudio.



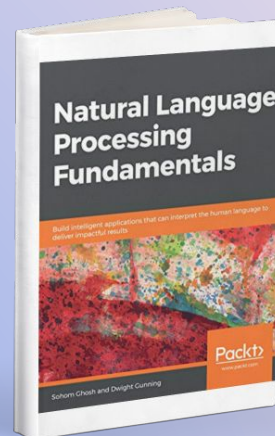
Para aprofundar ainda o seu conhecimento em NLP, recomendamos o artigo abaixo e os livros a seguir:

- Artigo:
BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding
Autores: Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova

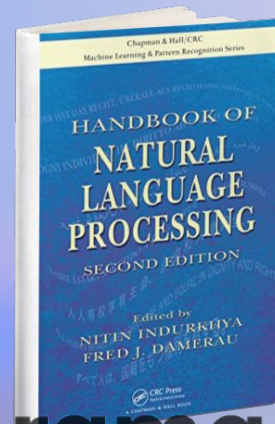


Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition

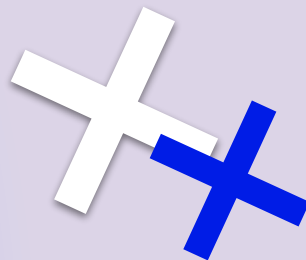
3



Natural Language Processing Fundamentals: Build intelligent applications that can interpret the human language to deliver impactful results



Handbook of Natural Language Processing



4