

Tarefas de NLP

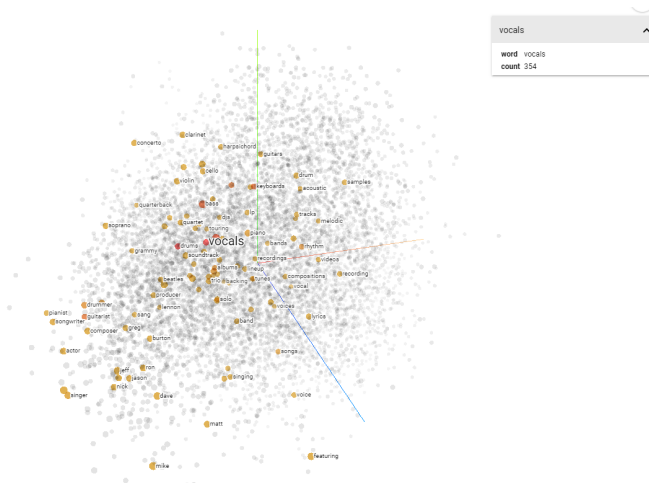
Advanced Institute for Artificial Intelligence – AI2

<https://advancedinstitute.ai>

Para uma determinada língua, em tese os embeddings aprendidos devem ser semelhantes para a maioria das tarefas.

Por que retreinar os embeddings para toda nova tarefa?

<https://projector.tensorflow.org/>



- Muitas empresas e organizações já disponibilizaram Embeddings de propósito geral:
 - Word2Vec - <https://en.wikipedia.org/wiki/Word2vec>
 - GloVe - <https://nlp.stanford.edu/projects/glove/>
 - Fasttext - <https://fasttext.cc/>
- Novos projetos de NLP geralmente reusam algum desses embeddings exceto se um vocabulário muito específico é esperado

- ❑ Esses Embeddings podem ser preparados para diversas línguas
- ❑ Inclusive Português!
- ❑ [http://nilc.icmc.usp.br/nilc/index.php/
repositorio-de-word-embeddings-do-nilc](http://nilc.icmc.usp.br/nilc/index.php/repositorio-de-word-embeddings-do-nilc)

Categorização de texto

The Premier League says it has listened to the objections of some clubs to the use of neutral grounds if the season resumes and will support them in putting those views to the authorities. → sports

The coronavirus pandemic has been bad news for business. But even as some technology companies have been squeezed tight, others have actually seen a boost. → tech

Como resolver esse problema?

Life is better when you're laughing.

→ *A vida é melhor quando você está rindo.*

I like multi-tasking, killing two birds with one stone.

→ *Eu gosto de fazer várias coisas ao mesmo tempo, matando dois coelhos com uma cajadada só.*

Como resolver esse problema?

- ☐ Stegosaurus
- ☐ Triceratops
- ☐ Velociraptor
- ☐ Allosaurus
- ☐ Archaeopteryx

- Mango → Mangosaurus
- Brazil → Brazilopteryx

Como resolver esse problema?