



BERT – Entendendo o estado da arte em NLP na prática



Paulo Finardi

Cientista de Dados, Itaú

01 DEZ - 12:25

promo.thedevconf.com/future-aws



O PAPEL DA TECNOLOGIA NA
CONSTRUÇÃO DO AMANHÃ

Core Tasks

*Covered in
Chapters 3-7*



Text
Classification



Information
Extraction



Conversational
Agent



Information
Retrieval



Question
Answering Systems

General Applications

*Covered in
Chapters 4-7*



Spam
Classification



Calendar Event
Extraction



Personal
Assistants



Search
Engines

JEOPARDY!

Jeopardy!

Industry Specific

*Covered in
Chapters 8-10*



Social Media
Analysis



Retail Catalog
Extraction



Health Records
Analysis



Financial
Analysis



Legal Entity
Extraction

Entender texto

qual o valor da minha fratura



All

Shopping

News

Images

Videos

More

Settings

Tools

About 1,770,000 results (0.52 seconds)

www.doctoralia.com.br > em-media-... · [Translate this page](#)

Em media quanto custa a cirurgia de correção da tibia? Boa ...

é normal essa dor e inchaço? Você quer enviar sua pergunta? Nossos especialistas responderam a 481 perguntas sobre Fraturas Da Tibia.

diegoariel.com.br > fratura-da-tibia · [Translate this page](#)

Fratura da Tibia - Dr. Diego Ariel

Basicamente são 5 causas principais de fratura da tibia: Quedas (própria altura, escada ou lugares altos); Lesões ... Após a cirurgia, em quanto tempo retorno às minhas atividades? ... Eu gostaria de saber o preço da cirurgia tibia fibula ...

www.scielo.br > pdf > rbgg [PDF](#)

O Custo Direto da fratura de fêmur por quedas em ... - SciELO

by ÂBM Arndt · 2011 · Cited by 37 — 17 O tratamento da fratura tem por finalidade restaurar a anatomia óssea, a função do membro e reabilitar o paciente efetivamente, devolvendo sua ...

www.aig.com.br > brasil > documents > brochure [PDF](#)

Fratura de Ossos - CONDIÇÕES GERAIS

Indenização desta cláusula, aplicados sobre o valor do capital segurado ... Contratuais. A fratura de ossos, deve estar relacionada exclusivamente à ocorrência de ... seguro, ainda que sua manifestação ocorra durante a vigência do seguro;

qual o valor da minha fratura do cartão de credito



All

News

Shopping

Videos

Images

More

Settings

Tools

About 5,570,000 results (0.88 seconds)

Showing results for qual o valor da minha **fatura** do cartão de credito

Search instead for qual o valor da minha fratura do cartão de credito

pagoquandopuder.com.br > fatura-ca... · [Translate this page](#)

[GUIA COMPLETO] Entenda A Fatura Do Seu Cartão De Crédito

Uma opção é escolher o parcelamento proposto pelo banco. Funciona assim: o valor da fatura atual é parcelado em algumas vezes (à sua escolha) e você paga ...

People also ask

Quanto é a fatura do cartão de crédito?



O que acontece se não pagar o valor total da fatura do cartão de crédito?



Como faço para ver a fatura do cartão de crédito?



Como funciona o pagamento de fatura de cartão de crédito?



[Feedback](#)

jurosbaixos.com.br > conteudo > apr... · [Translate this page](#)

Aprenda a ler a sua Fatura do Cartão de Crédito - Juros Baixos

Total da Fatura. É o valor total de todas as compras realizadas no cartão de crédito, no período da data de fechamento do cartão. Em suma, é o ...


Entender não é tão simples


Google


me mostre fotos de raças cachorro exceto beagles


Tudo Imagens Compras Vídeos Notícias Mais Definições Ferramentas


filhote raças caninas



Beagle (Beagle Inglês ou famoso Snoopy...
cachorroгато.com.br



Beagle – Raças Caninas - Raças de Cac...
petvale.com.br



476 x 238
Beagle (Beagle Inglês ou famoso Snoopy) - Raças de...
cachorroгато.com.br



Beagle | Raças de cães | Royal Canin
royalcanin.pt








Beagle - Blog do Cachorro
blogdocachorro.com.br


Nós, os Cachorros - N...
nososcachorros.blogspot...


Beagle: O cachorro cantor | Au au au!!!
caninablog.wordpress.com


Beagle: tudo sobre a raça em um guia co...
arbolez.com


Raças: Beagle | BitCã
bitcao.com.br



Entender não é tão simples

"Eu vi um homem na montanha
com um telescópio"



Entender não é tão simples

"Eu vi um homem na montanha
com um telescópio"



1

*Eu vi um homem. O homem estava na
montanha. Eu estava com o telescópio.*

Entender não é tão simples

"Eu vi um homem na montanha
com um telescópio"



1

Eu vi um homem. O homem estava na montanha. Eu estava com o telescópio.

2

Eu vi um homem. Eu estava na montanha. O homem estava com o telescópio.

Entender não é tão simples

"Eu vi um homem na montanha
com um telescópio"



1

Eu vi um homem. O homem estava na montanha. Eu estava com o telescópio.

2

Eu vi um homem. Eu estava na montanha. O homem estava com o telescópio.

3

Eu vi um homem. O homem estava na montanha. O homem estava com o telescópio.

Entender não é tão simples

"Eu vi um homem na montanha
com um telescópio"



1

Eu vi um homem. O homem estava na montanha. Eu estava com o telescópio.

2

Eu vi um homem. Eu estava na montanha. O homem estava com o telescópio.

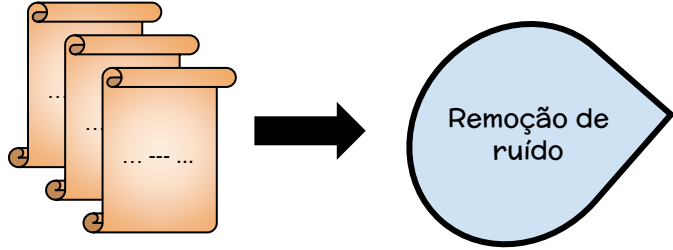
3

Eu vi um homem. O homem estava na montanha. O homem estava com o telescópio.

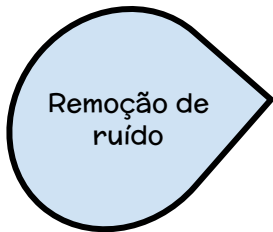
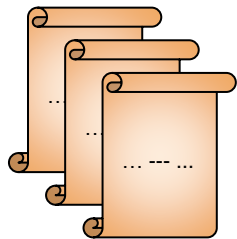
4

Eu vi um homem. Eu estava na montanha. Eu estava com o telescópio.

Dataprep NLP



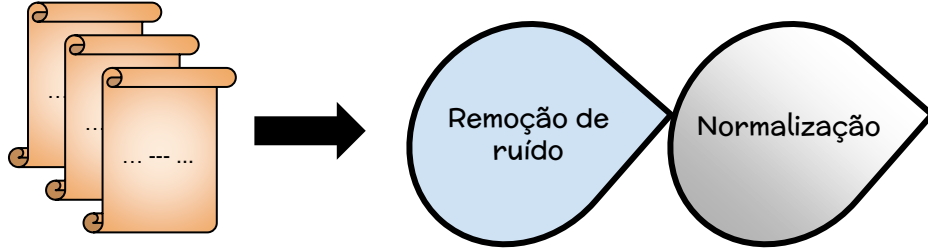
Dataprep NLP



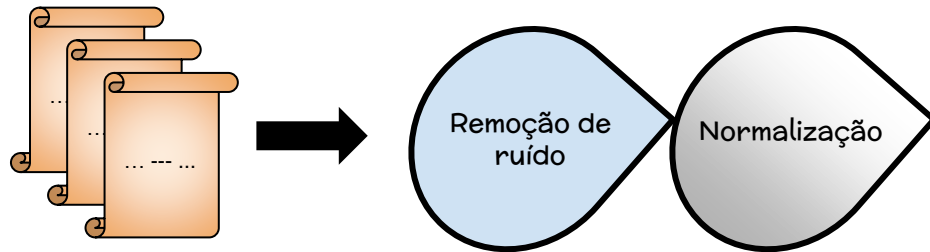
artigos, preposições...

{de, para a, e, em, ... etc}

Dataprep NLP



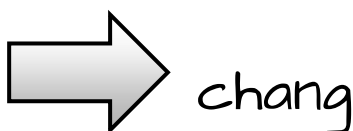
Dataprep NLP



Stemming é o processo de reduzir a inflexão das palavras às suas formas de raiz.

Lematização reduz as palavras flexionadas com garantia que a palavra raiz pertença ao idioma.

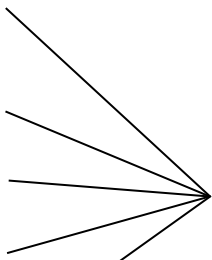
change
changing
changes
changed
changer



chang

The diagram shows the words 'change', 'changing', 'changes', 'changed', and 'changer' listed vertically. A vertical red dashed line is positioned between the 'e' and 'n' in each word. A large grey arrow points from this line to the word 'chang', illustrating the process of removing the inflectional suffixes to return the words to their root form.

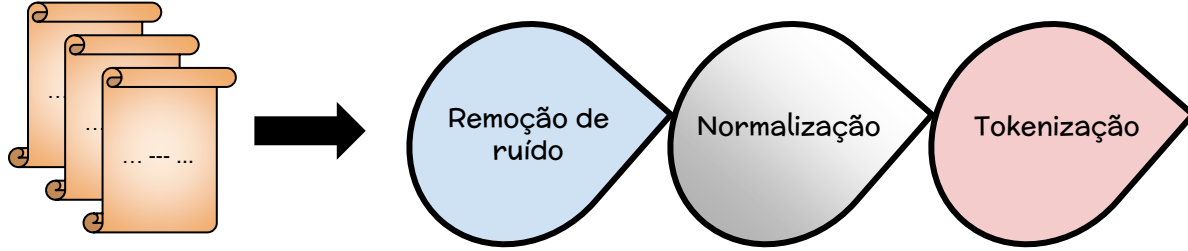
change
changing
changes
changed
changer



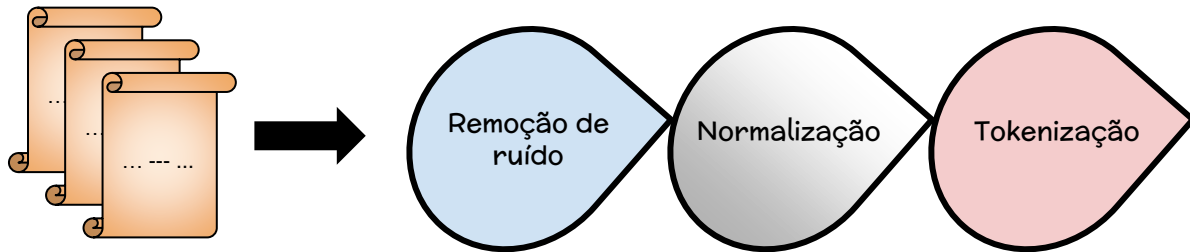
change

The diagram shows the same five words as the stemming example. Five lines originate from the right side of each word and converge towards a single point, which then points to the word 'change'. This illustrates the process of reducing all inflected forms of a word to its single, contextually appropriate lemma.

Dataprep NLP



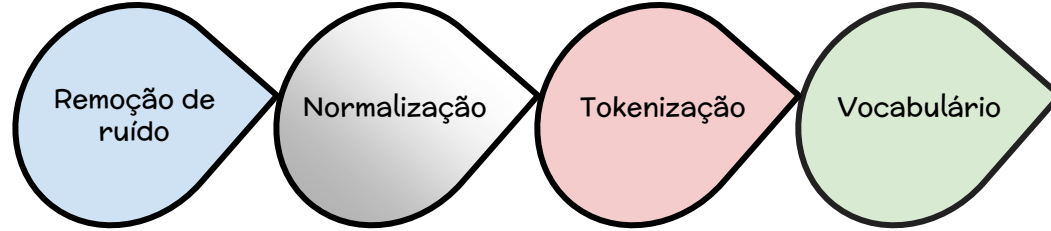
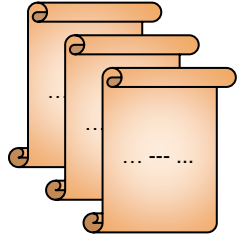
Dataprep NLP



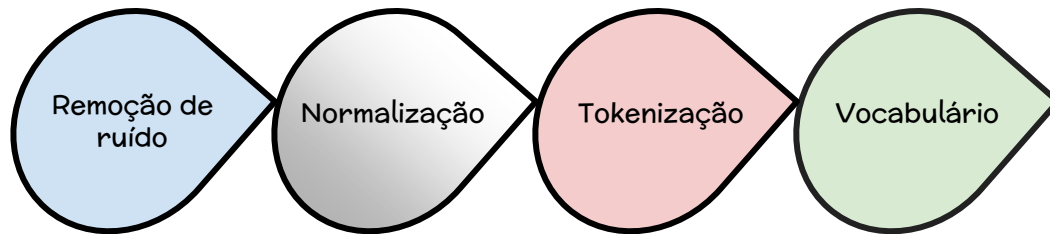
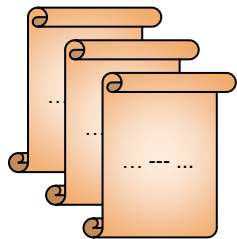
sentença: Quero um cartão adicional

Tokenização := [Quero][um][cartão][adicional]

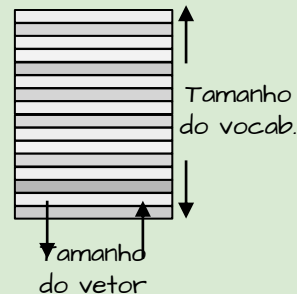
Dataprep NLP



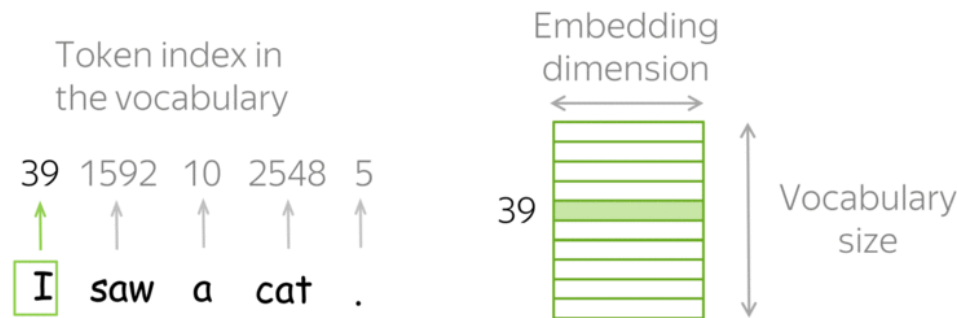
Dataprep NLP



Todos os Tokens
conhecidos e "especiais"



Como a máquina "lê" texto



fonte: https://lena-voita.github.io/nlp_course/word_embeddings.html



Uma máquina consegue "ler" textos de duas formas:

- Representação esparsa (BoW / One hot encoding / TF-IDF)
- Representação densa (Embeddings)

Representação Esparsa: Bag of Words (BoW) e One Hot Encoding



Good price! Quality not bad! I'm happy I bought it.

Representação Esparsa: Bag of Words (BoW) e One Hot Encoding



Good price! Quality not bad! I'm happy I bought it.

BoW

good	card	price	quality	bad	not	I	am	it	bought	return	happy	sad	will
1	0	1	1	1	1	1	1	1	1	0	1	0	0

Representação Esparsa: Bag of Words (BoW) e One Hot Encoding



Good price! Quality not bad! I'm happy I bought it.



Bad quality! I'm sad! I bought it I will return it.

BoW

good	card	price	quality	bad	not	I	am	it	bought	return	happy	sad	will
1	0	1	1	1	1	1	1	1	1	0	1	0	0
0	0	0	1	1	0	1	1	1	1	1	0	1	1

Representação Esparsa: Bag of Words (BoW) e One Hot Encoding



Good price! Quality not bad! I'm happy I bought it.



Bad quality! I'm sad! I bought it I will return it.

Price not good. Quality bad! I'm not happy I bought it.

BoW

good	card	price	quality	bad	not	I	am	it	bought	return	happy	sad	will
1	0	1	1	1	1	1	1	1	1	0	1	0	0
0	0	0	1	1	0	1	1	1	1	1	0	1	1
1	0	1	1	1	1	1	1	1	1	0	1	0	0

Representação Esparsa: Bag of Words (BoW) e One Hot Encoding



Good price! Quality not bad! I'm happy I bought it.



Bad quality! I'm sad! I bought it I will return it.



Price not good. Quality bad! I'm not happy I bought it.

BoW

good	card	price	quality	bad	not	I	am	it	bought	return	happy	sad	will
1	0	1	1	1	1	1	1	1	1	0	1	0	0
0	0	0	1	1	0	1	1	1	1	1	0	1	1
1	0	1	1	1	1	1	1	1	1	0	1	0	0

Representação Esparsa: Bag of Words (BoW) e One Hot Encoding



Good price! Quality not bad! I'm happy I bought it.



Bad quality! I'm sad! I bought it I will return it.



Price not good. Quality bad! I'm not happy I bought it.

BoW

good	card	price	quality	bad	not	I	am	it	bought	return	happy	sad	will
1	0	1	1	1	1	1	1	1	1	0	1	0	0
0	0	0	1	1	0	1	1	1	1	1	0	1	1
1	0	1	1	1	1	1	1	1	1	0	1	0	0

One Hot Encoding

good	card	price	quality	bad	not	I	am	it	bought	return	happy	sad	will
1	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	0	0	0	0

1 ...

Representação Esparsa: Bag of Words (BoW) e One Hot Encoding

Principais características da representação de texto esparsa

- Não retém informação sobre a gramática das sentenças nem sobre a ordem das palavras no texto.
- Se novas frases contiverem novas palavras, o tamanho do vocabulário e do vetor aumentará.
- É difícil de obter similaridade entre palavras/sentenças (grande quantidade de zeros).

tf - idf: Medida estatística que avalia a importância de uma palavra para um documento em uma coleção de documentos.

- **Motivação**: tf-idf é um dos métodos mais utilizados em sistemas de busca e NLP clássicos.
- **Intuição**: palavras que ocorrem muito (artigos, preposições) não são tão importante quanto as palavras menos frequentes.
- Estima a importância de uma palavra pela frequência com que ela ocorre em um conjunto de documentos.

$$\text{tf-idf}[t, d] = \text{tf}[t, d] * \text{idf}[t, D]$$

$\text{tf}[t, d]$ = número de vezes que o termo t aparece no documento d .

$\text{idf}[t, D] = \log(|D|/n_t)$,

n_t : número de documentos onde t aparece na coleção D -documentos

tf - idf: Exemplo

Documentos D:

- a) Essa é a primeira sentença
- b) Aqui é a segunda sentença
- c) E, por fim a última frase

Essa é a ¹segunda sentença

1 2 3 4 5

tf('segunda')

tf[t, d] = número de vezes que o termo *t* aparece no documento *d*.

$$\begin{aligned} &= 1/5 \\ &= 0.20 \end{aligned}$$

tf-idf('segunda') = **tf*idf**

$$0.20 * 0.48$$

idf('segunda')

$$\begin{aligned} \text{idf}[t, D] &= \log(|D|/n_t) \\ \text{idf}['segunda', D] &= \log(|3|/1) \\ &= 0.48 \end{aligned}$$

=

$$= 0.096$$

b) $\text{tf}('a')=0.20$, $\text{idf}('a')=0 \rightarrow \text{tf-idf}('a')=0$

Como a é um termo frequente em D, o tf-idf é menor do que a palavra segunda

“

Representação Densa

Como representar contexto/significado das palavras

Você sabe qual o significado da palavra **tezgüino**?



Como representar contexto/significado das palavras

Observe a palavra **tezgüino** em diferentes contextos:



Como representar contexto/significado das palavras

Observe a palavra **tezgüino** em diferentes contextos:

- Uma garrafa de **tezgüino** está sobre a mesa.



Como representar contexto/significado das palavras

Observe a palavra **tezgüino** em diferentes contextos:

- Uma garrafa de **tezgüino** está sobre a mesa.
- Todo mundo gosta de **tezgüino**.



Como representar contexto/significado das palavras

Observe a palavra **tezgüino** em diferentes contextos:

- Uma garrafa de **tezgüino** está sobre a mesa.
- Todo mundo gosta de **tezgüino**.
- Você pode ficar bêbado com **tezgüino**.



Como representar contexto/significado das palavras

Observe a palavra **tezgüino** em diferentes contextos:

- Uma garrafa de **tezgüino** está sobre a mesa.
- Todo mundo gosta de **tezgüino**.
- Você pode ficar bêbado com **tezgüino**.
- **Tezgüino** é feito de milho.



Como representar contexto/significado das palavras

Observe a palavra **tezgüino** em diferentes contextos:

- Uma garrafa de **tezgüino** está sobre a mesa.
- Todo mundo gosta de **tezgüino**.
- Você pode ficar bêbado com **tezgüino**.
- **Tezgüino** é feito de milho.

Consegue entender o que é **tezgüino**?



Como representar contexto/significado das palavras

Observe a palavra **tezgüino** em diferentes contextos:

- Uma garrafa de **tezgüino** está sobre a mesa.
- Todo mundo gosta de **tezgüino**.
- Você pode ficar bêbado com **tezgüino**.
- **Tezgüino** é feito de milho.

Com o contexto, conseguimos identificar do que se refere a palavra **tezgüino**.

Tezgüino:= é uma bebida alcoólica feita a base de milho.



Como representar contexto/significado das palavras

Como o cérebro faz isso?



Como representar contexto/significado das palavras

Quais outras palavras se "*encaixam*" nos slots das perguntas 1 até 4?

1. Uma garrafa de ----- está sobre a mesa.



Como representar contexto/significado das palavras

Quais outras palavras se "*encaixam*" nos slots das perguntas 1 até 4?

1. Uma garrafa de _____ está sobre a mesa.
2. Todo mundo gosta de _____.



Como representar contexto/significado das palavras

Quais outras palavras se "*encaixam*" nos slots das perguntas 1 até 4?

1. Uma garrafa de _____ está sobre a mesa.
2. Todo mundo gosta de _____.
3. Você pode ficar bêbado com _____.



Como representar contexto/significado das palavras

Quais outras palavras se "*encaixam*" nos slots das perguntas 1 até 4?

1. Uma garrafa de _____ está sobre a mesa.
2. Todo mundo gosta de _____.
3. Você pode ficar bêbado com _____.
4. _____ é feito de milho.



Como representar contexto/significado das palavras

Inserindo contexto de forma manual...

1. Uma garrafa de _____ está sobre a mesa.
2. Todo mundo gosta de _____.
3. Você pode ficar bêbado com _____.
4. _____ é feito de milho.

	(1)	(2)	(3)	(4) ← contextos
tezgüino	1	1	1	1
som	0	0	0	0
suco de laranja	1	1	0	0
vinho	1	1	1	0



Como representar contexto/significado das palavras

Inserindo contexto de forma manual...

1. Uma garrafa de _____ está sobre a mesa.
2. Todo mundo gosta de _____.
3. Você pode ficar bêbado com _____.
4. _____ é feito de milho.

	(1)	(2)	(3)	(4)	← contextos
tezgüino	1	1	1	1	Vetores similares
som	0	0	0	0	
suco de laranja	1	1	0	0	
vinho	1	1	1	0	



Como representar contexto/significado das palavras

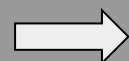
Inserindo contexto de forma manual...

1. Uma garrafa de _____ está sobre a mesa.
2. Todo mundo gosta de _____.
3. Você pode ficar bêbado com _____.
4. _____ é feito de milho.



	(1)	(2)	(3)	(4)	← contextos
tezgüino	1	1	1	1	
som	0	0	0	0	
suco de laranja	1	1	0	0	
vinho	1	1	1	0	

Vetores
Similares

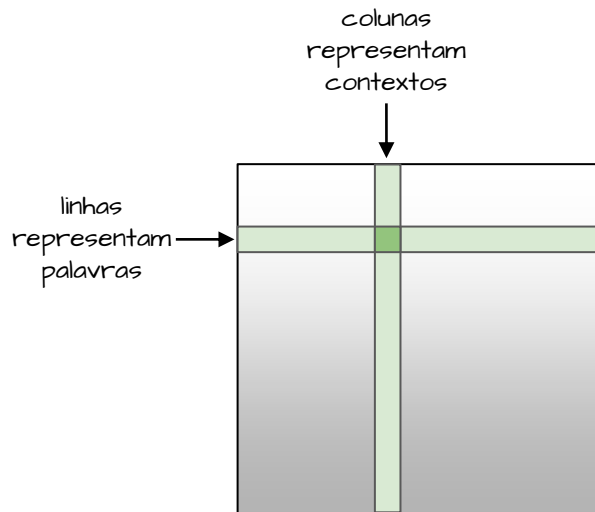


contexto similar

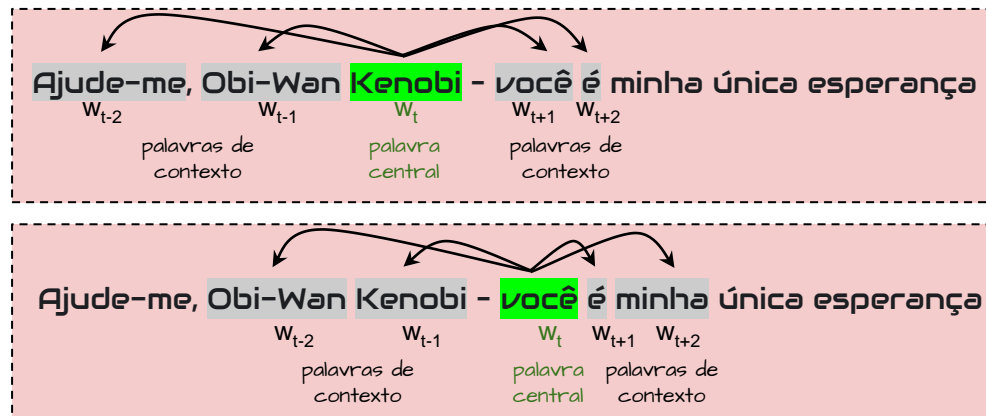
Como representar significado das palavras

Idéia principal: Precisamos colocar informações de contexto no vetor de palavra

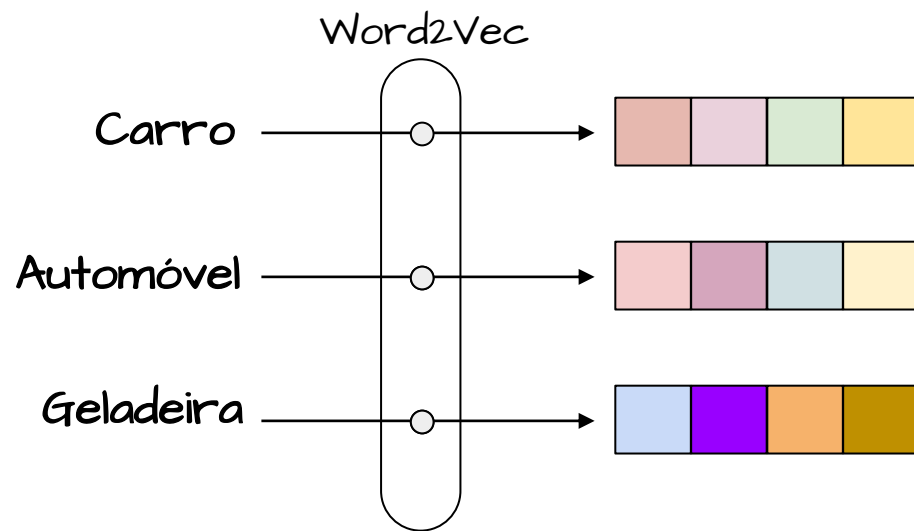
de forma manual... igual fizemos no exemplo da palavra *tezgüino*



de forma automática: Com um modelo que insere contexto na palavra baseado nas palavras vizinhas: AKA **Word2Vec**



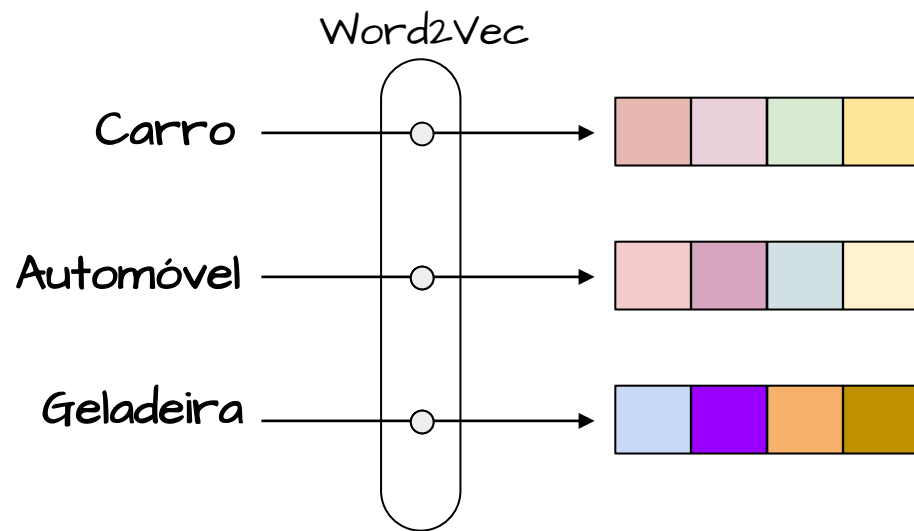
Word2Vec: intuição embedding



Intuição: Suponha que você precise se descrever em uma escala de 0 a 100 sobre:

1. **Você É Feliz?**

Word2Vec: intuição embedding

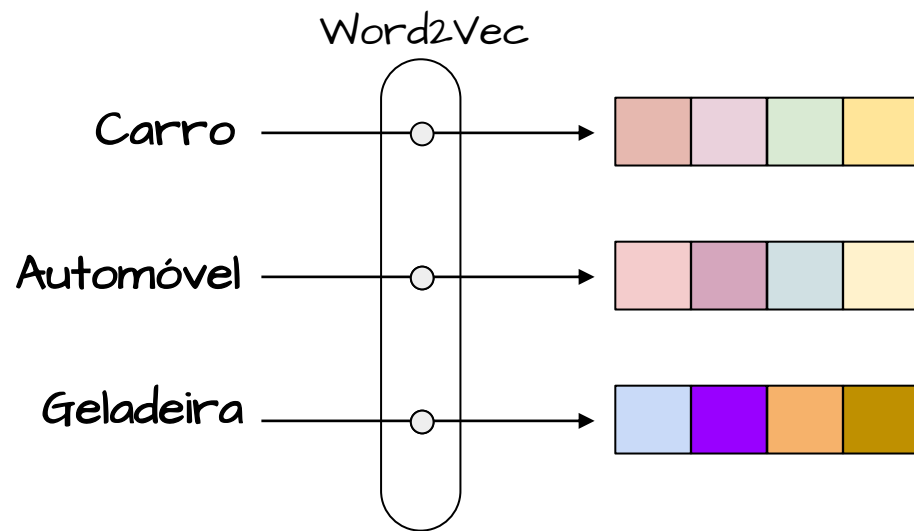


Intuição: Suponha que você precise se descrever em uma escala de 0 a 100 sobre:

1. **Você É Feliz?**



Word2Vec: intuição embedding

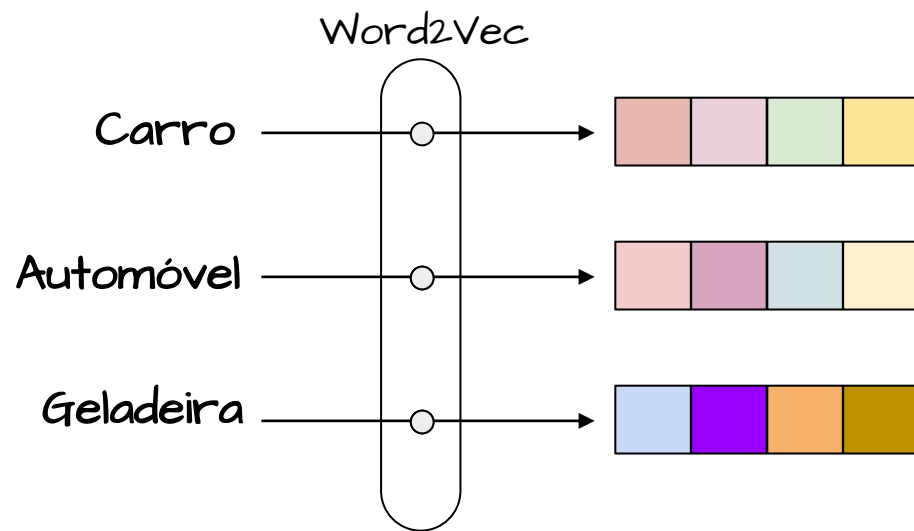


Intuição: Suponha que você precise se descrever em uma escala de 0 a 100 sobre:

1. **Você É Feliz?**
2. **Se alimenta bem?**

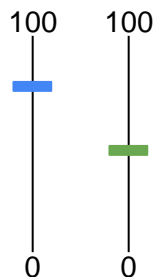


Word2Vec: intuição embedding

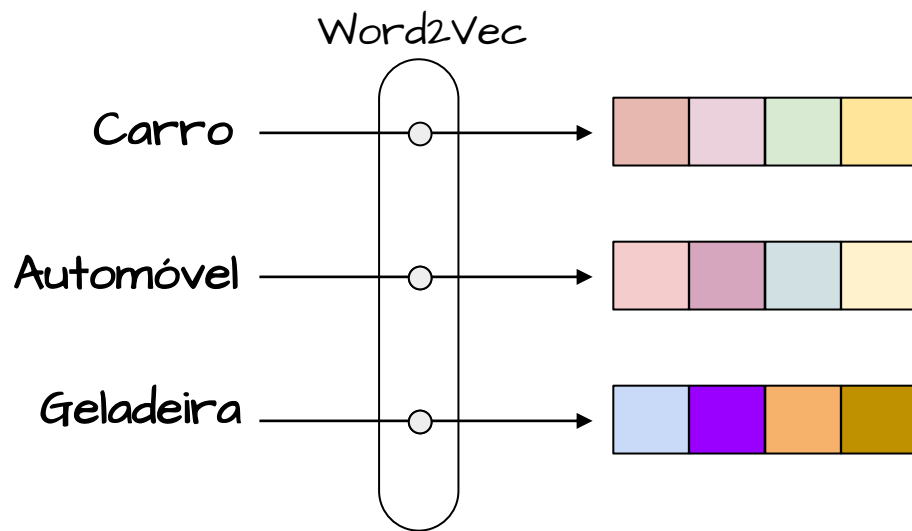


Intuição: Suponha que você precise se descrever em uma escala de 0 a 100 sobre:

1. **Você É Feliz?**
2. **Se alimenta bem?**

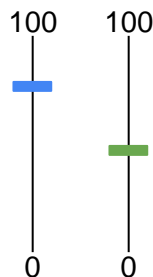


Word2Vec: intuição embedding

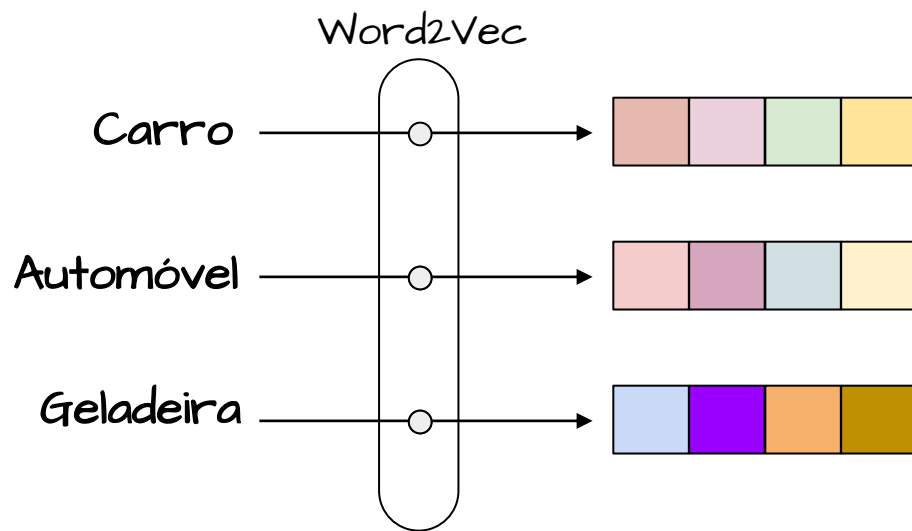


Intuição: Suponha que você precise se descrever em uma escala de 0 a 100 sobre:

1. *Você É Feliz?*
2. *Se alimenta bem?*
3. *Gosta de acordar cedo?*

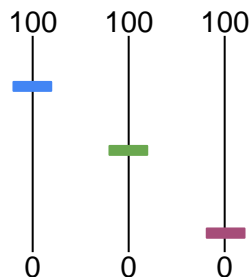


Word2Vec: intuição embedding

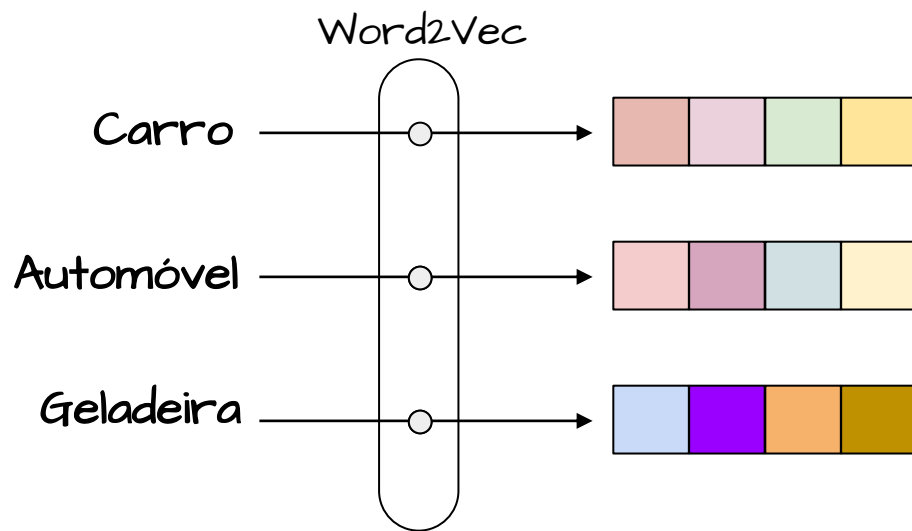


Intuição: Suponha que você precise se descrever em uma escala de 0 a 100 sobre:

1. *Você É Feliz?*
2. *Se alimenta bem?*
3. *Gosta de acordar cedo?*

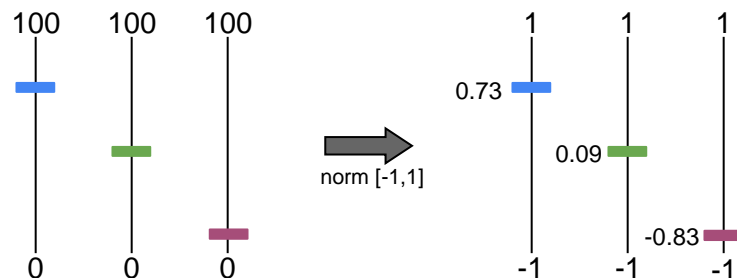


Word2Vec: intuição embedding

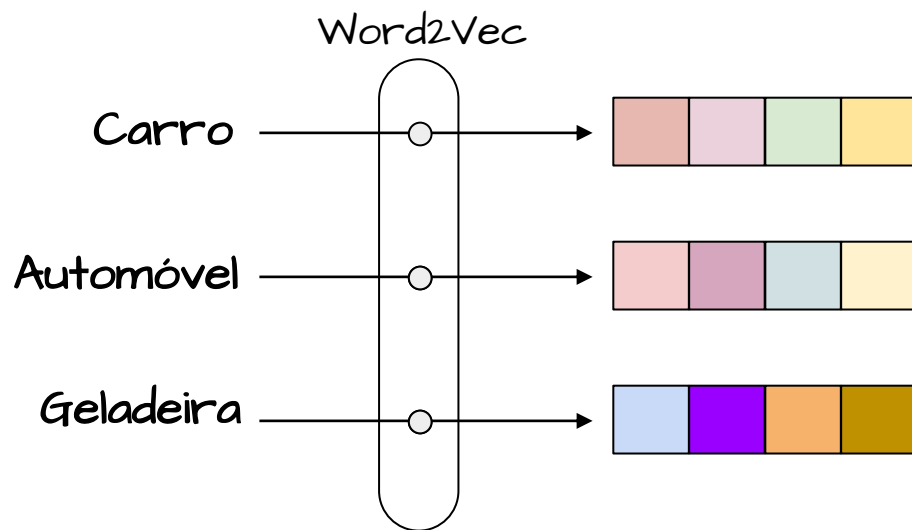


Intuição: Suponha que você precise se descrever em uma escala de 0 a 100 sobre:

1. Você É Feliz?
2. Se alimenta bem?
3. Gosta de acordar cedo?

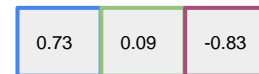
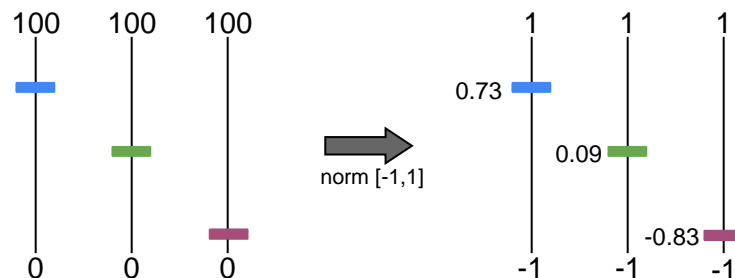


Word2Vec: intuição embedding



Intuição: Suponha que você precise se descrever em uma escala de 0 a 100 sobre:

1. *Você É Feliz?*
2. *Se alimenta bem?*
3. *Gosta de acordar cedo?*



Representação densa (Embedding)
com 3 dimensões

DataPrep Word2Vec

O número de vetores criados pelo Word2Vec é igual ao número de palavras do vocabulário. A redução de palavras semelhantes (inflexão), stopwords e lower_case geram melhores resultados:

1. Remoção de stop words (em ptbr: {de, para a, e, em, ... etc})
2. Texto para caixa baixa (lower_case)
3. Stemming ou Lematização

Exemplo Simpsons dataset

```
prep(lematização, lower, stopwords)
```

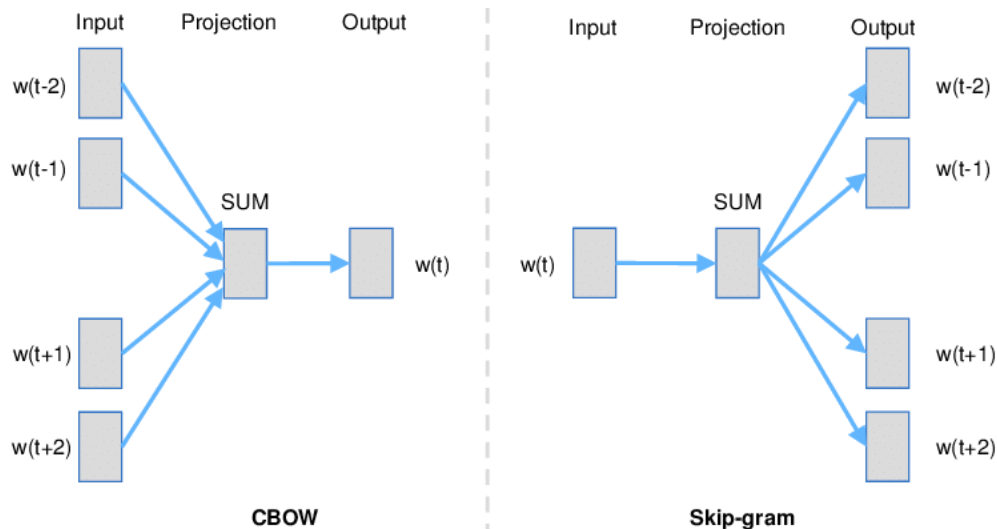
Raw: That life is worth living.

Processed: life worth live

Redução de 5 (raw) para 3
(processed) vetores

Word2Vec: CBOW e Skip-gram

- O CBOW prevê uma palavra-alvo aproveitando todas as palavras em sua vizinhança.
- O Skip-gram, aprende a prever uma palavra com base em uma palavra vizinha.



Ponto fraco do Word2Vec:
Não trata palavras O.O.V.

Qual é o melhor *Skip-gram* ou *CBOW*?

De acordo com o artigo original, Mikolov et al., O *Skip-gram* funciona bem em pequenos conjuntos de dados e pode representar melhor palavras menos frequentes. No entanto, o *CBOW* treina mais rápido do que *Skip-gram* e pode representar melhor as palavras mais frequentes.

FastText e ELMo

FastText: ao invés de treinar o modelo word2vec com palavras, n-gramas (sub-palavras) são utilizados. Exemplo, para 3-grama da palavra **carro**:

$['<ca', 'car', 'arr', 'rro', 'ro>']$

e o word embedding da palavra **carro** será a soma desses trigramas.

Dessa forma, uma palavra desconhecida, exemplo: **carroça** terá alguns n-gramas presentes em outras palavras conhecidas e a soma dos n-gramas darão a representação de **carroça**.

Embedding from Language Model ELMo: utiliza uma bi-LSTM para produzir embeddings contextuais, isto é, embeddings não possuem posição fixa. Exemplo:

sent_1 : **banco** da praça e sent_2 : app do **banco**

a palavra **banco** no word2vec sempre terá a mesma representação densa, o que não acontece no ELMo (devido ao treinamento bi-LSTM).



ELMo, qual é o embedding da palavra **banco**?

Existem muitas possibilidades, me forneça uma sentença

Estou sentado em um **banco**.

Ok, nesse caso é:
(-0.027, 1.83, ... etc)

Os últimos Jedi Language Models

6 Dec 2017

Attention Is All You Need

citações: **+31k**

Ashish Vaswani*
Google Brain
avaswani@google.com

Noam Shazeer*
Google Brain
noam@google.com

Niki Parmar*
Google Research
nikip@google.com

Jakob Uszkoreit*
Google Research
usz@google.com

Llion Jones*
Google Research
llion@google.com

Aidan N. Gomez* †
University of Toronto
aidan@cs.toronto.edu

Lukasz Kaiser*
Google Brain
lukaszkaizer@google.com

Illia Polosukhin* ‡
illia.polosukhin@gmail.com

Após a publicação do artigo *Attention Is All You Need* que apresentou a arquitetura Transformers, embeddings contextuais tiveram grande avanço comparado com o algoritmo ELMo.

2018

GPT

2019

GPT-2

XLNet

BERT

2020

T5

ALBERT

RoBERTa

XLNet

BART

DistilBERT

2021

GPT-3

ELECTRA

DeBERTa

Longformer

M2M100

LUKE

fonte: hugging face