

# Processamento Natural de Linguagem

## Aula 5

Modelos Deep Learning



# Semantix<sup>®</sup>

All about data

# Quem sou eu?



## Filipe Theodoro

Cientista de dados da Semantix  
Instrutor do Semantix Academy

### Contatos

[filipe.theodoro@semantix.com.br](mailto:filipe.theodoro@semantix.com.br)  
[linkedin.com/in/filipe-theodoro](https://linkedin.com/in/filipe-theodoro)



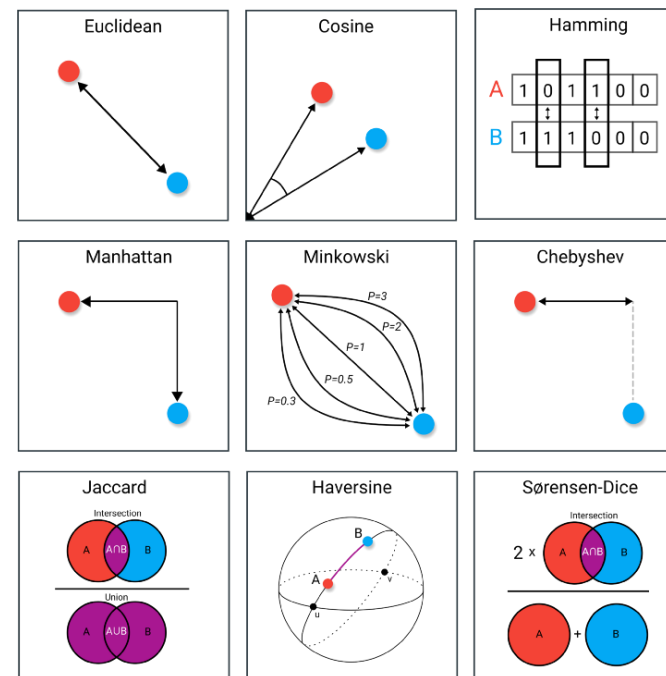
# Ementa

- Word embedding
- Word2vec
- RNN e LSTM
- Transformers

# Word embedding

# Word embedding

- Transformar um **token** em um **vetor denso**
- Captura informações do **contexto** da palavra
- Permite agrupar palavras similares por **proximidade** utilizando medidas de distancia



[0, 0, ... 0, 1, 0, 0, ... 0, 0, 0 ] -> cachorro

[1.4342, 2.1233, 0.9822, -1.0884, 3.1231] -> cachorro

[0, 0, ... 1, 0, 0 ] -> gato

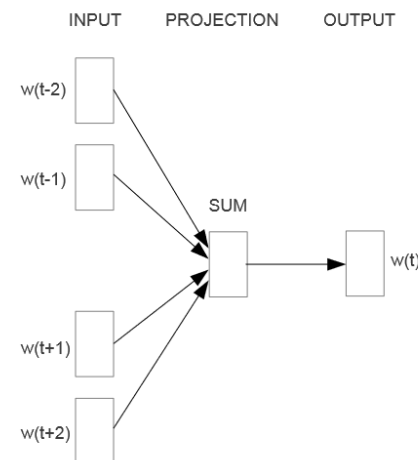
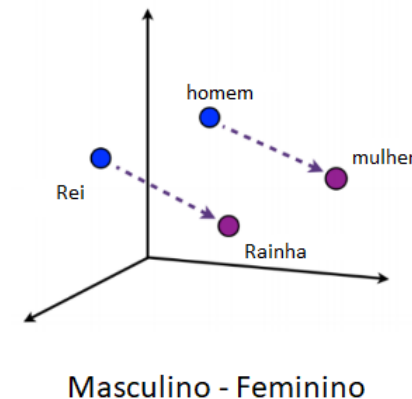
[1.1233, 2.6443, 0.0233, -1.5884, 3.9724] -> gato

# word2vec

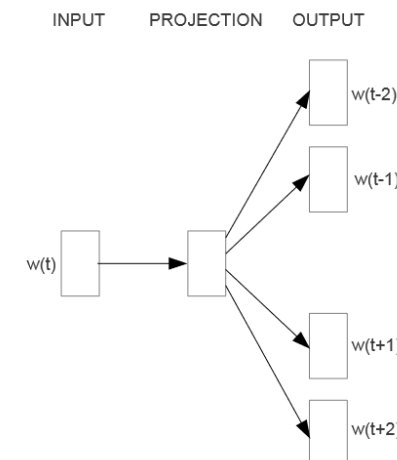


# word2vec

- Criado pelo time do Google em 2013
- Leva em consideração o contexto que o token aparece
- Nos permite extrair relacionamento entre os tokens
- Existem 2 formas de treinar um word2vec o CBOW e o Skip-gram



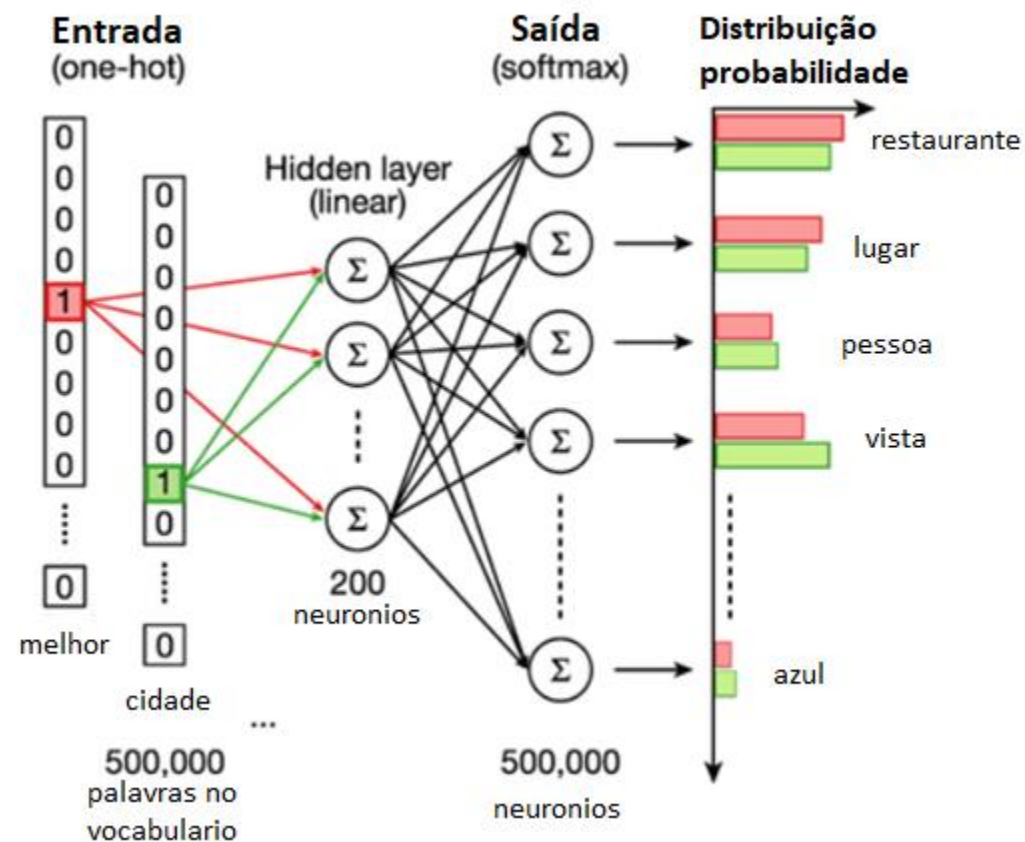
CBOW



Skip-gram

# word2vec

- É um modelo de redes neurais
- Método **semi supervisionado**
- Para uma boa **representação** é necessário um grande volume de dados
- A camada do **hidden layer** é o nosso **word embedding**



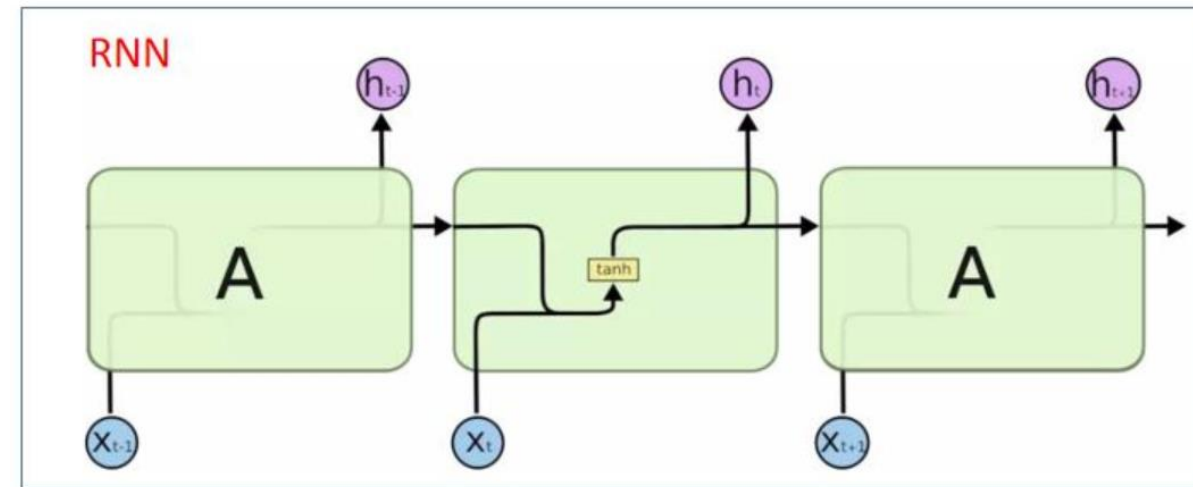
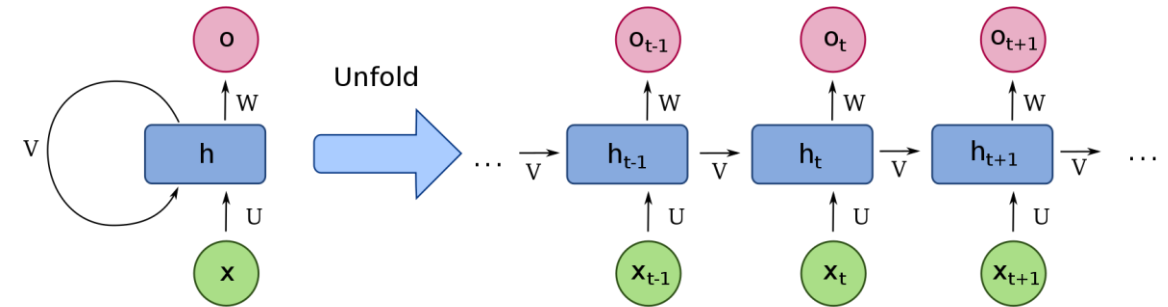


# RNN e LSTM



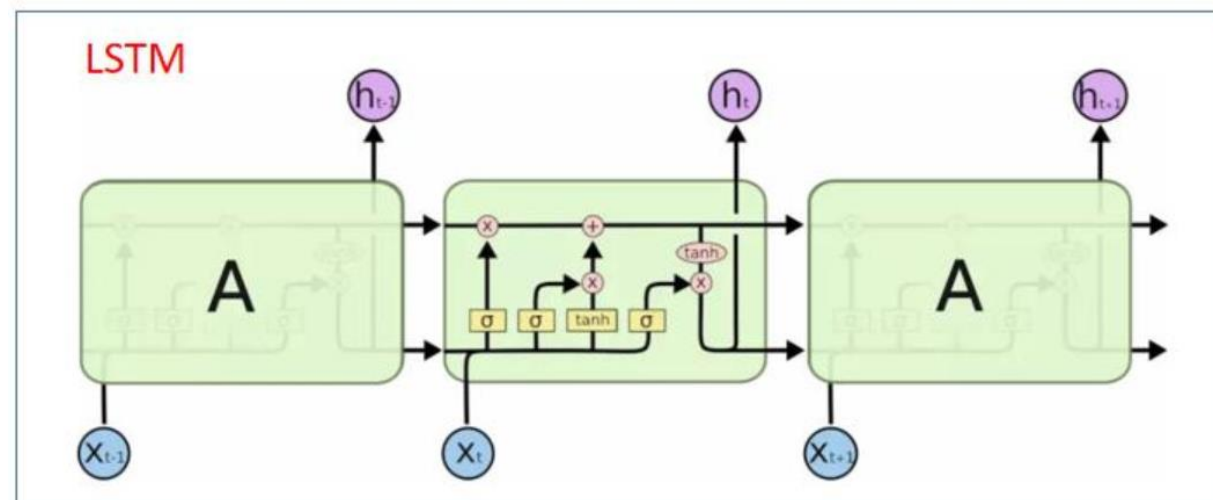
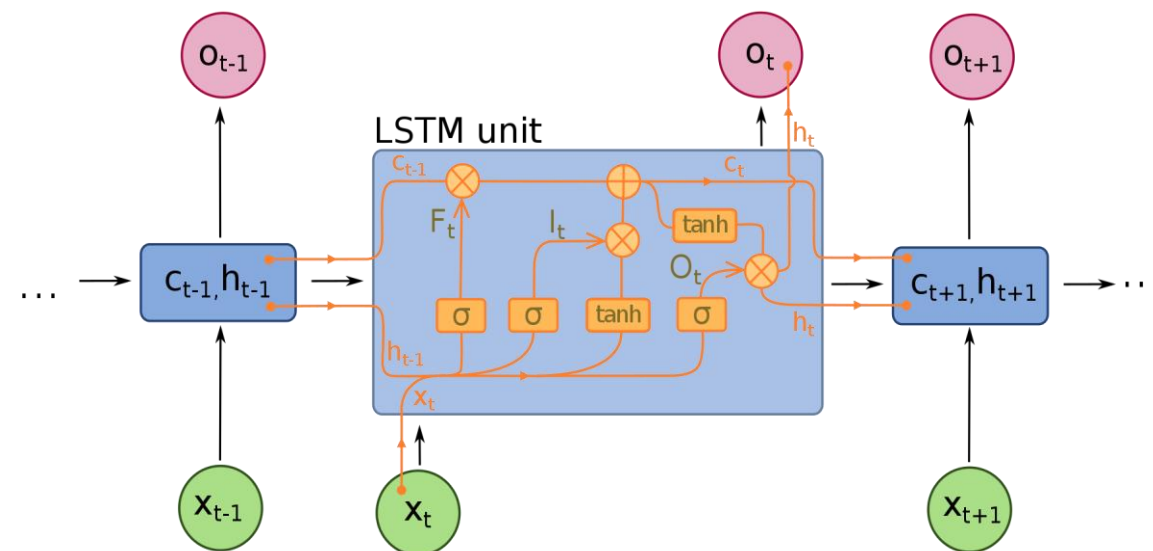
# RNN

- Rede neural recorrente
- Utilizado para modelar **sequencia** de dados no tempo
- Método **supervisionado**
- Foi desenvolvido para tarefa de **tradução de texto**
- Permite processar sequencias de **tamanhos diferentes**



# LSTM

- Memória de Curto e Longo prazo (Long Short term memory)
- Mecanismo de **esquecimento**
- Alternativa contra a dissipação do gradiente
- Aplicado em **reconhecimento de voz** e para transformar **texto em voz**.

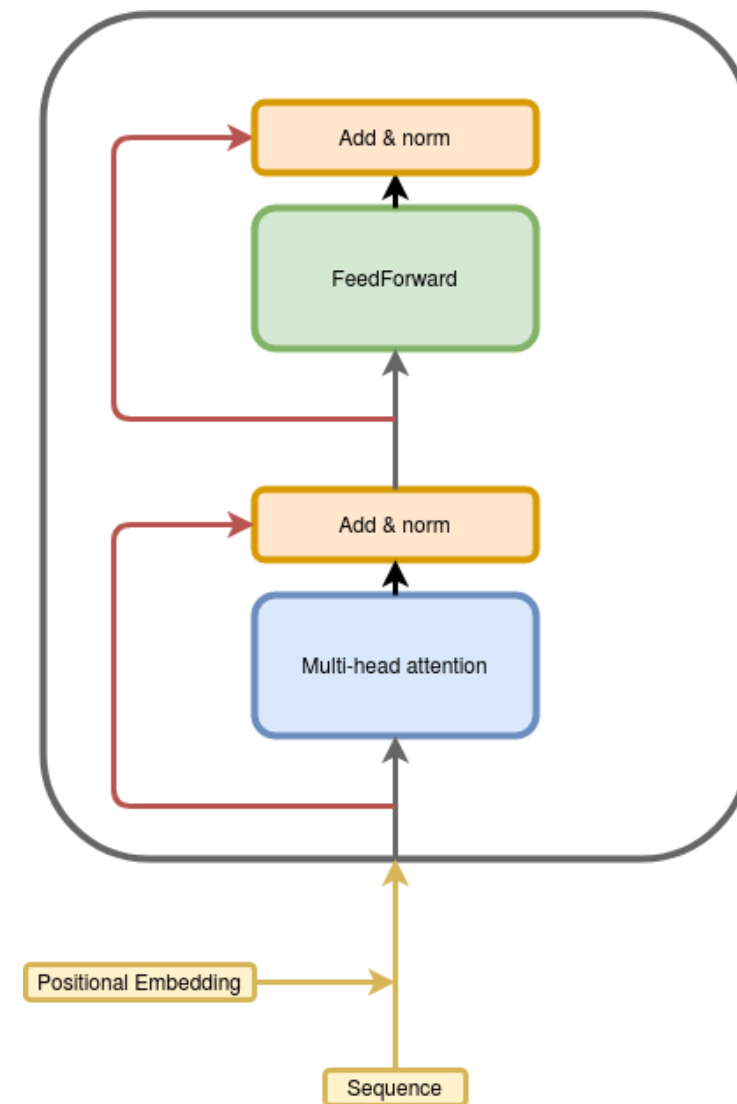


# Transformers



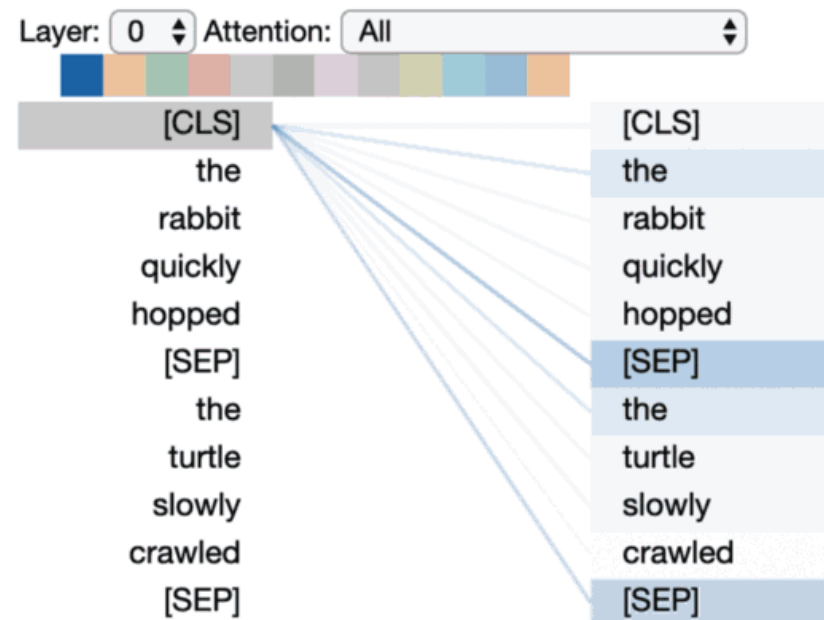
# Transformers

- Introdução do mecanismo de **atenção**
- Não precisa processar o dado em sequencia
- É a base para arquiteturas como **BERT** e **GPT**



# Transformers

- Leva em consideração a **posição do token** no texto
- Permite encontrar diferentes **relacionamentos** entre os tokens
- Podemos colocar **varias camadas** de atenção, onde cada uma vai fazer um **relacionamento diferente**.
- É a base para arquiteturas como **BERT** e **GPT**



# Bibliografia

- Word2vec - <https://arxiv.org/pdf/1301.3781.pdf>
- RNN – [https://en.wikipedia.org/wiki/Recurrent\\_neural\\_network](https://en.wikipedia.org/wiki/Recurrent_neural_network)
- LSTM – [https://en.wikipedia.org/wiki/Long\\_short-term\\_memory](https://en.wikipedia.org/wiki/Long_short-term_memory)
- Attention is all you need - <https://arxiv.org/abs/1706.03762>
- BERT - <https://jalammar.github.io/illustrated-bert/>



# Semantix<sup>®</sup>

All about data

[filipe.theodoro@semantix.com.br](mailto:filipe.theodoro@semantix.com.br)

[www.semantix.com.br](http://www.semantix.com.br)