

Processamento Natural de Linguagem

Aula 1

Introdução



Semantix[®]

All about data

Quem sou eu?



Filipe Theodoro

Cientista de dados da Semantix
Instrutor do Semantix Academy

Contatos

filipe.theodoro@semantix.com.br
linkedin.com/in/filipe-theodoro



Ementa

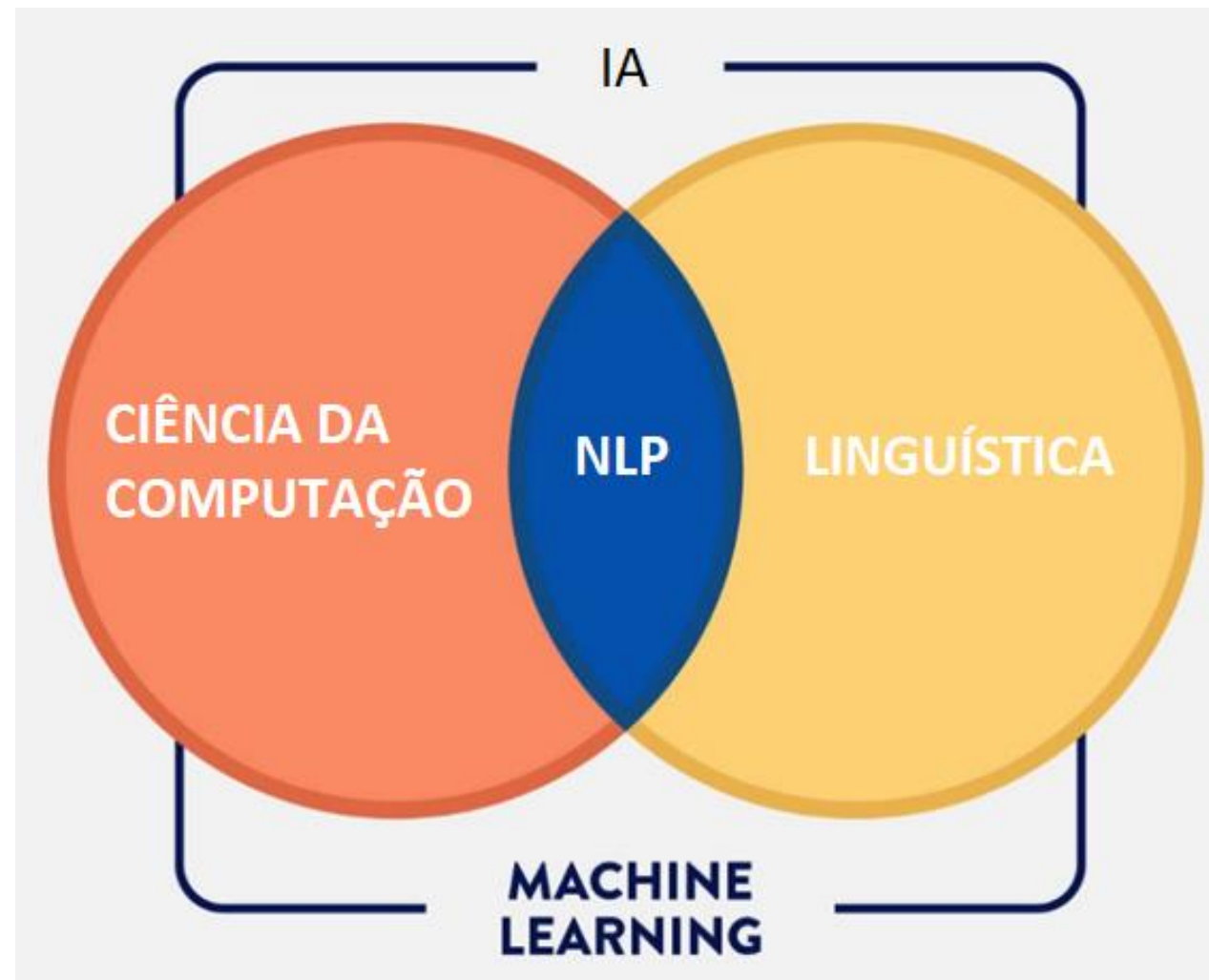
- O que é?
- História
- Aplicações práticas

O que é?

- **Definição**
- **Objetivo**
- **Tipos de análise da linguagem**

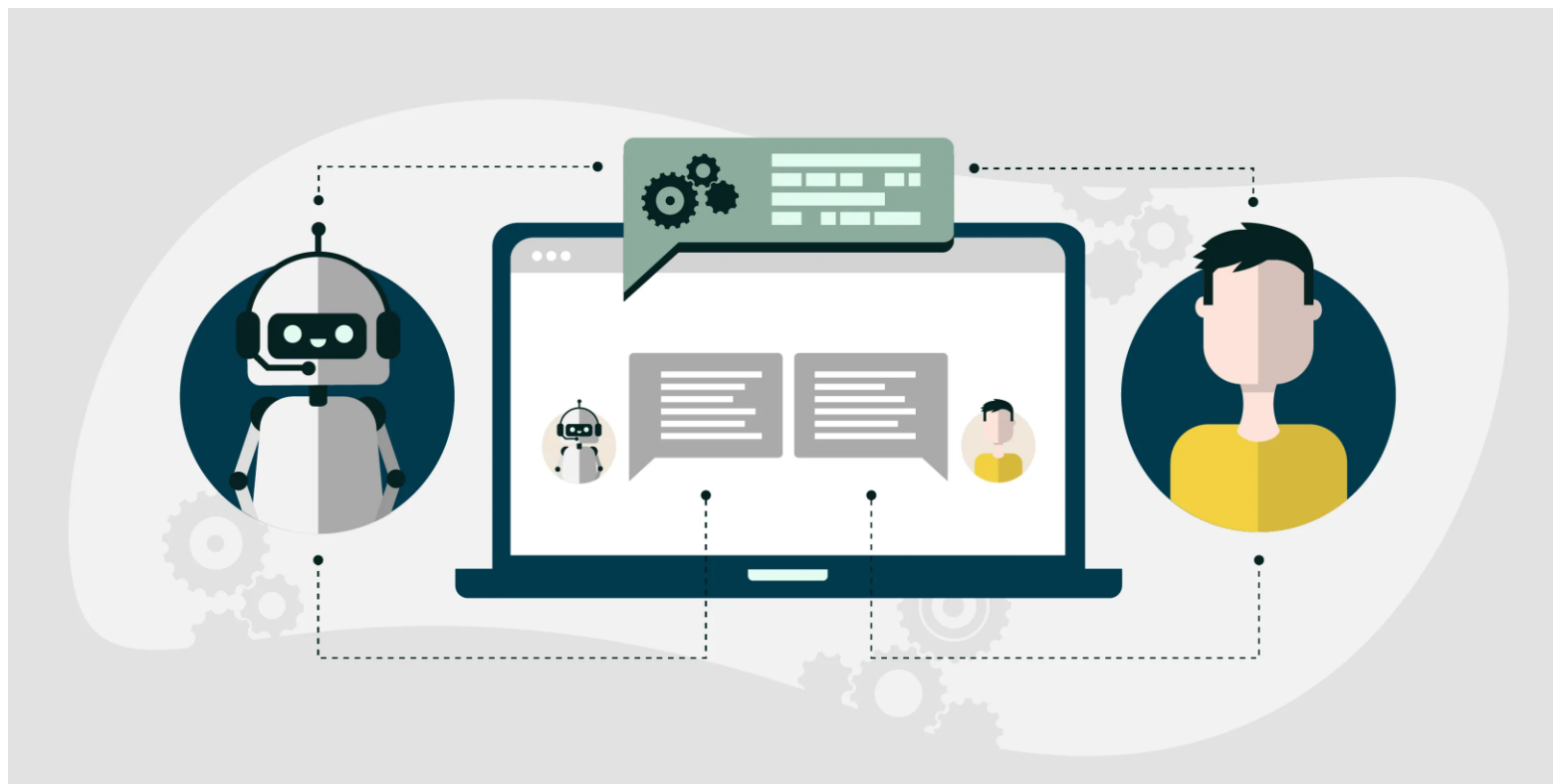
Definição

- Processamento Natural de Linguagem (PNL ou NLP)
- Permitir que máquinas entendam a linguagem humana
- Entender, interpretar e gerar resposta
- Simular conversação entre duas pessoas



Objetivo

- Reconhecer o **contexto** de um texto.
- Fazer uma análise **morfológica**, **semântica** e **sintática**, analisar **sentimentos** e interpretar **sentidos**.
- Com base nisso, a máquina obtém a capacidade de **entender** e de **criar respostas** para atender uma determinada tarefa.



Tipos de análise da linguagem

- Fonológica
 - Parte da linguagem voltada para a interpretação do som das palavras.
 - Leva em consideração a maneira que os sons são pronunciados.
 - Aplicado no desenvolvimento de aplicativos de voz.

Palavra	Categoria gramatical	Fonética
com•pu•ta• ção	nome feminino	kõ.pu.tẽ.s'ẽw

Dígrafo Consonantal	Fonema	Exemplos
ch	ʃ	chave - ficha - ch ocalho
lh	ɫ	telha - lhama - col her
nh	ɲ	rainha - vinho - moinho
gu	g	gu erra - gu itarra - sang ue
qu	k	qu erida - qu ase - qu ociente
sc	s	cô ns cio - pisc in a - discipl in a
sç		des sç a - cres sç a - flores sç a
xc		ex ceç ão - exc et o - exc ên trico
ss		ces s ão - sess ão - ass ado
rr	R	carro ça - carret a - carrocer ia

Tipos de análise da linguagem

- Morfológica
 - Remete a composição da palavra, que é dividida em morfemas.
 - Esses fragmentos podem representar um significado, porém não representam uma palavra.
 - Substantivos, adjetivos, artigo, adverbio, verbo.

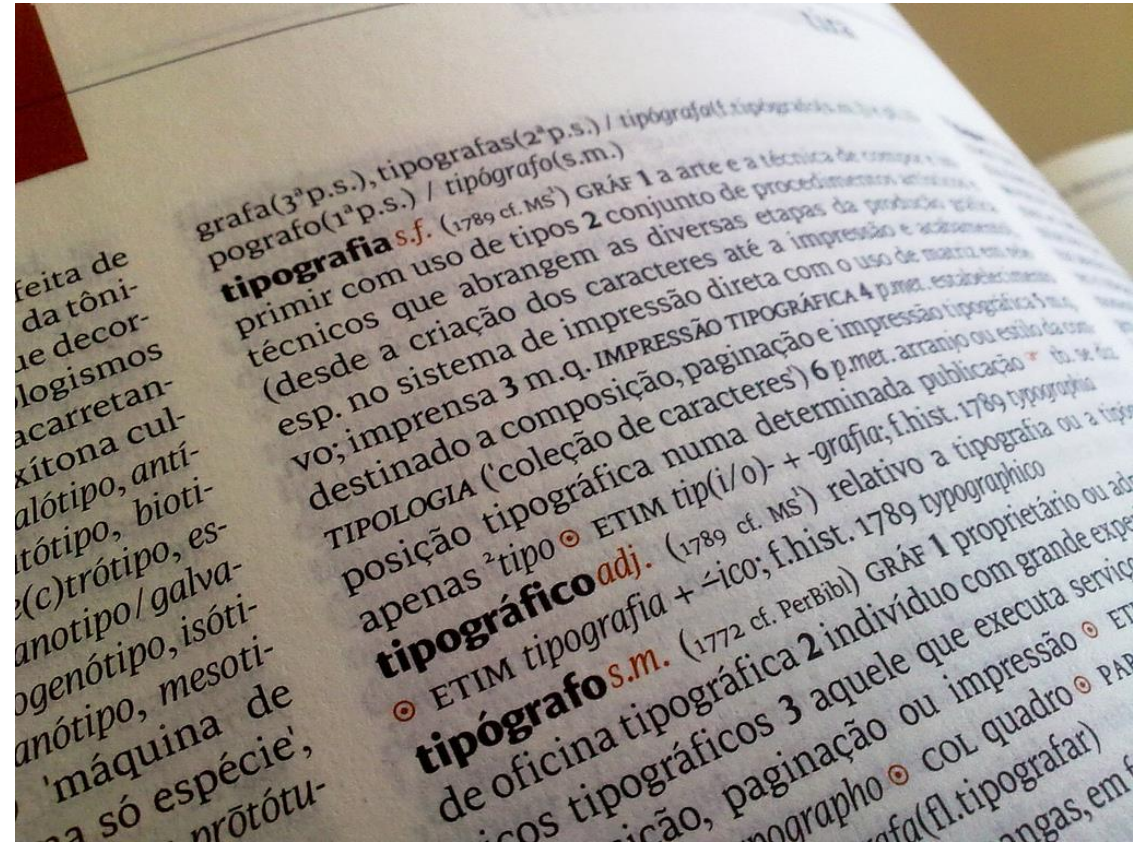
AM - AR
AM - OR
AM - ADO

radical morfema

FELIZ
IN – FELIZ
IN – FELIZ - MENTE

Tipos de análise da linguagem

- Léxica
 - Tem como função interpretar o significado individual de cada palavra ou expressão.
 - Compõe o vocabulário de uma língua.
 - Dicionário.



Tipos de análise da linguagem

- Sintática
 - Organização e relacionamento das palavras em uma oração.
 - PREDICADO, SUJEITO, VERBO

OS ALUNOS ESTUDAM NLP FELIZES.

SUJEITO

VERBO

OBJETO
DIRETO

PREDICATIVO

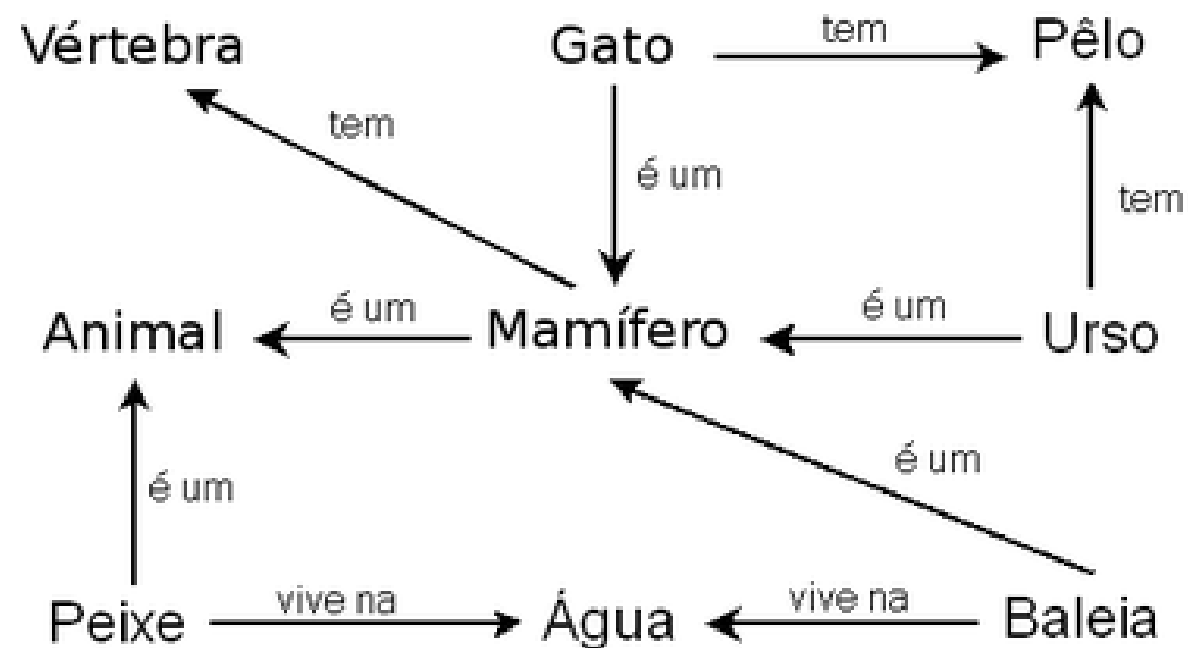
Tipos de análise da linguagem

- Semântica
 - Busca compreender o significado da palavra.
 - SINONIMO, ANTONIMO

Esse chapéu é muito bom.

Esse chapéu é muito ruim.

Esse chapéu é horrível.



História

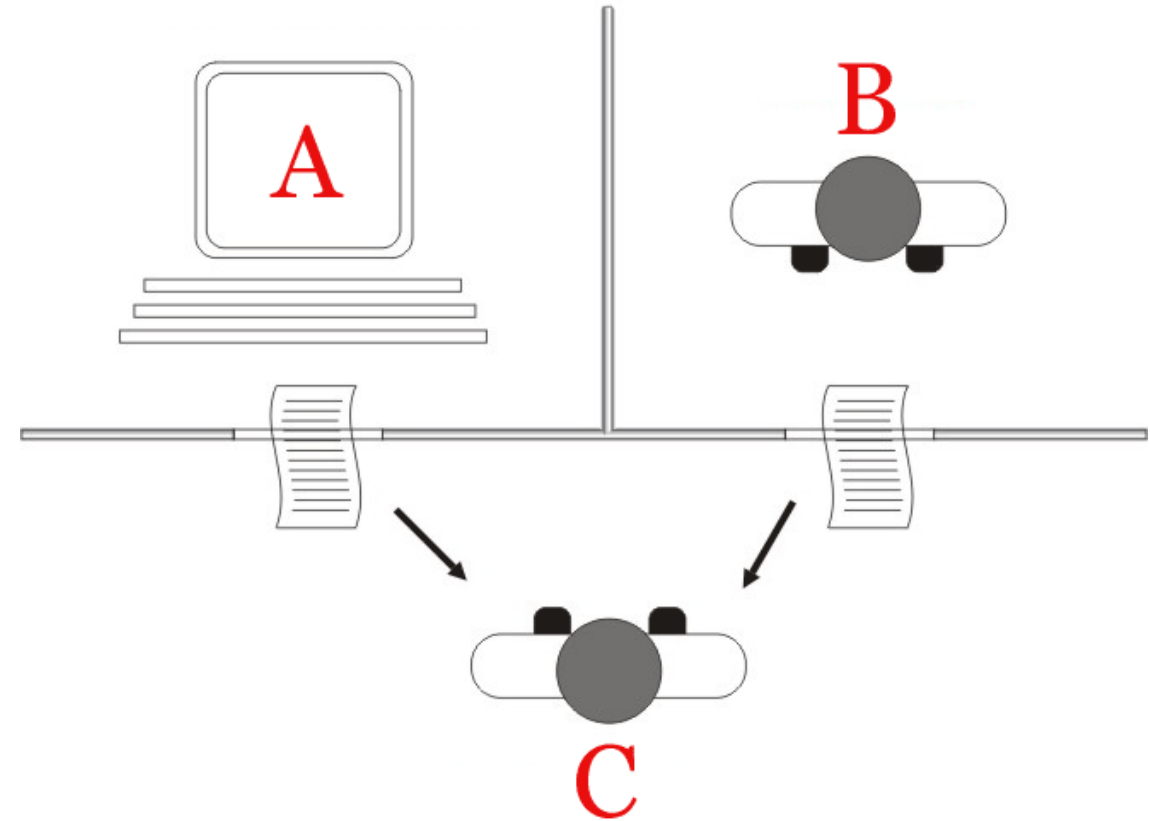
- Machine Translation
- Teste de Turing
- ELIZA

Machine Translation

- A primeira máquina de NLP foi desenvolvida em 1946.
- Tinha como objetivo decifrar as mensagens interceptadas dos inimigos.
- O primeiro estudo foi feito por Warren Weaver em "The memorandum" que a partir daí deu início a pesquisa na área de NLP.

Teste de Turing

- Jogo proposto por Alan Turing em 1950
- É possível ou não máquina apresentar comportamento inteligente?
- Sessões de perguntas e respostas entre humanos e máquinas.
- Identificar a máquina e a pessoa a partir de uma conversa.



ELIZA

- Baseado na teoria de gramática gerativa proposta por Chomsky.
- Primeiro chat bot.
- Simular um psicólogo.
- Composta de regras bem definidas.

```
=====
EEEEEEEE L      IIIIIII ZZZZZZZ AAA
E         L      I        Z      A  A
E         L      I        Z      A  A
EEEEEE    L      I        Z      A  A
E         L      I        Z      AAAAAA
E         L      I        Z      A  A
EEEEEEEE LLLLLLLL IIIIIII ZZZZZZ  A  A
=====
ELIZA > (PLEASE TYPE IN ALL CAPS) WHAT'S YOUR NAME DEAR ?
AMIT
ELIZA > HI, I'M ELIZA. WHAT DO YOU WANT TO TALK ABOUT ?
AMIT > CAN I TALK ABOUT MY PROBLEM ?
ELIZA > SURE... !
AMIT > DUE TO ON GOING EXAMS IN COLLEGE, I HAVE A LOT OF STRESS.
ELIZA > PLEASE RELAX A LITTLE. SLEEP WELL.
AMIT > THANKS FOR YOUR ADVICE.
ELIZA > NO MENTION.
AMIT > BYE
ELIZA > BYE AND KEEP IN TOUCH...
=====
```

Exemplo de aplicações

- Tradução
- Transcrição
- Chat-bot
- Classificação
- Clusterização
- Busca de informação
- Correção ortográfica
- NER

Tradução

- A tarefa de transformar um texto de uma língua para outra.


Português



↔

Inglês

Até hoje não encontramos a solução para o aquecimento global.

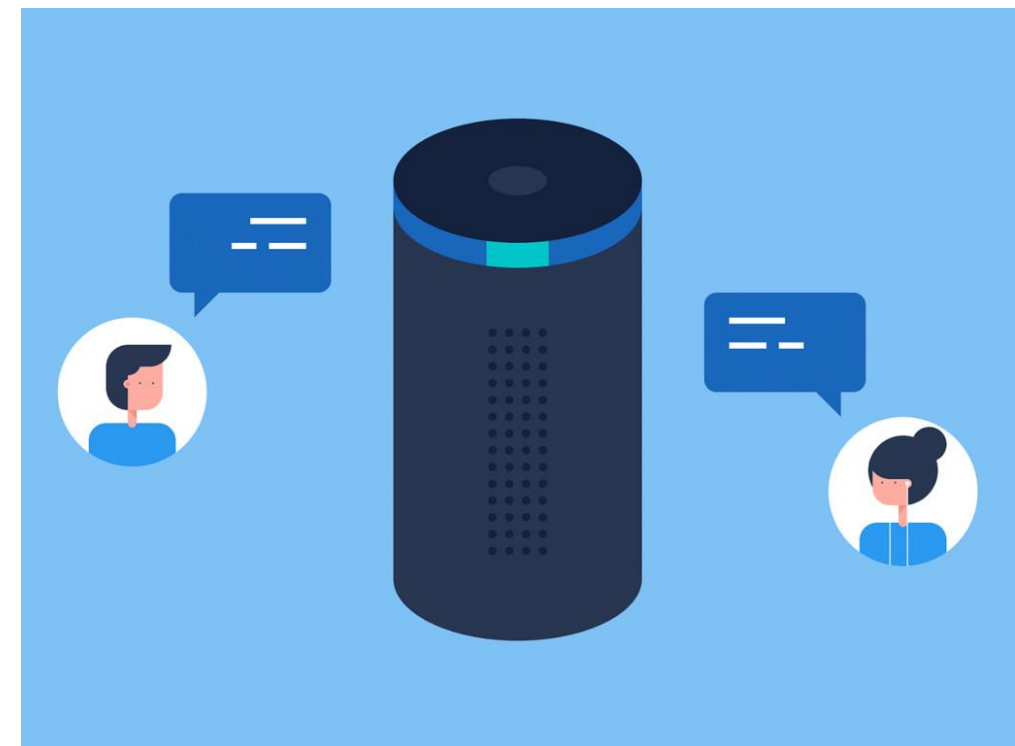
To date, we have not found a solution to global warming.



Transcrição

- A partir de um áudio transformar o som em texto.



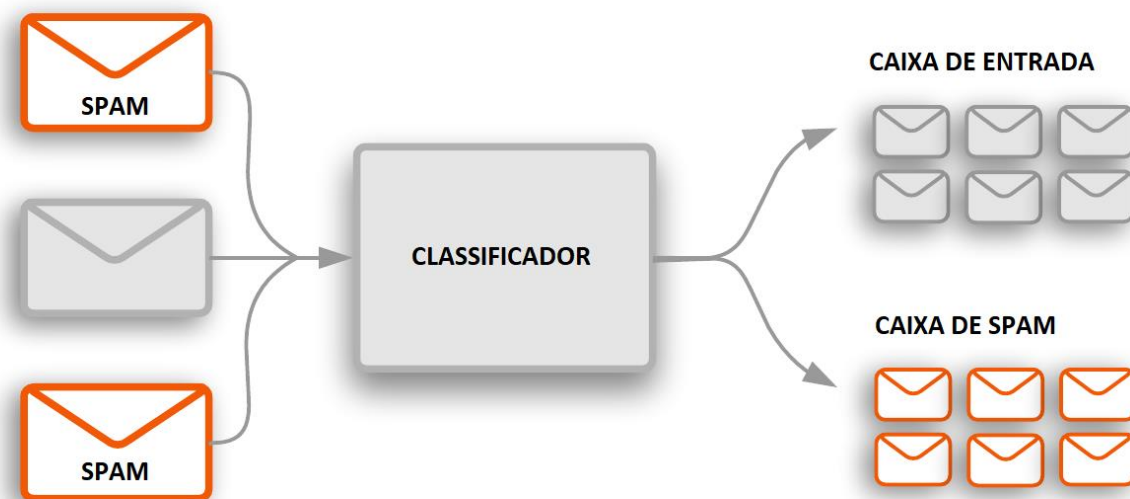
Chat bot

- Sistema de conversação.
- Pode ser um assistente virtual, ou um sistema de perguntas e respostas.



Classificação

- Determinar a classe de um texto.
- Detecção de spam, análise de sentimento.



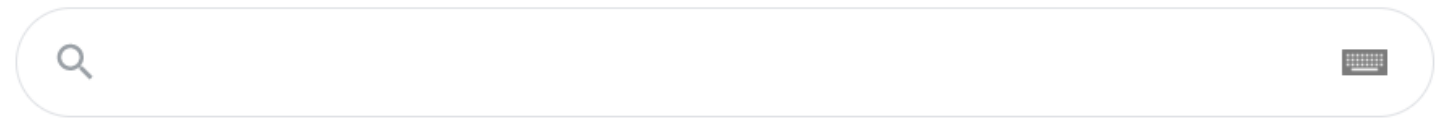
Clusterização

- Agrupar textos similares em grupos ou por proximidade.
- Modelagem de tópico.



Busca de informação

- Procurar em um conjunto de documentos a resposta para determinada solicitação.
- Google

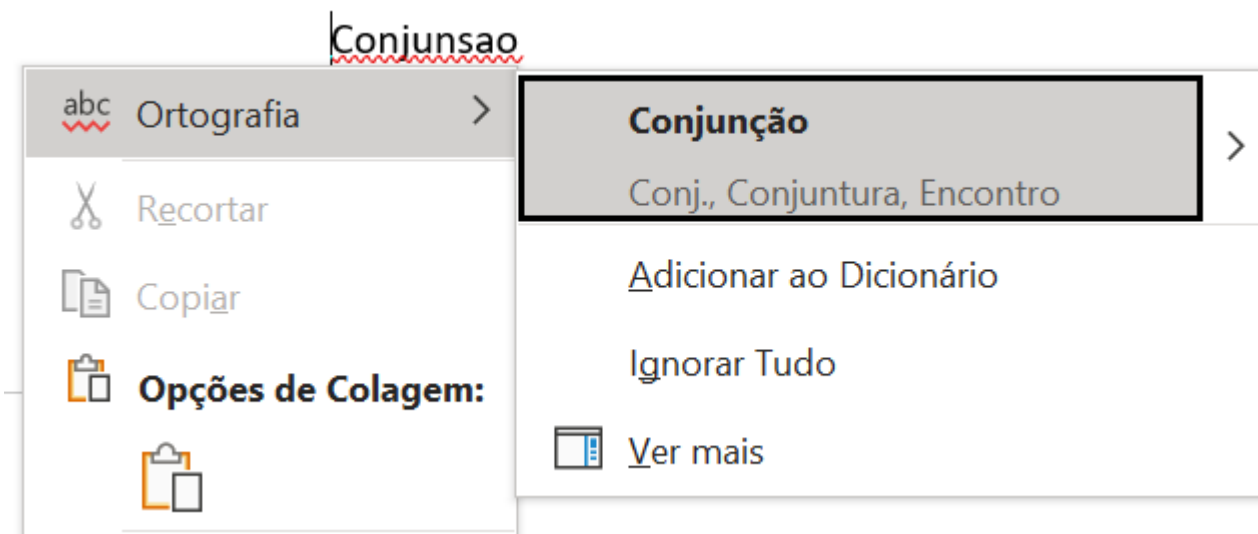


Pesquisa Google

Estou com sorte

Correção Ortográfica

- Substituir uma palavra errada por uma correta.



conjunsao



 Todas

 Imagens

 Vídeos

 Notícias

 Livros

 Mais

Ferramentas

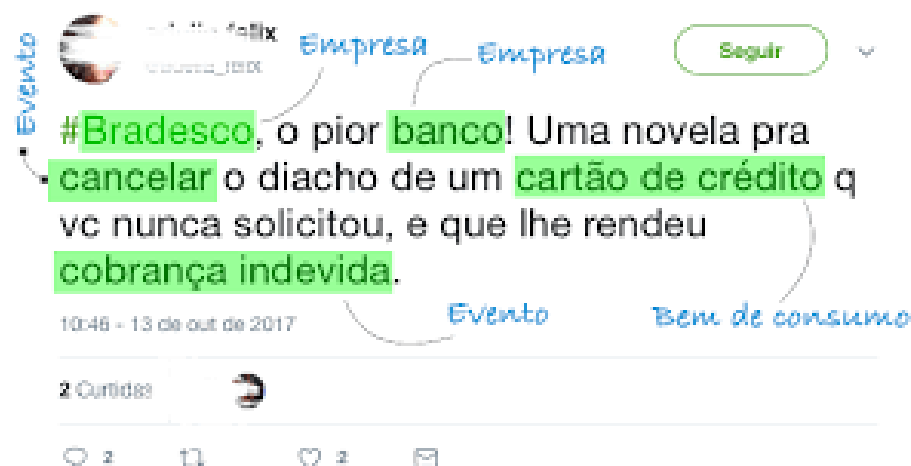
Aproximadamente 2.420.000 resultados (0,37 segundos)

Exibindo resultados para **conjunção**

Em vez disso, pesquisar por **conjunsao**

NER

- Identificar no texto algumas entidades.
- Por exemplo, identificar nome de cidades, nome de pessoas, datas.



Albert Einstein **PER** Albert Einstein was born in **Ulm LOC** in **Germany LOC** on March 14, 1879. Six weeks later the family moved to **Munich LOC**, where he later on began his schooling at the **Luitpold Gymnasium ORG**. In 1896 he entered the **Swiss Federal Polytechnic School ORG** in **Zurich LOC** to be trained as a teacher in physics and mathematics.

Bibliografia

- Machine Translation - <https://web.stanford.edu/class/linguist289/weaver001.pdf>
- Teste Turing - <https://academic.oup.com/mind/article/LIX/236/433/986238>
- ELIZA - <https://sites.google.com/view/elizagen-org/the-original-eliza>
- Gramática Gerativa - https://pt.wikipedia.org/wiki/Gram%C3%A1tica_gerativa



Semantix[®]

All about data

filipe.theodoro@semantix.com.br

www.semantix.com.br