

Fundamentos do Processamento de Linguagem Natural



Flávio Belizário da Silva Mota
Isabela Neves Drummond

O que é a linguagem natural?

Linguagem que usamos para nos expressar, baseada em um conjunto de protocolos acordados mutuamente envolvendo palavras e sons que usamos para nos comunicarmos



A CHEGADA. Direção: Denis Villeneuve. Produção de Paramount Pictures. Canadá: Sony Pictures, 2016. (116 min.)

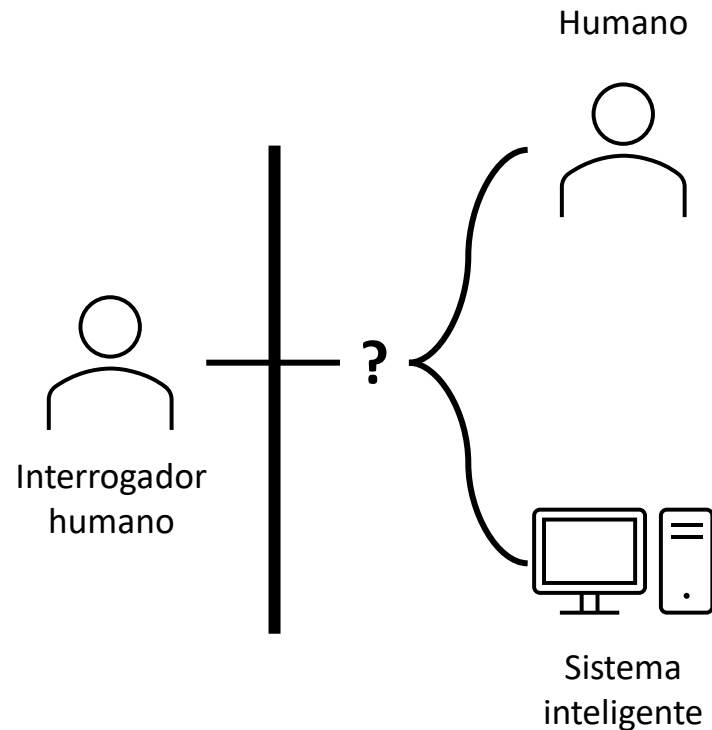
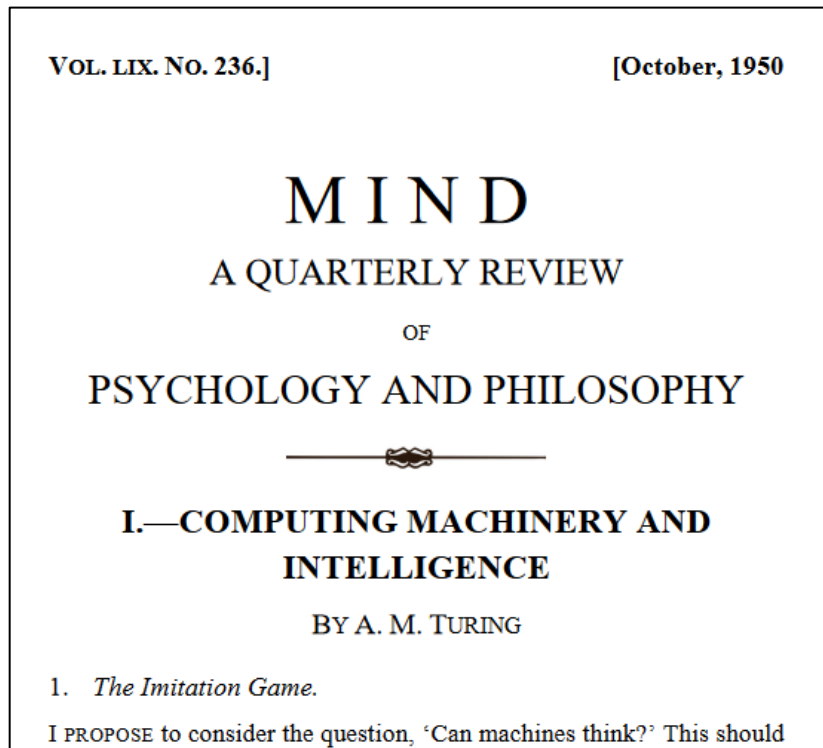
Processamento de Linguagem Natural (PLN)

“- Abra as portas da nave, HAL.
- Sinto muito, Dave, mas temo não
ser capaz de abri-las.”



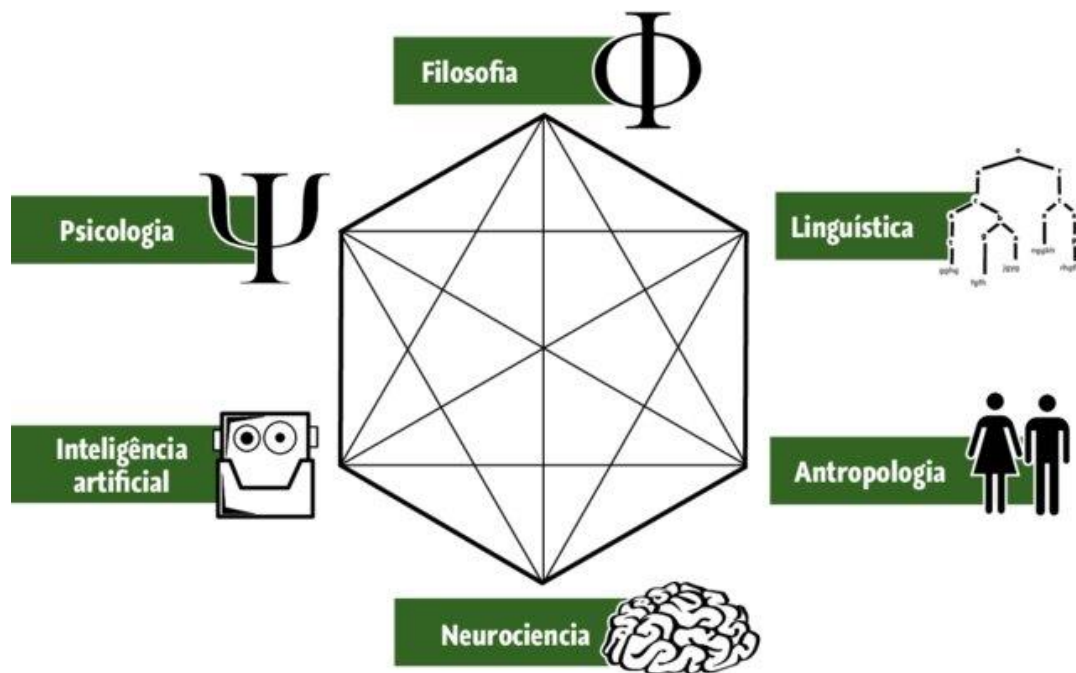
2001: UMA ODISSEIA NO ESPAÇO. Direção: Stanley Kubrick. Produção de Metro-Goldwyn-Mayer.
Estados Unidos: Metro-Goldwyn-Mayer, 1968. (148 min.)

O teste de Turing



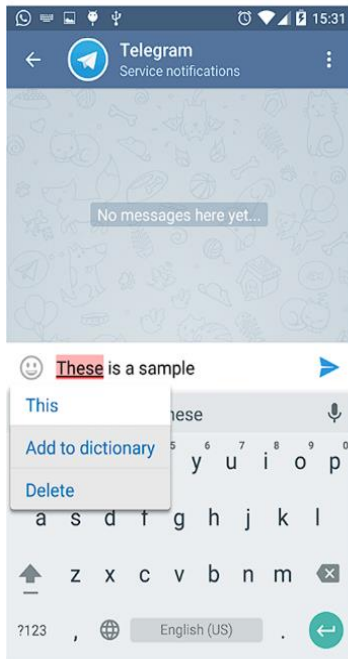
Computing Machinery and Intelligence

Uma ciência cognitiva

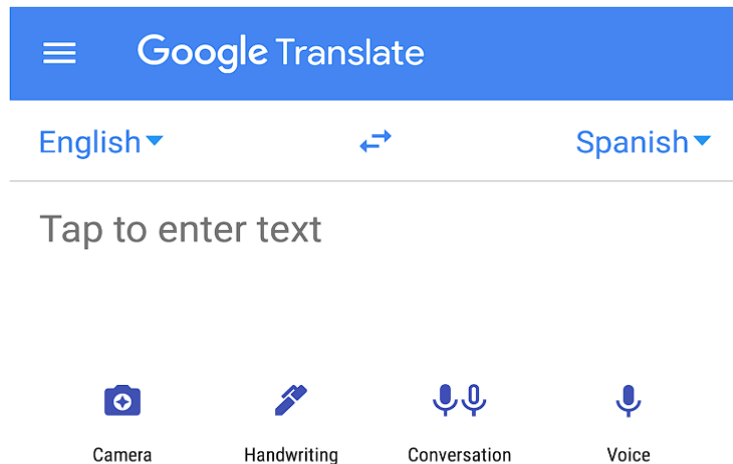


Aplicações

Corretores Automáticos e Auto-Complete



Tradução

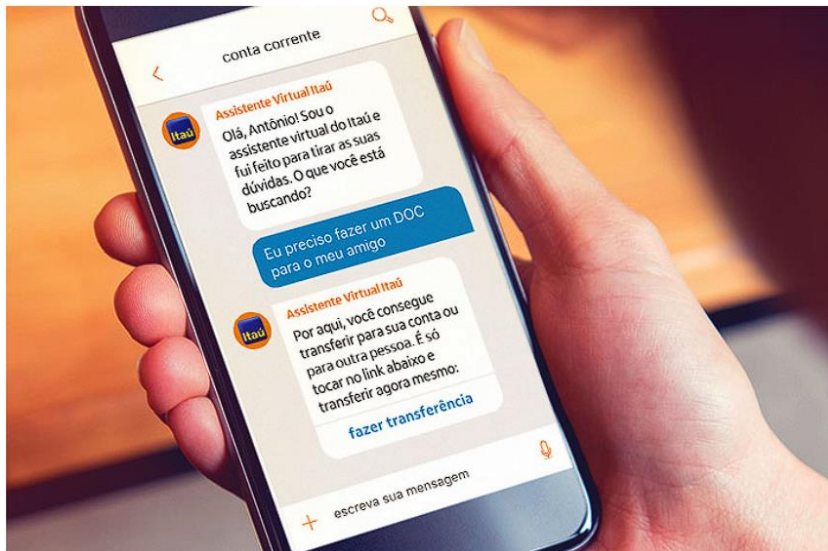


Aplicações

Bots de Mensagens Assistentes Pessoais



"Alexa, quando é a minha próxima reunião?"



Aplicações

ChatGPT



INTELIGÊNCIA ARTIFICIAL

Por que a GPT-3 é o melhor e o pior da IA atualmente

A linguagem de Inteligência Artificial da Open AI impressionou o público com seu aparente domínio do inglês — mas será tudo uma ilusão?

by MIT Technology Review

Abril 1, 2021

<https://mittechreview.com.br/por-que-a-gpt-3-e-o-melhor-e-o-pior-da-ia-atualmente/>

Aplicações

HUMANOS E TECNOLOGIA

IA e ChatGPT são criações inevitáveis do nosso tempo, assim como os benefícios da sua democratização

Notícias sobre IA e chats costumam acender discussões sobre o futuro do emprego, assim como aconteceu diversas vezes na sociedade: da invenção da lâmpada elétrica à revolução industrial. IA é a revolução do nosso tempo, e seus benefícios para a sociedade podem ser muito maiores do que os riscos apontados por alarmistas.

by **Fernando Teixeira**

Junho 22, 2023

Sabendo que a IA e suas formas generativas são a inevitável revolução do nosso tempo, precisamos focar em como extrair valor, incluir pessoas, criar profissões e, assim, beneficiar a sociedade. O mundo está cheio de problemas para resolver e não faltarão algoritmos para ajudar.

<https://mittechreview.com.br/ia-e-chatgpt-sao-criacoes-inevitaveis-do-nosso-tempo-assim-como-os-beneficios-da-sua-democratizacao/>

Fases de um projeto de PLN



Prática - Google Colab



[Introdução ao Processamento de Linguagem Natural](#)

Extração de características

Bag of Words

Tem como objetivo construir uma matriz onde cada coluna representa uma palavra do vocabulário de um documento e cada linha um texto desse documento

Texto 1: “Eu gosto do detetive Sherlock Holmes.”

Texto 2: “Sherlock Holmes não é um detetive, ele é um buscador da verdade.”

	Eu	gosto	do	detetive	Sherlock	Holmes	não	é	um	ele	buscador	da	verdade
Texto 1	1	1	1	1	1	1	0	0	0	0	0	0	0
Texto 2	0	0	0	1	1	1	1	2	2	1	1	1	1

Extração de características

TF-IDF - Term Frequency - Inverse Document Frequency

- **TF = Frequência do Termo:**

- Se um documento tem 100 palavras e a palavra "cachorro" aparece 5 vezes, então o TF para "cachorro" é $5/100 = 0.05$.

- **IDF = Frequência Inversa em Documentos**

- Se há um total de 1.000.000 de documentos e a palavra "cachorro" aparece em 1.000 documentos, então o IDF para "cachorro" é $\log(1.000.000 / 1.000) = \log(1000) \approx 3$.

- **TF-IDF = TF x IDF**

- TF-IDF para a palavra "cachorro" no exemplo acima seria
 $0.05 \text{ (TF)} * 3 \text{ (IDF)} \approx 0.15$

Extração de características

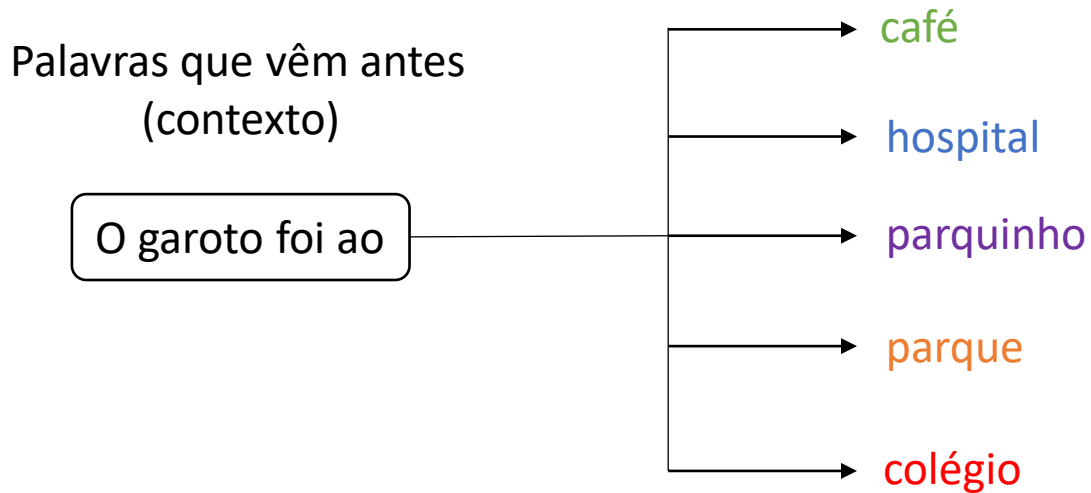
TF-IDF - Term Frequency - Inverse Document Frequency

Texto 1: “Eu gosto do detetive Sherlock Holmes.”

Texto 2: “Sherlock Holmes não é um detetive, ele é um buscador da verdade.”

	Eu	gosto	do	detetive	Sherlock	Holmes	não	é	um	ele	buscador	da	verdade
Frase 1	0.116	0.116	0.116	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Frase 2	0.000	0.000	0.000	0.000	0.000	0.000	0.063	0.063	0.126	0.063	0.063	0.063	0.063

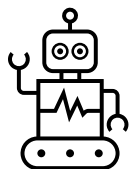
E os LLMs (Large Language Models)?



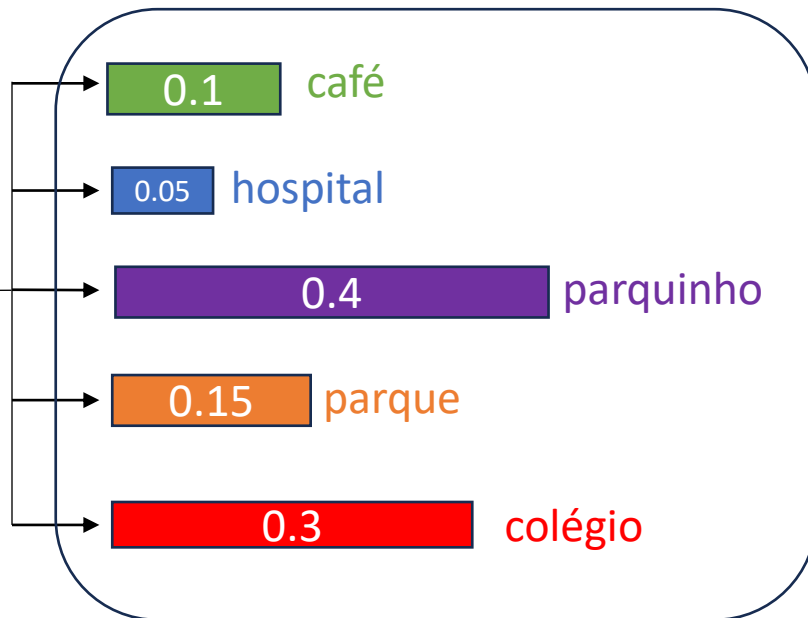
E os LLMs (Large Language Models)?

Palavras que vêm antes
(contexto)

O garoto foi ao



LLM
(GPT, Llama, Gemini)



Palavra com a maior probabilidade
é escolhida

Indicações e Referências

