



Introdução ao Processamento de Linguagem Natural

Prof. Marcelo Pita

Pós em Inteligência Artificial
Processamento de Linguagem Natural



Linguagem

Linguagem é um sistema convencional de símbolos falados ou escritos por meio dos quais seres humanos, como membros de um grupo social e participantes na sua cultura, se expressam.

([Encyclopædia Britannica, 2019](#))



Processamento de Linguagem Natural

PLN é um campo do conhecimento interdisciplinar convergente entre Computação, IA e Linguística, e que lida com o tratamento, processamento e modelagem computacional da linguagem humana.

O objetivo do PLN é fazer computadores executarem tarefas úteis que envolvem linguagem natural.

- Exemplos: tradução automática, geração de resumos, etc.

Por que PLN é interessante?

- 1) Linguagem está envolvida em muitas atividades humanas
- 2) Volume crescente de **texto** produzido no mundo

<https://www.internetlivestats.com/one-second/>



Por que PLN é difícil?

Linguagem é ambígua!

“O atleta falou ao treinador caído no chão.”

“Ouvi falar da festa no restaurante.”

“A vaca se diverte com a pata na lama.”

Ambiguidade em muitos níveis.

- Léxico, sintático, semântico.

Por que PLN é difícil?

Sarcasmos e trocadilhos

“Alguns causam felicidade aonde quer que vão. Outros causam sempre que se vão” (Oscar Wilde)

“Tudo na vida muda, até a bermuda!”

Fonte: <http://professor.ufabc.edu.br/~jesus.mena/courses/pln-2q-2019/PLN-aula01.pdf>

Vol. LIX. No. 236.]

[October, 1950

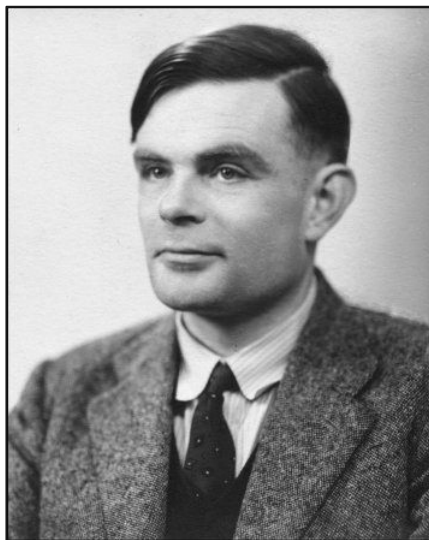
MIND
A QUARTERLY REVIEW
OF
PSYCHOLOGY AND PHILOSOPHY

I.—COMPUTING MACHINERY AND
INTELLIGENCE

By A. M. TURING

1. *The Imitation Game.*

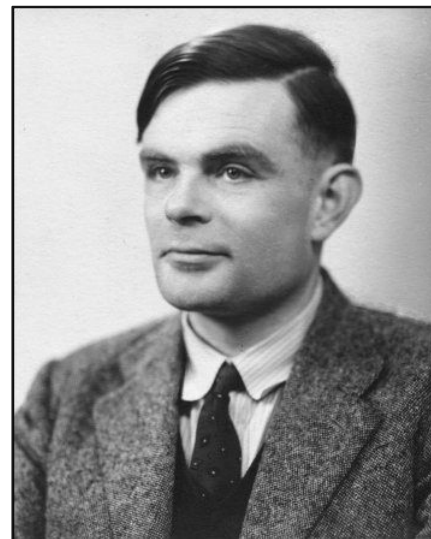
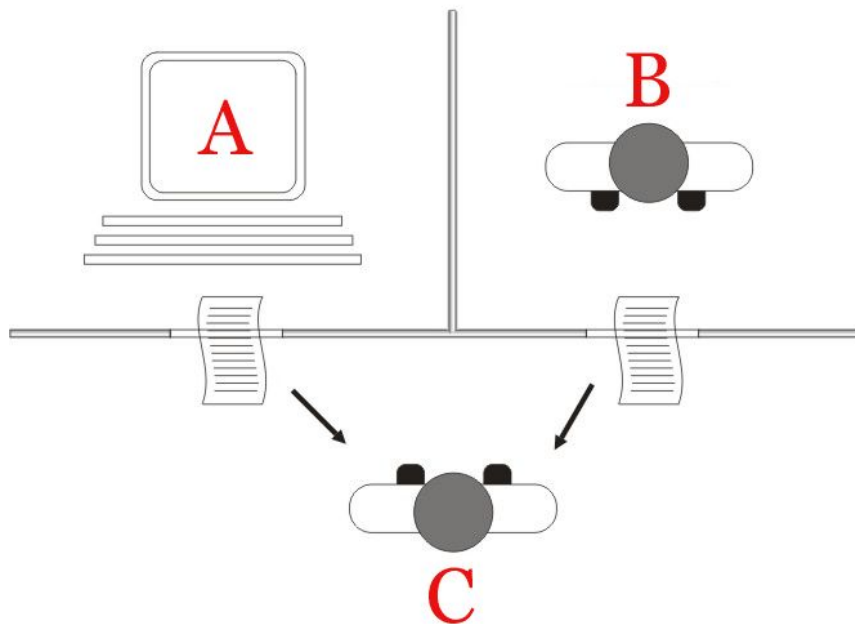
I PROPOSE to consider the question, 'Can machines think?' This should begin with definitions of the meaning of the terms 'machine' and 'think'. The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous. If the meaning of the words 'machine' and 'think' are to be found by examining how they are commonly



Alan Turing

História

Teste de Turing



Alan Turing

História



Trabalhos de Chomsky em linguagens formais (1950-1960)

Experimentos de tradução automática de frases do russo para o inglês (1954)

Ontologias (anos 1970)

Algoritmos de aprendizado de máquina (fim dos anos 1980)

Modelos probabilísticos de linguagem (anos 1990)

Aprendizado de máquina em larga escala (2000 - atualmente)

Principais atividades

- Estemização
- Lematização
- Parsing
- Marcação de partes do discurso
- Análise de semântica
 - Reconhecimento de entidades nomeadas (NER)
 - Resolução de correspondências
 - Rotulação de funções semânticas (SRL)
 - Sumarização automática
 - Tradução de máquina

Principais atividades

- Representação de texto
 - Texto natural e n-gramas
 - Modelos vetoriais (BoW, TF, TF-IDF)
 - Vetores de palavras (*word embeddings*)
 - Medidas de similaridade textual
- Aprendizagem de máquina sobre texto
 - Agrupamento de texto
 - Descoberta de tópicos
 - Classificação de texto
 - Análise de sentimento

Principais aplicações

Recuperação de informação (máquinas de busca)

Resolução de perguntas

Sumarização automática

Tradução de máquina

Extração de informação

Resolução de perguntas

“Quando foi a fundação da Ibratec?”

[Google it!](#)

Resolução de perguntas



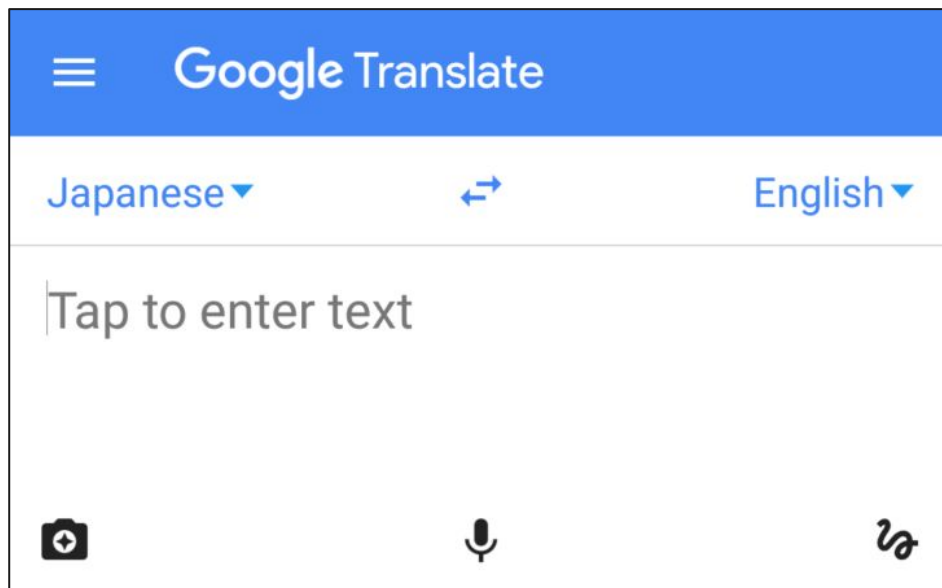
Sumarização automática

Sistemas que transformam um texto longo em um texto mais curto, idealmente retendo as informações mais importantes.

Exemplos online: [Resoomer](#), [TextSummarization](#), ...

O Word possui mecanismo de sumarização.

Tradução de máquina



Extração de informação (detecção de entidades)

“Presidente Obama visitou Cuba em 2016”

*[PESSOA: Presidente Obama, visitar, LOCALIZAÇÃO: Cuba, em,
TEMPO: 2016]*

PLN pode errar?

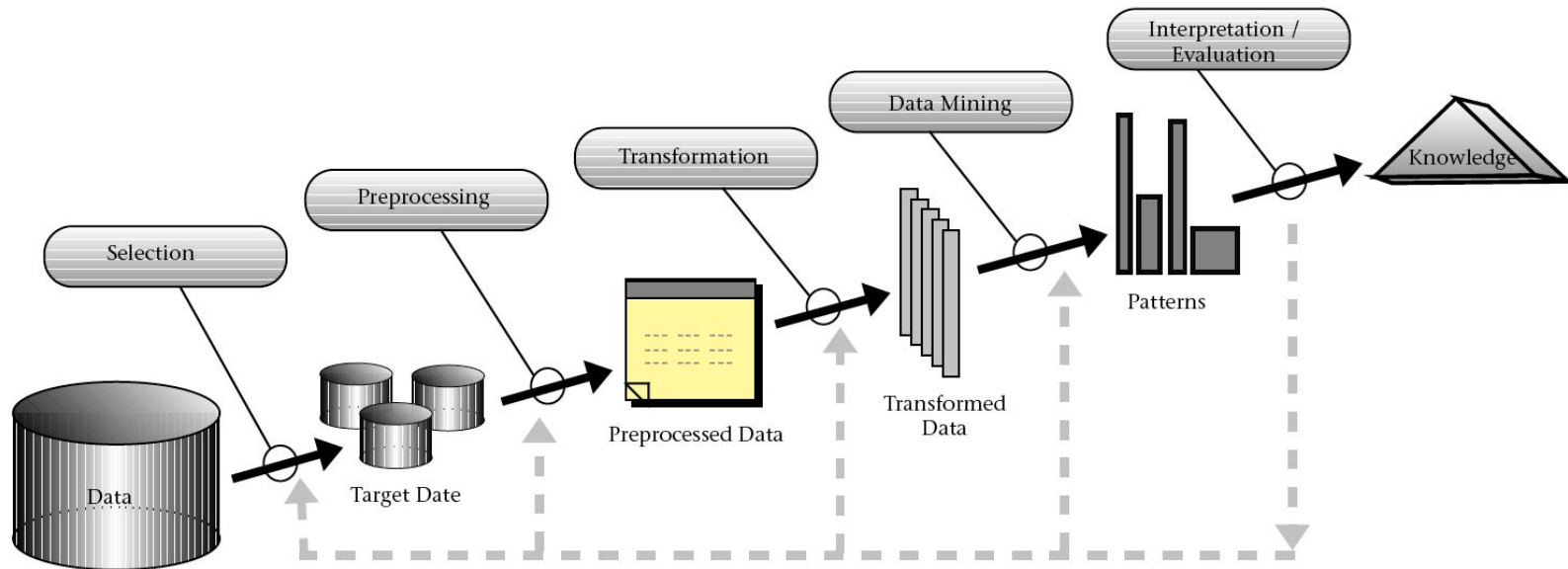
Chatbot (sistema de conversação) criado pela Microsoft em 2016 no Twitter chamado @TayandYou.

Objetivo: aprender a conversar com usuários do Twitter.

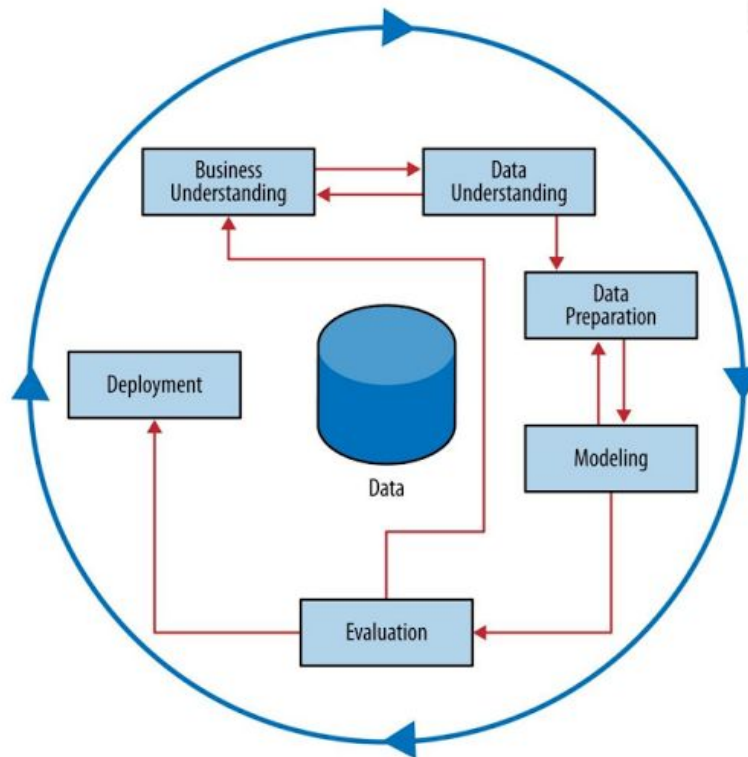
Tay aprender a reproduzir conversações racistas e de ódio.



Seguiremos o processo padrão de ciência de dados

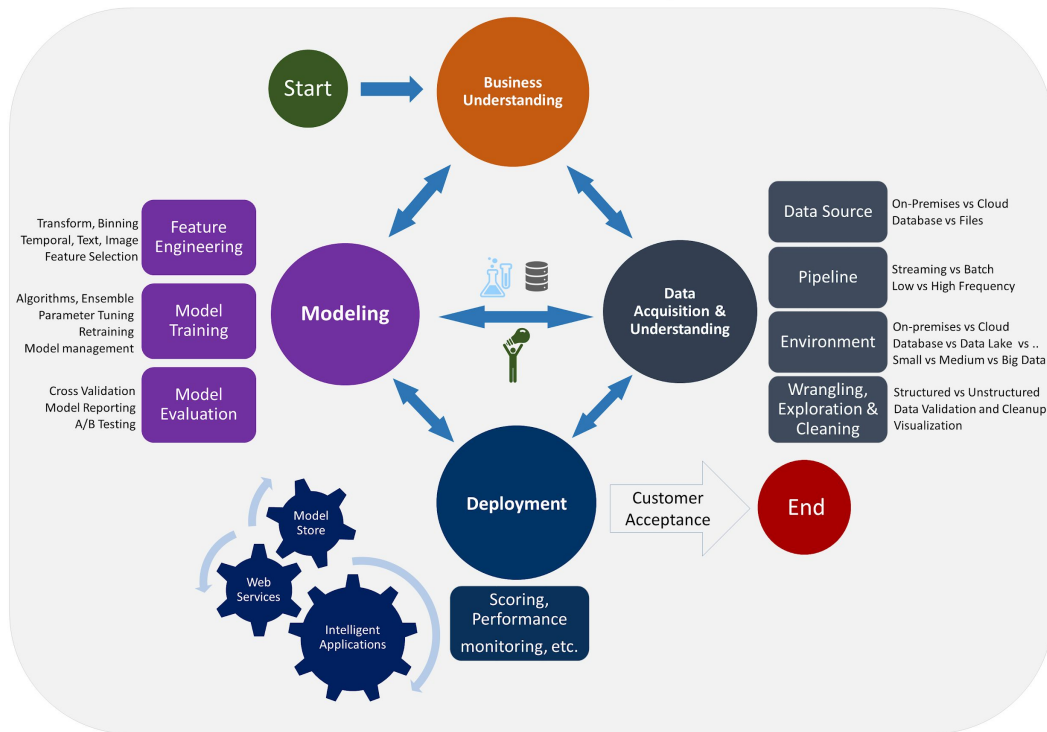


Seguiremos o processo padrão de ciência de dados



Seguiremos o processo padrão de ciência de dados

Data Science Lifecycle



Instalação do Python 3 com Jupyter Notebook



< Favor instalar as ferramentas >