

# Mineração de Texto e Web: Introdução ao Curso

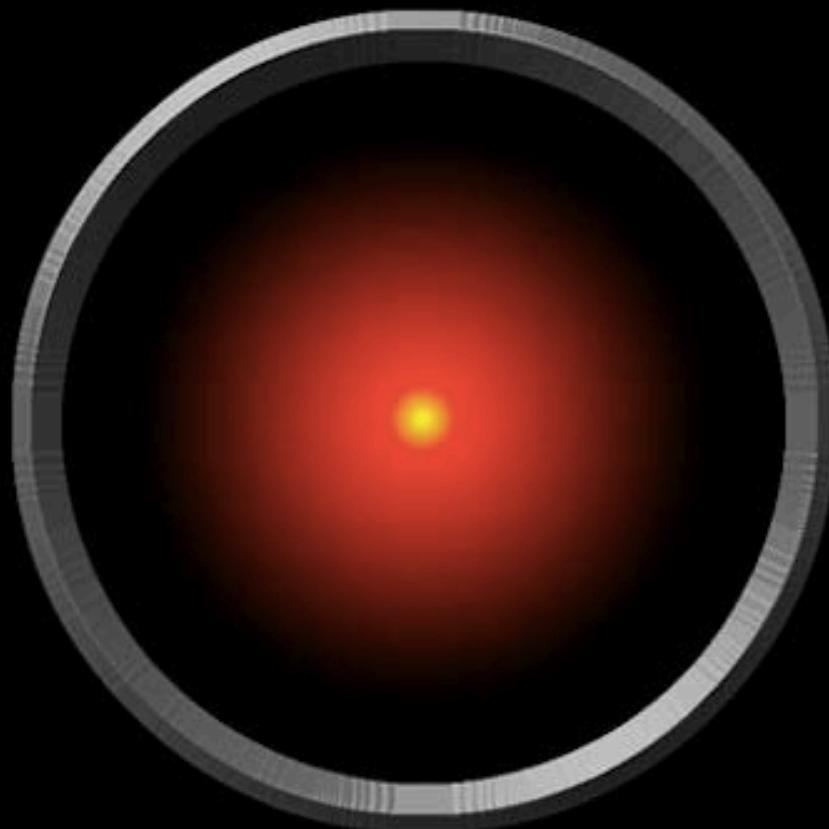
Luciano Barbosa

# O que é PLN

"Natural language processing (NLP) is a field of computer science, artificial intelligence (also called machine learning), and linguistics concerned with the interactions between computers and human (natural) languages. Specifically, the process of a computer extracting meaningful information from natural language input and/or producing natural language output "

Wikipedia

I'm sorry Dave,  
I'm afraid I can't do that.



# Por que é PLN Interessante

- Linguagem envolve várias atividades humanas
  - Leitura, escrita, fala, audição
- Extrair conhecimento de textos
  - Artigos científicos, notícias etc
- Muitas das tarefas de PLN são difíceis

# A Disciplina

- Objetivo:
  - Aprender principais técnicas, tarefas e aplicações de Processamento de Linguagem Natural
- Avaliação: projeto
- Página: [https://profluciano.github.io/mineracao\\_texto](https://profluciano.github.io/mineracao_texto)

# Projeto: Construção de um Sistema de Monitoramento de Reviews

- Seleção de um produto a ser monitorado
- Coleta de reviews de sites de e-commerce desse produto
  - Manual ou usando um parser HTML (ex., beautifulsoup)
  - Campos a serem extraídos das páginas do produto:
    - Conteúdo do review
    - Estrelas de avaliação
    - Data do review
- Criação de classificador de sentimentos
  - Random Forest com bag-of-words
  - Redes: CNN, LSTM e BERT
  - Pré-processamento: com e sem stemming
  - Métricas de avaliação: F1, precision e recall
- Construção de frontend (ex., <https://streamlit.io/>)
  - Plot de reviews positivos e negativos no tempo
  - Word cloud dos reviews
  - Reviews mais representativos

# Projeto: Construção de um Sistema de Monitoramento de Reviews

- Exemplo



<https://streamlit.io/>

# Bibliografia

- Speech and Language Processing (3rd ed. draft). Dan Jurafsky and James H. Martin
- Foundations of Statistical Natural Language Processing. Christopher D. Manning and Hinrich Schütze
- Natural Language Processing. Jacob Eisenstein

# Tópicos

- Processamento de texto
- Deep learning
- Classificação de texto
- Representação de palavras
- Information extraction
- Language modeling
- Modelos contextuais (Transformers)

# Processamento de Texto: Segmentação de Palavras

- Primeira tarefa em um sistema de PLN

毎年うちの研究室の学生が1-2名国語研でアルバイトさせてもらっているので、今日は新しくアルバイトするB4学生の紹介である。

# Classificação de Texto: Análise de Sentimentos

*"I am happy with this water bottle."*



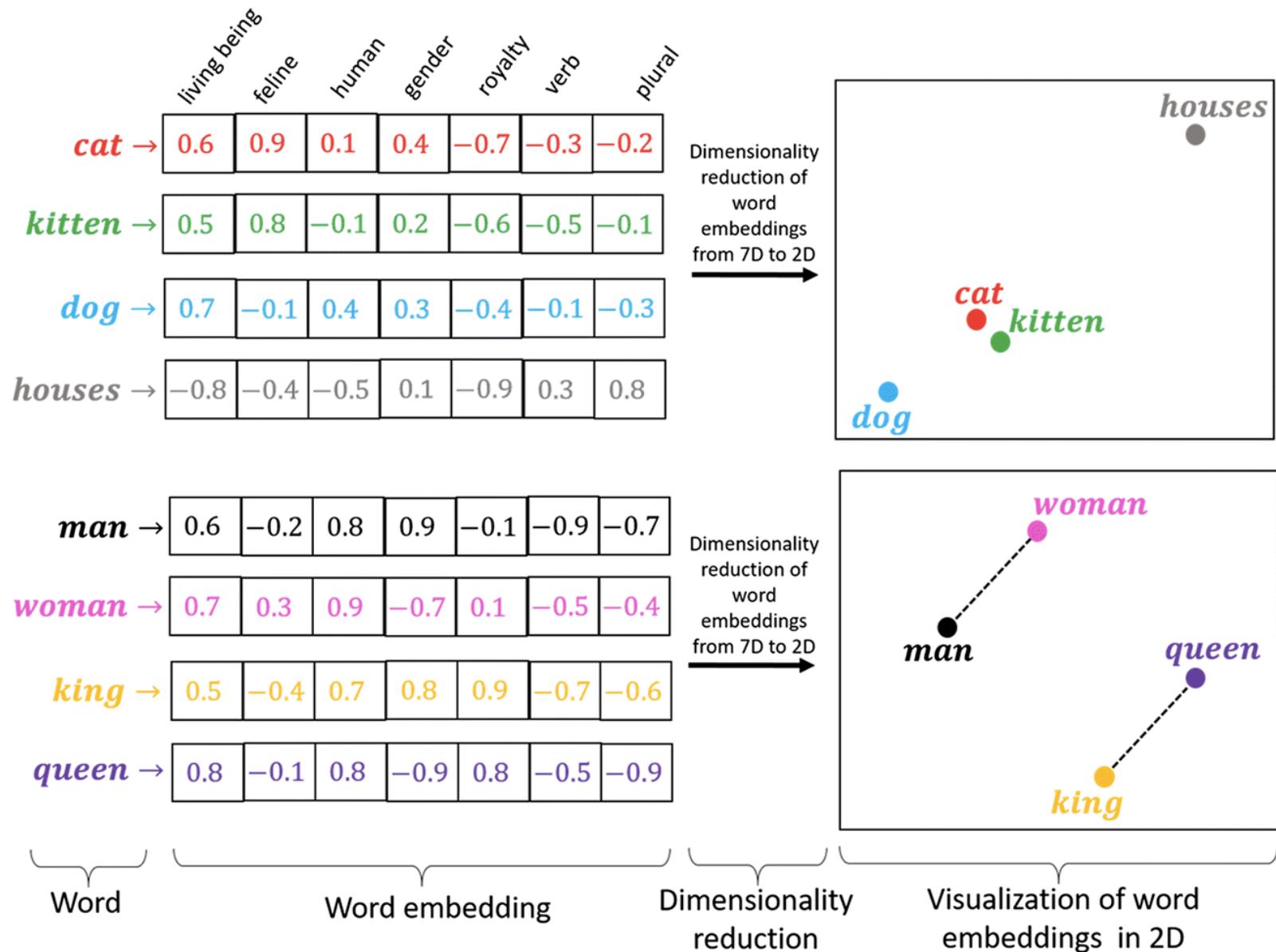
*"This is a bad investment."*



*"I am going to walk today."*



# Representação de Palavras: Embeddings



# Information Extraction

We have decided to meet tomorrow at 10:00am in the lab.

To do : meeting

Time : 10:00 am, 22/3/2012

Venue : Lab

# Language Model



natural language processing

natural language processing **algorithms**

natural language processing **tutorial**

natural language processing **python**

natural language processing **machine le**

natural language processing **course**

natural language processing **examples**

natural language processing **application**

natural language processing **pdf**

# Modelos Contextuais: Self-Attention

