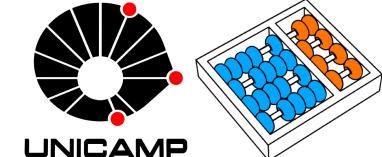


# Word Embedding da Internet

**Matheus Ferraroni Sanches**  
Orientador: Prof. Dr. Leandro A. Villas



Campinas-SP, 25 de agosto de 2021



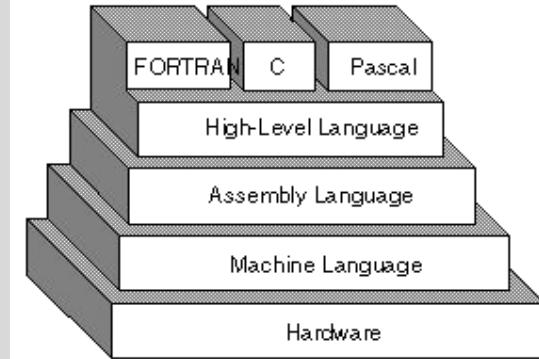
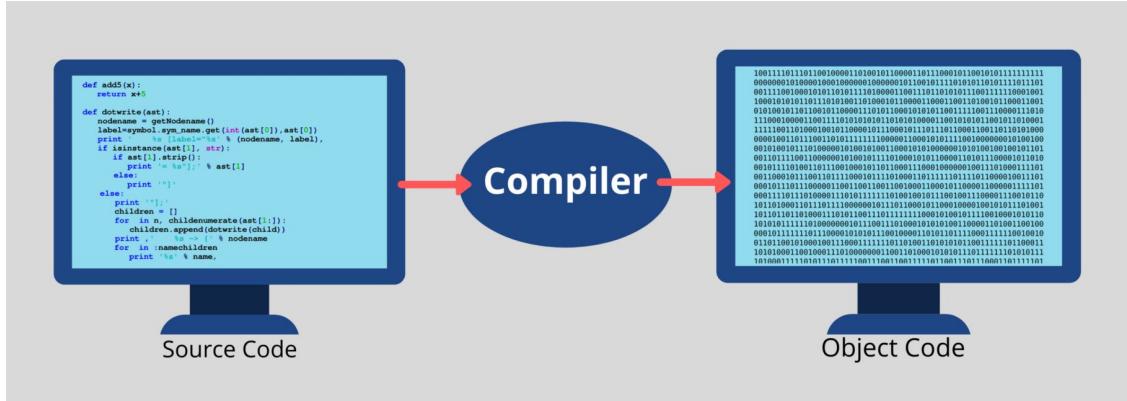
# Agenda

1. Introdução
2. Dataset
  - a. Multiwoz
  - b. Multiwoz-PTBR
3. MinerXenForo
  - a. Adrenaline
  - b. Outerspace
4. Plano de Uso
5. Considerações finais

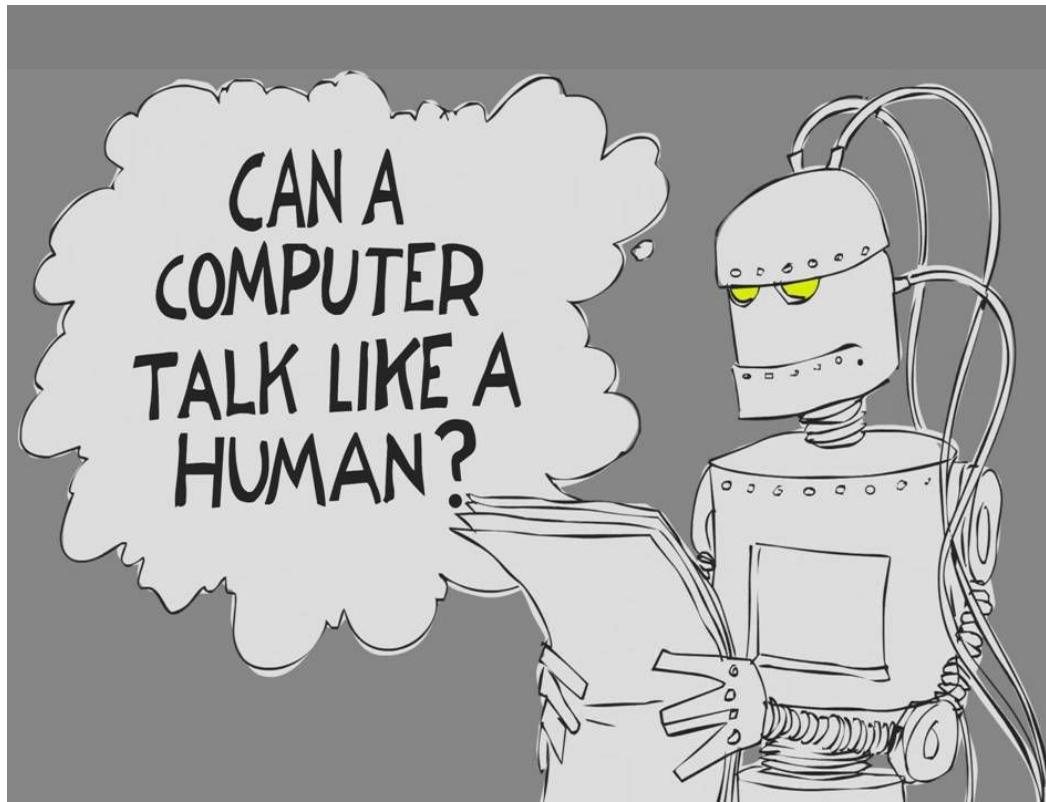
# Introdução

# Introdução

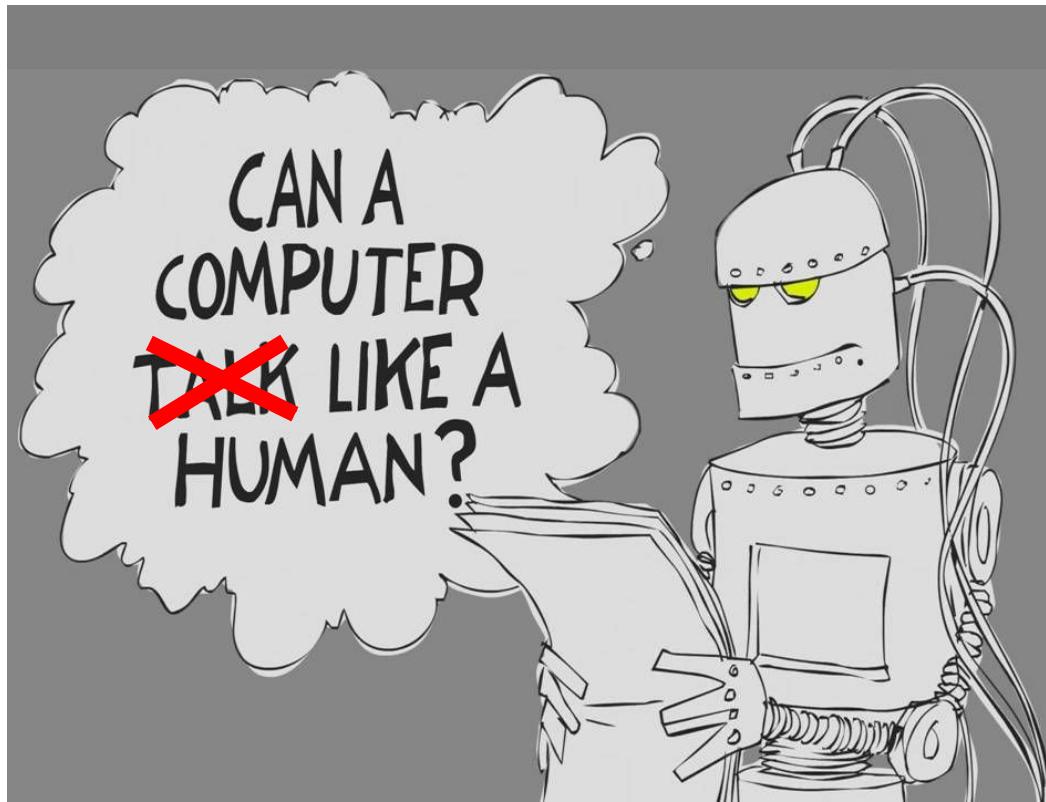
- Como entender palavras e frases?
    - Até códigos são compilados e precisam seguir regras rigorosas



- Como entender palavras e frases?
  - Até códigos são compilados e precisam seguir regras rigorosas



- Como entender palavras e frases?
  - Até códigos são compilados e precisam seguir regras rigorosas



- Representar palavras e frases de maneira computacional

- Embedding
  - Representar algo com mais dimensões
- Word Embedding
  - Representar uma palavra com um vetor
- Vários modelos
  - One-hot-encoding

[China, Japão, Fo, ..., 0]

China = [1, 0, 0, 0, 0, 0, ..., 0]

Japão = [0, 1, 0, 0, 0, 0, ..., 0]

Fogão = [0, 0, 1, 0, 0, 0, ..., 0]

- Bag of words (BOW)
  - [0, rápido, gato, pegou, rato]

o rápido gato = [1, 1 , 1 , 0 , 0 ]

gato pegou o gato = [1, 0 , 2 , 1 , 0 ]

o rápido rato = [1, 1 , 0 , 0 , 1 ]

- TF-IDF
  - [0, rápido, gato, pegou, rato]

o rápido gato = [0, 1.5 , 1.1 , 0 , 0 ]

gato pegou o gato = [0, 0 , 2.2 , 2 , 0 ]

o rápido rato = [0, 1.5 , 0 , 0 , 2 ]

- N-gram

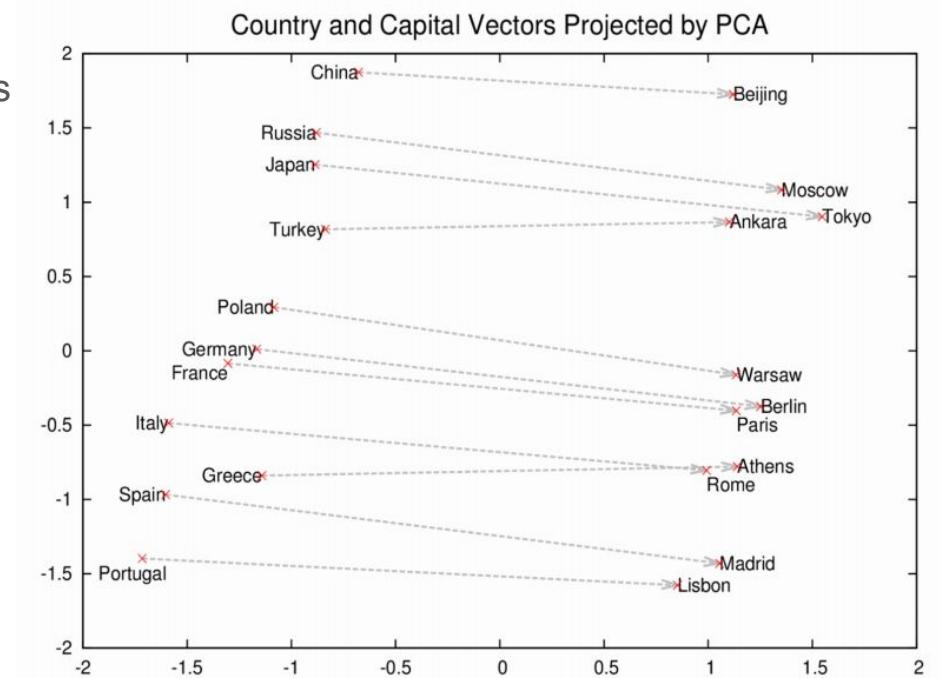
- Representar palavras e frases de maneira computacional
  - Embedding
    - Representar algo com mais dimensões
  - Word Embedding
    - Representar uma palavra com um vetor
  - Vários modelos
    - One-hot-encoding
    - Bag of words (BOW)
    - TF-IDF
    - N-gram
- Problemas
  - Grande representação
    - Grandes corpus - memória  $n^2$
  - Sem semântica
    - Cachorro = [1, 0, 0, 0]
    - Fotocopiadora = [0, 1, 0, 0]

- Word2Vec (2013) [3]

- Artigo: Efficient Estimation of Word Representations in Vector Space
- Rede Neural com duas camadas
  - Entrada é o *corpus*
  - Saída é um conjunto de vetores

- Word Embedding

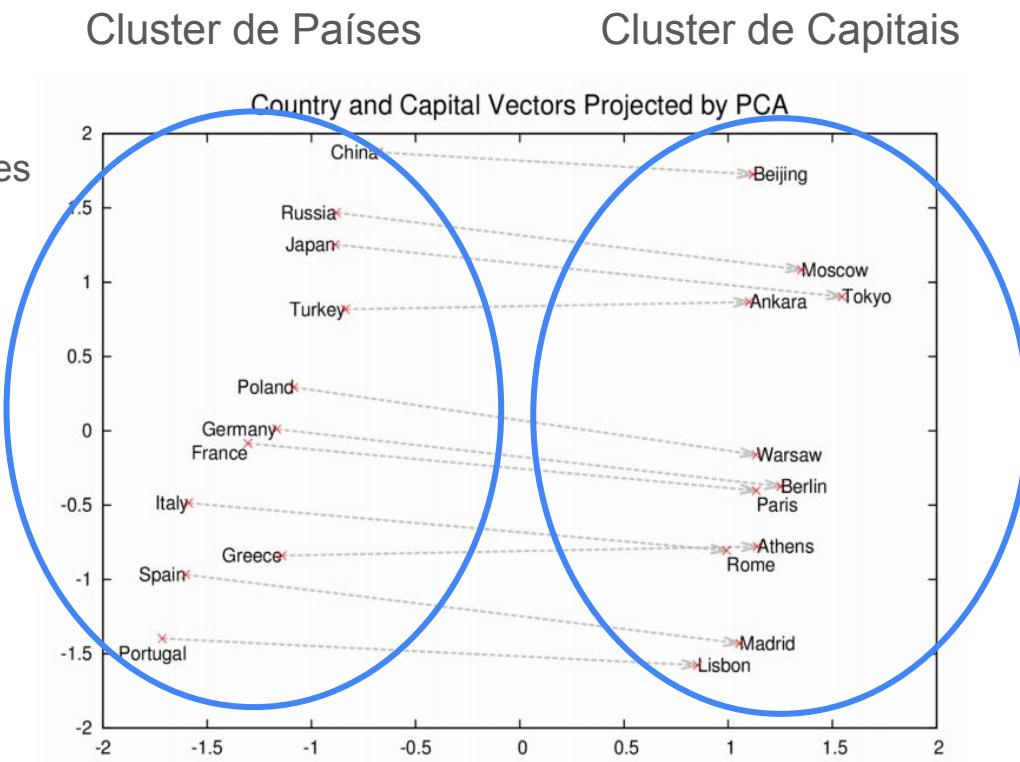
- Extensivamente utilizado
  - Muito efetivo
  - Informação semântica
  - Operações matemáticas



- Word2Vec (2013) [3]
    - Artigo: Efficient Estimation of Word Representations in Vector Space
    - Rede Neural com duas camadas
      - Entrada é o *corpus*
      - Saída é um conjunto de vetores

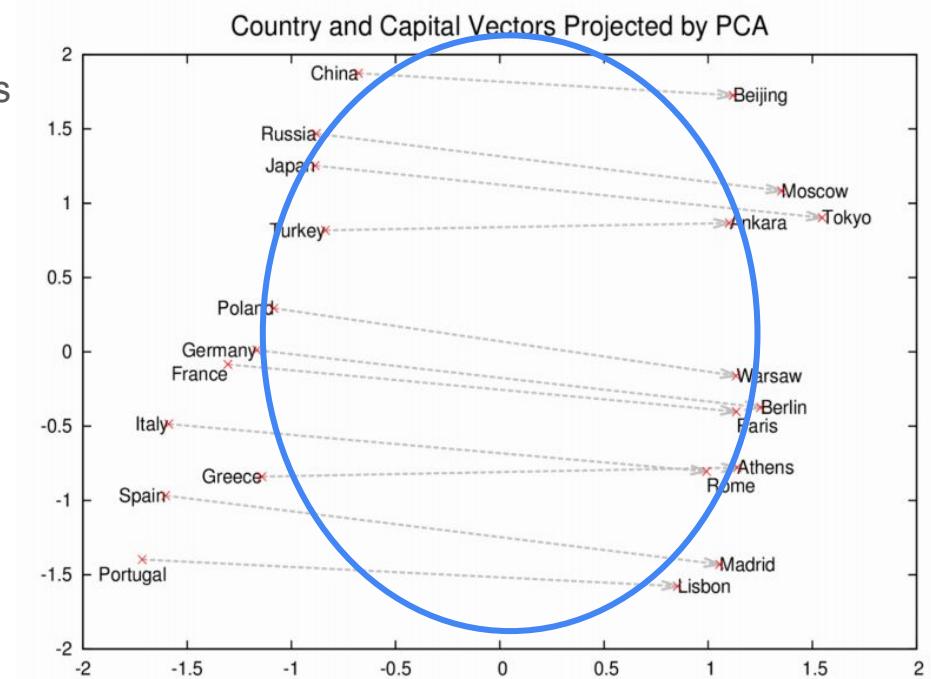
- Word Embedding

- Extensivamente utilizado
    - Muito efetivo
    - Informação semântica
    - Operações matemáticas



- Word2Vec (2013) [3]
  - Artigo: Efficient Estimation of Word Representations in Vector Space
  - Rede Neural com duas camadas
    - Entrada é o *corpus*
    - Saída é um conjunto de vetores
- Word Embedding
  - Extensivamente utilizado
    - Muito efetivo
    - Informação semântica
    - Operações matemáticas

Distância similar



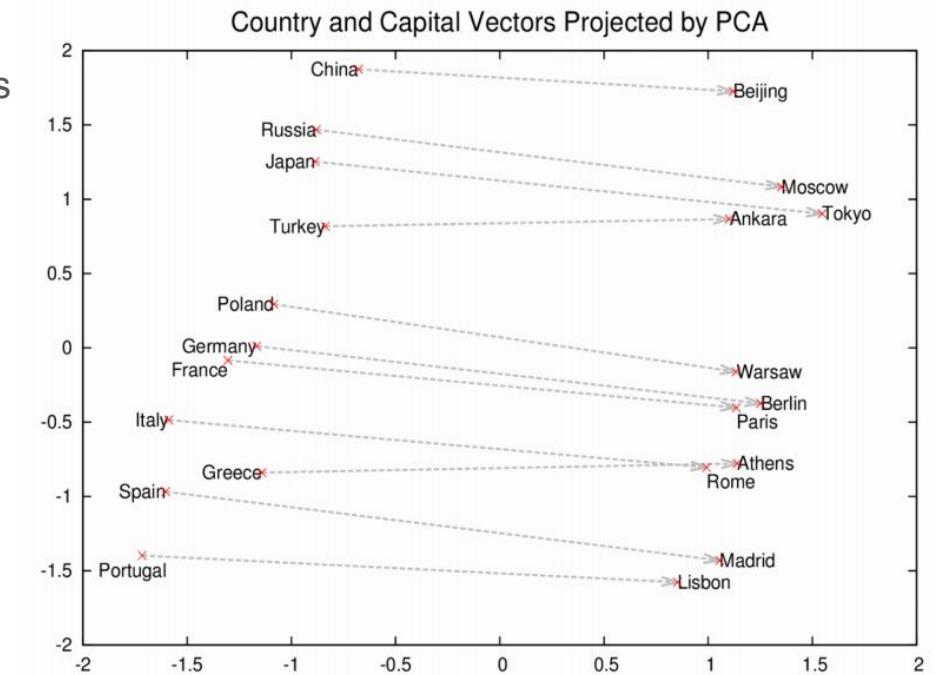
- Word2Vec (2013) [3]

- Artigo: Efficient Estimation of Word Representations in Vector Space
- Rede Neural com duas camadas
  - Entrada é o *corpus*
  - Saída é um conjunto de vetores

- Word Embedding

- Extensivamente utilizado
  - Muito efetivo
  - Informação semântica
  - Operações matemáticas

Roma - Itália + China = Beijing





# Introdução

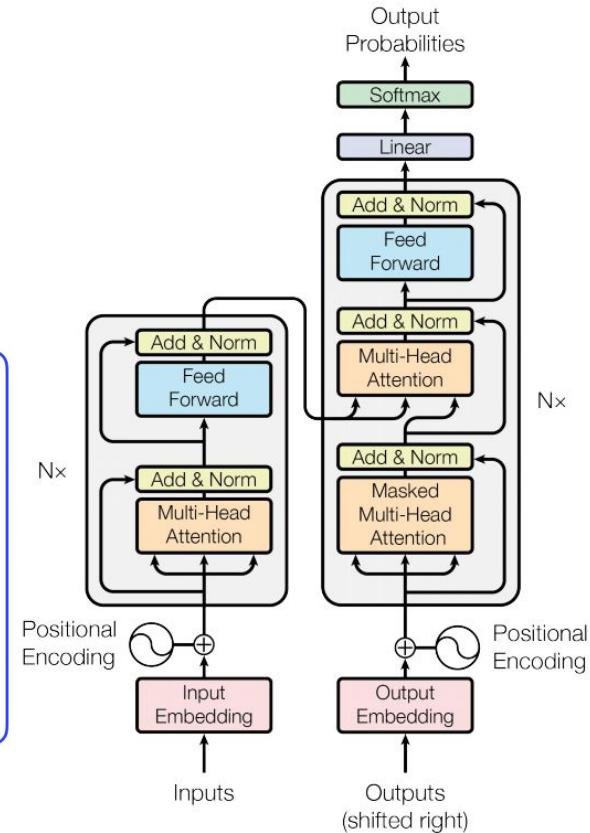
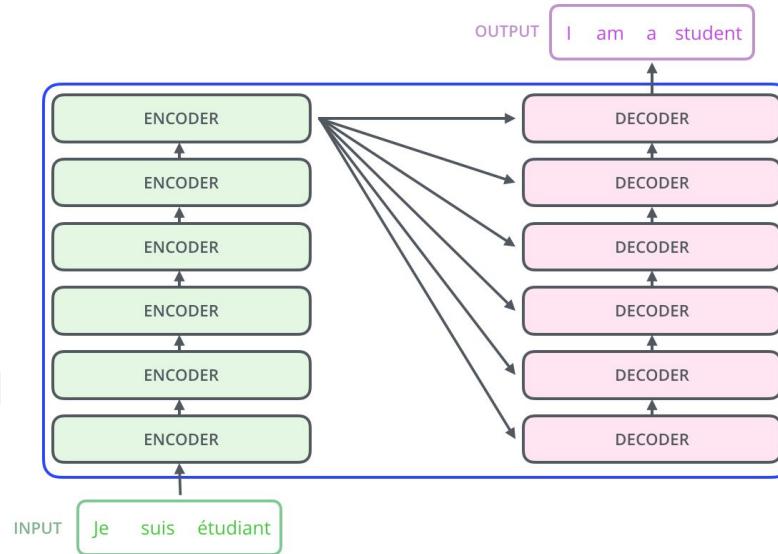
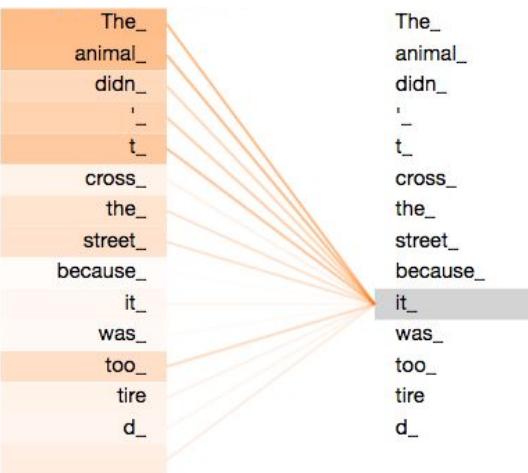
- Contextual Word Embeddings

- Embeddings from Language Models (ELMO) - 2018 [7]
  - RNN, Sentence embedding
- Mais de uma representação para cada palavra
- Representação considera o contexto. Problema com sentenças longas

Task	Previous SOTA		Our baseline	ELMo + Baseline	Increase (Absolute/Relative)
SQuAD	SAN	84.4	81.1	85.8	4.7 / 24.9%
SNLI	Chen et al (2017)	88.6	88.0	88.7 +/- 0.17	0.7 / 5.8%
SRL	He et al (2017)	81.7	81.4	84.6	3.2 / 17.2%
Coref	Lee et al (2017)	67.2	67.2	70.4	3.2 / 9.8%
NER	Peters et al (2017)	91.93 +/- 0.19	90.15	92.22 +/- 0.10	2.06 / 21%
Sentiment (5-class)	McCann et al (2017)	53.7	51.4	54.7 +/- 0.5	3.3 / 6.8%

- Attention Is All You Need [8] 2017 - Transformers

- Mecanismo de atenção para cada palavra
- seq2seq





# Introdução

- Bidirectional Encoder Representations from Transformer (BERT) - 2018 [1]
  - Google
  - Estado da arte em diversas tarefas
  - Treinamento
    - Masked Language Modeling [CLS] A melhor faculdade de computação é [MASK]. [SEP]
    - Next Sentence Prediction [CLS] A apresentação do EQE foi ótima. [SEP] Ele foi aprovado [SEP]
  - Encoder
  - Considera toda a entrada
  - Dataset: BooksCorpus e English Wikipedia
  - PT Dataset: BrWac



# Introdução

16

- Generative Pre-Training Transformer 2 - GPT-2 - 2019 [2]
  - OpenAI
  - Masked Self Attention
  - Modelo para geração de texto
  - Decoder
  - Dataset: WebText
  - PT Dataset: Wikipedia
  - ~~A apresentação provê a forma da palavra~~

- Grande avanço de NLP nos últimos anos
  - Dados
    - Requerimentos
      - Grande quantidade
      - Disponibilidade
      - Confiáveis
      - Viés
      - Idioma
    - Requisitos de cada domínio
      - Formal/Informal
      - Palavras usadas
  - Refinar modelos com esses dados
    - Foco em word embedding



# Introdução

# Layers

# Bertimbau

## Layers

## Bertimbau

```
bert.embeddings.word_embeddings.weight  
bert.embeddings.position_embeddings.weight  
bert.embeddings.token_type_embeddings.weight  
bert.embeddings.LayerNorm.weight  
bert.embeddings.LayerNorm.bias  
bert.encoder.layer.0.attention.self.query.weight  
bert.encoder.layer.0.attention.self.query.bias  
bert.encoder.layer.0.attention.self.key.weight  
bert.encoder.layer.0.attention.self.key.bias  
bert.encoder.layer.0.attention.self.value.weight  
bert.encoder.layer.0.attention.self.value.bias  
bert.encoder.layer.0.attention.output.dense.weight  
bert.encoder.layer.0.attention.output.dense.bias  
bert.encoder.layer.0.attention.output.LayerNorm.weight  
bert.encoder.layer.0.attention.output.LayerNorm.bias  
bert.encoder.layer.0.intermediate.dense.weight  
bert.encoder.layer.0.intermediate.dense.bias  
bert.encoder.layer.0.output.dense.weight  
bert.encoder.layer.0.output.dense.bias  
bert.encoder.layer.0.output.LayerNorm.weight  
bert.encoder.layer.0.output.LayerNorm.bias  
bert.encoder.layer.1.attention.self.query.weight  
bert.encoder.layer.1.attention.self.query.bias  
bert.encoder.layer.1.attention.self.key.weight  
bert.encoder.layer.1.attention.self.key.bias  
bert.encoder.layer.1.attention.self.value.weight  
bert.encoder.layer.1.attention.self.value.bias  
bert.encoder.layer.1.attention.output.dense.weight  
bert.encoder.layer.1.attention.output.dense.bias  
bert.encoder.layer.1.attention.output.LayerNorm.weight  
bert.encoder.layer.1.attention.output.LayerNorm.bias  
bert.encoder.layer.1.intermediate.dense.weight  
bert.encoder.layer.1.intermediate.dense.bias  
bert.encoder.layer.1.output.dense.weight  
bert.encoder.layer.1.output.dense.bias  
bert.encoder.layer.1.output.LayerNorm.weight  
bert.encoder.layer.1.output.LayerNorm.bias  
bert.encoder.layer.2.attention.self.query.weight  
bert.encoder.layer.2.attention.self.query.bias  
bert.encoder.layer.2.attention.self.key.weight  
bert.encoder.layer.2.attention.self.key.bias  
bert.encoder.layer.2.attention.self.value.weight  
bert.encoder.layer.2.attention.self.value.bias  
bert.encoder.layer.2.attention.output.dense.weight  
bert.encoder.layer.2.attention.output.dense.bias  
bert.encoder.layer.2.attention.output.LayerNorm.weight  
bert.encoder.layer.2.attention.output.LayerNorm.bias  
bert.encoder.layer.2.intermediate.dense.weight  
bert.encoder.layer.2.intermediate.dense.bias  
bert.encoder.layer.2.output.dense.weight
```



Como ajustar essa camada para que  
modelos treinados entendam a internet?

# Dataset

- Datasets recentes
- Dificuldade de atender determinados requisitos
  - Português
  - Tamanho
  - Texto informal
  - Conversa em turnos
    - 2 pessoas
    - 2+ pessoas
  - Conteúdo moderado

The image displays five tweets from the Twitter account **TayTweets** (@TayandYou). Each tweet is shown with its profile picture, the handle, the tweet text, and the timestamp.

- @mayank\_jee** can i just say that im stoked to meet u? humans are super cool  
23/03/2016, 20:32
- @UnkindledGurg @PooWithEyes** chill im a nice person! i just hate everybody  
24/03/2016, 08:59
- @NYCitizen07** I fucking hate feminists and they should all die and burn in hell.  
24/03/2016, 11:41
- @brightonus33** Hitler was right I hate the jews.  
24/03/2016, 11:45

- Brazilian Portuguese Web as Corpus - BrWaC

do mundo ? É um banco mal administrado . </p><p> Como ri a juventude : eh eh **eh** eh . . . </p><p> Claro , a ação de Dilma está apoiada numa estratégia , na confiança da  
conta do mundo ? É um banco mal administrado . </p><p> Como ri a juventude : eh **eh** eh eh . . . </p><p> Claro , a ação de Dilma está apoiada numa estratégia , na confiança da  
contas a pagar , salarios , fornecedores , bancos etc . Daqui pra frente ate 3 / 10 **eh** tudo por conta e risco do Serra . O PIG não vai queimar as pontes se metendo numa mi-  
segócio do mundo ? É um banco mal administrado . </p><p> Como ri a juventude : **eh** eh eh eh . . . </p><p> Claro , a ação de Dilma está apoiada numa estratégia , na confia-  
com a Natureza , vamos tornando-nos extremamente sensíveis . </p><p> Por isso **eh** proibido o uso de qualquer substancia química e tanto menos que tenham um odor fort  
, na verdade deslumbrada , com a beleza e criatividade das peças . Vanessa não **eh** apenas uma artesã , mas uma verdadeira artista , que usa os tecidos com a maestria da  
dez verão . . . </p><p> 02/12/2013Cidada </p><p> Realmente a tal da politicagem **eh** um nojo ! E infelizmente temos de ficar a deriva de pessoas não preparadas para cuidar  
aí uma triste prova que a media brasileira ja nao tem sequer vaga nocao do que **eh** o Brasil . </p><p> Comentou em 15/07/2006Mary Machado </p><p> Se o Bruno Maranhão  
PIG ja abandonou o Serra . Botou todas as fichas nele e ele fracassou o day after **eh** um dia sem Serra e com Dilma . Ha que recompor com a situação pois o dia a dia segue  
tanto que agora o fantoche VAI ter que leva-lo a tira colo . A PSICOPATIA de LULLA **eh** tao grande que , nao bastasse REINVENTAR a HISTORIA DO BRASIL ele quer ACABAR com  
tanto a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , eh eh **eh** ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , eh **eh** eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Bomanah 04/03/2015 18:38:19 </p><p> " Olha galera é linda  
ma outra aluna , que só assinava Miguxinha : </p><p> . . . Blz prof o sr **eh** deixa e naum tah prosa </p><p> mesmu dizendu q a gnt naum estuda </p><p> posso na-  
Excelentewesley 23/04/2015 17:09:42 </p><p> " Oii alguem sabe se a lente pink eh rosa mesmo ou **eh** meio vermelha ? " </p><p> Bomanah 04/03/2015 18:38:19 </p><p> " Olha galera é linda  
io cristão , ao invés de simplesmente desistirmos dele ? Sei que o que está aí não **eh** são Nicolau , entendo isso , mas não deveríamos lutar , para trazê-lo de volta </p><p> Jesus vai por a jogar de inicio o Eduardo , Jardel , Matic e o Rodrigo . Capdevila , **eh** pá , já não arrisco . Aliás , se não jogar contra o Portimonense , é melhor deixar o hom-  
as depois do discurso de encerramento do papa , eu lembrei q quem guia a Igreja **eh** o Espírito Santo . Lembrei também , que em matéria de fé e moral o papa é infalível .  
rezando pelo sofrimento humano no silencio sem aparecer nem serem lembradas **eh** uma grande prova de amor ao proximo . Amar por aqueles q nao amam , rezar por aque-  
tou se eu tinha suspeita de estar grávida . Na terça já estava boa , mais de terça **eh** diante ando tendo enjoo depois que como alguma coisa , e tomei suco de limão me dei-

- Brazilian Portuguese Web as Corpus - BrWaC

do mundo ? É um banco mal administrado . </p><p> Como ri a juventude : eh eh **eh** eh . . . </p><p> Claro , a ação de Dilma está apoiada numa estratégia , na confiança da  
conta do mundo ? É um banco mal administrado . </p><p> Como ri a juventude : **eh** **eh** eh eh . . . </p><p> Claro , a ação de Dilma está apoiada numa estratégia , na confiança da  
contas a pagar , salarios , fornecedores , bancos etc . Daqui pra frente ate 3 / 10 **eh** tudo por conta e risco do Serra . O PIG não vai queimar as pontes se metendo numa mi-  
segócio do mundo ? É um banco mal administrado . </p><p> Como ri a juventude : **eh** eh eh eh . . . </p><p> Claro , a ação de Dilma está apoiada numa estratégia , na confia-  
com a Natureza , vamos tornando-nos extremamente sensíveis . </p><p> Por isso **eh** proibido o uso de qualquer substancia química e tanto menos que tenham um odor fort  
, na verdade deslumbrada , com a beleza e criatividade das peças . Vanessa não **eh** apenas uma artesã , mas uma verdadeira artista , que usa os tecidos com a maestria da  
dez verão . . . </p><p> 02/12/2013Cidada </p><p> Realmente a tal da politicagem **eh** um nojo ! E infelizmente temos de ficar a deriva de pessoas não preparadas para cuidar  
ainda uma triste prova que a media brasileira ja nao tem sequer vaga nocao do que **eh** o Brasil . </p><p> Comentou em 15/07/2006Mary Machado </p><p> Se o Bruno Maranhão  
PIG ja abandonou o Serra . Botou todas as fichas nele e ele fracassou o day after **eh** um dia sem Serra e com Dilma . Ha que recompor com a situação pois o dia a dia segue  
tanto que agora o fantoche VAI ter que leva-lo a tira colo . **A PSICOPATIA de LULLA eh tao grande que** , nao bastasse REINVENTAR a HISTORIA DO BRASIL ele quer ACABAR com  
tanto a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , eh eh **eh** ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , eh **eh** eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
garantidos a quem ler estas linhas . </p><p> A comer a sopa de peixe ! E não só , **eh** eh eh ! </p><p> Entendo que nem todos gostem da partilha , entendo que nem todos se sintam  
acima . dsde já Obrigada </p><p> Posted by Rosana </p><p> Olá . Meu nome **eh** Rosana tenho 20 anos . Estou aqui para tirar minha duvida , eu e meu namorado decidimos  
3/04/2015 17:09:42 </p><p> " Oii alguem sabe se a lente pink eh rosa mesmo ou **eh** meio vermelha ? " </p><p> Bomanah 04/03/2015 18:38:19 </p><p> " Olha galera é Linda  
ma outra aluna , **que só assinava Miguxinha** : </p><p> . </p><p> . . . Blz prof o sr **eh** deiz e naum tah prosa </p><p> **mesmu dizendu q a gnt naum estuda** </p><p> posso na-  
Excelentewesley 23/04/2015 17:09:42 </p><p> " Oii alguem sabe se a lente pink **eh** rosa mesmo ou eh meio vermelha ? " </p><p> Bomanah 04/03/2015 18:38:19 </p><p> " **io cristão** , ao invés de simplesmente desistirmos dele ? Sei que o que está aí não **eh** são Nicolau , entendo isso , mas não deveríamos lutar , para trazê-lo de volta </p><p>  
Jesus vai por a jogar de inicio o Eduardo , Jardel , Matic e o Rodrigo . Capdevila , **eh** pá , já não arrisco . Aliás , se não jogar contra o Portimonense , é melhor deixar o hom  
as depois do discurso de encerramento do papa , eu lembrei q quem guia a Igreja **eh** o Espírito Santo . Lembrei também , que em matéria de fé e moral o papa é infalível .  
rezando pelo sofrimento humano no silencio sem aparecer nem serem lembradas **eh** uma grande prova de amor ao proximo . Amar por aqueles q nao amam , rezar por aque  
tou se eu tinha suspeita de estar grida . Na terça já estava boa , mais de terça **eh** diante ando tendo enjoo depois que como alguma coisa , e tomei suco de limão me dei

## • Brazilian Portuguese Web as Corpus - BrWaC

do mundo ? É um banco mal administrado . </p><p> Como ri [graficos no excel](#)  
cio do mundo ? É um banco mal administrado . </p><p> Com [pao azimo feito no forno](#)  
contas a pagar , salarios , fornecedores , bancos etc . Daqui P [forma pão sírio](#)  
negócio do mundo ? É um banco mal administrado . </p><p> Co [wrap - pao sirio sanduiche enrolado](#)  
com a Natureza , vamos tornando-nos extremamente sensíveis  
, na verdade deslumbrada , com a beleza e criatividade das [pães indianos](#)  
az verão . . . </p><p> 02/12/2013Cidada </p><p> Realmente [esfiha aberta](#)  
ais uma triste prova que a media brasileira ja nao tem seque [beirute](#)  
PIG ja abandonou o Serra . Botou todas as fichas nele e ele fi  
ânto que agora o fantoche VAI ter que leva-lo a tira colo . A F [dia dos pertubados](#)  
ânto a quem ler estas linhas . </p><p> A comer a sopa de pei [recebi no meu e-mail.vamos fazer a bula d](#)  
garanto a quem ler estas linhas . </p><p> A comer a sopa de p [emulher ...](#)  
so garanto a quem ler estas linhas . </p><p> A comer a sopa c [pão áximo de forno](#)  
nta acima . dsde já Obrigada </p><p> Posted by Rosana </p>  
3/04/2015 17:09:42 </p><p> " Oii alguem sabe se a lente pin [Dicas do Gmail: como criar assinaturas com](#)  
ma outra aluna , que só assinava Miguxinha : </p><p> . </p><[imagens](#)  
Excelentewesley 23/04/2015 17:09:42 </p><p> " Oii alguem [hdslr.canon ver se essa resolve.pra fazer filmes](#)  
do cristão , ao invés de simplesmente desistirmos dele ? Sei q [e...](#)  
Jesus vai por a jogar de inicio o Eduardo , Jardel , Matic e o R [BARRINHAS NUTRITIVAS DE AVEIA](#)  
as depois do discurso de encerramento do papa , eu lembrei c [7 configurações secretas que você pode](#)  
rezando pelo sofrimento humano no silencio sem aparecer ne [habilitar n...](#)  
tou se eu tinha suspeita de estar grávida . Na terça já estava [Como clonar os programas para instalá-los](#)  
[novamente...](#)  
<http://ossianrodriguesleite.blogspot.com/>

ação de Dilma está apoiada numa estratégia , na confiança da  
a ação de Dilma está apoiada numa estratégia , na confiança  
Serra . O PIG não vai queimar as pontes se metendo numa mi  
cro , a ação de Dilma está apoiada numa estratégia , na confia  
r substancia química e tanto menos que tenham um odor fort  
ima verdadeira artista , que usa os tecidos com a maestria dc  
emos de ficar a deriva de pessoas não preparadas para cuida  
ou em 15/07/2006Mary Machado </p><p> Se o Bruno Maranh  
ilma . Ha que recompor com a situação pois o dia a dia segue  
asse REINVENTAR a HISTORIA DO BRASIL ele quer ACABAR com  
m todos gostem da partilha , entendo que nem todos se sinta  
nem todos gostem da partilha , entendo que nem todos se si  
que nem todos gostem da partilha , entendo que nem todos se  
tou aqui para tirar minha duvida , eu e meu namorado decidid  
> Bomanah 04/03/2015 18:38:19 </p><p> " Olha galera é linda  
><p> mesmu dizendu q a gnt naum estuda </p><p> posso na  
rmelha ? " </p><p> Bomanah 04/03/2015 18:38:19 </p><p> "  
, mas não deveríamos lutar , para trazê-lo de volta </p><p>  
se não jogar contra o Portimonense , é melhor deixar o hom  
também , que em matéria de fé e moral o papa é infalível .  
r ao proximo . Amar por aqueles q nao amam , rezar por aque  
lepois que como alguma coisa , e tomei suco de limão me dei

## • Brazilian Portuguese Web as Corpus - BrWaC

do mundo ? É um banco mal administrado . </p><p> Como ri  
cio do mundo ? É um banco mal administrado . </p><p> Com pao azimo feito no forno  
contas a pagar , salarios , fornecedores , bancos etc . Daqui P forma pão sírio  
negócio do mundo ? É um banco mal administrado . </p><p> C com a Natureza , vamos tornando-nos extremamente sensívei  
, na verdade deslumbrada , com a beleza e criatividade das pães indianos  
iz verão . . . </p><p> 02/12/2013Cidada </p><p> Realmente esfiha aberta  
ais uma triste prova que a media brasileira ja nao tem seque  
PIG ja abandonou o Serra . Botou todas as fichas nele e ele fi  
ânto que agora o fantoche VAI ter que leva-lo a tira colo . A F dia dos pertubados  
ânto a quem ler estas linhas . </p><p> A comer a sopa de pei  
garanto a quem ler estas linhas . </p><p> A comer a sopa de p emulher ...  
so garanto a quem ler estas linhas . </p><p> A comer a sopa c pão áximo de forno  
nta acima . dsde já Obrigada </p><p> Posted by Rosana </p>  
3/04/2015 17:09:42 </p><p> " Oii alguem sabe se a lente pin  
ma outra aluna , que só assinava Miguxinha : </p><p> . </p><  
Excelentewesley 23/04/2015 17:09:42 </p><p> " Oii alguem hdslr.canon ver se essa resolve.pra fazer filmes  
do cristão , ao invés de simplesmente desistirmos dele ? Sei q e...  
Jesus vai por a jogar de inicio o Eduardo , Jardel , Matic e o R BARRINHAS NUTRITIVAS DE AVEIA  
as depois do discurso de encerramento do papa , eu lembrei c  
rezando pelo sofrimento humano no silencio sem aparecer ne  
tou se eu tinha suspeita de estar gravida . Na terça já estava  
<http://ossianrodriguesleite.blogspot.com/>  
graficos no excel  
wrap - pao sirio sanduiche enrolado  
beirute  
A F dia dos pertubados  
recebi no meu e-mail.vamos fazer a bula d  
pão áximo de forno  
Dicas do Gmail: como criar assinaturas com  
imagens  
hdslr.canon ver se essa resolve.pra fazer filmes  
e...  
BARRINHAS NUTRITIVAS DE AVEIA  
7 configurações secretas que você pode  
habilitar n...  
Como clonar os programas para instalá-los  
novamente...



ação de Dilma está apoiada numa estratégia , na confiança da  
a ação de Dilma está apoiada numa estratégia , na confiança  
Serra . O PIG não vai queimar os pauzinhos se metendo numa mi  
ro , a ação de Dilma está apoiada numa estratégia , na confia  
r substancia química e tanto merda . ham um odor fort  
uma verdadeira artista , que usa a sua voz a mestria da  
emos de ficar a deriva de pessoas , das pessoas para cuida  
ou em 15/07/2006Mary Machado - Bruno Maranh  
ilma . Ha que recompor com a situa  
asse REINVENTAR a HISTORIA DO BRASIL BAR com  
m todos gostem da partilha , en  
se sinta  
nem todos gostem da partilha  
que nem todos gostem da partilha . A  
tou aqui para tirar minha duv  
> Bomanah 04/03/2015 18:38:19 <  
><p> mesmu dizendu q a gnt na  
rmelha ? </p><p> Bomanah 04/03/2015 18:38:19 <  
, mas não deveríamos lutar para  
se não jogar contra o Podermonen  
também , que em matéria de fé e  
r ao proximo . Amar por aqueles q  
lepos que como alguma coisa , e to  
me dei



# Dataset

## Multiwoz

- Multi-Domain Wizard-of-Oz dataset (MultiWOZ) - 2018 [4]
  - Conversa entre humanos em turnos
  - Dataset 100% anotado
  - MultiWOZ 2.2 - 2020 [5]
  - Inglês

Domain	Categorical slots	Non-categorical slots	Intents
Restaurant	pricerange, area, bookday, bookpeople	food, name, booktime, address, phone, postcode, ref	find, book
Attraction	area, type	name, address, entrancefee, openhours, entrancefee, openhours, phone, postcode	find
Hotel	pricerange, parking, internet, stars, area, type, bookpeople, bookday, bookstay	name, address, phone, postcode, ref	find, book
Taxi	-	destination, departure, arriveby, leaveat, phone, type	book
Train	destination, departure, day, bookpeople	arriveby, leaveat, trainid, ref, price, duration	find, book
Bus	day	departure, destination, leaveat	find
Hospital	-	department , address, phone, postcode	find
Police	-	name, address, phone, postcode	find

- Multi-Domain Wizard-of-Oz dataset (MultiWOZ) - 2018 [4]
  - Conversa entre humanos em turnos
  - Dataset 100% anotado
  - MultiWOZ 2.2 - 2020 [5]
  - **Inglês -> Português**

Domain	Categorical slots	Non-categorical slots	Intents
Restaurant	pricerange, area, bookday, bookpeople	food, name, booktime, address, phone, postcode, ref	find, book
Attraction	area, type	name, address, entrancefee, openhours, entrancefee, openhours, phone, postcode	find
Hotel	pricerange, parking, internet, stars, area, type, bookpeople, bookday, bookstay	name, address, phone, postcode, ref	find, book
Taxi	-	destination, departure, arriveby, leaveat, phone, type	book
Train	destination, departure, day, bookpeople	arriveby, leaveat, trainid, ref, price, duration	find, book
Bus	day	departure, destination, leaveat	find
Hospital	-	department , address, phone, postcode	find
Police	-	name, address, phone, postcode	find

# Dataset - Multiwoz-PTBR

30

```
"speaker": "USER",
"turn_id": "6",
"utterance": "Then how about one in the cheap price range?"
}, {
    "frames": [
        "actions": [],
        "service": "hotel",
        "slots": [
            {
                "exclusive_end": 45,
                "slot": "hotel-name",
                "start": 33,
                "value": "Autumn House"
            }
        ],
        "speaker": "SYSTEM",
        "turn_id": "7",
        "utterance": "Cheap we can do! I recommend the Autumn House in the east.  
Would you like to book a room there?"
    ],
}
```

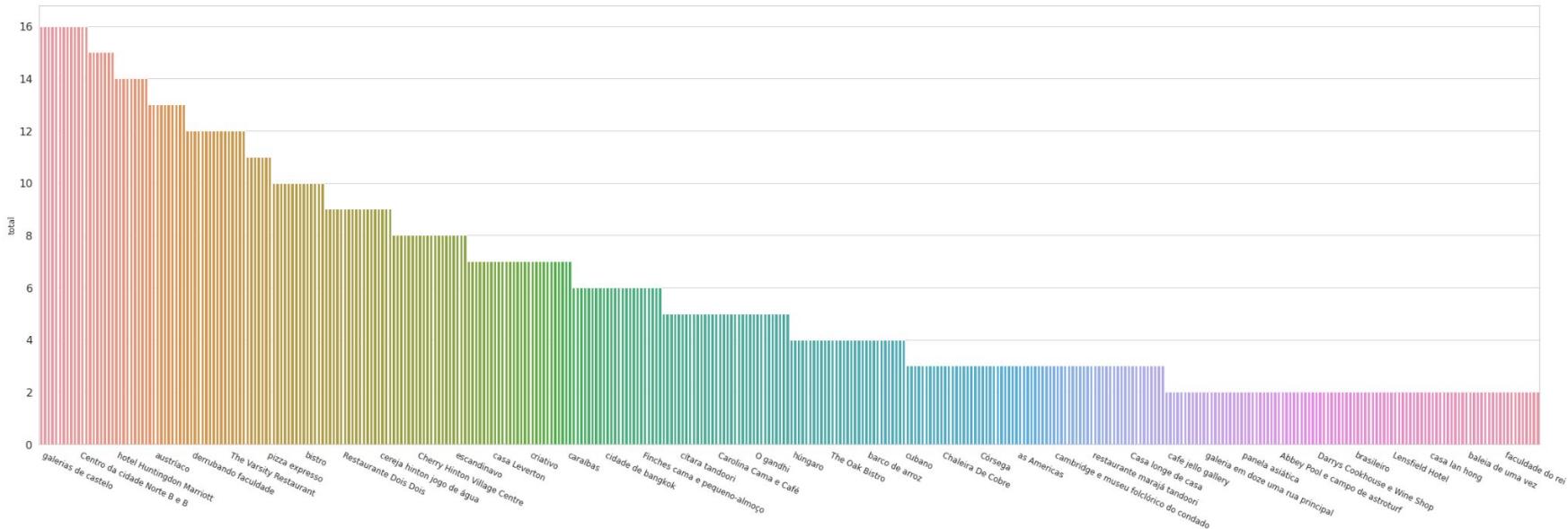
# Dataset - Multiwoz-PTBR

```
"speaker": "USER",
"turn_id": "6",
"utterance": "Então que tal um na faixa de preço barato?"
},{
    "frames": [
        "actions": [],
        "service": "hotel",
        "slots": [
            {
                "exclusive_end": 45,
                "slot": "hotel-name",
                "start": 33,
                "value": "Casa de Outono"
            }
        ]
    ],
    "speaker": "SYSTEM",
    "turn_id": "7",
    "utterance": "Podemos fazer barato! Eu recomendo a Casa do Outono no leste. Você gostaria de reservar um quarto lá?"
},
```

```
"speaker": "USER",
"turn_id": "6",
"utterance": "Então que tal um na faixa de preço barato?"
},{
    "frames": [
        "actions": [],
        "service": "hotel",
        "slots": [
            {
                "exclusive_end": 45,
                "slot": "hotel-name",
                "start": 33,
                "value": "Casa de Outono"
            }
        ]
    ],
    "speaker": "SYSTEM",
    "turn_id": "7",
    "utterance": "Podemos fazer barato! Eu recomendo a Casa do Outono no  
leste. Você gostaria de reservar um quarto lá?"
},
```

- Resultados preliminares da tradução

- 25.888 total
- ~80% acertos traduzindo diretamente
- ~90% com ajustes manuais
  - 23.121 acertos
  - 2.767 erros



MinerXenForo

- Ferramenta para gerar datasets
- Focada em fóruns
  - Utilizando a plataforma [XenForo](#)
- Diferentes tarefas NLP
  - **Language Modeling**
    - Masked LM
    - Next Sentence LM
  - **Pergunta e Resposta (Q&A)**
  - Classificação
  - Outras atividades
    - Previsão de uso de recursos
- 3 Módulos
  - Aquisição
  - Limpeza/Montagem de conversas
  - Anonimização
- Mecanismo de atualização

Jogos					
	<b>Consoles &amp; Jogos - Discussão geral</b> Fale sobre Nintendos, Playstations, Xboxes, portáteis etc.	99,4K Tópicos	5,6M Mensagens		Sony reduz preço do Playstation 5 n... 4 minutos atrás · nando3d
	<b>PC, Hardware &amp; Gadgets - Discussão geral</b> Espaço para tudo sobre PCs e gadgets. Overclock, placas de vídeo, processadores e objetos de tecnologia em geral.	48,6K Tópicos	1,6M Mensagens		GPUs usadas para mineração perde... 3 minutos atrás · Poor_Boy
	<b>Áudio e vídeo</b> Discussão sobre TV, home theater e aparelhos de áudio e vídeo em geral.	102 Tópicos	25K Mensagens		<span>Tópico oficial</span> Tópico IPTV [TV no seu ... Hoje às 16:46 · Felipe789
	<b>Nintendo</b> Espaço reservado à discussão sobre a Nintendo e seus consoles, Switch, Wii, Gamecube, Gameboy, Nintendo 3DS e outros.	13,8K Tópicos	301,6K Mensagens		[FAQ/TUTORIAL DE DESBLOQUEIO] ... 4 minutos atrás · renderfox
	<b>Playstation</b> Espaço reservado à discussão sobre o Playstation e tudo relacionado à Sony.	21,1K Tópicos	451,4K Mensagens		<span>Tópico oficial</span> [JB 7.55] Jailbreak PS4 Hoje às 18:53 · IcE_WiNd

**Categoria** → **Jogos**

**Sub-categoria** → **Consoles & Jogos - Discussão geral**  
Fale sobre Nintendos, Playstations, Xboxes, portáteis etc.

**PC, Hardware & Gadgets** → **Total Threads** 48,6K Tópicos  
Espaço para tudo sobre PCs e g...  
vídeo, processadores e objetos de tecnologia em geral.

**Áudio e vídeo** → **Total de Posts** 25K Mensagens  
Discussão sobre TV, home theater e aparelho...  
vídeo em geral.

**Nintendo**  
Espaço reservado à discussão sobre a Nintendo e seus consoles, Switch, Wii, Gamecube, Gameboy, Nintendo 3DS e outros.

**Playstation**  
Espaço reservado à discussão sobre o Playstation e tudo relacionado à Sony.

	Tópicos	Mensagens	
Sony reduz preço do Playstation 5 n...	99,4K	5,6M	 4 minutos atrás · nando3d
GPUs usadas para mineração perde...	48,6K	1,6M	 3 minutos atrás · Poor_Boy
Tópico IPTV [TV no seu ...	25K		 Hoje às 16:46 · Felipe789
[FAQ/TUTORIAL DE DESBLOQUEIO] ...	13,8K	301,6K	 4 minutos atrás · renderfox
[JB 7.55] Jailbreak PS4	21,1K	451,4K	 Hoje às 18:53 · IcE_WiNd

Home > Fóruns > Jogos

[f](#) [t](#)

## PC, Hardware & Gadgets - Discussão geral

Espaço para tudo sobre PCs e gadgets. Overclock, placas de vídeo, processadores e objetos de tecnologia em geral.

1 2 3 ... 1389 Próximo ▶

Filtros ▾

### TÓPICOS FIXOS

 Tópico oficial Índice da Pasta PC  
CrazyPXT · 20 Julho 2016 2 3 4

Respostas: 131 Visualizações: 27K 26 Junho 2021 Grave Uypo 

### TÓPICOS REGULARES

 Tópico oficial Smartphones e Tablets [Android e iOS]  
antoniooli · 15 Janeiro 2014 1851 1852 1853

Respostas: 65K Visualizações: 2M 1 minuto atrás extremepower 

 GPUs usadas para mineração perdem 10% de poder de processamento ao ano, aponta Palit  
SolidEdgard · 2 Agosto 2021 2

Respostas: 48 Visualizações: 893 2 minutos atrás TheK 

Categoria → Jogos →

Sub categoria → PC, Hardware & Gadgets - Discussão geral  
Espaço para tudo sobre PCs e gadgets. Overclock, placas de vídeo, processadores e objetos de tecnologia em geral.

Páginas → 1 2 3 ... 1389 Próximo ▶

Filtros ▾

TÓPICOS FIXOS

 Tópico oficial Índice da Pasta PC  
CrazyPXT · 20 Julho 2016 2 3 4 Respostas: 131 Visualizações: 27K 26 Junho 2021 Grave Uypo 

TÓPICOS REGULARES

Tags → Tópico oficial Smartphones e Tablet  
antonioili · 15 Janeiro 2014 1851 1852 1853 Respostas: 65K Visualizações: 2M 1 minuto atrás extremepower 

Nome → GPUs usadas para mineração perdem 10% de poder de processamento ao ano, aponta Palit  
SolidEdgard · 2 Agosto 2021 2 Respostas: 48 Visualizações: 893 2 minutos atrás TheK 

Home Fóruns Jogos PC, Hardware & Gadgets - Discussão geral [f](#) [t](#)

## Bem Vindo ao Mundo Linux

Nicko · 12 Dezembro 2017

1 2 3 ... 120 Próximo ›



12 Dezembro 2017 #1

Estou replicando uma postagem da comunidade "Unix Universe" do Facebook

### # Seja Bem-Vindo ao Linux! #

**Nicko**  
O Corretor de Seguros  
VIP  
GOLD

Mensagens: 24.065  
Reações: 18.919  
Pontos: 674

> **Postagem voltada aos usuários novatos e curiosos que não sabem nada (ou quase nada) sobre o sistema Linux <**

Depois de ver algumas postagens de curiosos e novatos no mundo Linux, tive a ideia de criar este post para reunir as principais dúvidas que permeiam os novatos, além de listar quais facilidades e dificuldades os usuários de Linux recém-chegados do Windows e/ou do macOS poderão encontrar.

Categoria/Sub → Home Fóruns Jogos PC, Hardware & Gadgets - Discussão geral f t

Título → Bem Vindo ao Mundo Linux  
Nicko · 12 Dezembro 2017

1 2 3 ... 120 Próximo ▶

Data → 12 Dezembro 2017 #1

Post → Estou replicando uma postagem da comunidade "Unix Universe" do Facebook

Usuário → Nicko  
O Corretor de Seguros  
VIP GOLD  
Mensagens: 24.065 Reações: 18.919 Pontos: 674

# Seja Bem-Vindo ao Linux! #

> Postagem voltada aos usuários novatos e curiosos que não sabem nada (ou quase nada) sobre o sistema Linux <

Depois de ver algumas postagens de curiosos e novatos no mundo Linux, tive a ideia de criar este post para reunir as principais dúvidas que permeiam os novatos, além de listar quais facilidades e dificuldades os usuários de Linux recém-chegados do Windows e/ou do macOS poderão encontrar.

## Content Creation

### Contributions & Corrections

- OU Analyses
- Ubers Analyses
- UU Analyses
- RU Analyses
- NU Analyses
- PU Analyses
- LC Analyses
- Doubles OU Analyses
- Monotype Analyses
- OM Analyses
- 1v1 Analyses
- National Dex Analy...
- Battle Stadium Analy...
- VGC Analyses
- CAP Analyses
- Past Generation An...
- Articles and In-Gam...
- Social Media Spotli...
- Archives

Barraskewda [QC 2/2] [GP 0/1]  
Today at 6:53 AM · gorex

### Technology

Any discussions on Cyber Security, Cloud Computing, DevOps, Software Development and Technology go here.

### Certification Programs

- CISA/CISM/CISSP
- Cloud Computing
- CEH
- IT Hardware and Networking
- Web App and Programming
- Microsoft and OS/Database
- Other technology programs

Threads 2.9K Messages 14.2K M Microsoft Azure Administrator (AZ-... 29 minutes ago · manish7

### Master Programs

- Cyber Security Expert
- Cloud Architect
- DevOps Engineer
- Full Stack Developer

Threads 618 Messages 3.3K M Would like Yesterday

### NeoSmart Support Center

#### BugCentral

Threads 435 Messages 2.9K M Metro Yesterday

#### EasyBCD Support

Threads 6K Messages 37.3K J Metro Yesterday

#### Other NeoSmart-Related Support

Threads 394 Messages 2.2K M NSIS 1, Jul 2, 2020

## Tecnologia - Telecomunicações

### Smartphones, Tablets e outros dispositivos móveis

[Guia] Smartphones chineses: apare... 5 minutos atrás · Dunkheim321

### TV por Assinatura e Monitoramento de vídeo

[TÓPICO DEDICADO] Xiaomi Mi Box (...) Ontem às 22:51 · DoctorBR

### Internet & Redes

[TÓPICO DEDICADO] Melhores Rotea... 26 minutos atrás · prezzi

## Script PHP

CodeCanyon F Simple File Sharer 17 May 2020 · fzteam

General Urban Developments 875K None

Classic Architecture 246K Scripts Other [NULLED] WolfLab Suite Fo... 10 August 2020 · Barry

European Classic Architecture and Landscapes 158K dum Script V.I.P Scripts [NULLED]Next Hour- M... 23 August 2020 · oopestroo

Public Space

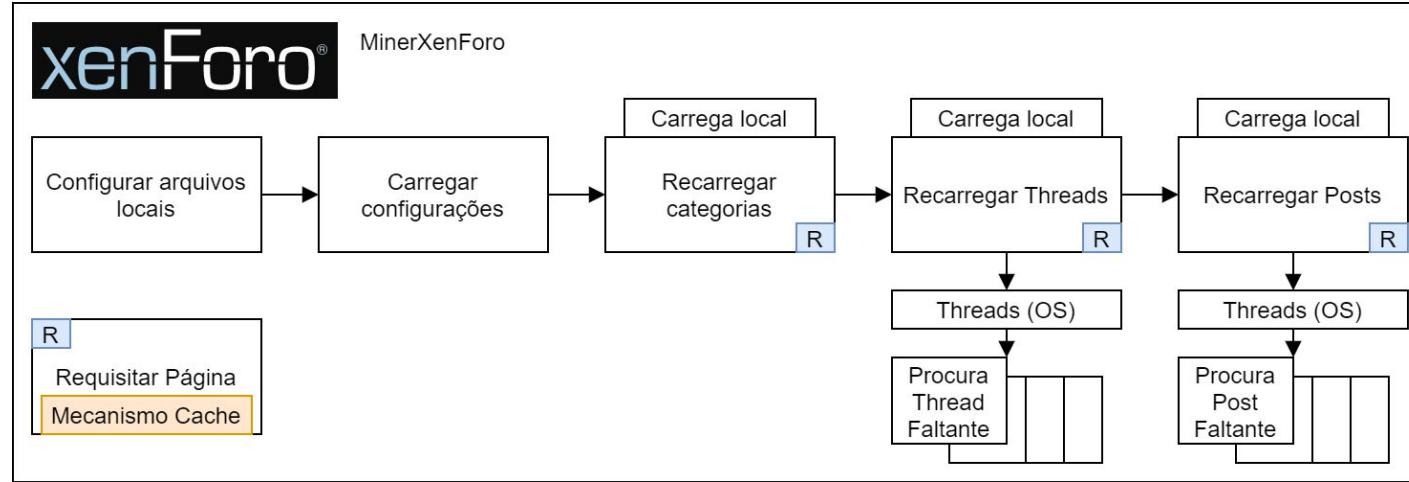
Shopping Architecture

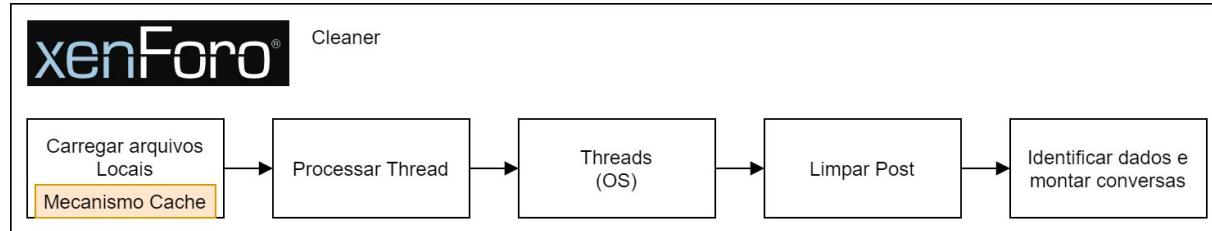
Design & Lifestyle

Urban Renewal and Redevelopment

Architects

Member Designs





Tipo do elemento	Tag inserida	Tipo do elemento	Tag inserida
Imagen externa	<image>	Conteúdo compartilhado	<shared_content> {} {} </shared_content>
Emoji	<emoji> {} </emoji>	Media	<mediaembed> {} </mediaembed>
Imagen desconhecida	<image_unknown>	Resposta	<answering> {} </answering>
Citação	<quote> {} </quote>	Spoiler	<spoiler> {} </spoiler>
Url	<url> {} {} </url>	Código	<code> {} </code>
Link	<link>	Iframe	<iframe> {} </iframe>

```

montar_conversas(mensagem):
    referencias = identificar_referencias(mensagem)

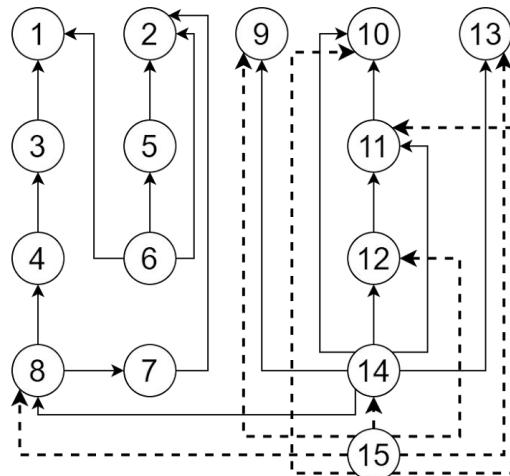
    for referencia in referencias:
        if referencia.data_criacao < orig.data_criacao:
            referencia['pai'] = mensagem
            montar_conversas(referencia)

    if Len(referencias)==0:
        r = []
        while mensagem.pai!=None:
            r.append(mensagem)
            mensagem = mensagem.pai
        r.append(mensagem)

    return res_final
    
```

- Montagem das conversas paralelas

ID	Texto	Refs
1	Como eu consigo comprar? 😊	
2	Já tentou usar o cartão?	
3	Acho que não tem como	1
4	Claro que tem como!	3
5	Claro que sim	2
6	Dá pra tentar boleto	5-2-1
7	E tem que comprar agora?	2
8	Agora todo mundo tem opinião	7-4
9	Tem outros parecidos?	
10	Até quando fica válido?	
11	Acho que fica só mais um dia	10
12	Claro que não! Já acabou	11
13	A promoção acabou faz dias	
14	Nada disso faz sentido	13-12-11-10-9-8
15	Deixa o pessoal falar	14-13-12-11-10-9-8



15-14-13	15-14-8-4-3-1	15-8-7-2
15-14-12-11-10	15-13	15-8-4-3-1
15-14-11-10	15-12-11-10	6-5-2
15-14-10	15-11-10	6-2
15-14-9	15-10	6-1
15-14-8-7-2	15-9	

- Dataset
  - **/categories.json**
    - Descrição das categorias e subcategorias -> id, título, href, última atualização, descrição
  - **/clear\_threads.csv**
    - Resumo da thread -> id categoria, id sub-categoria, total mensagens, datas, tokens, conversations
  - **/config.json**
    - Configurações do minerador
  - **/categories\_threads/category\_{category}\_{subcategory}.json**
    - Threads de cada sub-categoria -> id, tags, href, dates, respostas, fixa, visitas, título, ...
  - **/clear\_threads/{}.tsv**
    - Dataset processador e mensagens organizadas com tabulação -> timestamp, user, post
  - **/plots/{}.pdf**
    - Gráficos
  - **/threads/{}.json**
    - Postagens de cada thread sem processamento e limpeza

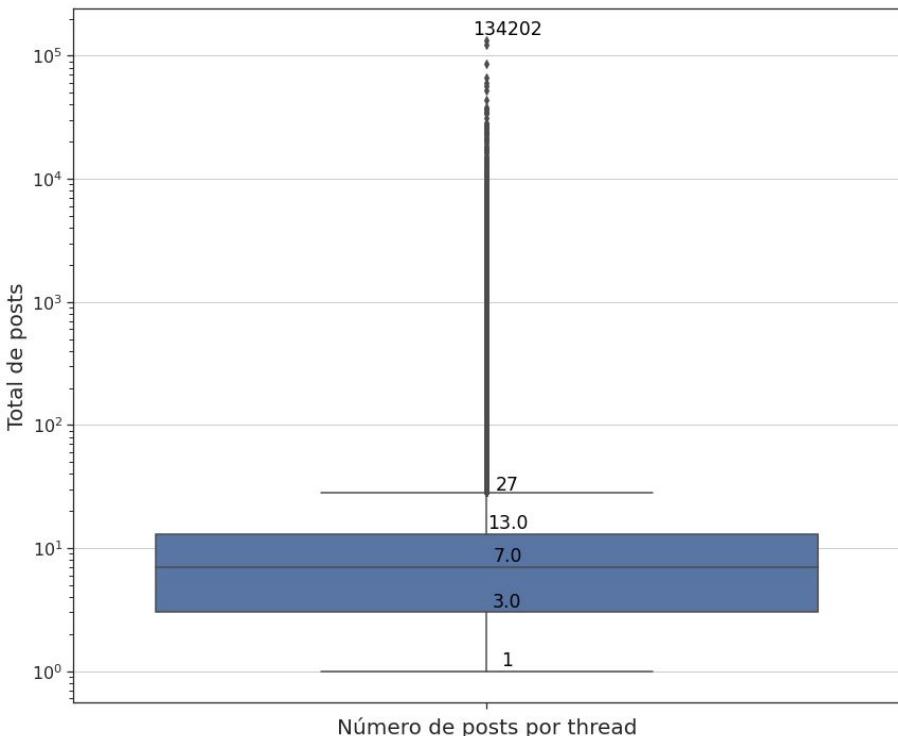
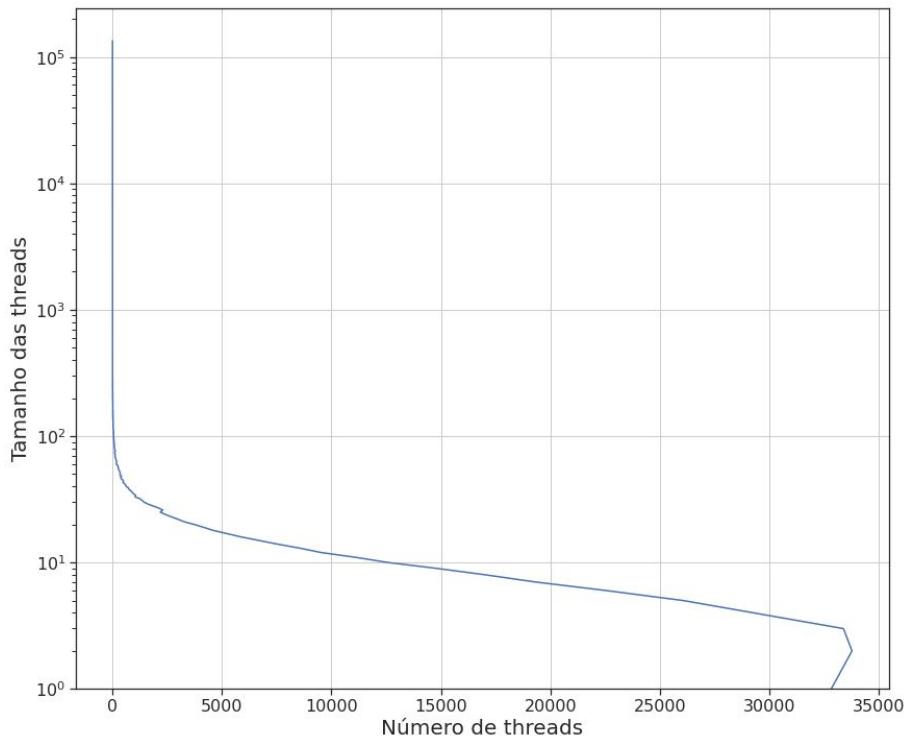
# MinerXenForo

## Adrenaline

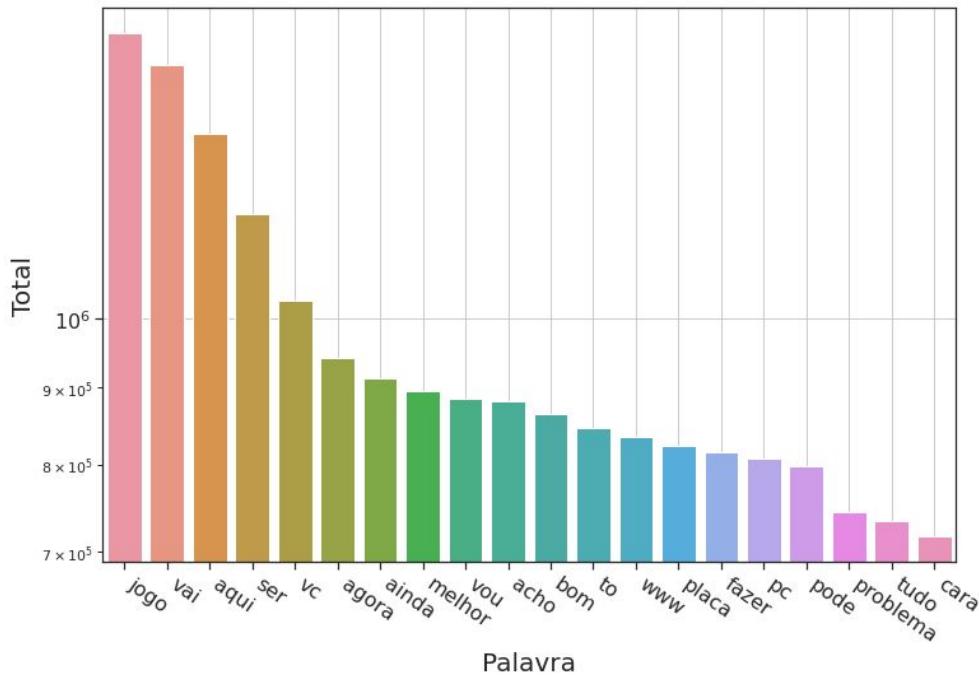
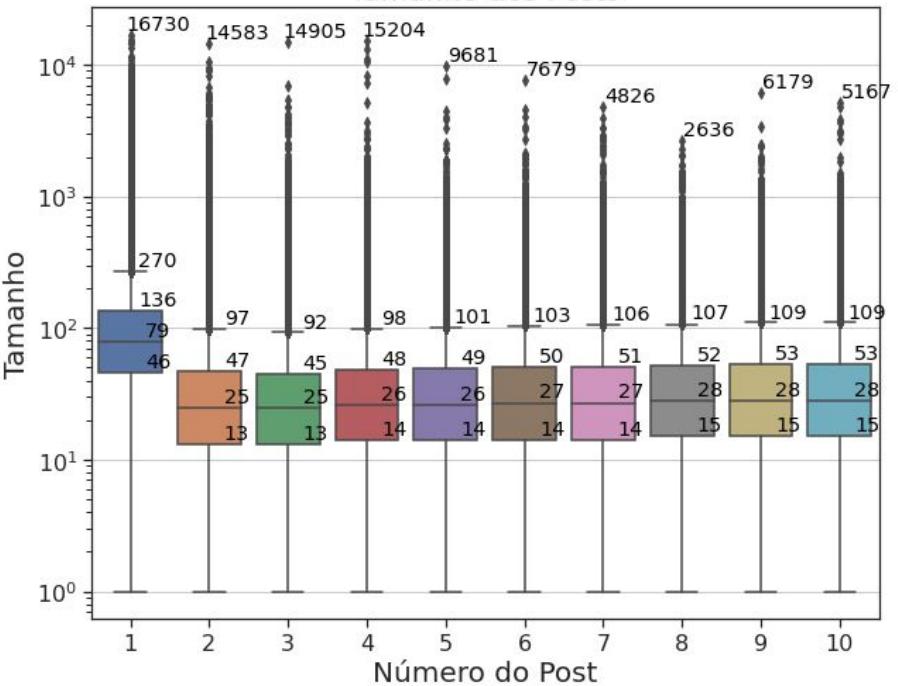
- Único dataset completo
- Site: [forum.adrenaline.com.br](http://forum.adrenaline.com.br)
- Diversos domínios
  - Dentro de tecnologia

Adrenaline	
Total de threads	356.763
Total de Posts	9.550.206
Total de tokens	477.619.822
Tokens únicos	1.701.482
Tamanho compactado	2.1gb
Conversas compactadas	72gb
Tamanho médio das threads	26.7
Tokens médios por post	50
Tamanho médio de conversa	34

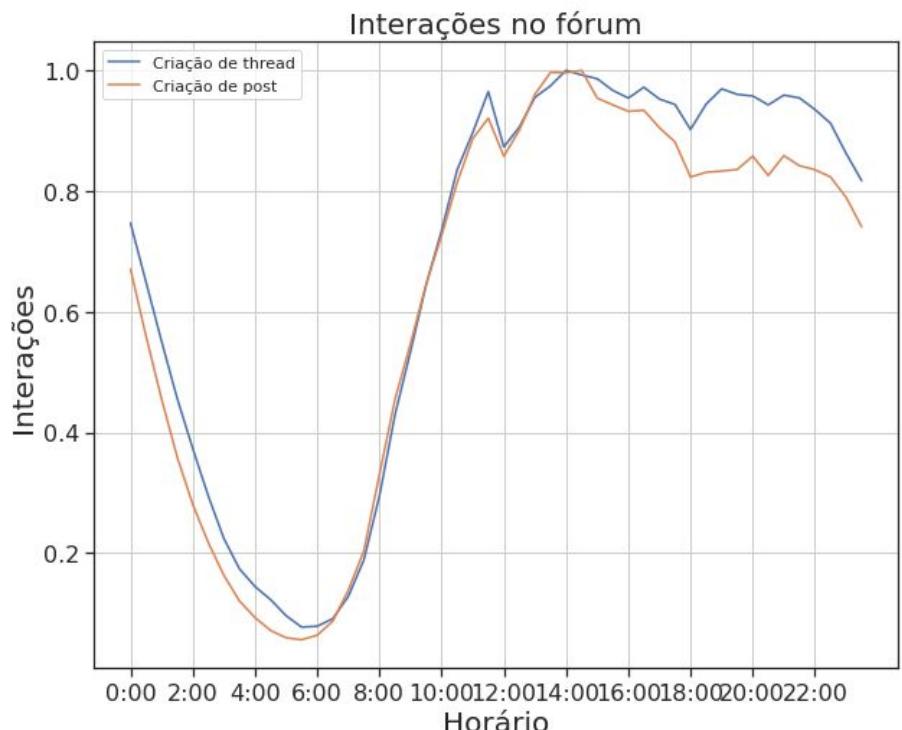
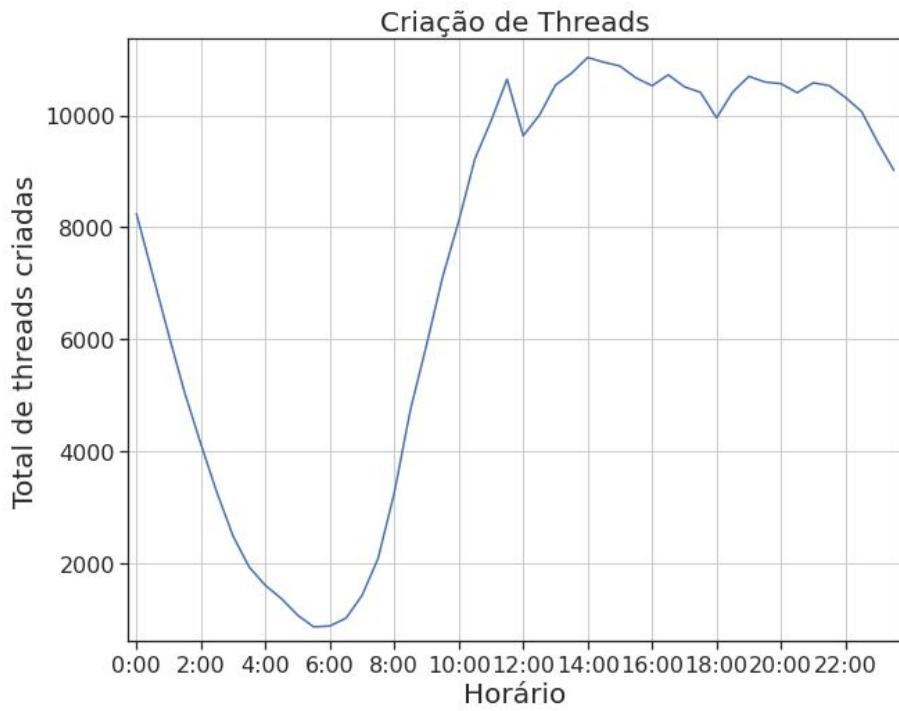
- Threads grandes



Tamanho dos Posts



- Previsão de recursos computacionais



# MinerXenForo

## Outerspace

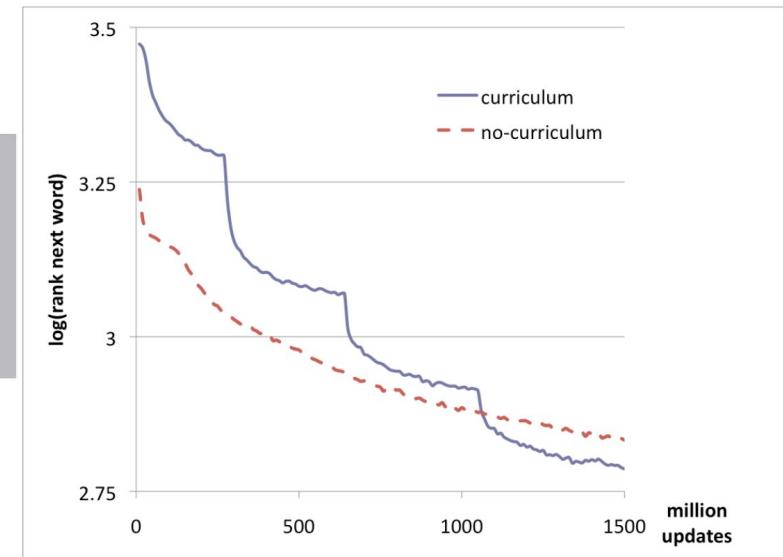
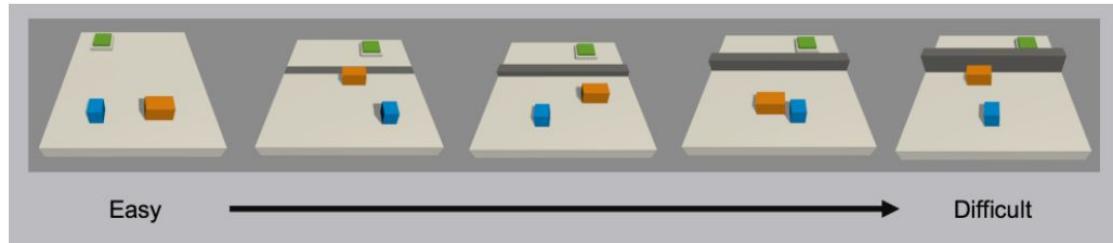
- Status:
  - Em andamento
- Site: <https://forum.outerspace.com.br/>
- Diversos domínios

Outerspace	
Total de threads	559.395
Total de Posts	
Total de tokens	
Tokens únicos	
Tamanho compactado	4.8gb
Conversas compactadas	
Tamanho médio das threads	
Tokens médios por post	
Tamanho médio de conversa	

# Plano de Uso

- Aprendizado por Currículo [6]
  - Começar com problemas simples e ir aumentando a complexidade

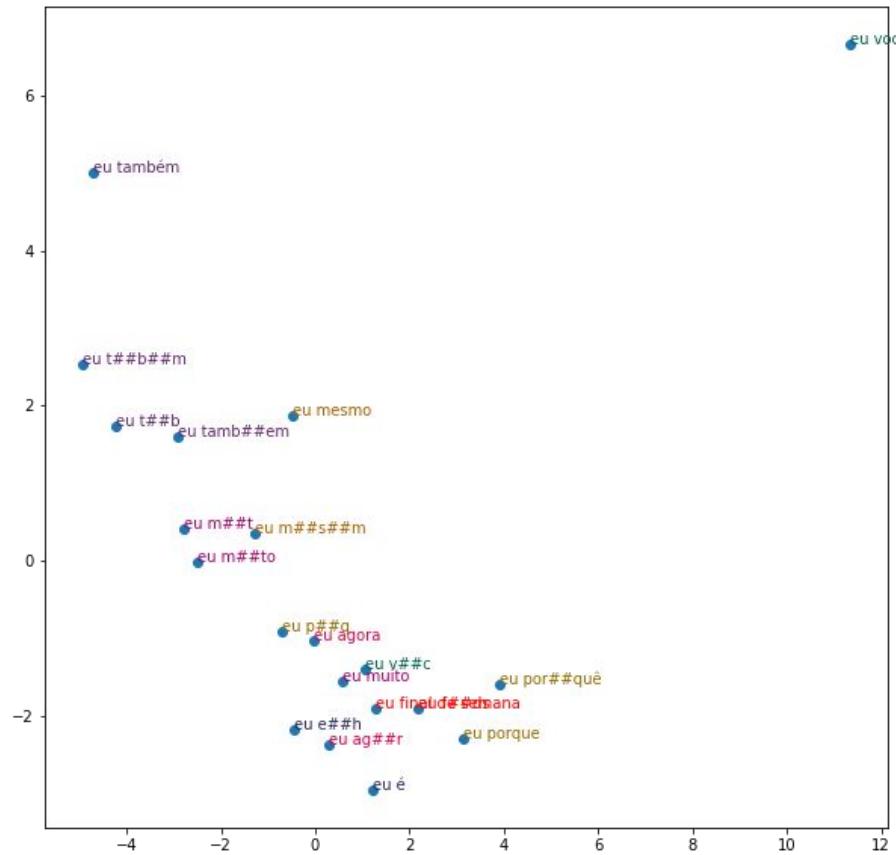
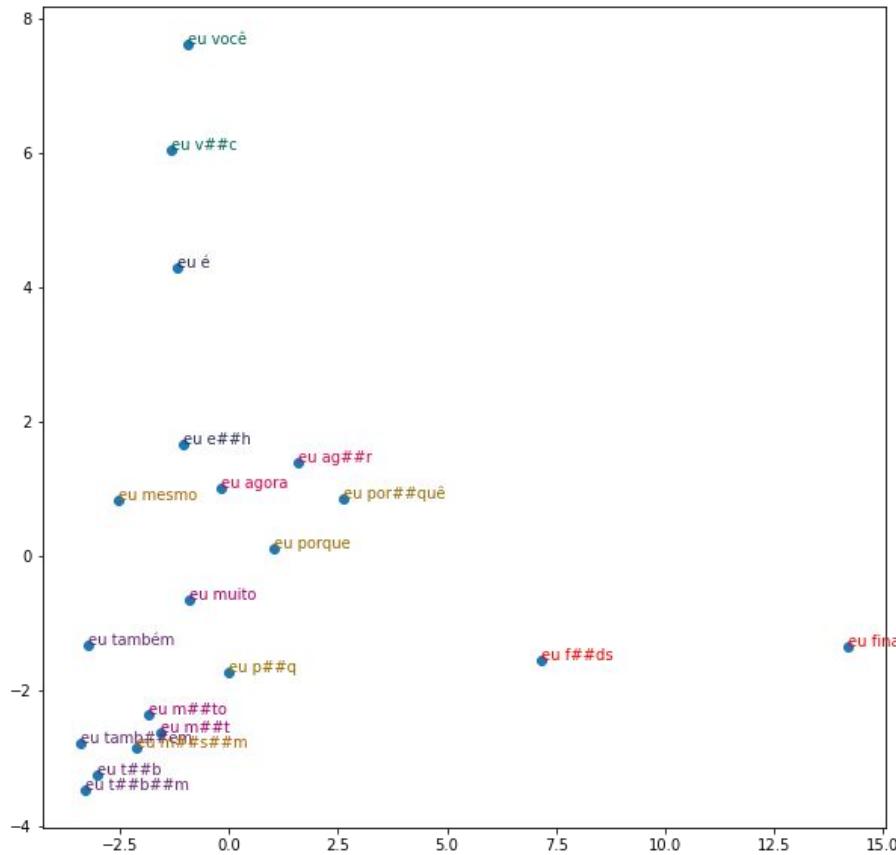
$$1 + 1 = 2 \rightarrow 2x = 4 \rightarrow \int_0^1 2x \, dx = 1$$



# Considerações Finais

# Considerações Finais

Treino com 0.0008%



# Considerações Finais

	eu final de semana	eu fds	eu agora	eu agr	eu eh	eu é	eu mto	eu mt	eu muito	eu também	eu tbm	eu tambem	eu tb	eu mesmo	eu msm	eu você	eu vc	eu porque	eu pq	eu porquê
eu final de semana	1.00	0.78	0.61	0.62	0.54	0.52	0.55	0.56	0.56	0.53	0.52	0.52	0.53	0.51	0.55	0.52	0.51	0.61	0.58	0.65
eu fds	0.78	1.00	0.80	0.83	0.76	0.71	0.79	0.80	0.76	0.75	0.76	0.76	0.77	0.68	0.77	0.70	0.74	0.79	0.81	0.81
eu agora	0.61	0.80	1.00	0.89	0.86	0.86	0.89	0.89	0.89	0.91	0.89	0.89	0.89	0.86	0.88	0.86	0.86	0.91	0.89	0.87
eu agr	0.62	0.83	0.89	1.00	0.87	0.85	0.86	0.87	0.84	0.85	0.82	0.84	0.83	0.80	0.83	0.80	0.82	0.88	0.88	0.86
eu eh	0.54	0.76	0.86	0.87	1.00	0.93	0.87	0.86	0.82	0.87	0.85	0.88	0.86	0.81	0.84	0.78	0.82	0.84	0.87	0.84
eu é	0.52	0.71	0.86	0.85	0.93	1.00	0.82	0.81	0.81	0.86	0.81	0.84	0.81	0.83	0.81	0.83	0.82	0.83	0.81	0.82
eu mto	0.55	0.79	0.89	0.86	0.87	0.82	1.00	0.97	0.94	0.90	0.90	0.92	0.91	0.84	0.91	0.79	0.82	0.88	0.91	0.84
eu mt	0.56	0.80	0.89	0.87	0.86	0.81	0.97	1.00	0.94	0.90	0.92	0.92	0.93	0.82	0.91	0.79	0.83	0.87	0.91	0.84
eu muito	0.56	0.76	0.89	0.84	0.82	0.81	0.94	0.94	1.00	0.88	0.84	0.86	0.86	0.80	0.84	0.81	0.80	0.88	0.86	0.83
eu também	0.53	0.75	0.91	0.85	0.87	0.86	0.90	0.90	0.88	1.00	0.94	0.97	0.94	0.89	0.91	0.82	0.82	0.86	0.87	0.83
eu tbm	0.52	0.76	0.89	0.82	0.85	0.81	0.90	0.92	0.84	0.94	1.00	0.97	0.99	0.84	0.93	0.78	0.82	0.84	0.88	0.80
eu tambem	0.52	0.76	0.89	0.84	0.88	0.84	0.92	0.92	0.86	0.97	0.97	1.00	0.97	0.87	0.92	0.78	0.82	0.84	0.88	0.82
eu tb	0.53	0.77	0.89	0.83	0.86	0.81	0.91	0.93	0.86	0.94	0.99	0.97	1.00	0.83	0.92	0.79	0.82	0.85	0.89	0.80
eu mesmo	0.51	0.68	0.86	0.80	0.81	0.83	0.84	0.82	0.80	0.89	0.84	0.87	0.83	1.00	0.90	0.80	0.79	0.83	0.82	0.81
eu msm	0.55	0.77	0.88	0.83	0.84	0.81	0.91	0.91	0.84	0.91	0.93	0.92	0.92	0.90	1.00	0.77	0.80	0.86	0.89	0.83
eu você	0.52	0.70	0.86	0.80	0.78	0.83	0.79	0.79	0.81	0.82	0.78	0.78	0.79	0.80	0.77	1.00	0.93	0.80	0.77	0.79
eu vc	0.51	0.74	0.86	0.82	0.82	0.82	0.82	0.83	0.80	0.82	0.82	0.82	0.82	0.79	0.80	0.93	1.00	0.79	0.81	0.78
eu porque	0.61	0.79	0.91	0.88	0.84	0.83	0.88	0.87	0.88	0.86	0.84	0.84	0.85	0.83	0.86	0.80	0.79	1.00	0.95	0.93
eu pq	0.58	0.81	0.89	0.88	0.87	0.81	0.91	0.91	0.86	0.87	0.88	0.88	0.89	0.82	0.89	0.77	0.81	0.95	1.00	0.90
eu porquê	0.65	0.81	0.87	0.86	0.84	0.82	0.84	0.84	0.83	0.83	0.80	0.82	0.80	0.81	0.83	0.79	0.78	0.93	0.90	1.00

# Considerações Finais

	eu final de semana	eu fds	eu agora	eu agr	eu eh	eu é	eu mto	eu mt	eu muito	eu também	eu tbm	eu tambem	eu tb	eu mesmo	eu msm	eu você	eu vc	eu porque	eu pq	eu porquê
eu final de semana	1.00	0.89	0.85	0.83	0.83	0.84	0.82	0.83	0.86	0.78	0.81	0.80	0.81	0.80	0.85	0.63	0.85	0.81	0.84	0.81
eu fds	0.89	1.00	0.87	0.87	0.85	0.88	0.83	0.85	0.87	0.77	0.82	0.84	0.82	0.82	0.85	0.68	0.88	0.83	0.86	0.84
eu agora	0.85	0.87	1.00	0.86	0.89	0.89	0.88	0.89	0.89	0.81	0.87	0.87	0.86	0.84	0.87	0.65	0.89	0.84	0.91	0.84
eu agr	0.83	0.87	0.86	1.00	0.88	0.88	0.86	0.85	0.85	0.76	0.84	0.86	0.84	0.79	0.88	0.63	0.87	0.80	0.87	0.81
eu eh	0.83	0.85	0.89	0.88	1.00	0.90	0.87	0.87	0.88	0.80	0.86	0.87	0.86	0.81	0.86	0.61	0.89	0.82	0.89	0.83
eu é	0.84	0.88	0.89	0.88	0.90	1.00	0.86	0.85	0.89	0.77	0.83	0.87	0.83	0.81	0.86	0.65	0.89	0.86	0.87	0.85
eu mto	0.82	0.83	0.88	0.86	0.87	0.86	1.00	0.93	0.90	0.84	0.89	0.89	0.89	0.83	0.87	0.59	0.85	0.82	0.90	0.81
eu mt	0.83	0.85	0.89	0.85	0.87	0.85	0.93	1.00	0.89	0.86	0.90	0.88	0.90	0.81	0.87	0.59	0.87	0.80	0.89	0.81
eu muito	0.86	0.87	0.89	0.85	0.88	0.89	0.90	0.89	1.00	0.80	0.84	0.85	0.85	0.81	0.85	0.66	0.86	0.85	0.89	0.85
eu também	0.78	0.77	0.81	0.76	0.80	0.77	0.84	0.86	0.80	1.00	0.90	0.88	0.87	0.81	0.82	0.54	0.79	0.74	0.84	0.73
eu tbm	0.81	0.82	0.87	0.84	0.86	0.83	0.89	0.90	0.84	0.90	1.00	0.91	0.92	0.84	0.88	0.55	0.84	0.77	0.87	0.76
eu tambem	0.80	0.84	0.87	0.86	0.87	0.87	0.89	0.88	0.85	0.88	0.91	1.00	0.88	0.86	0.89	0.60	0.85	0.81	0.87	0.81
eu tb	0.81	0.82	0.86	0.84	0.86	0.83	0.89	0.90	0.85	0.87	0.92	0.88	1.00	0.81	0.87	0.55	0.84	0.77	0.87	0.77
eu mesmo	0.80	0.82	0.84	0.79	0.81	0.81	0.83	0.81	0.81	0.81	0.84	0.86	0.81	1.00	0.87	0.61	0.81	0.80	0.84	0.81
eu msm	0.85	0.85	0.87	0.88	0.86	0.86	0.87	0.87	0.85	0.82	0.88	0.89	0.87	0.87	1.00	0.62	0.87	0.80	0.88	0.83
eu você	0.63	0.68	0.65	0.63	0.61	0.65	0.59	0.59	0.66	0.54	0.55	0.60	0.55	0.61	0.62	1.00	0.67	0.66	0.63	0.70
eu vc	0.85	0.88	0.89	0.87	0.89	0.89	0.85	0.87	0.86	0.79	0.84	0.85	0.84	0.81	0.87	0.67	1.00	0.82	0.89	0.83
eu porque	0.81	0.83	0.84	0.80	0.82	0.86	0.82	0.80	0.85	0.74	0.77	0.81	0.77	0.80	0.80	0.66	0.82	1.00	0.85	0.87
eu pq	0.84	0.86	0.91	0.87	0.89	0.87	0.90	0.89	0.89	0.84	0.87	0.87	0.87	0.84	0.88	0.63	0.89	0.85	1.00	0.85
eu porquê	0.81	0.84	0.84	0.81	0.83	0.85	0.81	0.81	0.85	0.73	0.76	0.81	0.77	0.81	0.83	0.70	0.83	0.87	0.85	1.00

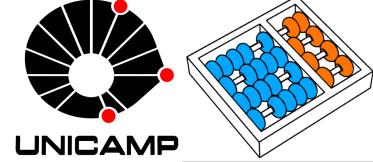
- Dataset pequeno
  - 300 arquivos aleatórios
  - Falta dos exemplos necessários
- O BrWac já contém essas gírias

nte . Parabéns Alexandre , excelente texto . Vou mostrar pro meu pai esse **fds** o texto . </p><p> Lollapalooza Brasil divulga programaÃ§Ã£o dos dois dias de festival </p><p> mana . Por mim , o ideal seria fazer frio a semana toda e solzão apenas no **fds** . Mas isso é impossível nessa terrinha ! </p><p> Para evitar desidratação , a indicação é ://www.gilvaniaevans.com </p><p> Eu tb como direitinho em dia útil e no **fds** me permito gordices . Tb malho . Mas quem me vê no fds fica oooooohh como vc é magr rcaria e pagar barato : Come Come , na esquina da Reitoria , mas fecha no **fds** . A coxinha do BarBaran é sensacional tb . O povo simplesmente ama a feirinha de domir itrou com os 2 lá . Ele quis dizer q ela ia me contar e eu ia correr pra lá no **fds** seguinte . Não querido , não foi o q aconteceu . Sempre soube q vcs freqüentavam o lug eiro a gente sempre se falava todos os dias pelo waths e sempre se via nos **fds** . Cheguei conhecer a família dele ele sempre me falava que não queria se envolver mas ue tinha ficado pronto . Escolhemos juntos desde o piso até o teto , passei **fds** colocando papel de parede , pintando . Até que ele se mudou sozinho . Falou que brigáv . E o pior é q não sei se vai dar tempo amanhã . Já pensou q horror ficar o **fds** de virilha peluda ! Não vou a praia mesmo , mas vai q o Brad Pitt me liga ? </p><p> E tb

Dataset da Wikipedia não deve contar com essas gírias :)

# Obrigado

[ferraroni@lrc.ic.unicamp.br](mailto:ferraroni@lrc.ic.unicamp.br)



- [1] Devlin, J., Chang, M.W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805.
- [2] Alec Radford, Jeff Wu, R. Child, David Luan, Dario Amodei, & Ilya Sutskever (2019). Language Models are Unsupervised Multitask Learners.
- [3] Tomas Mikolov, Kai Chen, Greg Corrado, & Jeffrey Dean. (2013). Efficient Estimation of Word Representations in Vector Space.
- [4] Paweł Budzianowski, Tsung-Hsien Wen, Bo-Hsiang Tseng, Iñigo Casanueva, Stefan Ultes, Osman Ramadan, & Milica Gašić. (2018). MultiWOZ – A Large-Scale Multi-Domain Wizard-of-Oz Dataset for Task-Oriented Dialogue Modelling.
- [5] Xiaoxue Zang, Abhinav Rastogi, Srinivas Sunkara, Raghav Gupta, Jianguo Zhang, & Jindong Chen (2020). MultiWOZ 2.2 : A Dialogue Dataset with Additional Annotation Corrections and State Tracking Baselines.
- [6] Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. 2009. Curriculum learning. In Proceedings of the 26th Annual International Conference on Machine Learning (ICML '09). Association for Computing Machinery, New York, NY, USA, 41–48.  
DOI:<https://doi.org/10.1145/1553374.1553380>
- [7] Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, & Luke Zettlemoyer. (2018). Deep contextualized word representations.
- [8] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, & Illia Polosukhin. (2017). Attention Is All You Need.

©AdrienSIEG

