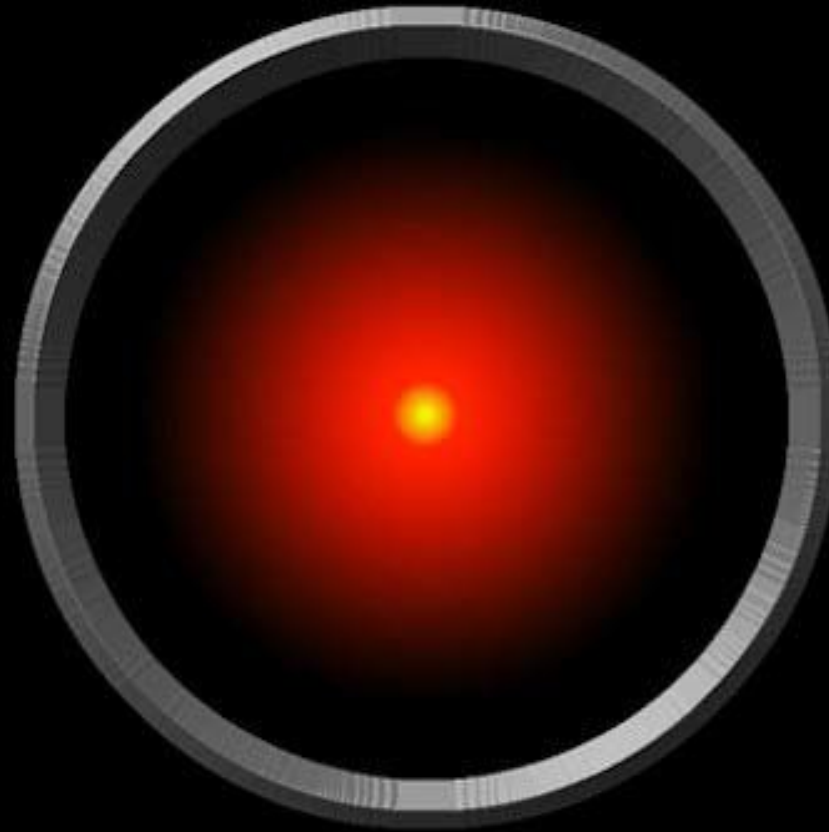


Text Mining: Overview

Luciano Barbosa

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



ChatGPT



Examples

"Explain quantum computing in simple terms"

"Got any creative ideas for a 10 year old's birthday?"

"How do I make an HTTP request in Javascript?"



Capabilities

Remembers what user said earlier in the conversation

Allows user to provide follow-up corrections

Trained to decline inappropriate requests



Limitations

May occasionally generate incorrect information

May occasionally produce harmful instructions or biased content

Limited knowledge of world and events after 2021



Free Research Preview: ChatGPT is optimized for dialogue. Our goal is to make AI systems more natural to interact with, and your feedback will help us improve our systems and make them

Applications




Applications: Translation

Portuguese ▼

↔

English ▼

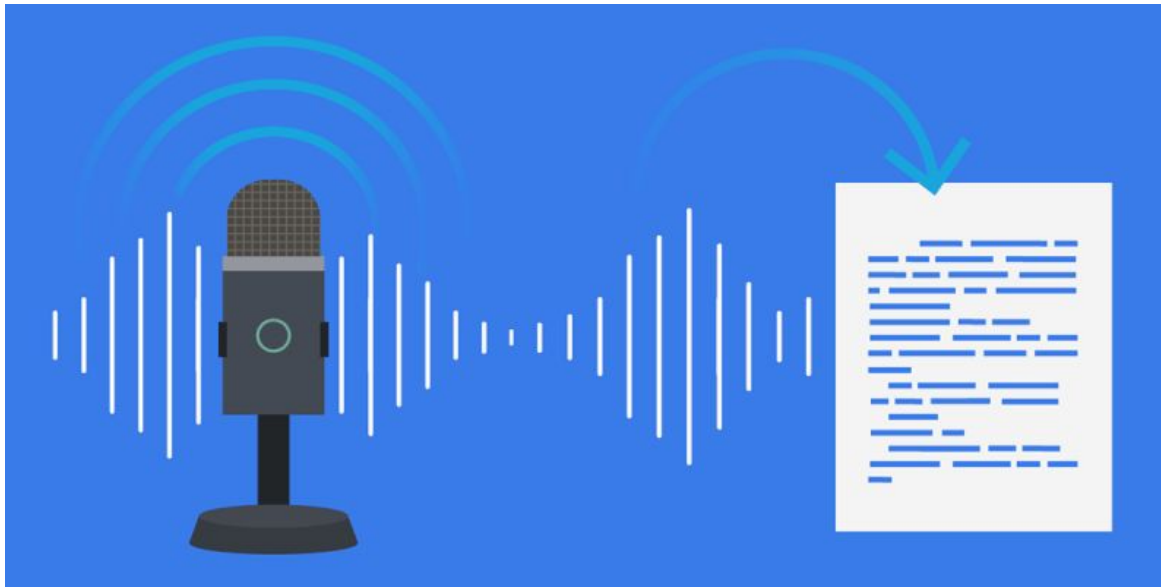
Enter text



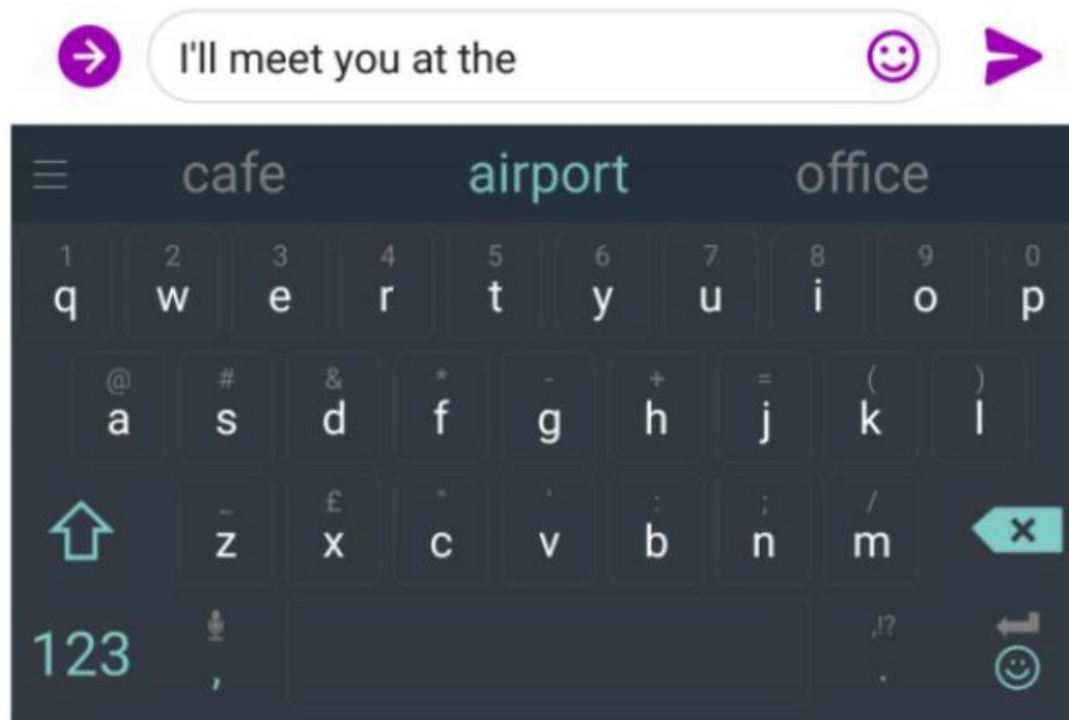
Translation

[Open in Google Translate](#) • [Feedback](#)

Applications: Speech Recognition



Applications: Word Prediction

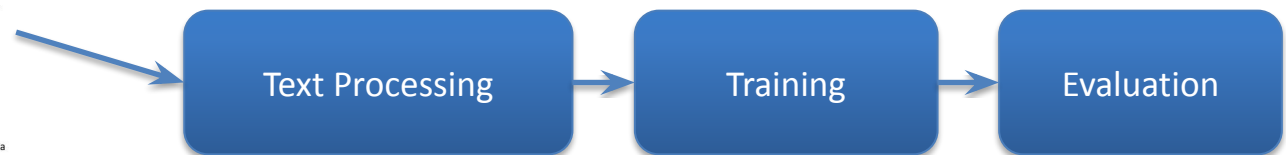
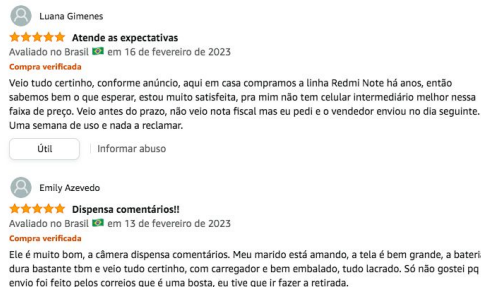


The Course

- Goal:
 - To learn to apply advanced text mining techniques to classify real-world text data, such as social media posts and customer reviews.
- Evaluation: project
- Home page: https://profluciano.github.io/text_mining/

Project: Sentiment Analysis on Product Reviews

- Collect reviews of a product: text and rating
- Train a classifier: SVM + bow, SVM + embeddings, BERT
- Make a presentation reporting the models' building and results (F1 and accuracy) e analysis (slides + 15 min video)



Avaliações

Topics

- Text processing
- Word representation
- Text classification
- Contextual models: Transformers

Text Processing: Tokenization

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。

Text Classification

"I am happy with this water bottle."



"This is a bad investment."



"I am going to walk today."

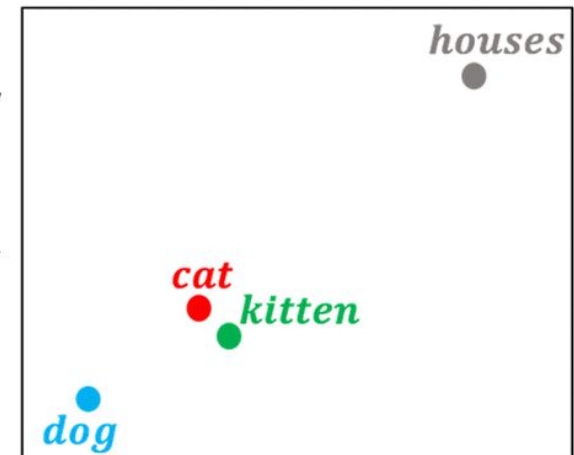


Word Representation

	living being	feline	human	gender	royalty	verb	plural
<i>cat</i> →	0.6	0.9	0.1	0.4	-0.7	-0.3	-0.2
<i>kitten</i> →	0.5	0.8	-0.1	0.2	-0.6	-0.5	-0.1
<i>dog</i> →	0.7	-0.1	0.4	0.3	-0.4	-0.1	-0.3
<i>houses</i> →	-0.8	-0.4	-0.5	0.1	-0.9	0.3	0.8

Dimensionality
reduction of
word
embeddings
from 7D to 2D

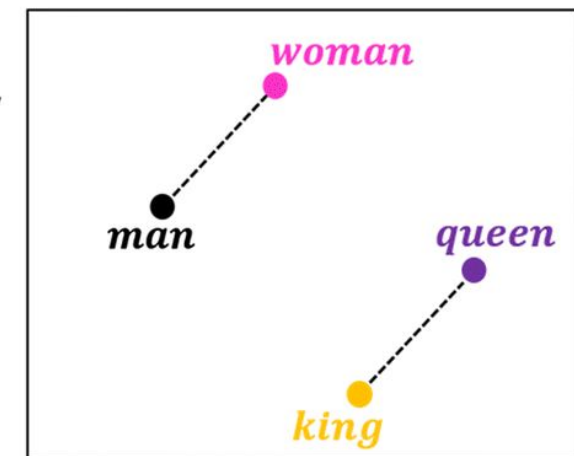
→



<i>man</i> →	0.6	-0.2	0.8	0.9	-0.1	-0.9	-0.7
<i>woman</i> →	0.7	0.3	0.9	-0.7	0.1	-0.5	-0.4
<i>king</i> →	0.5	-0.4	0.7	0.8	0.9	-0.7	-0.6
<i>queen</i> →	0.8	-0.1	0.8	-0.9	0.8	-0.5	-0.9

Dimensionality
reduction of
word
embeddings
from 7D to 2D

→



Word

Word embedding

Dimensionality
reduction

Visualization of word
embeddings in 2D

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

Chatbot: ELIZA

User1: You are like my father in some ways.

ELIZA1: WHAT RESEMBLANCE DO YOU SEE

User2: You are not very aggressive but I think you don't want me to notice that.

ELIZA2: WHAT MAKES YOU THINK I AM NOT AGGRESSIVE

User3: You don't argue with me.

ELIZA3: WHY DO YOU THINK I DON'T ARGUE WITH YOU

User4: You are afraid of me.

ELIZA4: DOES IT PLEASE YOU TO BELIEVE I AM AFRAID OF YOU

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**

Machine Translation

मेरा नाम रजत है | => My name is Rajat.

Grass is greener on the other side. => दूर के ढोल सुहावने ।

Google's Translation :

घास दूसरी तरफ हरियाली है ।

Q&A

Quora



who was the highest IQ

Who has the highest recorded IQ of all time?



Answer



Follow · 311



Request

3



Ad by Forge of Empires



Through the ages!

Develop a civilization from the stone age to the future in this award-winning game.



Learn More

81 Answers



Alan Muller, Self-taught

Answered December 16, 2016



William James Sidis is another individual whose exact IQ score is unknown. Experts have pegged his IQ somewhere in the range of 250 to 300, which would give him the highest IQ score in history.

William James Sidis was an american child prodigy with exceptional mathematical abilities and a claimed mastery of many languages. After his death, his sister made the unverifiable claim that his IQ was "the very highest that had ever been obtained", but any records of any IQ testing that Sidis actually took have been lost to history.

He entered Harward at age 11 and, as an adult, was claimed to be conversant in over 40 languages and dialects. It was later acknowledged, however, that some of the claims made were exaggerations, with a researcher stating "I have been researching the veracity of primary sources of various subjects for about twenty-eight years, and never before have I found a topic so satiated with lies, myths, half-truths, exaggerations, and other forms of misinformation as is in the history behind William Sidis"

Text Generation

Instruction: Given an input question, respond with syntactically correct PostgreSQL. Be creative but the SQL must be correct.

Input: how many users signed up in the past month?

GPT-3 Response: SELECT COUNT(*) FROM users
WHERE signup_time > now() - interval '1 month'