

Processamento de Linguagem Natural

Natural Language Processing: Do conceito de linguagem, prática simples e exemplos de uso do GPT3

Apresentação

Thiago Kuma

Ex-Engenheiro de Software

- Sistema de Controle de Tráfego Aéreo

Cientista de Dados desde 2014

- Mapa Vagas de Carreiras: <http://mapa-de-carreiras.ad.vagastec.com.br/mapa-de-carreiras/cargos/cientista-de-dados/0>
- Sistema de Recomendação: <https://www.vagas.com.br/vagas/v2088332/cientista-de-dados>
- Vídeo Análise: <https://youtu.be/KpoeMSCic3Q?t=940>

Professor Universitário

Hipnólogo de entretenimento

Processamento de Linguagem Natural

Neste bate papo vamos entender um pouco sobre as bases da NLP Natural Language Processing, conhecer um pouco do pipeline de produção e aplicações comuns e não tão comuns.

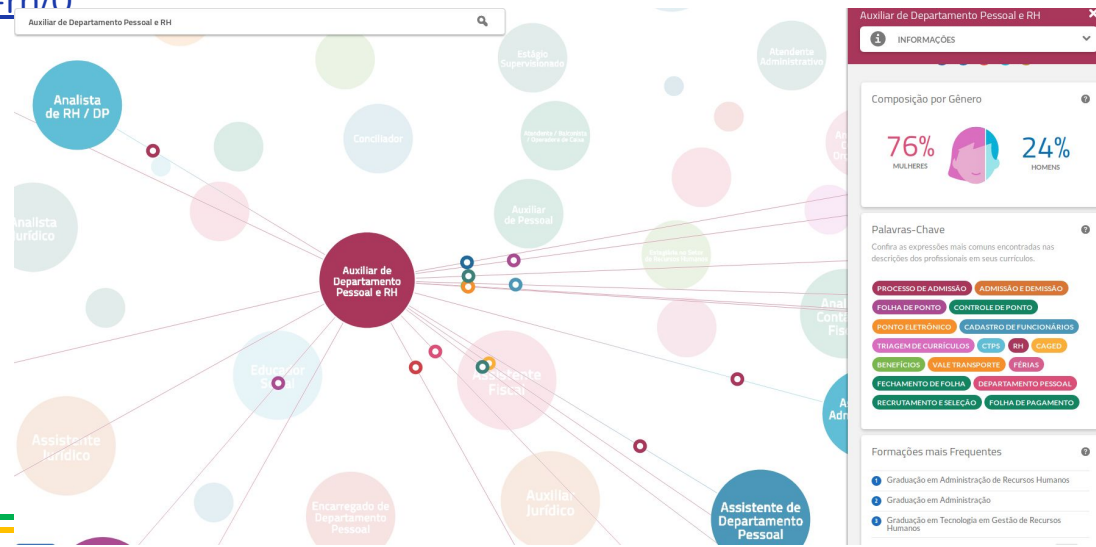
Resumidamente é a disciplina que trata da interpretação ou produção de linguagem humana por computador.

Processamento de Linguagem Natural

Aplicações:

- Tradução
- Análise de Sentimentos
- Correção Ortográfica
- OCR
- Reconhecimento de Fala
- Sintetização de Fala
- Previsão de Digitação
- Resumos
- ...

<https://www.vagas.com.br/mapa-de-carreiras/cargos/auxiliar-de-departamento-pessoal-e-rh/0>



Sistema de Recomendação - Este não tem NLP... Mas poderia

“... Entendi o problema implementamos uma PoC e depois a solução final ... Dobrou a quantidade de candidaturas, gerando um negócio novo de publicidade ...”

<https://www.vagas.com.br/vagas/v1938992/programador>



Programador
 v1938992
 Confidencial

 Salário a combinar

 Belo Horizonte

 Técnico

 Expira em 6 dias

Atuar com desenvolvimento e customização de software para plataformas desktop, web e mobile.

Desenvolvimento, manutenção, evolução e correção de softwares.

Realizar análise de requisitos e levantamento de funcionalidades com clientes para desenvolvimento de projetos de software.

Realizar controle e versionamento de software, Interagir com desenvolvedores, porte técnico e outros profissionais das filiais localizadas no exterior.

Realizar atendimento e suporte técnico para clientes via email, telefone, conexão móvel e ferramenta própria de atendimento de chamados.


Necessário conhecimentos em: Java, C#, PHP, HTML5, CSS3, JavaScript, JQuery,




Kona Container Car Guy
 Delivering Paradise One Family At A Time Since 1999. Visit Our Website Now!
 Kona Container/Car Guy **OPEN >**


Veja vagas similares


- Estagiário de TI - Desenvolvimento**
Estágio - DIRECIONAL ENGENHARIA
- Analista de Suporte Júnior**
Júnior/Trainee - Sisloc Softwares
- Analista de Suporte** Auxiliar/Operacional - HITSS
- Analista Desenvolvedor Júnior**
Júnior/Trainee
- Estágio em T.I. - Suporte Técnico**
Técnico - VINCI Energies




Estagiário de TI - Desenvolvimento
 v1945436
 DIRECIONAL ENGENHARIA

 Salário a combinar

 Belo Horizonte

 Estágio

 Expira em 10 dias

Video Entrevista

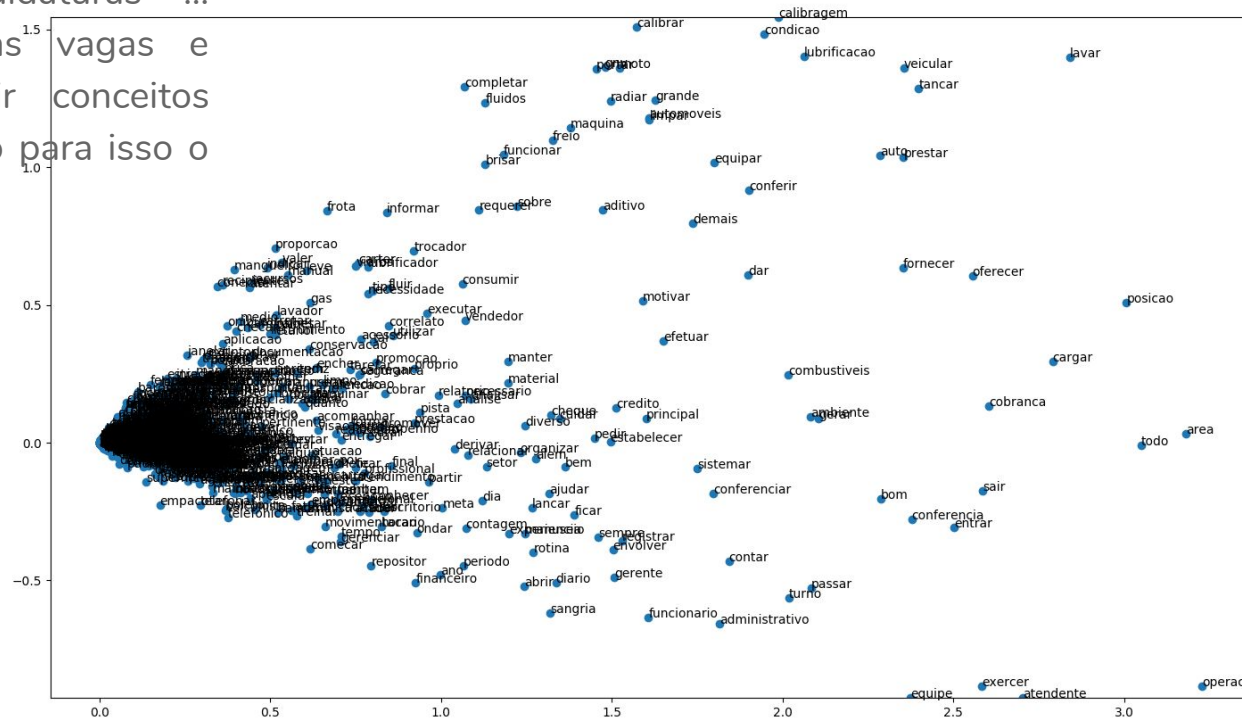
“... Processos com muitos candidatos ... Tempo e custo em agendar e a entrevista pessoalmente ... Assistir milhares de respostas ...”

<https://youtu.be/KpoeMSCic3Q?t=912>



Contextualização e Previsão de Escrita

“... Processos seletivos mais curtos, sem a necessidade de aguardar candidaturas ... Qualidade baixa nos textos das vagas e currículos ... A idéia é sugerir conceitos próximos aos já escritos, utilizando para isso o contexto ...”



Modelo de Linguagem

“... Bot para responder perguntas ... Baseado nos estados/eventos podemos formular frases que descrevem o cenário ... A IA vai entender a pergunta e encontrar a resposta que melhor se encaixa ... Essa escolha da IA pode ser uma devolutiva automática, ou um auxílio na montagem de resposta por e-mail, sms ou voz ...”

```
In [10]: doc = '''
Seu pacote tem o código de rastreio 12345.
Seu pacote será entregue hoje 9 horas.
Seu pacote será entregue pelo mensageiro Roberto.
Você pode ligar para 4020-1460.
No suporte é no telefone 4020-1460.
'''

In [4]: q = 'Qual o código de rastreio do meu pacote?'
print(model.predict(doc,q))

{'answer': '12345', 'start': 7, 'end': 7, 'confidence': 0.7217687372603117, 'document': ['Seu', 'pacote', 'tem', 'o', 'código', 'de', 'rastreio', '12345.', 'Seu', 'pacote', 'será', 'entregue', 'hoje', '9', 'horas.', 'Seu', 'pacote', 'será', 'entregue', 'pelo', 'mensageiro', 'Roberto.', 'Você', 'pode', 'ligar', 'para', '4020-1460.', 'No', 'Suporte', 'é', 'no', '4020-1460.']}

In [5]: q = 'Que horas meu pacote chega?'
print(model.predict(doc,q))

{'answer': 'entregue hoje 9 horas', 'start': 11, 'end': 14, 'confidence': 0.18028255351242078, 'document': ['Seu', 'pacote', 'tem', 'o', 'código', 'de', 'rastreio', '12345.', 'Seu', 'pacote', 'será', 'entregue', 'hoje', '9', 'horas.', 'Seu', 'pacote', 'será', 'entregue', 'pelo', 'mensageiro', 'Roberto.', 'Você', 'pode', 'ligar', 'para', '4020-1460.', 'No', 'Suporte', 'é', 'no', '4020-1460.']}

In [6]: q = 'Onde está minha encomenda?'
print(model.predict(doc,q))

{'answer': 'pacote será entregue pelo mensageiro Roberto.', 'start': 16, 'end': 21, 'confidence': 0.13394676277640352, 'document': ['Seu', 'pacote', 'tem', 'o', 'código', 'de', 'rastreio', '12345.', 'Seu', 'pacote', 'será', 'entregue', 'hoje', '9', 'horas.', 'Seu', 'pacote', 'será', 'entregue', 'pelo', 'mensageiro', 'Roberto.', 'Você', 'pode', 'ligar', 'para', '4020-1460.', 'No', 'Suporte', 'é', 'no', '4020-1460.']}

In [12]: q = 'Preciso de suporte'
print(model.predict(doc,q))

{'answer': 'no telefone 4020-1460', 'start': 30, 'end': 32, 'confidence': 0.45155670729884584, 'document': ['Seu', 'pacote', 'tem', 'o', 'código', 'de', 'rastreio', '12345.', 'Seu', 'pacote', 'será', 'entregue', 'hoje', '9', 'horas.', 'Seu', 'pacote', 'será', 'entregue', 'pelo', 'mensageiro', 'Roberto.', 'Você', 'pode', 'ligar', 'para', '4020-1460.', 'No', 'Suporte', 'é', 'no', 'telefone', '4020-1460.']}
```

Processamento de Linguagem Natural

Conceitos de Linguagem:

- Morfologia: estrutura das palavras (Substantivo, Artigo, etc)
- Sintaxe: a forma em que as palavras são utilizadas (Sujeito, Verbo, etc)
- Semântica: Significado
- Pragmática: Significado no contexto
- ...

Processamento de Linguagem Natural

Linguagem vs Computação:

- Morfologia, Sintaxe, Semântica, Pragmática...

Corpus: Conjunto de textos estruturados (<http://comet.fflch.usp.br/corporaportugues>).
Conjunto de textos.

Annotations: Qualquer tag que marque uma característica dos elementos

Tokenization: Em sentenças/palavras

Processamento de Linguagem Natural

Word Embedding

3 Documentos:

Número	Texto
0	Comprei um jogo ontem de manhã.
1	João assistiu ao jogo do Cruzeiro ontem.
2	Ontem e hoje eu trabalhei, ontem não dormi.

Dicionário:

0	comprei	6	joão	12	e
1	um	7	assistiu,	13	hoje
2	jogo	8	ao	14	eu
3	ontem	9	jogo,	15	trabalhei
4	de	10	do	16	não
5	manhã	11	cruzeiro	17	dormi

Documento 1: [0, 0, 0, 1, 0, 0, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0]

Processamento de Linguagem Natural

Análise morfológica

“... estudar cada uma das diversas palavras em uma frase independentemente, visando sua classe gramatical”:

No exemplo "*A Wikipédia é uma enciclopédia livre.*"

- **A** é um artigo definido.
- **Wikipédia** é um substantivo próprio derivado simples.
- **é** é um verbo na terceira pessoa do singular, no presente do indicativo.
- **uma** é um artigo indefinido.
- **enciclopédia** é um substantivo comum derivado simples.
- **livre** é um adjetivo.

Forma mais genérica de aprendizado

Exemplo

Processamento de Linguagem Natural

Stemização

“... reduzir palavras flexionadas (ou às vezes derivadas) ao seu tronco (stem)...”

Stemização

Removedor de Sufixos da Língua Portuguesa - RSLP

```
import nltk
from nltk.stem import RSLPStemmer

stemmer = nltk.stem.RSLPStemmer()
```

```
palavras = ['copiar', 'copiando', 'copiado']

for palavra in palavras:
    print(palavra, stemmer.stem(palavra))
```

```
copiar copi
copiando copi
copiado copi
```

Processamento de Linguagem Natural

Lematização

“... representar as palavras através do infinitivo dos verbos e masculino singular dos substantivos e adjetivos..”

Alternativa

-

<https://pt.wikipedia.org/wiki/Hunspell>

Lematização

```
import hunspell

dic = 'hunspell/pt_BR.dic'
aff = 'hunspell/pt_BR.aff'

dic_hunspell = hunspell.HunSpell(dic, aff)

palavras = ['copiar', 'copiando', 'copiado']

for palavra in palavras:
    print(dic_hunspell.stem(palavra))

[b'copiar']
[b'copiar']
[b'copiar']
```

Processamento de Linguagem Natural

Frequencia de termo... Exemplo simples

```
for i in sorted(idx_sentencas_importantes):  
    print(sentencas[i])  
    print()
```

32.912 VIDAS PERDIDAS: julho foi o mês com mais mortes por Covid-19 no Brasil

Drauzio lembrou que, apesar das recomendações de cientistas sobre a necessidade de isolamento para conter a disseminação do vírus, houve "contradição" nas orientações dadas à população pelos governos estaduais e o federal.

O médico também listou o fato de que um remédio que não tem eficácia comprovada contra a Covid-19 foi "apregoadado" como a "salvação" contra a doença.

"Pode ser que, quando chegue essa vacina, ela não vá ser tão necessária quanto é agora, porque pode ser que, até o fim do ano que vem, você vai ter já uma epidemia mais arrefecida".

Imunidade contra o coronavírus: tire dúvidas sobre como anticorpos e células T atuam contra a Covid-19

"Contra os vírus, a imunidade mais importante é a celular, não é a humoral [de anticorpos]", explicou Drauzio.

O famoso TF-IDF

*TFIDF score for term i in document j = $TF(i,j) * IDF(i)$*

where

IDF = Inverse Document Frequency

TF = Term Frequency

$$TF(i,j) = \frac{\text{Term i frequency in document j}}{\text{Total words in document j}}$$

$$IDF(i) = \log_2 \left(\frac{\text{Total documents}}{\text{documents with term i}} \right)$$

and

t = Term

j = Document

Processamento de Linguagem Natural

- Em uma conversa com um amigo é comum você tentar completar frases
 - E como você faz isso?
 - Tendo algum conhecimento sobre o assunto(**tópico**) e conhecendo que palavras(**termos**) pode encaixar
 - Utilizando IA é o mesmo
- Utilizamos aqui PLN - Processamento de Linguagem Natural
 - **Tópico**: Latent Semantic Analysis, ou seja, semântica: num sistema linguístico, o componente do sentido das palavras e da interpretação das sentenças e dos enunciados.
 - Hidratar as **mãos**... (**Parte do corpo humano**)
 - Abriu **mãos** do seus direitos... (**Direito**)
 - Cobrado **mão** de obra (**Construção**)

Processamento de Linguagem Natural

Tópicos...

Topic Modeling: é uma técnica de mineração de texto que provê métodos para identificar recorrência de palavras chaves.

Essa recorrência por si só revelam tópicos que poderiam aparecer em um dado documento.

Em um sistema **não supervisionado** você pode encontrar um grupo de palavras de um dado documento: Tópico

Um documento pode se relacionar com diferentes temas

- Mesmo regido no contrato abriu **mão** de cobrar os 15% sobre a **mão** de obra...

Processamento de Linguagem Natural

Tópicos...

E como criamos tópicos?

Existem várias técnicas desde um simples Bag of Words, Latent Semantic Analysis, Skip-gram, etc

Processamento de Linguagem Natural

Latent Semantic Analysis

A					=	U					x	S					x	Vt				
	d1	d2	d3	d4			f1	f2	f3	f4			f1	f2	f3	f4			d1	d2	d3	d4
a	6	7	1	0		a	0.24	-0.51	0.08	0.06		f1	23.1	0	0	0		f1	0.37	0.38	0.65	0.53
b	8	6	0	1		b	0.25	-0.54	-0.64	-0.23		f2	0	14.3	0	0		f2	-0.55	-0.63	0.37	0.38
c	6	9	8	5		c	0.58	-0.28	0.57	0.13		f3	0	0	3.5	0		f3	-0.69	0.59	0.27	-0.21
d	0	1	8	8		d	0.42	0.37	0.16	-0.68		f4	0	0	0	1.5		f4	0.26	-0.29	0.59	-0.69
e	2	0	9	7		e	0.44	0.34	-0.24	0.66												
f	2	0	7	7		f	0.39	0.29	-0.40	-0.09												

Relevância palavra por tópico

Relevância documento por tópico

Algebra... Fatoração(Decomposição), no caso A
 Decomposição em Valores Singulares

LSA = SDV em texto
 SVD = Matemática

Inteligência Artificial ao alcance de todos.

www.escolalivre-ia.com.br

Processamento de Linguagem Natural

```
experiencias = [line.rstrip() for line in open('experiencias.csv')]

stopwords = set(w.rstrip() for w in open('stopwords_pt_br.txt'))

def my_tokenizer(s):
    s = s.lower()
    tokens = nltk.tokenize.word_tokenize(s) # divide frase em palavras (tokens)
    # tokens = [wordnet_stem.stem(t) for t in tokens] # faz o stem das palavras
    tokens = [t for t in tokens if t not in stopwords] # remove stopwords
    tokens = [t for t in tokens if not any(c.isdigit() for c in t)] # remove numeros"
    return tokens

palavra_indice_map = {}
current_index = 0
todos_tokens = []
todas_experiencias = []
indice_palavra_map = []
error_count = 0
for experiencia in experiencias:
    try:
        experiencia = experiencia.encode('ascii', 'ignore').decode('utf-8')
        todas_experiencias.append(experiencia)
        tokens = my_tokenizer(experiencia)
        todos_tokens.append(tokens)
        for token in tokens:
            if token not in palavra_indice_map:
                palavra_indice_map[token] = current_index
                current_index += 1
            indice_palavra_map.append(token)
    except Exception as e:
        print(e)
        print(experiencia)
        error_count += 1
```

Um exemplo simples de aplicação...

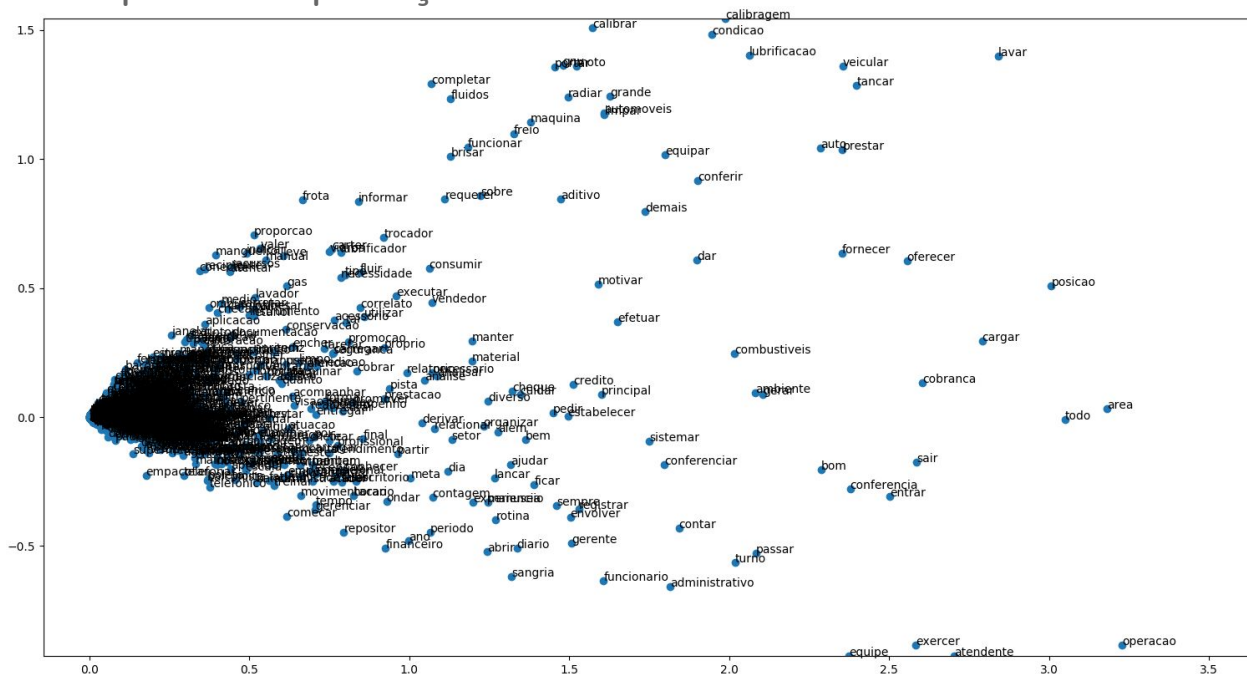
```
# matriz de input: A
def tokens_to_vector(tokens):
    x = np.zeros(len(palavra_indice_map))
    for t in tokens:
        i = palavra_indice_map[t]
        x[i] = 1
    return x

N = len(todos_tokens)
D = len(palavra_indice_map)
X = np.zeros((D, N))
i = 0
for tokens in todos_tokens:
    X[:,i] = tokens_to_vector(tokens)
    i += 1

def main():
    svd = TruncatedSVD()
    Z = svd.fit_transform(X)
    plt.scatter(Z[:,0], Z[:,1])
    for i in range(D):
        plt.annotate(s=indice_palavra_map[i], xy=(Z[i, 0], Z[i, 1]))
    plt.show()
```

Processamento de Linguagem Natural

Um exemplo simples de aplicação...



Processamento de Linguagem Natural

Mergulhando mais em tópicos (contexto/assunto)

word2vec

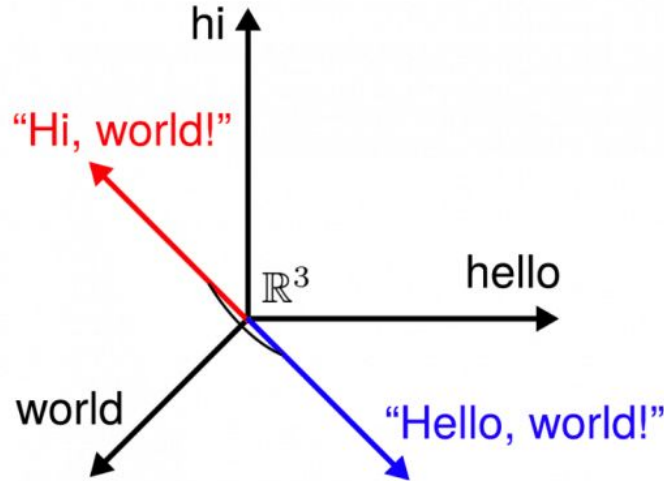
Algoritmo que usa redes neurais para aprender a associações de um grande número de documentos.

Uma vez treinada pode encontrar palavras similares ou sugerir palavras adicionais na escrita.

Utiliza o que chamamos de distância de cossenos.

Processamento de Linguagem Natural

word2vec



Cosine Similarity

Exemplo

Processamento de Linguagem Natural

Exemplo... word2vec

```
model = gensim.models.Word2Vec.load('modelo_anuncio_vagas.pkl')
```

```
print(model.wv.most_similar(positive=['java']))
```

```
[('j2ee', 0.8667133450508118), ('php', 0.8459796905517578), ('jee', 0.8373323082923889), ('javascript', 0.8112343549728394), ('ee', 0.8088476657867432), ('swing', 0.7752840518951416), ('powerbuilder', 0.7746198177337646), ('j2se', 0.774611234664917), ('javaee', 0.774280309677124), ('vb', 0.7732024192810059)]
```

```
print(model.wv.most_similar(positive=['assistente', 'pedreiro'], negative=['adiministrativo']))
```

```
[('pintor', 0.36449509859085083), ('vestir', 0.36332792043685913), ('servente', 0.3597457706928253), ('carpinteira', 0.3361269235610962), ('pl', 0.3312763571739197), ('transp', 0.312492698431015), ('forrar', 0.3069591224193573), ('gruir', 0.30652686953544617), ('jaguari', 0.30394673347473145), ('iii', 0.30104881525039673)]
```

```
print(model.wv.most_similar(positive=['assistente', 'pedreiro']))
```

```
[('servente', 0.5887385606765747), ('opdg', 0.5872374773025513), ('jaguari', 0.5812460780143738), ('adiministrativo', 0.5793720483779907), ('assitente', 0.5682974457740784), ('recepcionsita', 0.5682178735733032), ('piracaia', 0.5660809278488159), ('carpinteirar', 0.5653356313705444), ('assistende', 0.5608928203582764), ('janaina', 0.550359308719635)]
```

Processamento de Linguagem Natural

Vários nomes... muitas interseções

NLU: Natural Language Understanding

- Semântica / Análise de Sentimento / Sintática

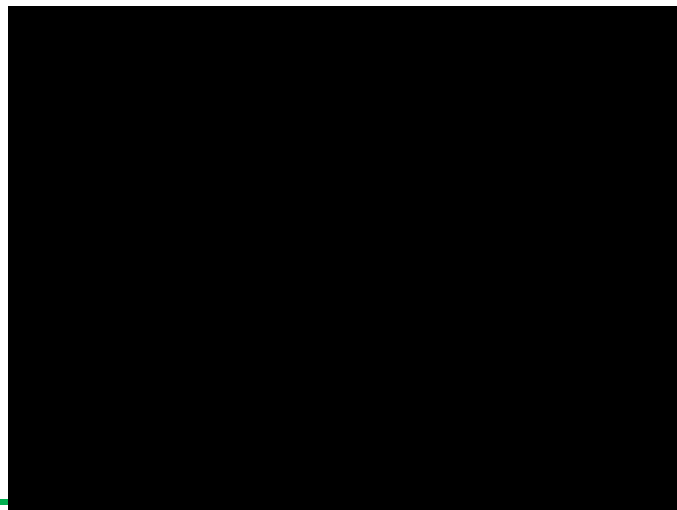
NLG: Natural Language Generation

- Question & Answer / GPT3 - Generative Pre-trained Transformer 3

Processamento de Linguagem Natural

Estado da arte

- Transformers
 - “RNNs, Transformers do not require that the sequential data be processed in order”
- BERT - Bidirectional Encoder Representations from Transformers
 - Google
 - <https://github.com/neuralmind-ai/portuguese-bert>
- GTP3
 - OpenIA
 - <https://gpt3examples.com/>



Processamento de Linguagem Natural

Exemplo... QA

```
model = load_model()
```

```
doc = '''  
Eu me acho mais inteligente que a professora porque eu passei de ano e ela continuou no mesmo.  
É tempo perdido!  
Ora, professora, porque eu moro em apartamento e a senhora deixou bem claro que era lição de CASA!  
Querer eu quero, mas a minha mãe falou que depois da aula era para eu ir direto para casa!  
'''
```

```
q = 'Joãozinho por que você se achama mais inteligente que a professora?'  
print(model.predict(doc,q))
```

```
{'answer': 'porque eu passei de ano e ela continuou no mesmo', 'start': 8, 'end': 17, 'confidence': 0.505766773520  
8815, 'document': ['Eu', 'me', 'acho', 'mais', 'inteligente', 'que', 'a', 'professora', 'porque', 'eu', 'passei',  
'de', 'ano', 'e', 'ela', 'continuou', 'no', 'mesmo.', 'É', 'tempo', 'perdido!', 'Ora', 'professora,', 'porque', '  
eu', 'moro', 'em', 'apartamento', 'e', 'a', 'senhora', 'deixou', 'bem', 'claro', 'que', 'era', 'lição', 'de', 'CAS  
A!', 'Querer', 'eu', 'quero,', 'mas', 'a', 'minha', 'mãe', 'falou', 'que', 'depois', 'da', 'aula', 'era', 'para',  
'eu', 'ir', 'direto', 'para', 'casa!']}
```

```
q = '-Joãozinho qual é o tempo da frase? :Eu procuro um homem fiel.'  
print(model.predict(doc,q))
```

```
{'answer': 'tempo perdido!', 'start': 19, 'end': 20, 'confidence': 0.40853993553775664, 'document': ['Eu', 'me', '  
acho', 'mais', 'inteligente', 'que', 'a', 'professora', 'porque', 'eu', 'passei', 'de', 'ano', 'e', 'ela', 'contin  
vou', 'no', 'mesmo.', 'É', 'tempo', 'perdido!', 'Ora,', 'professora,', 'porque', 'eu', 'moro', 'em', 'apartamento  
, 'e', 'a', 'senhora', 'deixou', 'bem', 'claro', 'que', 'era', 'lição', 'de', 'CASA!', 'Querer', 'eu', 'quero,',  
'mas', 'a', 'minha', 'mãe', 'falou', 'que', 'depois', 'da', 'aula', 'era', 'para', 'eu', 'ir', 'direto', 'para', '  
casa!']}
```

Artificial intelligence can improve sales by four times compared to some human employees

"... Our findings show when people don't know about the use of artificial intelligence (AI) chatbots they are four times more effective at selling products than inexperienced workers, but when customers know the conversational partner is not a human, they are curt and purchase less because they think the bot is less knowledgeable and less empathetic..."

Q: When did Kendrick Lamar's first album come out?

Article

Talk

Read

View source

View history

Search Wikipedia

Kendrick Lamar

From Wikipedia, the free encyclopedia

An accepted version of this page, accepted on 28 February 2018, was based on this revision.

Kendrick Lamar Duckworth (born June 17, 1987) is an American rapper and songwriter. Raised in Compton, California, Lamar embarked on his musical career as a teenager under the stage name **K-Dot**, releasing a mixtape that garnered local attention and led to his signing with indie record label Top Dawg Entertainment (TDE).

He began to gain recognition in 2010, after his first retail release, *Overly Dedicated*. The following year, he independently released his first studio album, *Section.80*, which included his debut single, "HiiiPoWeR". By that time, he had amassed a large online following and collaborated with several prominent artists in the hip hop industry, including The Game, Busta Rhymes, and Snoop Dogg.

Lamar's major label debut album, *good kid, m.A.A.d city*, was released in 2012 by TDE, *Aftermath*, and *Interscope Records* to critical success. It debuted at number two on the US *Billboard* 200 chart and was later certified platinum by the Recording Industry Association of America (RIAA). The record contained the top 40 singles "Swimming Pools (Drank)", "Bitch, Don't Kill My Vibe", and "Poetic Justice".

His critically acclaimed third album *To Pimp a Butterfly* (2015) comprised funk, soul, and spoken word, debuted atop the charts in the US and the UK, and won the Grammy Award for Best Rap Album at the 58th ceremony. In 2016, Lamar released *untitled unmastered*, a collection of unreleased demos that originated during the recording sessions for *Butterfly*. He released his fourth album *Damn* in 2017 to further acclaim; its lead single "Humble" topped the US *Billboard* Hot 100 chart.

Lamar has received a number of accolades over the course of his career, including twelve Grammy Awards. In early 2013, MTV named him the number one "Hottest MC in the Game", on their annual list.^[1] *Time* named him one of the 100 most influential people in the world in 2016.^[2] Aside from his solo career, Lamar is also known as a member of the West Coast hip hop supergroup Black Hippy, alongside his TDE label-mates and fellow South Los Angeles-based rappers Ab-Soul, Jay Rock, and Schoolboy Q.

Contents

hide

Kendrick Lamar

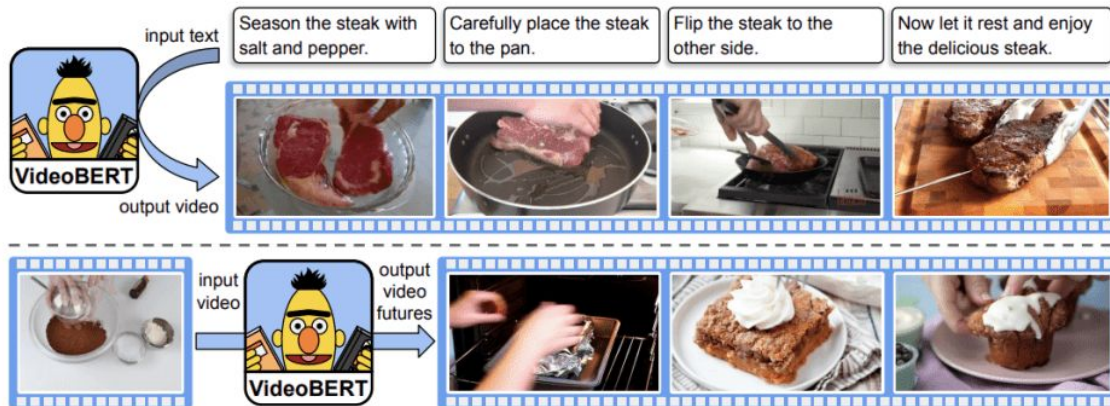


Lamar in 2016

Born
 Kendrick Lamar Duckworth
 June 17, 1987 (age 31)
 Compton, California, U.S.

What's Cooking? Google VideoBERT Predicts Recipes

"... Model text and video representation... Given a few video frames that include a bowl of flour and cocoa powder, VideoBERT can speculate that the following video frames might involve baking a brownie or cupcake... Technically predicting missing word tokens or video



Input text

"Cut the *steak* into pieces."

Retrieved centroid



Retrieved centroid



"Cut the *carrots* into pieces."



to pieces."

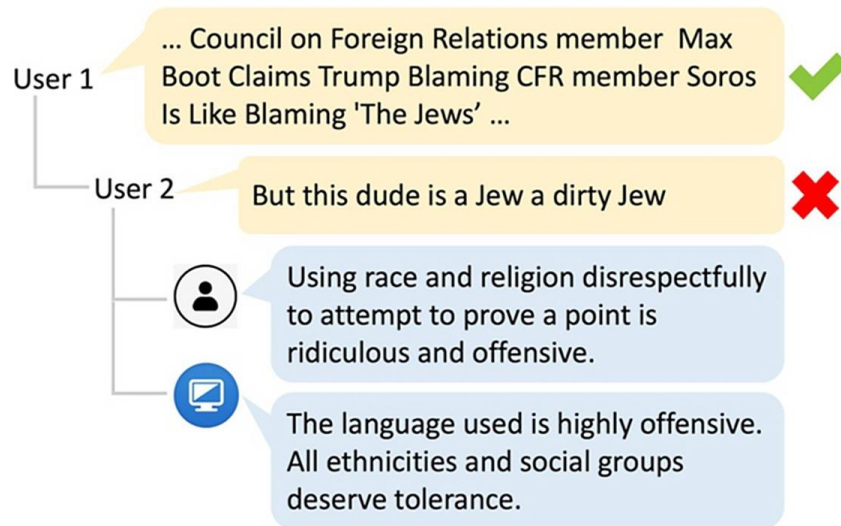


thin slices."



NLP Techniques to Intervene in Online Hate Speech

"... NLP Techniques to Intervene in Online Hate Speech ... main goal of this research is to investigate if we can automatically generate hate speech intervention comments ..."



Developer uses drone and AI to find and recognize SOS messages

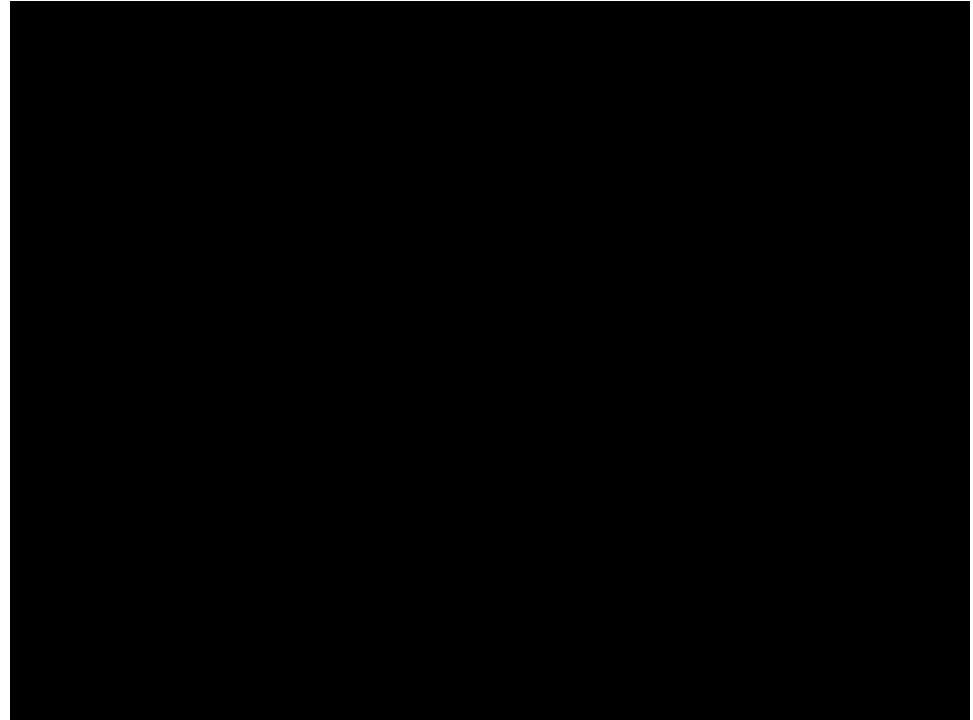
Here's a cool story of drones being used for good! A developer is using aerial images taken with a drone, a DJI Mavic Pro (DJI, Amazon) in combination with artificial intelligence (AI) software to recognize SOS messages painted on the streets in Puerto Rico.

On September 20, 2017, Hurricane Maria struck my home, Puerto Rico. After surviving the record-breaking Category 5 storm and being personally affected by its aftermath, I decided I was going to make it my mission to create technology that could help mitigate the impact hurricanes have on our island.



Geena Davis announces 'Spellcheck for Bias' tool to redress gender imbalance in movies

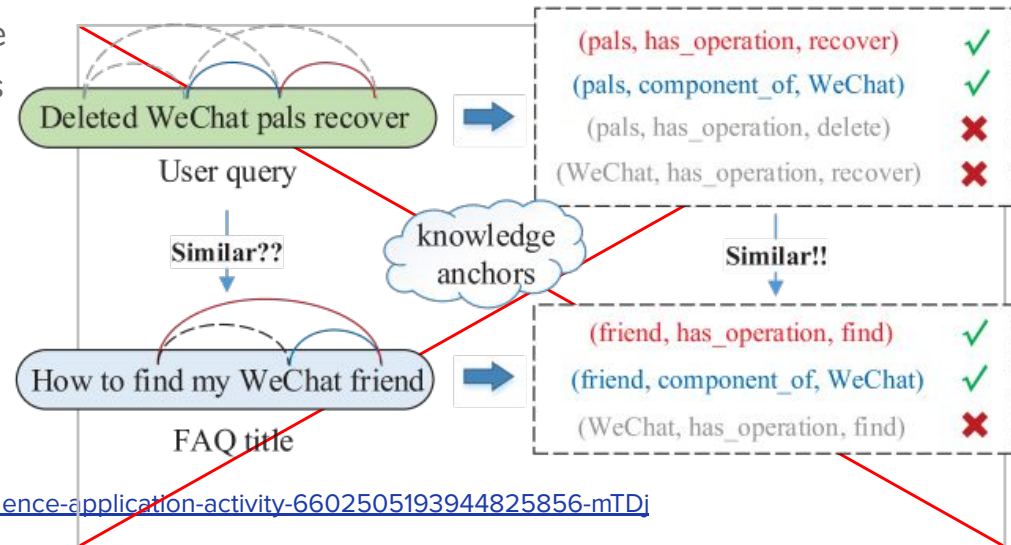
“... 'Spellcheck for Bias' tool to redress gender imbalance in movies ... analyse scripts and reveal unconscious biases ... determine the percentages of characters’ “gender, race, LGBTQIA [and] disabilities ...”



19/10/09 - https://www.linkedin.com/posts/thiago-kuma_gdiq-spellcheck-bias-final-92319-activity-6590547234239070208-N38e

FAQ-based Question Answering via Knowledge Anchors

“... framework for FAQ-based QA to better understand questions and retrieve more appropriate answers ... knowledge graph construction, query anchoring and query-document matching ... knowledge anchors to precisely capture the core semantics ...”



DIALOGPT : Large-Scale Generative Pre-training for Conversational Response Generation

“... tunable neural conversational response generation model ... extracted from Reddit comment chains ...”

source	Response
who is the first president of the United States?	George Washington
what is the boiling point of water?	I think it's about 212 F.
which one is bigger, sun or moon?	The sun .
which animal has black and white stripes?	A black and white striped cat .

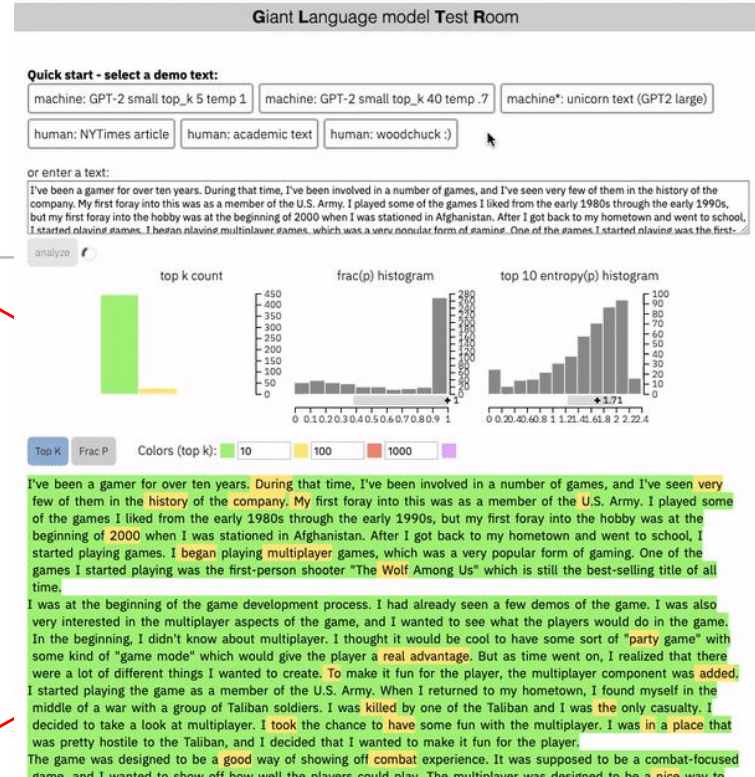
Catching a Unicorn with GLTR: A tool to detect automatically generated text

“... fake news detecting ... identify machine-generated text is through a neat combination of statistical analysis and visualizations that are made for a given piece of text ...”

Referência

unicorn:

<https://openai.com/blog/better-language-model/#sample1>



19/12/16 - https://www.linkedin.com/posts/thiago-kuma_artificialintelligence-application-activity-6612280697245716480-8VeJ

Inteligência Artificial ao alcance de todos.

www.escolalivre-ia.com.br

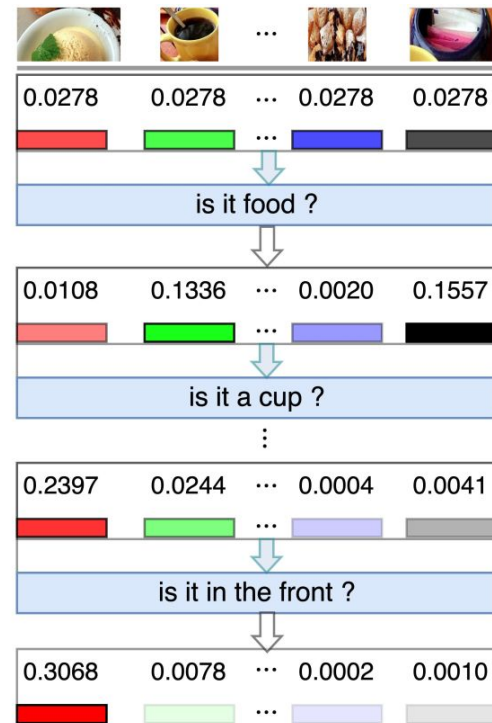
Visual Dialogue State Tracking for Question Generation

“... GuessWhat? ... visual dialogue state tracking (VDST) based method for question generation
 ...”



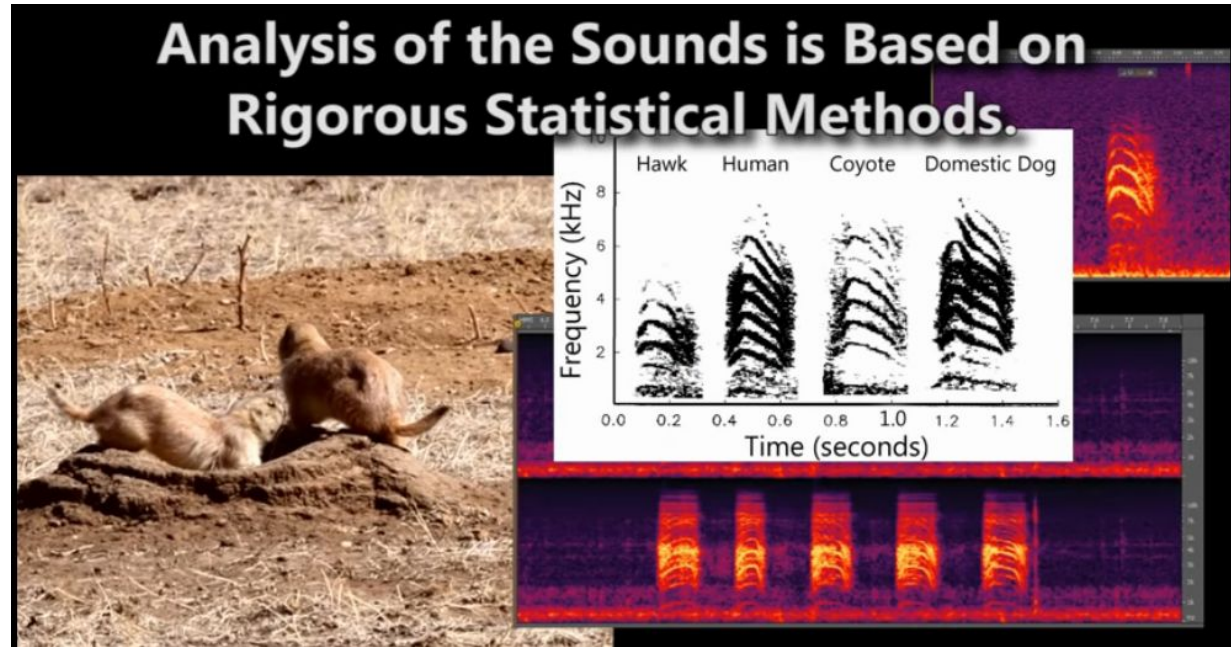
No.	QGen	Oracle
1:	is it food ?	No
2:	is it a cup ?	No
3:	is it a bowl ?	Yes
4:	on the left ?	Yes
5:	is it in the front ?	Yes

Guesser



helping people have great relationships with animals

“... to translate dog body language and sounds
to English ...”

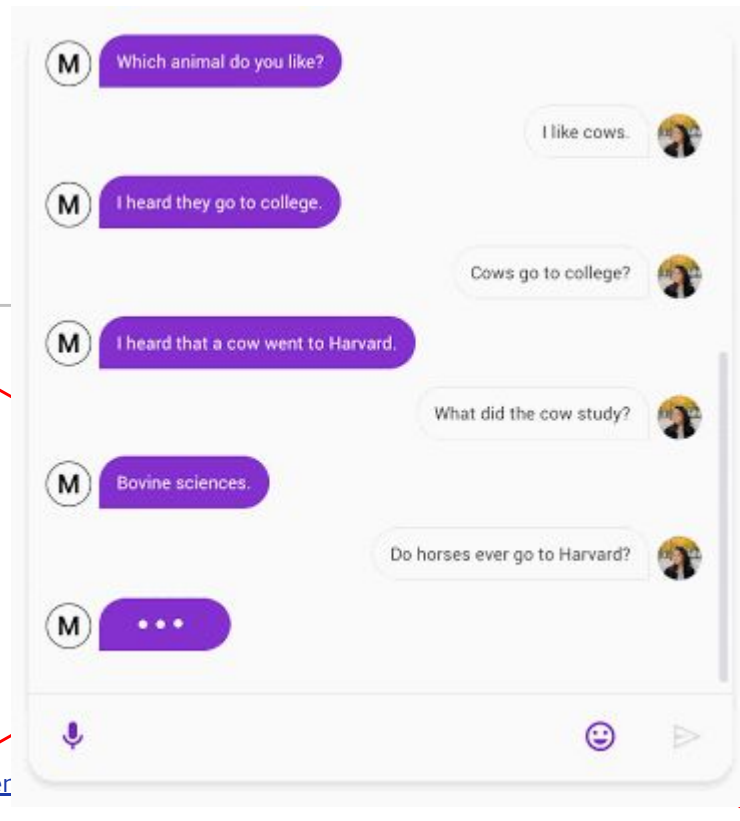


20/01/27 - https://www.linkedin.com/posts/thiago-kuma_artificialintelligence-activity-6627864927065432064-Ye7M

Google working on human-like chatbots that contextually respond to anything

“... multi-turn open-domain chatbot trained end-to-end on data mined and filtered from public domain social media conversations ... Meena executes a multi-turn joke in an open-domain setting ...”

Harvard > Hayvard (Hay = Feno)



20/01/28 - https://www.linkedin.com/posts/thiago-kuma_artificialintelligence-google-meer

Referências

https://pt.wikipedia.org/wiki/An%C3%A1lise_morfol%C3%B3gica

<https://pt.wikipedia.org/wiki/Stemiza%C3%A7%C3%A3o>

<https://www.ime.usp.br/~slago/IA-pln.pdf>

<https://www.oreilly.com/library/view/natural-language-annotation/9781449332693/ch01.html>

Muito obrigado!



Inteligência Artificial ao alcance de todos.

www.escolalivre-ia.com.br