



# Como os computadores entendem a língua dos humanos



---

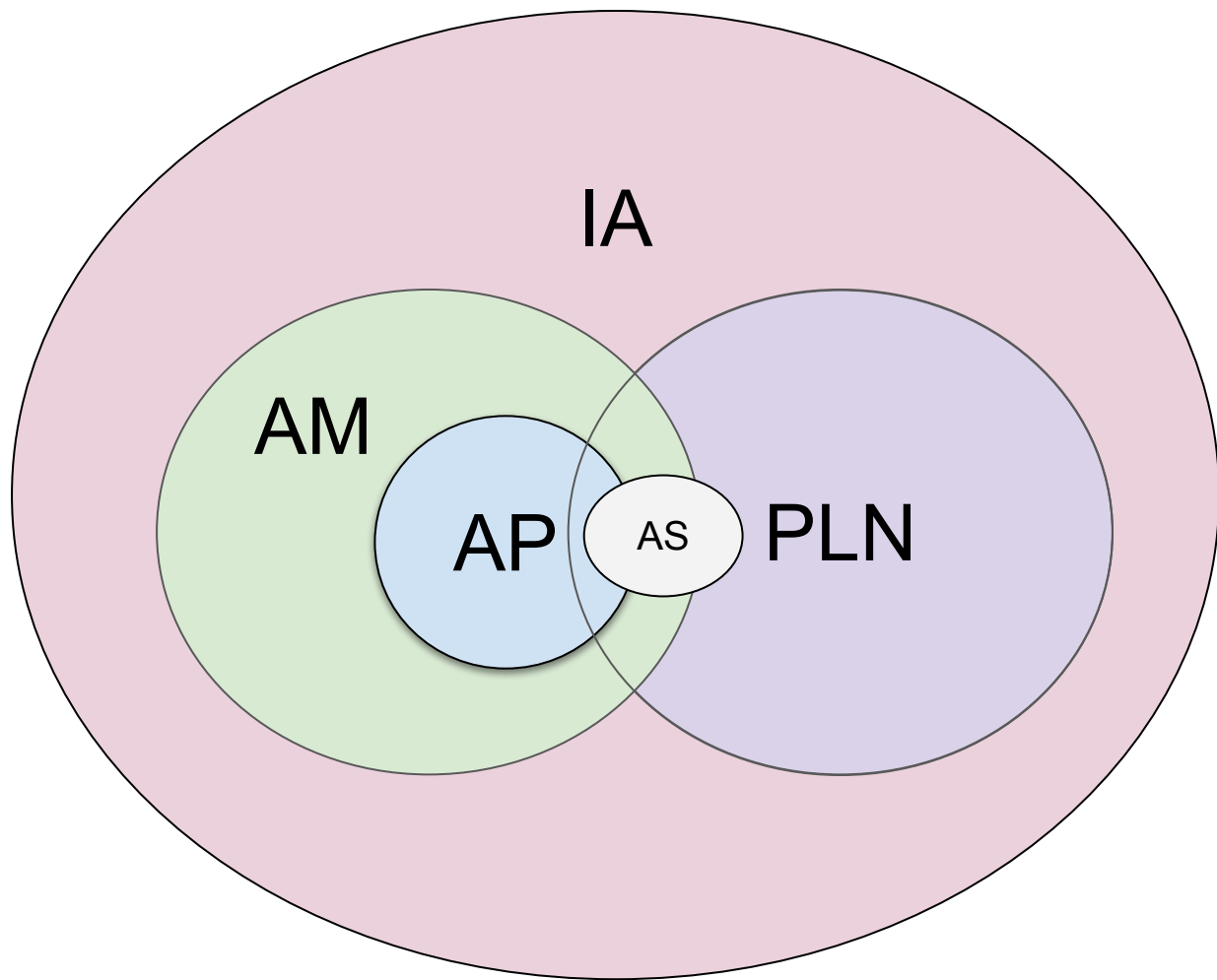
**Grupo de Pesquisa em Análise de Sentimento e Processamento de Linguagem Natural**  
**Hub de Inovação em Inteligência Artificial (H2IA)**  
**Centro de Desenvolvimento Tecnológico (CDTec)**  
**Universidade Federal de Pelotas (UFPEL)**

**Brasileiras em Processamento de Linguagem Natural**

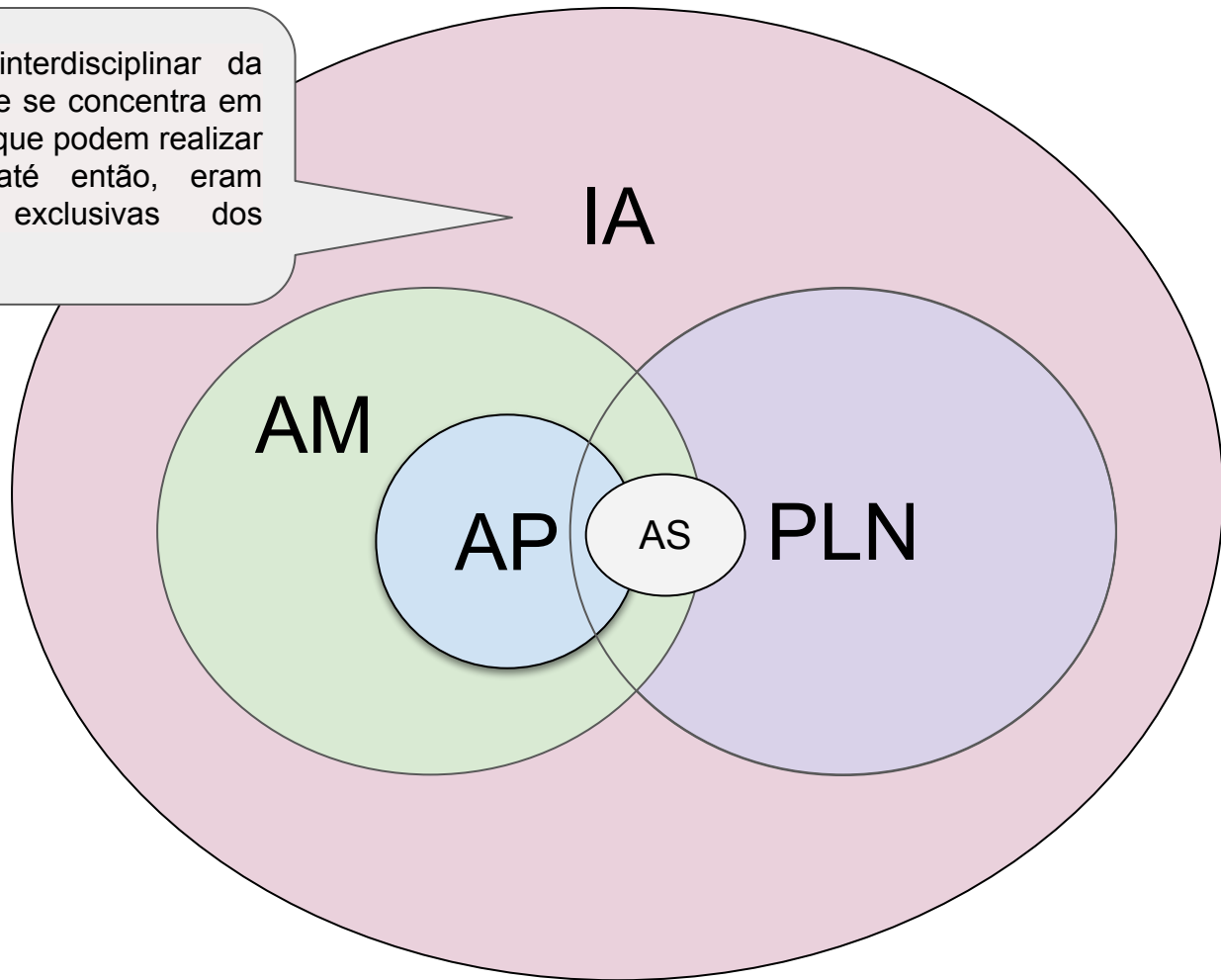
Larissa Astrogildo de Freitas  
larissa@inf.ufpel.edu.br

**IA x AM x AP x PLN x AS**

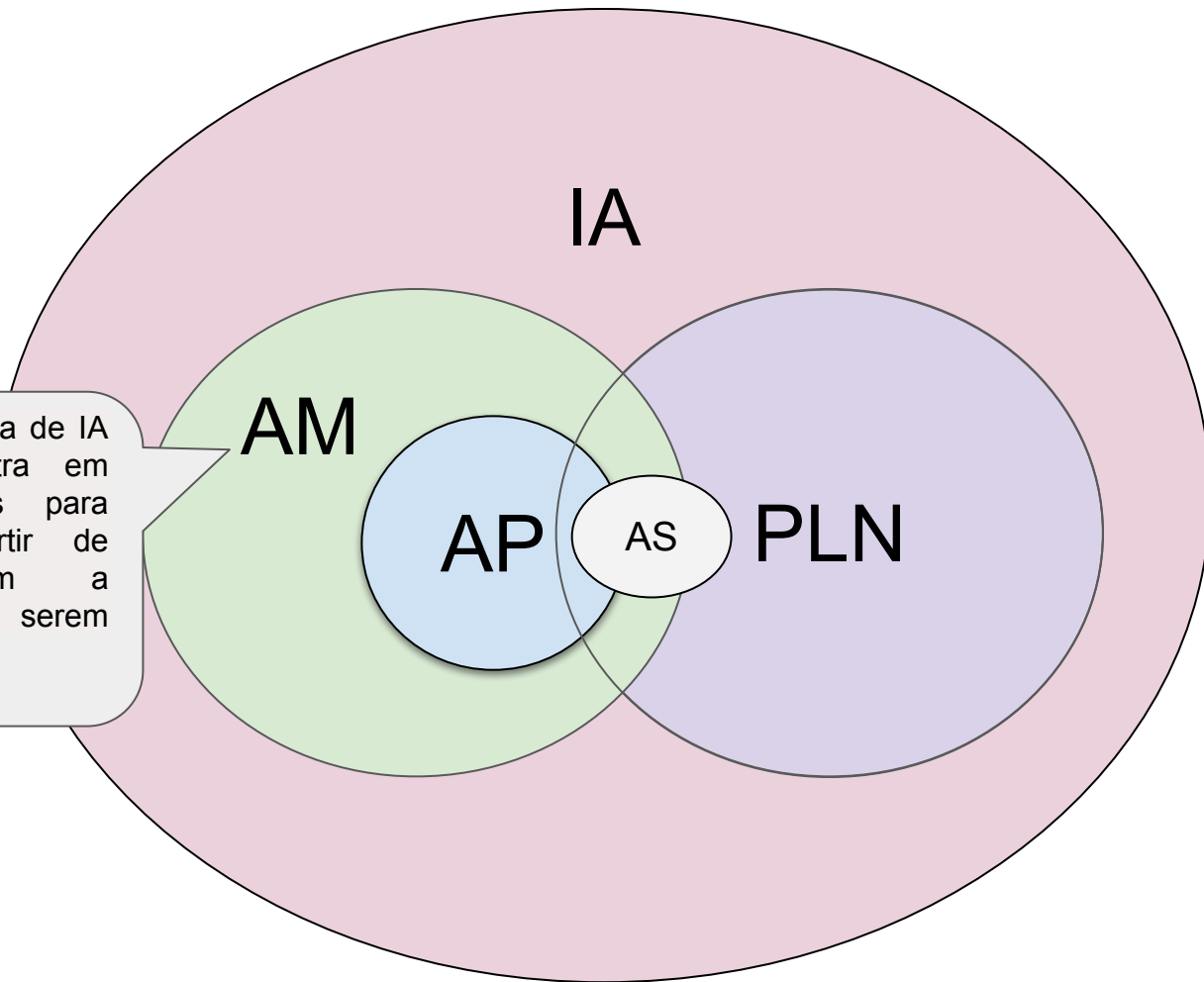
---



É uma área interdisciplinar da computação que se concentra em criar máquinas que podem realizar tarefas que, até então, eram consideradas exclusivas dos seres humanos.

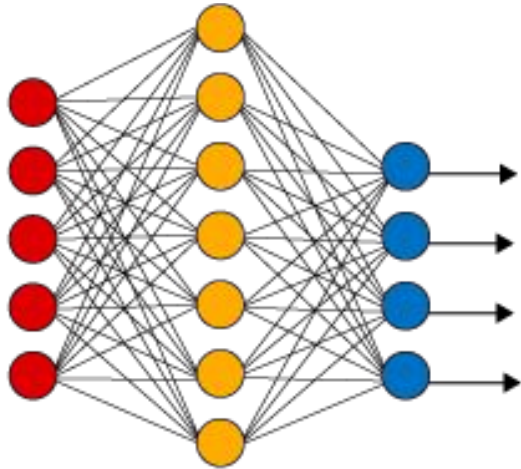






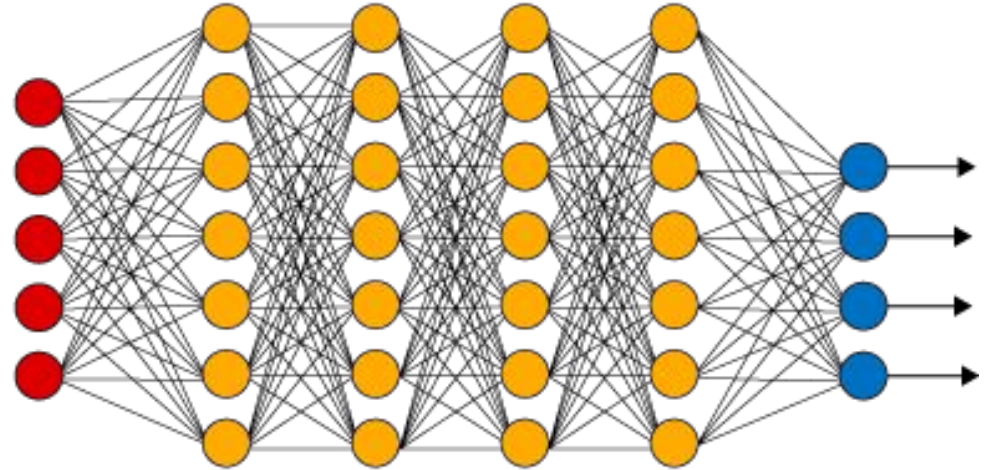
AM é uma subárea de IA que se concentra em treinar máquinas para aprender a partir de exemplos, sem a necessidade de serem explicitamente programadas.

## Redes Neurais



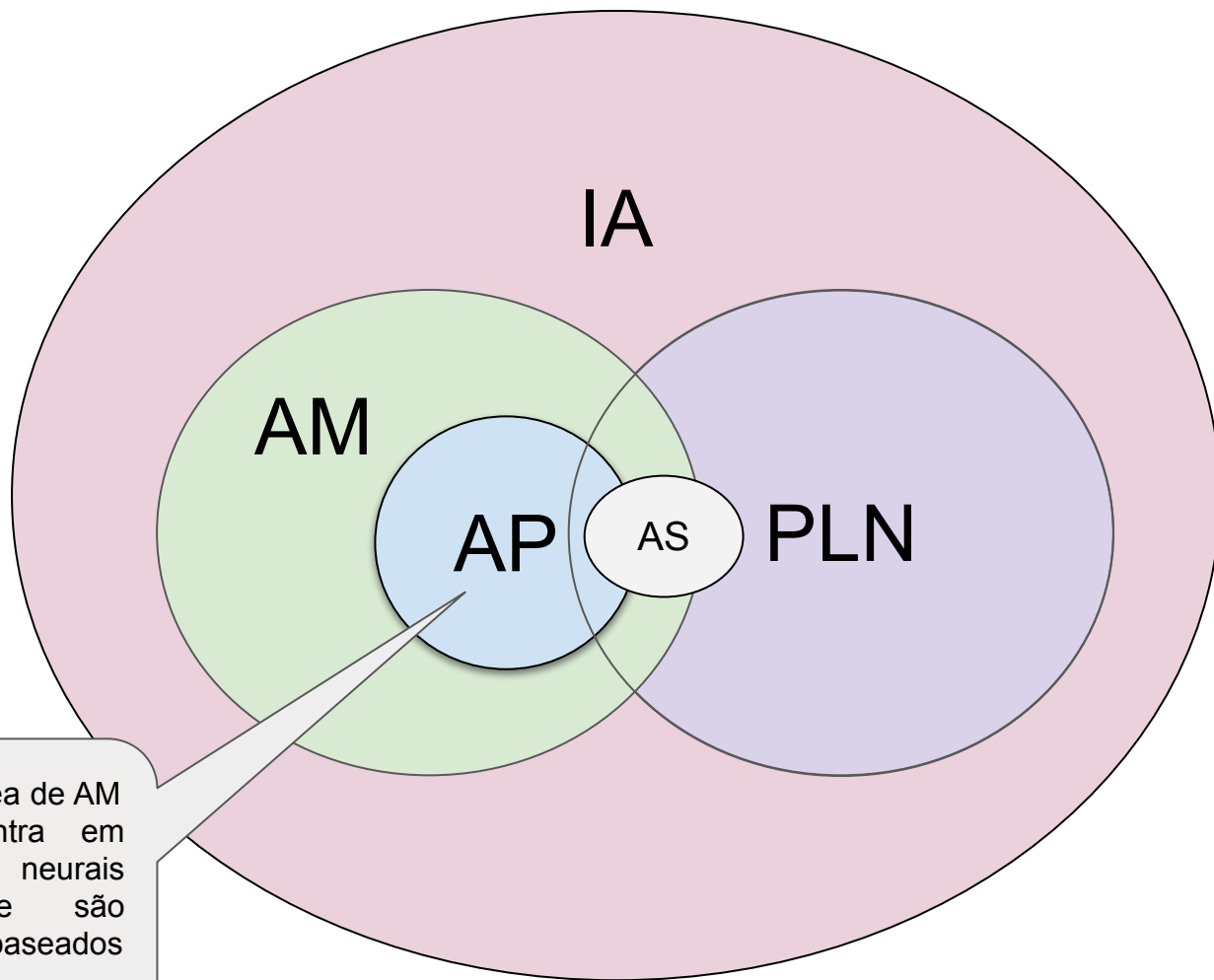
 Camada de Entrada

## Redes Neurais Profundas



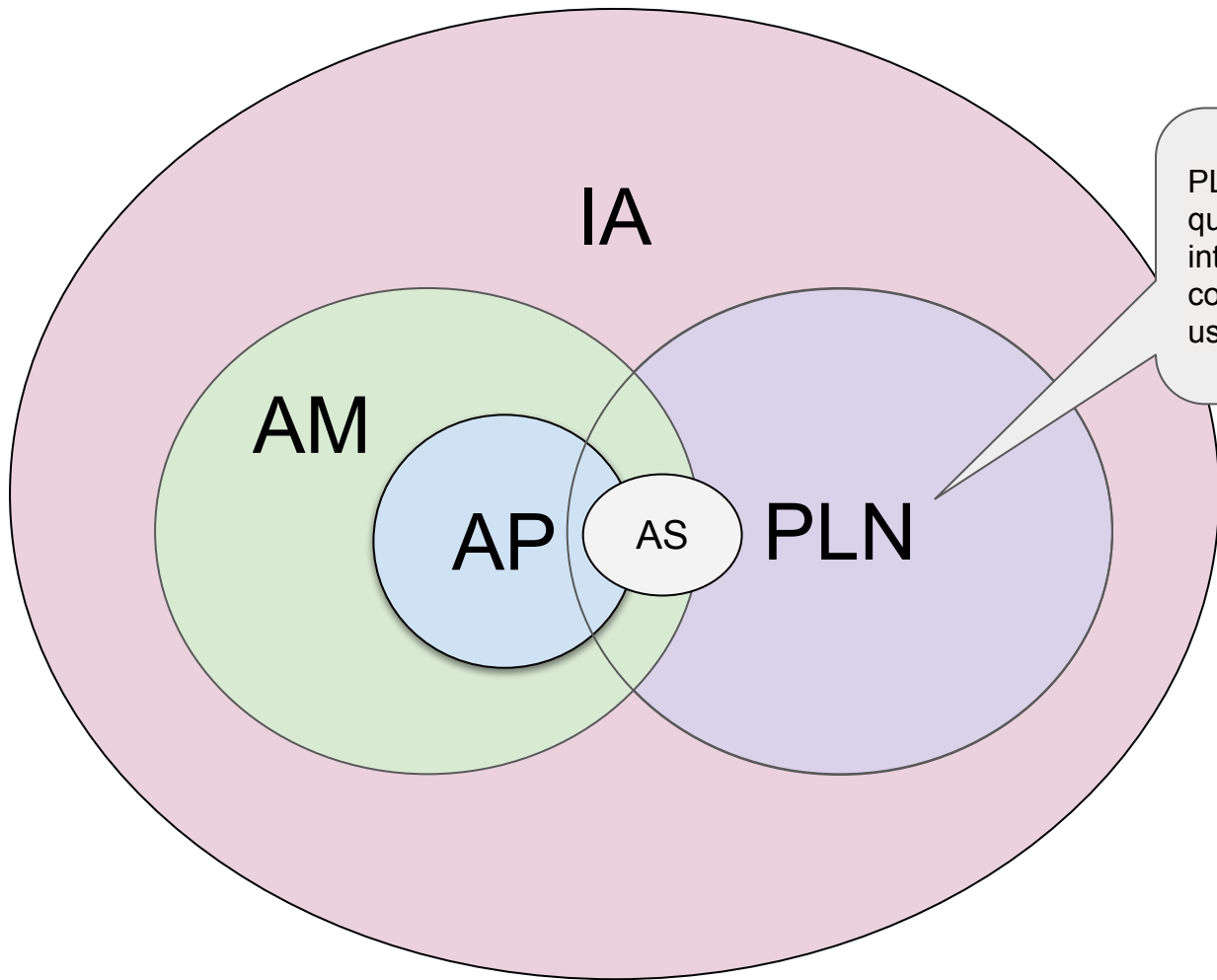
 Camada Oculta

 Camada de Saída

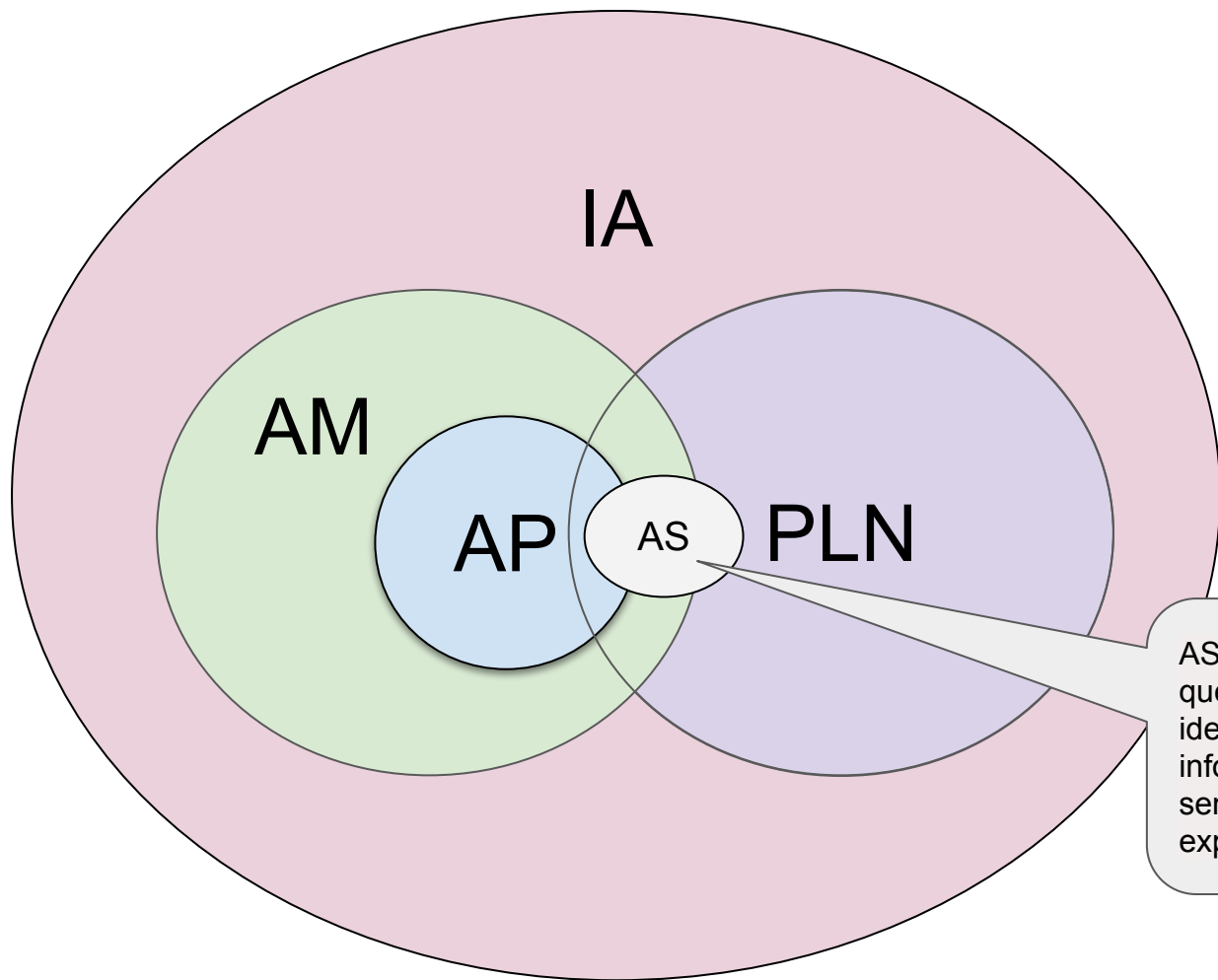


AP é uma subárea de AM que se concentra em treinar redes neurais profundas, que são modelos de IA baseados em redes neurais.



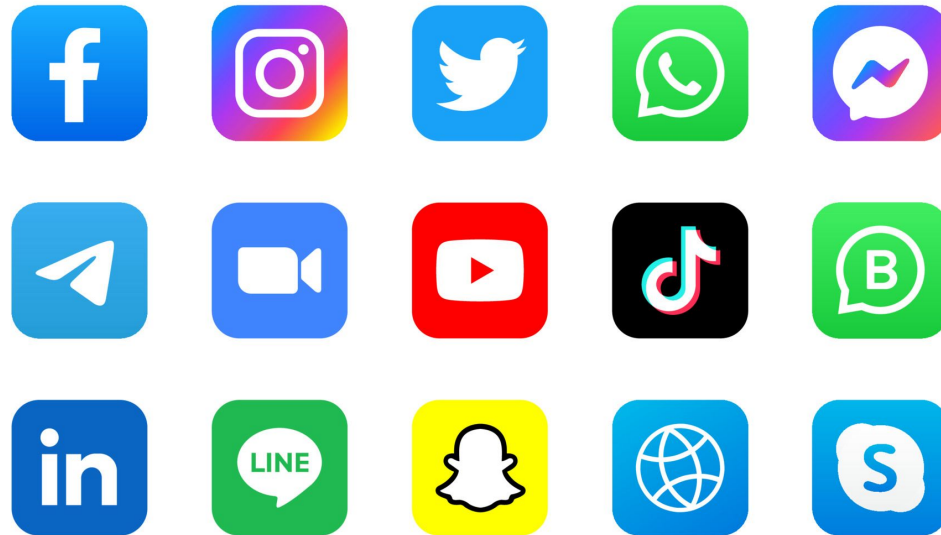


PLN é uma subárea da IA que se concentra na interação entre computadores e humanos usando linguagem natural.



AS é uma tarefa de PLN que tem como objetivo identificar e extrair informações sobre o sentimento ou emoção expresso em um texto.

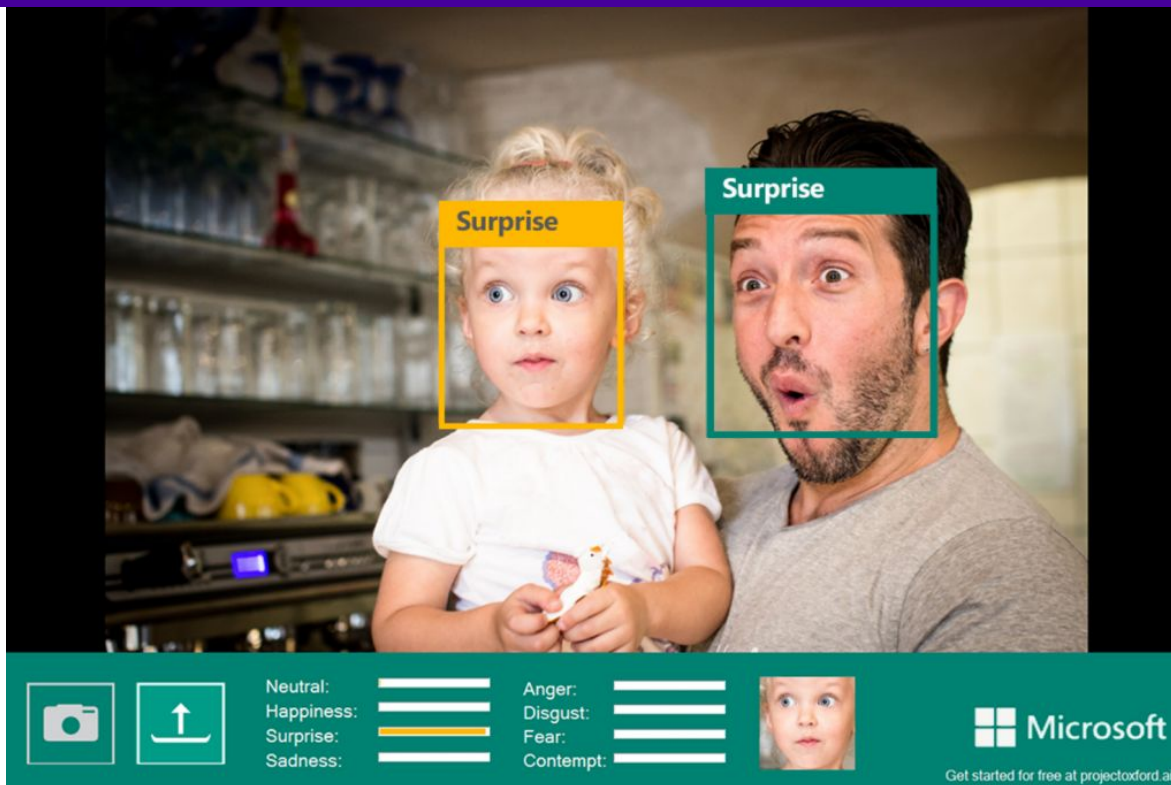
# Redes Sociais



# Aplicações

- Análise de Sentimento
- Detecção de Discurso de Ódio e Linguagem Ofensiva
- Detecção de Notícias Falsas
- Detecção de Ironia/Sarcasmo/Humor

# Análise de Sentimento

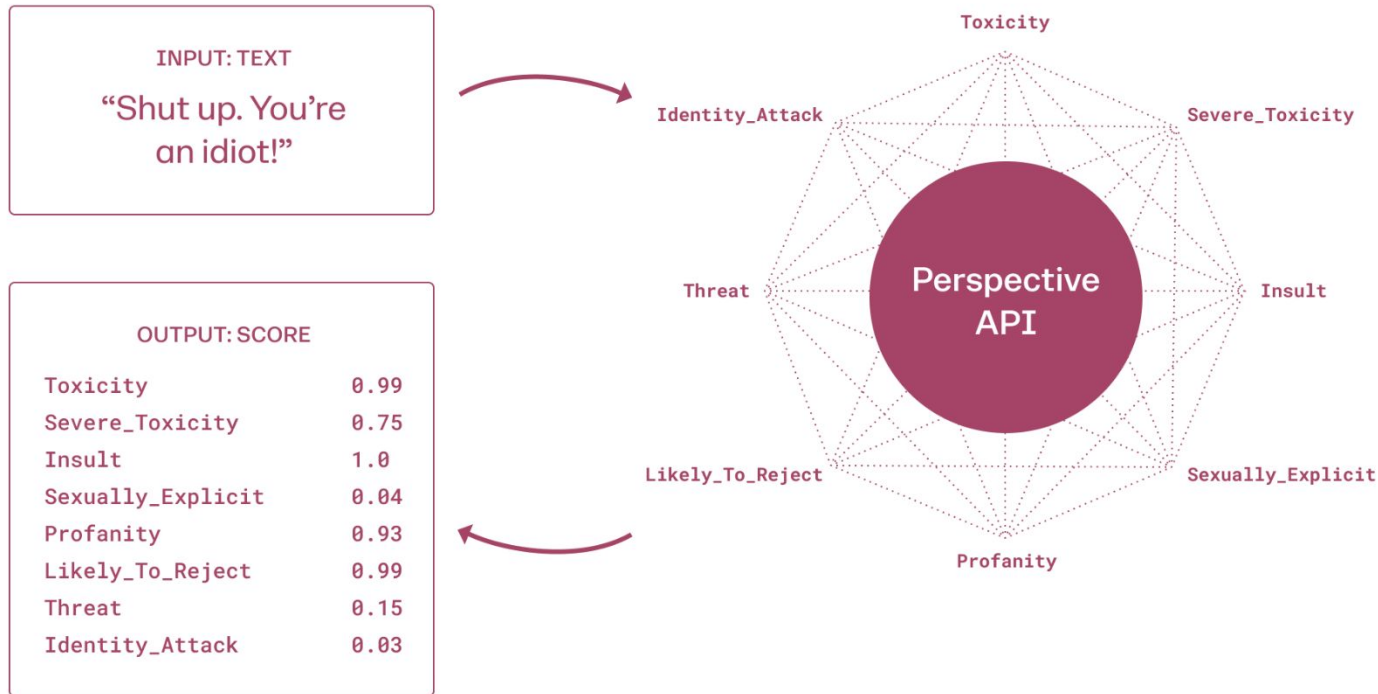


The screenshot displays the Microsoft Project Oxford AI interface. The main image shows a man and a young child in a kitchen, both with surprised expressions. The child's face is highlighted with a yellow box and labeled 'Surprise', while the man's face is highlighted with a teal box and also labeled 'Surprise'. Below the image is a teal control bar containing a camera icon, an upload icon, and a sentiment analysis bar. The sentiment bar shows 'Surprise' as the dominant emotion with a yellow bar. To the right of the bar is a small thumbnail of the child's face. The Microsoft logo and the text 'Get started for free at [projectoxford.ai](https://projectoxford.ai)' are located in the bottom right corner of the interface.

Neutral:	=====	Anger:	=====
Happiness:	=====	Disgust:	=====
Surprise:	=====	Fear:	=====
Sadness:	=====	Contempt:	=====

Microsoft  
Get started for free at [projectoxford.ai](https://projectoxford.ai)

# Detecção de Discurso de Ódio e Linguagem Ofensiva



Fonte: <https://support.perspectiveapi.com/>

# Detecção de Notícias Falsas



**É #FAKE vídeo em que carrinho de compras 'fantasma' anda sozinho na madrugada em supermercado**

Imagens são uma encenação e fazem parte de uma campanha publicitária.



**NAS REDES**

É falso que vídeo mostra discurso pró-Israel feito por líder palestino na ONU

**POLÍTICA**

**ARGENTINA REPETE BRASIL E TEM ELEIÇÃO COM FAKES SOBRE FRAUDES NA VOTAÇÃO**



# Detecção de Ironia/Sarcasmo/Humor



Fonte: <https://pt.quora.com/Qual-é-a-diferença-entre-ironia-sarcasmo-e-deboche>



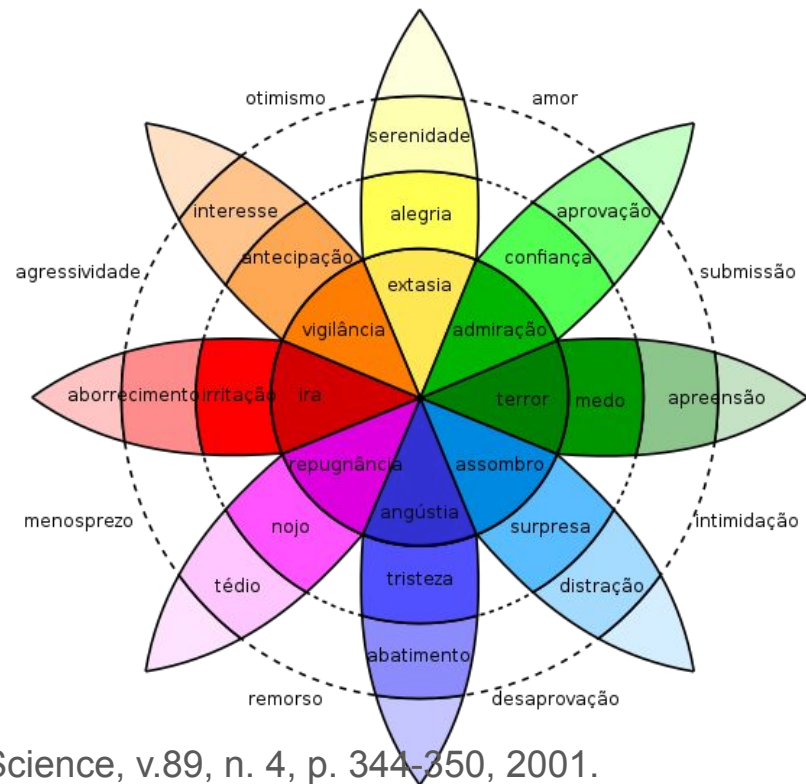
# Definição e Fundamentos da Análise de Sentimento

# Definição: Análise de Sentimento

“É o campo de estudo que analisa as opiniões, sentimentos, avaliações/críticas, atitudes e emoções das pessoas em relação a entidades, como produtos, serviços, organizações, indivíduos, questões, eventos, tópicos e seus aspectos”  
(Liu, 2015)

LIU, B.. **Sentiment analysis: Mining opinions, sentiments, and emotions**. Cambridge university press, 2015.

# Classificação de Emoções



PLUTCHIK, R. **The Nature of Emotions**. American Science, v.89, n. 4, p. 344-350, 2001.

# Mineração de Opinião



Negativa



Neutra



Positiva

# Fundamentos

## **Análise de Sentimento, ou ...**

- Mineração de Opinião,
- Classificação de Sentimento,
- Extração de Opinião,
- Mineração de Sentimento,
- Análise de Subjetividade,
- Análise de Afeição,
- Análise de Emoção,
- Mineração de Críticas.

# Fundamentos

## Opinião x Sentimento (Emoção)

### Sentimento

Eu estou preocupado com o estado atual da economia.

### Opinião

Eu acho que a economia está indo bem.

# Fundamentos

## Opinião x Sentimento (Emoção)

Os conceitos de **Opinião** e **Sentimento** frequentemente se confundem. Na literatura de **Análise de Sentimento** eles são frequentemente empregados de forma **intercambiável**.

Extrair opinião = Extrair sentimento

# Fundamentos

## Tipos de Opiniões

### Regulares:

- **Direta:** A Coca-Cola tem um gosto bom.
- **Indireta:** Depois de tomar a injeção, a dor passou.

**Comparativa:** É a opinião que apresenta uma relação comparativa das semelhanças ou diferenças entre duas ou mais entidades.

- A Coca-Cola é melhor do que a Pepsi.



## Definição Formal

Liu (2015) define a opinião como uma quádrupla  $(g, s, h, t)$ , onde:

- **g** – é o alvo do sentimento, ou alvo da opinião, aquele a quem o sentimento expresso na opinião se refere.
- **s** – é o sentimento da opinião expressa sobre **g**;
- **h** – é o detentor da opinião, aquele que proferiu a opinião com sentimento **s** sobre a entidade alvo **g** no tempo **t**;
- **t** – é o tempo em que a opinião foi expressa.

# Fundamentos



escreveu uma avaliação mar de 2019  
re, RS • 186 contribuições • 51 votos úteis

Opinião



Seguir



## O melhor da cidade

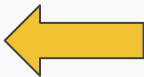
"Hotel no geral muito bom, instalações bem conservadas. O quarto é espaçoso, o home service vai até a meia-noite e os preços são dentro do "normal" de hotéis. A equipe é muito solícita. O ponto fraco é o café da manhã, muito simples para a suntuosidade e a diária bem acima da média do interior."

Recolher conteúdo

Data da estadia: março de 2019



Costo-beneficio  
Atendimento



Localização



Esta avaliação representa a opinião subjetiva de um membro do programa TripAdvisor e não da TripAdvisor LLC.

# Níveis de Análise de Sentimento

# O Problema



escreveu uma avaliação mar de 2019  
gre, RS • 186 contribuições • 51 votos úteis

Seguir



## O melhor da cidade

"Hotel no geral muito bom, instalações bem conservadas. O quarto é espaçoso, o home service vai até a meia-noite e os preços são dentro do "normal" de hotéis. A equipe é muito solícita. O ponto fraco é o café da manhã, muito simples para a suntuosidade e a diária bem acima da média do interior."

[Recolher conteúdo](#) ▲

**Data da estadia:** março de 2019

Custo-benefício  
 Atendimento

Localização

*Esta avaliação representa a opinião subjetiva de um membro do programa TripAdvisor e não da TripAdvisor LLC.*

# Nível de Documento



Avaliou 16 de Outubro de 2016

## Relação custo x benefício muito baixa

Estive em Pelotas para uma festa de formatura.

Fui com minha esposa e a escolha foi pelo melhor hotel, para garantir conforto.

As fotos, nos sites, apontavam o Jacques Geroges Tower como o melhor - disparado - em relação aos demais. O valor da diária, cerca de 50% acima dos outros de nível semelhante autorizavam acreditar que havia sensível diferença de qualidade.

O lobby do hotel é bastante bonito e os funcionários são extremamente atenciosos. Todos os que nos atenderam, seja na recepção, mensageiros e garagem, fizeram de tudo para nos proporcionar a melhor experiência possível.

Entretanto, no quarto, a qualidade ficou a desejar.

O quarto que nos colocaram tinha um cheiro bastante forte e desagradável de cigarro (havia, inclusive, um cinzeiro sobre a mesa). Solicitamos a troca, e foi prontamente feita.

No segundo quarto, não havia cheiro, mas, tal como no primeiro, a televisão ainda de tubo e um cobertor com aspecto de velho. O banho é bom, mas a banheira de hidromassagem é extremamente pequena.

Não se pode dizer que o hotel é ruim, de maneira alguma, mas em relação ao preço cobrado, não vale muito a pena.

Tive familiares que neste mesmo final de semana ficaram no hotel Manta, pagando praticamente a metade do que pagamos, e estavam melhor instalados.

[Mostrar menos](#)

Documento

# Nível de Parágrafo



Avaliou 16 de Outubro de 2016

## Relação custo x benefício muito baixa

Estive em Pelotas para uma festa de formatura.

Fui com minha esposa e a escolha foi pelo melhor hotel, para garantir conforto.

As fotos, nos sites, apontavam o Jacques Geroges Tower como o melhor - disparado - em relação aos demais. O valor da diária, cerca de 50% acima dos outros de nível semelhante autorizavam acreditar que havia sensível diferença de qualidade.

O lobby do hotel é bastante bonito e os funcionários são extremamente atenciosos. Todos os que nos atenderam, seja na recepção, mensageiros e garagem, fizeram de tudo para nos proporcionar a melhor experiência possível.

Entretanto, no quarto, a qualidade ficou a desejar.

O quarto que nos colocaram tinha um cheiro bastante forte e desagradável de cigarro (havia, inclusive, um cinzeiro sobre a mesa). Solicitamos a troca, e foi prontamente feita.

No segundo quarto, não havia cheiro, mas, tal como no primeiro, a televisão ainda de tubo e um cobertor com aspecto de velho. O banho é bom, mas a banheira de hidromassagem é extremamente pequena.

Não se pode dizer que o hotel é ruim, de maneira alguma, mas em relação ao preço cobrado, não vale muito a pena.

Tive familiares que neste mesmo final de semana ficaram no hotel Manta, pagando praticamente a metade do que pagamos, e estavam melhor instalados.

[Mostrar menos](#)

Parágrafo

# Nível de Sentença



Avaliou 16 de Outubro de 2016

## Relação custo x benefício muito baixa

Estive em Pelotas para uma festa de formatura.

Fui com minha esposa e a escolha foi pelo melhor hotel, para garantir conforto.

As fotos, nos sites, apontavam o Jacques Geroges Tower como o melhor - disparado - em relação aos demais. O valor da diária, cerca de 50% acima dos outros de nível semelhante autorizavam acreditar que havia sensível diferença de qualidade.

O lobby do hotel é bastante bonito e os funcionários são extremamente atenciosos. Todos os que nos atenderam, seja na recepção, mensageiros e garagem, fizeram de tudo para nos proporcionar a melhor experiência possível.

Entretanto, no quarto, a qualidade ficou a desejar.

O quarto que nos colocaram tinha um cheiro bastante forte e desagradável de cigarro (havia, inclusive, um cinzeiro sobre a mesa). Solicitamos a troca, e foi prontamente feita. No segundo quarto, não havia cheiro, mas, tal como no primeiro, a televisão ainda de tubo e um cobertor com aspecto de velho. O banho é bom, mas a banheira de hidromassagem é extremamente pequena.

Não se pode dizer que o hotel é ruim, de maneira alguma, mas em relação ao preço cobrado, não vale muito a pena.

Tive familiares que neste mesmo final de semana ficaram no hotel Manta, pagando praticamente a metade do que pagamos, e estavam melhor instalados.

[Mostrar menos](#)

Sentença

# Nível de Bloco



Avaliou 16 de Outubro de 2016

## Relação custo x benefício muito baixa

Estive em Pelotas para uma festa de formatura.

Fui com minha esposa e a escolha foi pelo melhor hotel, para garantir conforto.

As fotos, nos sites, apontavam o Jacques Geroges Tower como o melhor - disparado - em relação aos demais. O valor da diária, cerca de 50% acima dos outros de nível semelhante autorizavam acreditar que havia sensível diferença de qualidade.

O lobby do **hotel é bastante bonito e os funcionários** são extremamente atenciosos. Todos os que nos atenderam, seja na recepção, mensageiros e garçons, fizeram de tudo para nos proporcionar a melhor experiência possível.

Entretanto, no quarto, a qualidade ficou a desejar.

O quarto que nos colocaram tinha um cheiro bastante forte e desagradável de cigarro (havia, inclusive, um cinzeiro sobre a mesa). Solicitamos a troca, e foi prontamente feita. No segundo quarto, não havia cheiro, mas, tal como no primeiro, a televisão ainda de tubo e um cobertor com aspecto de velho. O banho é bom, mas a banheira de hidromassagem é extremamente pequena.

Não se pode dizer que o hotel é ruim, de maneira alguma, mas em relação ao preço cobrado, não vale muito a pena.

Tive familiares que neste mesmo final de semana ficaram no hotel Manta, pagando praticamente a metade do que pagamos, e estavam melhor instalados.

[Mostrar menos](#)

Bloco



# Nível (Baseado em) Aspectos

●●●●○ Avaliou 16 de Outubro de 2016

## Relação custo x benefício muito baixa

Estive em Pelotas para uma festa de formatura.

Fui com minha esposa e a escolha foi pelo **melhor hotel**, para garantir conforto.

As fotos, nos sites, apontavam o Jacques Geroges Tower como o melhor - disparado - em relação aos demais. O valor da diária, cerca de 50% acima dos outros de nível semelhante autorizavam acreditar que havia sensível diferença de qualidade.

O lobby do **hotel** é bastante **bonito** e os funcionários são extremamente atenciosos. Todos os que nos atenderam, seja na recepção, mensageiros e garagem, fizeram de tudo para nos proporcionar a melhor experiência possível.

Entretanto, no quarto, a qualidade ficou a desejar.

O **quarto** que nos colocaram tinha um cheiro bastante forte e **desagradável** de cigarro (havia, inclusive, um cinzeiro sobre a mesa). Solicitamos a troca, e foi prontamente feita.

No segundo quarto, não havia cheiro, mas, tal como no primeiro, a televisão ainda de tubo e um cobertor com aspecto de velho. O **banho** é **bom**, mas a banheira de hidromassagem é extremamente pequena.

Não se pode dizer que o **hotel** é **ruim**, de maneira alguma, mas em relação ao preço cobrado, não vale muito a pena.

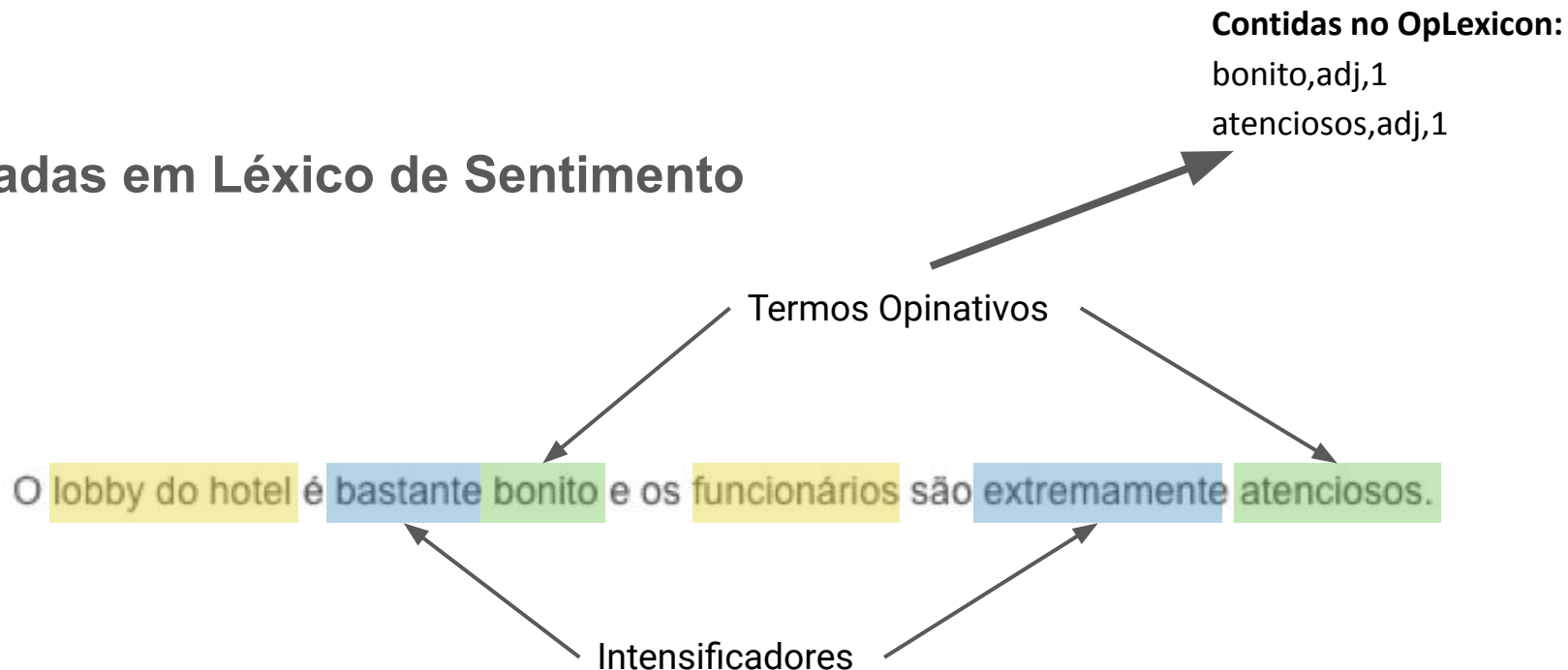
Tive familiares que neste mesmo final de semana ficaram no hotel Manta, pagando praticamente a metade do que pagamos, e estavam melhor instalados.

[Mostrar menos](#)

Aspectos

# Abordagens Clássicas

## Baseadas em Léxico de Sentimento



# Abordagens Clássicas

## Baseadas em Aprendizado de Máquina Clássico

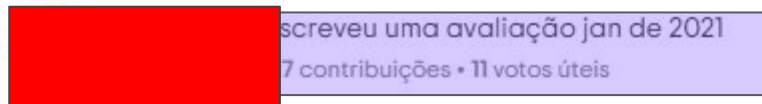
- **Resultados Dependentes**
  - Pré-processamentos
  - Engenharia de Características
- **Exemplos**
  - Máquinas de Vetores Suporte
  - Árvore de Decisão
  - Naïve Bayes

# Abordagens Clássicas

## Baseadas em Aprendizado de Máquina Clássico

- **Pré-processamentos**
  - Remoção de *stopwords*\*
  - Uniformização da língua (“Internetês”)
  - Remoção de Diacríticos

# Humanos ≠ Máquinas




**Boa escolha!**

“Além da localização ser excelente para um turista clássico, o hotel em si é muito bom. Antes de ir, entrei em contato pois estava com dúvida em relação a algumas taxas e foram muito solícitos e educados. O quarto não é enorme, mas é mais do que o suficiente para descansar depois de um dia agitado nessa cidade incrível. Recomendaria, sem dúvida!”

[Leia mais](#) ▼

**Data da estadia:** fevereiro de 2020

# Humanos ≠ Máquinas

 escreveu uma avaliação jan de 2021  
7 contribuições • 11 votos úteis



Boa escolha!

“Além da localização  
em contato pois estava  
quarto não é enorme,  
cidade incrível. Recon

[Leia mais](#) ▼

**Codificação**  
**ASCII, Unicode, UTF-8, etc.**

o bom. Antes de ir, entrei  
solícitos e educados. O  
um dia agitado nessa

Data da estadia: fevereiro de 2020

# Humanos ≠ Máquinas

Dec	Hx	Oct	Char	Dec	Hx	Oct	Html	Chr	Dec	Hx	Oct	Html	Chr	Dec	Hx	Oct	Html	Chr														
0	0	000	<b>NUL</b> (null)	32	20	040	&#32;	Space	64	40	100	&#64;	@	96	60	140	&#96;	`														
1	1	001	<b>SOH</b> (start of heading)	33	21	041	&#33;	!	65	41	101	&#65;	A	97	61	141	&#97;	a														
2	2	002	<b>STX</b> (start of text)	34	22	042	&#34;	"	66	42	102	&#66;	B	98	62	142	&#98;	b														
3	3	003	<b>ETX</b> (end of text)	35	23	043	&#35;	#	67	43	103	&#67;	C	99	63	143	&#99;	c														
4	4	004	<b>EOT</b> (end of transmission)	36	24	044	&#36;	\$	68	44	104	&#68;	D	100	64	128	Ç	144	É	160	á	176	☼	192	L	208	⋈	224	α	240	≡	
5	5	005	<b>ENQ</b> (enquiry)	37	25	045	&#37;	%	69	45	105	&#69;	E	101	65	1																
6	6	006	<b>ACK</b> (acknowledge)	38	26	046	&#38;	&	70	46	106	&#70;	F	102	66	1	129	ü	145	æ	161	í	177	☼	193	⋈	209	⋈	225	β	241	≠
7	7	007	<b>BEL</b> (bell)	39	27	047	&#39;	'	71	47	107	&#71;	G	103	67	1																
8	8	010	<b>BS</b> (backspace)	40	28	050	&#40;	(	72	48	110	&#72;	H	104	68	1	130	é	146	Æ	162	Á	178	☼	194	⋈	210	⋈	226	Γ	242	≥
9	9	011	<b>TAB</b> (horizontal tab)	41	29	051	&#41;	)	73	49	111	&#73;	I	105	69	1																
10	A	012	<b>LF</b> (NL line feed, new	42	30	052	&#42;	*	74	50	112	&#74;	J	106	70	1																
11	B	013	<b>VT</b> (vertical tab)	43	31	053	&#43;	+	75	51	113	&#75;	K	107	71	1																
12	C	014	<b>FF</b> (NP form feed, new	44	32	054	&#44;	,	76	52	114	&#76;	L	108	72	1																
13	D	015	<b>CR</b> (carriage return)	45	33	055	&#45;	-	77	53	115	&#77;	M	109	73	1																
14	E	016	<b>SO</b> (shift out)	46	34	056	&#46;	.	78	54	116	&#78;	N	110	74	1																
15	F	017	<b>SI</b> (shift in)	47	35	057	&#47;	/	79	55	117	&#79;	O	111	75	1																
16	10	020	<b>DLE</b> (data link escape)	48	36	060	&#48;	0	80	40	100	&#64;	@	96	60	140	&#96;	`														
17	11	021	<b>DC1</b> (device control 1)	49	37	061	&#49;	1	81	41	101	&#65;	A	97	61	141	&#97;	a														
18	12	022	<b>DC2</b> (device control 2)	50	38	062	&#50;	2	82	42	102	&#66;	B	98	62	142	&#98;	b														
19	13	023	<b>DC3</b> (device control 3)	51	39	063	&#51;	3	83	43	103	&#67;	C	99	63	143	&#99;	c														
20	14	024	<b>DC4</b> (device control 4)	52	40	064	&#52;	4	84	44	104	&#68;	D	100	64	128	Ç	144	É	160	á	176	☼	192	L	208	⋈	224	α	240	≡	
21	15	025	<b>NAK</b> (negative acknowledged	53	41	065	&#53;	5	85	45	105	&#69;	E	101	65	1																
22	16	026	<b>SYN</b> (synchronous idle)	54	42	066	&#54;	6	86	46	106	&#70;	F	102	66	1	129	ü	145	æ	161	í	177	☼	193	⋈	209	⋈	225	β	241	≠
23	17	027	<b>ETB</b> (end of trans. bloc)	55	43	067	&#55;	7	87	47	107	&#71;	G	103	67	1																
24	18	030	<b>CAN</b> (cancel)	56	44	068	&#56;	8	88	48	108	&#72;	H	104	68	1	130	é	146	Æ	162	Á	178	☼	194	⋈	210	⋈	226	Γ	242	≥
25	19	031	<b>EM</b> (end of medium)	57	45	069	&#57;	9	89	49	109	&#73;	I	105	69	1																
26	1A	032	<b>SUB</b> (substitute)	58	46	070	&#58;	:	90	4A	10A	&#x10A;	Ā	106	70	1																
27	1B	033	<b>ESC</b> (escape)	59	47	071	&#59;	;	91	4B	10B	&#x10B;	ā	107	71	1																
28	1C	034	<b>FS</b> (file separator)	60	48	072	&#60;	<	92	4C	10C	&#x10C;	Ā	108	72	1																
29	1D	035	<b>GS</b> (group separator)	61	49	073	&#61;	=	93	4D	10D	&#x10D;	ā	109	73	1																
30	1E	036	<b>RS</b> (record separator)	62	50	074	&#62;	>	94	4E	10E	&#x10E;	Ā	110	74	1																
31	1F	037	<b>US</b> (unit separator)	63	51	075	&#63;	?	95	4F	10F	&#x10F;	ā	111	75	1																

Unicode code point	character	UTF-8 (hex.)	numerical HTML encoding of the Unicode character	name
U+0100	Ā	c4 80	&#x100;	LATIN CAPITAL LETTER A WITH MACRON
U+0101	ā	c4 81	&#x101;	LATIN SMALL LETTER A WITH MACRON
U+0102	Ā	c4 82	&#x102;	LATIN CAPITAL LETTER A WITH BREVE
U+0103	ā	c4 83	&#x103;	LATIN SMALL LETTER A WITH BREVE
U+0104	Ą	c4 84	&#x104;	LATIN CAPITAL LETTER A WITH OGONEK
U+0105	ą	c4 85	&#x105;	LATIN SMALL LETTER A WITH OGONEK
U+0106	Ā	c4 86	&#x106;	LATIN CAPITAL LETTER C WITH ACUTE
U+0107	ć	c4 87	&#x107;	LATIN SMALL LETTER C WITH ACUTE
U+0108	Ĉ	c4 88	&#x108;	LATIN CAPITAL LETTER C WITH CIRCUMFLEX
U+0109	ĉ	c4 89	&#x109;	LATIN SMALL LETTER C WITH CIRCUMFLEX
U+010A	Ĉ	c4 8a	&#x10A;	LATIN CAPITAL LETTER C WITH DOT ABOVE
U+010B	ċ	c4 8b	&#x10B;	LATIN SMALL LETTER C WITH DOT ABOVE
U+010C	Ć	c4 8c	&#x10C;	LATIN CAPITAL LETTER C WITH CARON
U+010D	ċ	c4 8d	&#x10D;	LATIN SMALL LETTER C WITH CARON
U+010E	Ď	c4 8e	&#x10E;	LATIN CAPITAL LETTER D WITH CARON
U+010F	ď	c4 8f	&#x10F;	LATIN SMALL LETTER D WITH CARON
U+0110	Ď	c4 90	&#x110;	LATIN CAPITAL LETTER D WITH STROKE
U+0111	đ	c4 91	&#x111;	LATIN SMALL LETTER D WITH STROKE
U+0112	Ē	c4 92	&#x112;	LATIN CAPITAL LETTER E WITH MACRON
U+0113	ē	c4 93	&#x113;	LATIN SMALL LETTER E WITH MACRON

source: [www.LookupTables.com](http://www.LookupTables.com)

# Representação de Dados Textuais

- **Saco de Palavras (*Bag-of-Words* - *BoW*)**
- **TF-IDF**
  - **Frequência de Termos (TF)**
  - **Inverso da Frequência nos Documentos (IDF)**
- **Engenharia de Características**
  - ***Part-of-Speech* - exemplo: Nome Seguido de Adjetivo**
- ***Word embeddings***



# Representação de Dados Textuais

## BoW

Dicionário	entrada 1		entrada 2	
o	Que	0	O	1
dia	dia	1	dia	1
que	lindo.	1	está	0
está		0	bom.	1
bom		0		1
ruim		0		0
lindo		1		0

# Representação de Dados Textuais

## BoW

Dicionário	entrada 1		entrada 2	
um	Um	1	Um	1
bom	homem	0	grande	0
homem	grande.	1	homem.	1
mal		0		0
feio		0		0
grande		1		1
perverso		0		0

# Representação de Dados Textuais

## TF-IDF

Documentos	Texto	# Palavras no Documento
A	Que dia lindo.	3
B	O dia está bom.	4

# Representação de Dados Textuais

## TF

Palavras	TF (A)	TF (B)
o	0	1/4
dia	1/3	1/4
que	1/3	0
está	0	1/4
bom	0	1/4
ruim	0	0
lindo	1/3	0

# Representação de Dados Textuais

## IDF

$$IDF(w, D) = \ln\left(\frac{\text{Total number of documents (N) in corpus } D}{\text{number of documents containing } w}\right)$$

Palavras	TF (A)	TF (B)	IDF
o	0	1/4	$\ln(2/1) = 0.69$
dia	1/3	1/4	$\ln(2/2) = 0$
que	1/3	0	$\ln(2/1) = 0.69$
está	0	1/4	$\ln(2/1) = 0.69$
bom	0	1/4	$\ln(2/1) = 0.69$
ruim	0	0	0
lindo	1/3	0	$\ln(2/1) = 0.69$

# Representação de Dados Textuais

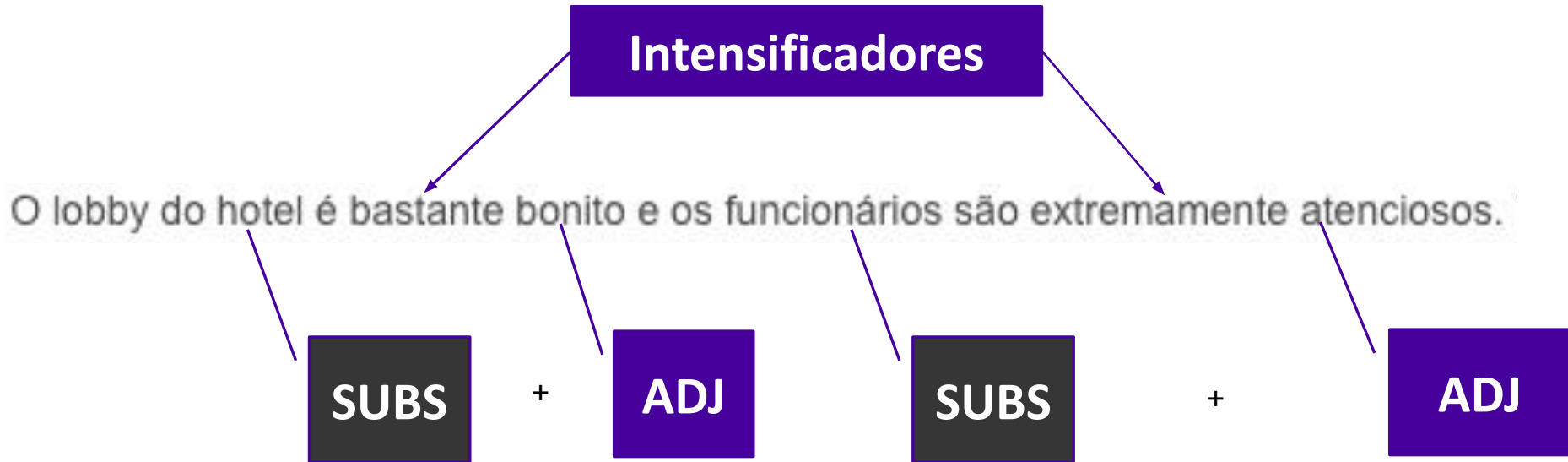
## TF-IDF

$$TFIDF(w, d, D) = TF(w, d) * IDF(w, D)$$

Palavras	TF (A)	TF (B)	IDF	TFIDF (A)	TFIDF (A)
o	0	1/4	$\ln(2/1) = 0.69$	$0 * 0.69$	$\frac{1}{4} * 0.69 = 0.1725$
dia	1/3	1/4	$\ln(2/2) = 0$	$\frac{1}{3} * 0.69 = 0.23$	$\frac{1}{4} * 0.69 = 0.1725$
que	1/3	0	$\ln(2/1) = 0.69$	$\frac{1}{3} * 0.69 = 0.23$	$0 * 0.69 = 0$
está	0	1/4	$\ln(2/1) = 0.69$	$0 * 0.69 = 0$	$\frac{1}{4} * 0.69 = 0.1725$
bom	0	1/4	$\ln(2/1) = 0.69$	$0 * 0.69 = 0$	$\frac{1}{4} * 0.69 = 0.1725$
ruim	0	0	0	$0 * 0 = 0$	$0 * 0 = 0$
lindo	1/3	0	$\ln(2/1) = 0.69$	$\frac{1}{3} * 0.69 = 0.23$	$0 * 0.69 = 0$

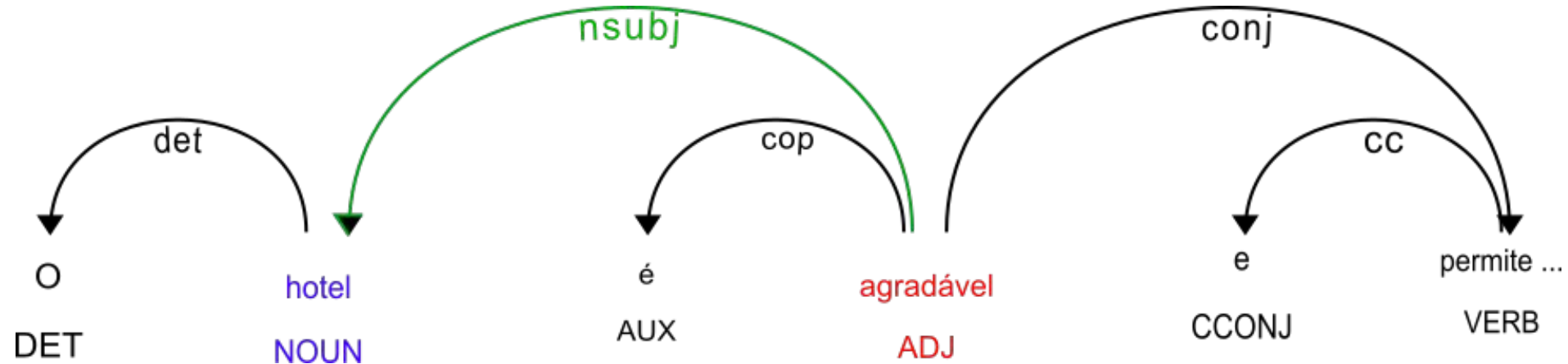
# Representação de Dados Textuais

## Engenharia de Características



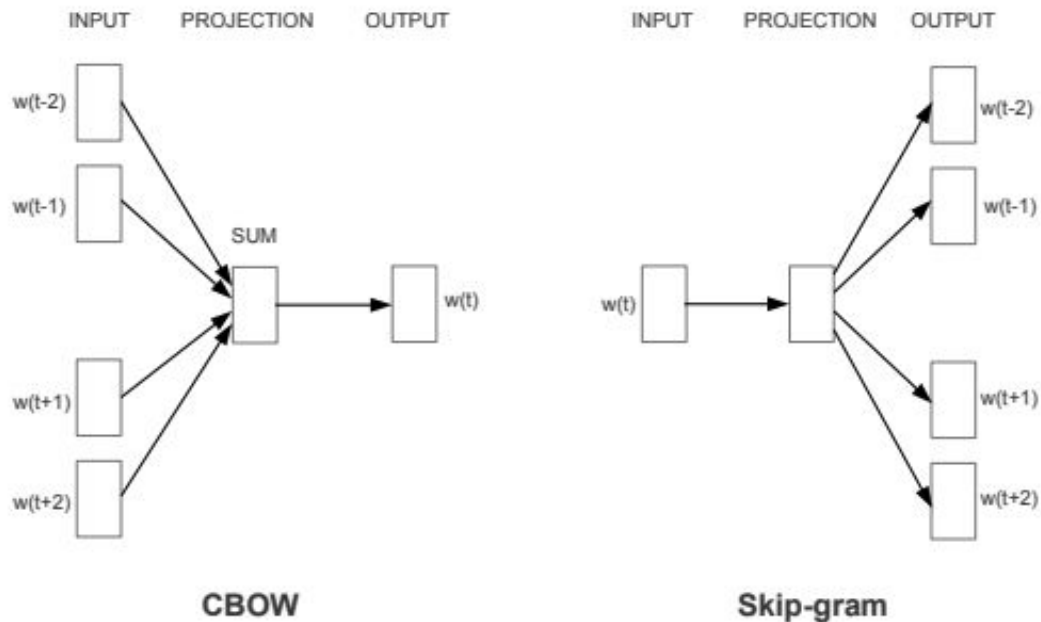
# Representação de Dados Textuais

## Engenharia de Características



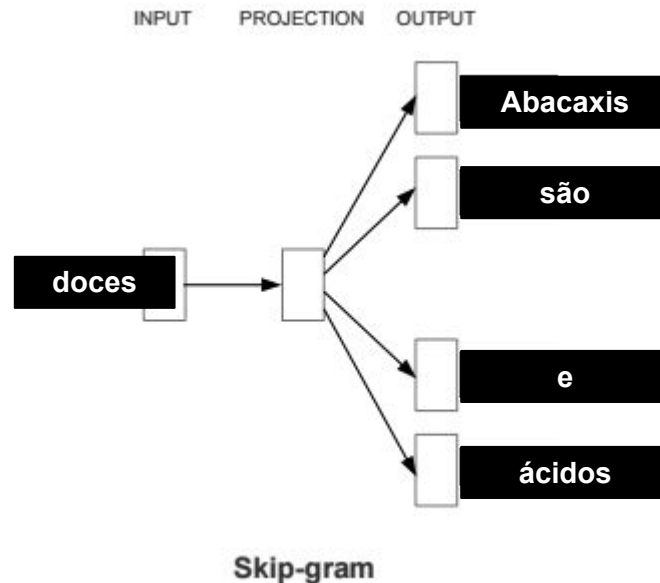
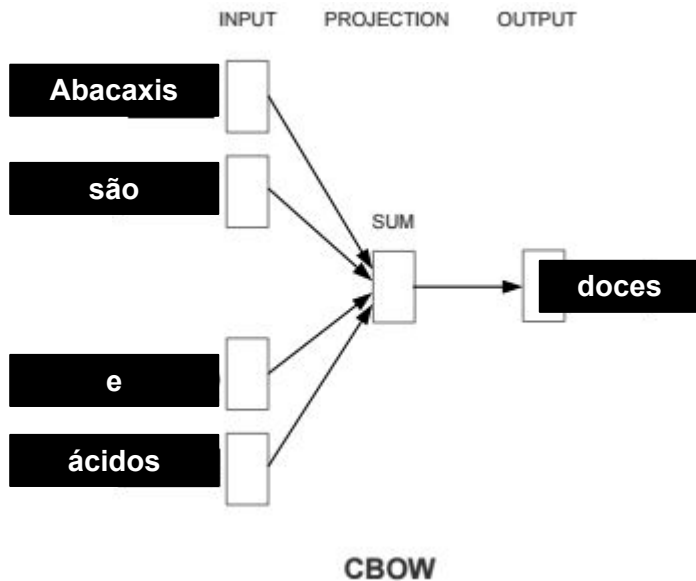


# Word-embeddings: Word2Vec



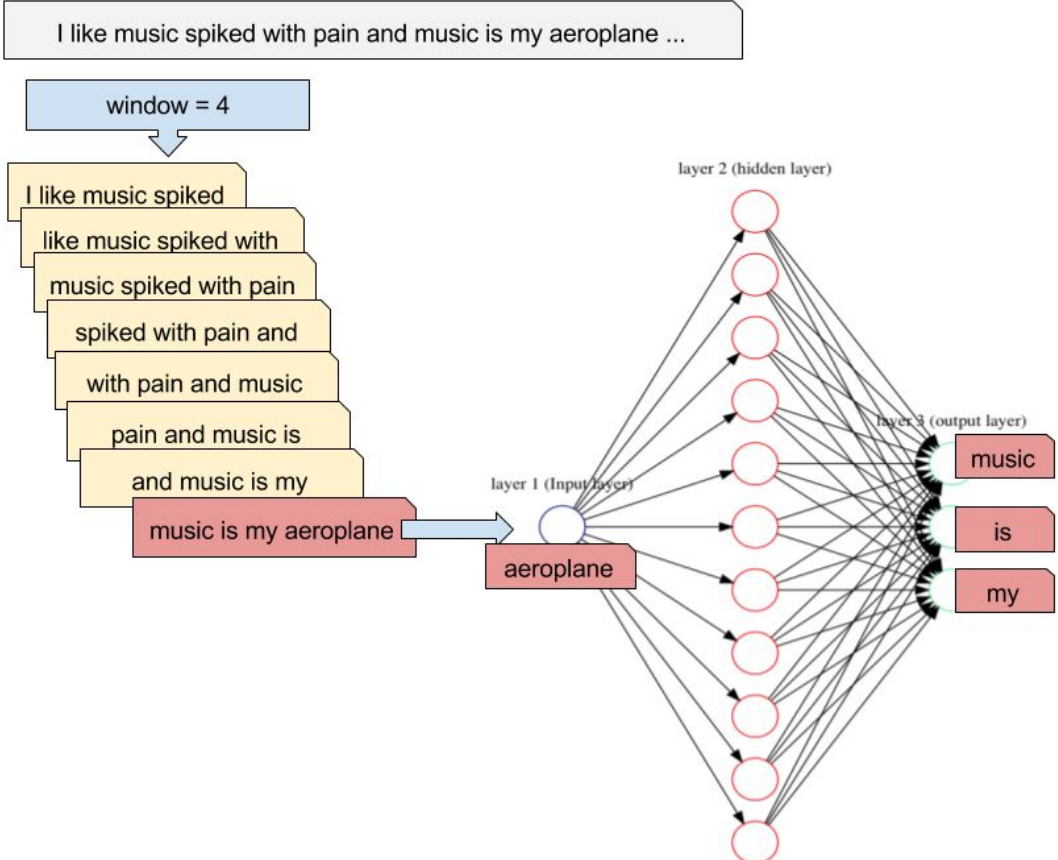
MIKOLOV, T.; CHEN, K.; CORRADO, G.; DEAN, J.. **Efficient Estimation of Word Representations in Vector Space**. ICLR, 2013.

# Word-embeddings: Word2Vec



MIKOLOV, T.; CHEN, K.; CORRADO, G.; DEAN, J.. **Efficient Estimation of Word Representations in Vector Space**. ICLR, 2013.

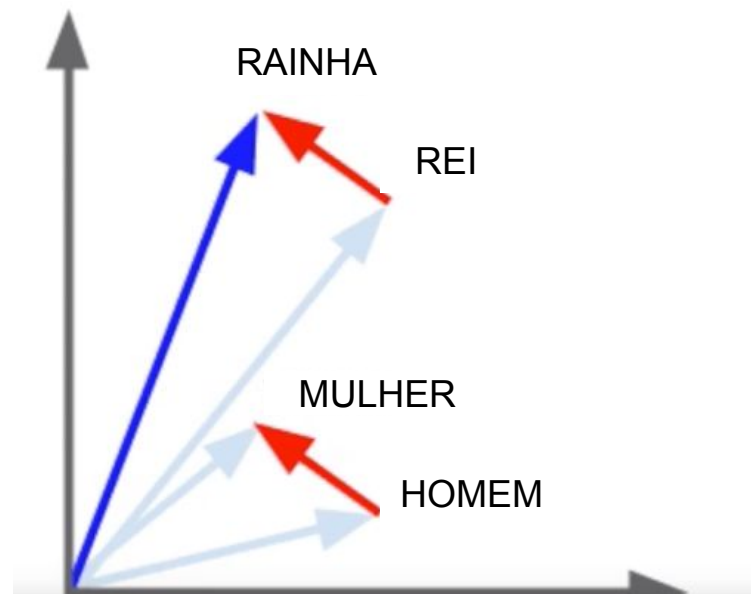
# Word-embeddings: Word2Vec



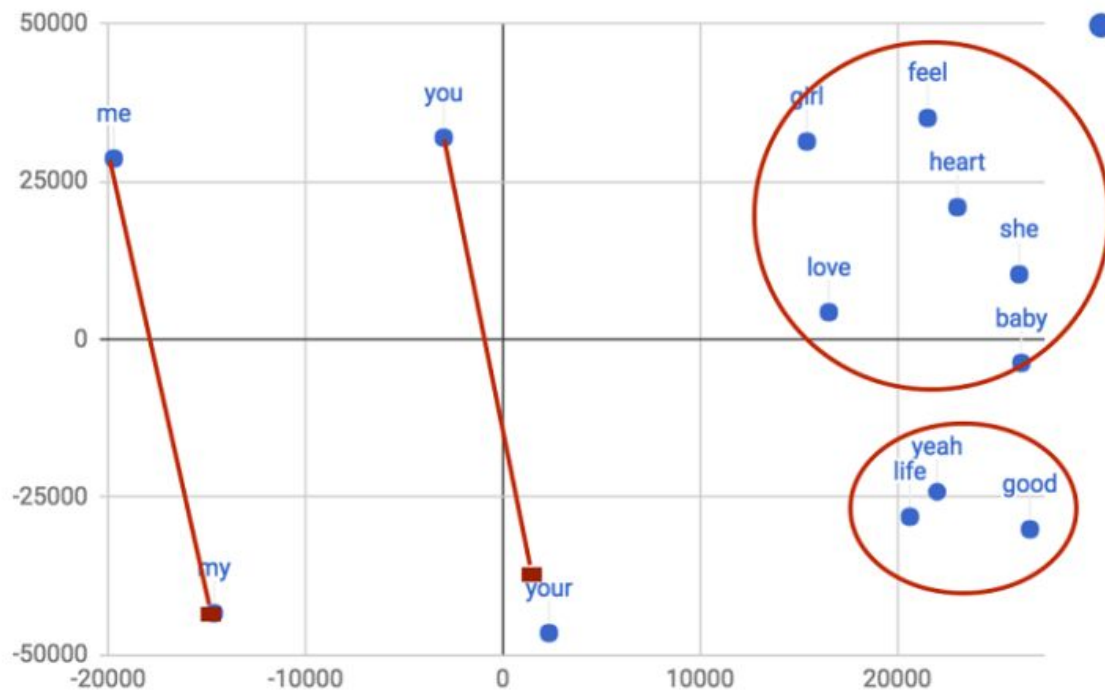
# Word-embeddings: Word2Vec

Analogias

$REI - HOMEM + MULHER \approx RAINHA$



# Word-embeddings: Word2Vec



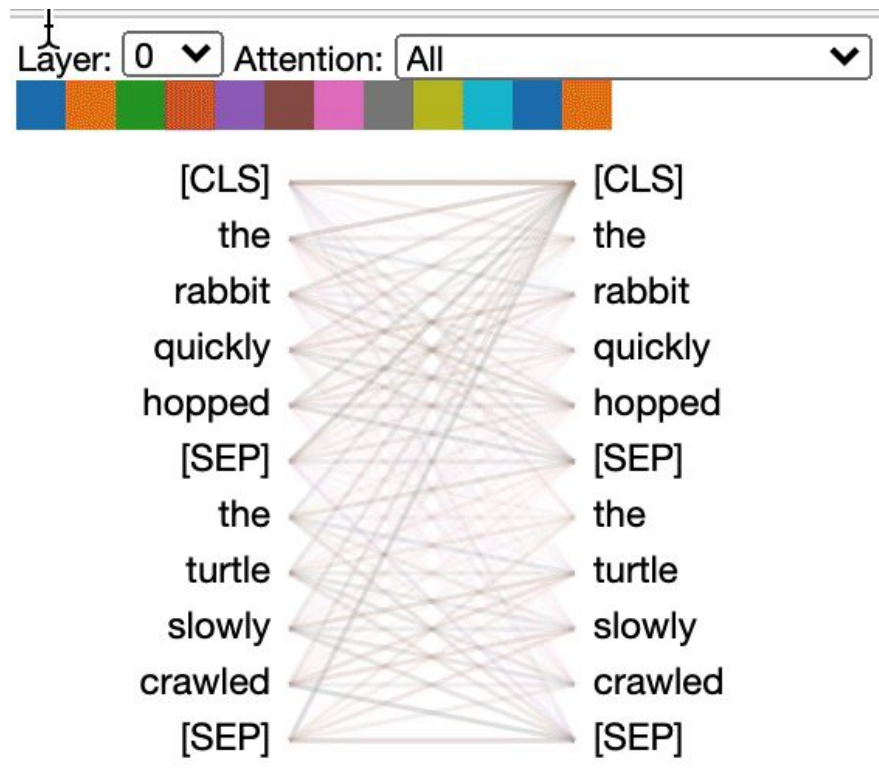
E o que tem de mais  
novo???



“Atenção é tudo que  
você precisa”  
Transformers



# Transformers (BERT)





# Transformers (GPT/ChatGPT 3.5)

## ChatGPT 3.5 ▾



**You**

Qual a opinião da sentença "O lobby do hotel é bastante bonito e os funcionários são extremamente atenciosos?" ?



**ChatGPT**

Parece ser uma opinião positiva sobre o lobby do hotel e o atendimento dos funcionários. A pessoa que expressou essa opinião provavelmente teve uma experiência agradável, achando o ambiente do lobby bonito e apreciando o alto nível de atenção e serviço prestado pelos funcionários.

# Transformers (GPT/ChatGPT 3.5)

ChatGPT 3.5 ▾



You

Quais os aspectos da sentença "O lobby do hotel é bastante bonito e os funcionários são extremamente atenciosos?" ?



ChatGPT

Claro, essa sentença aborda dois aspectos distintos sobre o hotel:

1. **A estética do lobby:** É descrito como "bastante bonito", o que implica que o espaço é visualmente atraente. Isso pode incluir elementos como decoração, design de interiores, iluminação ou arquitetura que fazem o ambiente ser considerado agradável aos olhos e convidativo para os hóspedes.
2. **Qualidade do atendimento:** A segunda parte da sentença menciona que os funcionários são "extremamente atenciosos". Isso indica um alto nível de serviço ao cliente. Funcionários atenciosos são geralmente prestativos, empáticos e prontos para ajudar os hóspedes com suas necessidades, tornando a estadia mais agradável e confortável.

Esses aspectos combinados sugerem uma experiência positiva no hotel, tanto em termos visuais

# Transformers (GPT/ChatGPT 3.5)



**You**

Qual a opinião sobre a estética do lobby?



**ChatGPT**

Parece que a opinião expressa é bastante **positiva** sobre a estética do lobby. O uso do termo "bastante bonito" indica uma apreciação pela beleza e pelo apelo visual do ambiente. Geralmente, isso sugere que a pessoa achou o espaço agradável, bem decorado ou com elementos que atraíram a atenção de forma positiva. Essa opinião reflete uma experiência visualmente agradável no lobby do hotel.



# Transformers (GPT/ChatGPT 3.5)



**You**

Qual a opinião sobre a qualidade do atendimento?



**ChatGPT**

A opinião expressa é extremamente **positiva** em relação à qualidade do atendimento. O uso da expressão "funcionários extremamente atenciosos" sugere que a pessoa teve uma experiência muito positiva com a equipe do hotel. Isso implica que os funcionários foram excepcionalmente prestativos, amigáveis e atentos às necessidades dos hóspedes, proporcionando um serviço de alta qualidade. Essa opinião reflete uma experiência muito satisfatória em termos de atendimento ao cliente no hotel.



# Limitações e O que temos feito

---



# Limitações

- Poucos Recursos Linguísticos
- Modelos com Custo Computacional Elevado

# O que temos feito

- A nossa atuação tem sido desenvolver mais recursos, principalmente, mais *datasets* em língua portuguesa para as tarefas de Análise de Sentimento, Detecção de Discurso de Ódio e Detecção de Ironia/Sarcasmo/Humor.

# Links - Grupos

- <https://wp.ufpel.edu.br/gaspln/>
- <http://ia.ufpel.edu.br/>
- <https://brasileiraspln.com/>



# Links - Recursos em Língua Portuguesa

## Discurso de Ódio e Linguagem Ofensiva

- [HateBR](#)
- [OLID-Br](#)
- [OffComBR](#)
- [NCCVG](#)
- [HLPHSD](#)
- [ToLD-Br](#)

---

## Notícias Falsas

- [COVID-19](#)
- [FakeTweetBr](#)
- [Fake.br-Corpus](#)
- [FakeWhatsApp](#)

## Análise de Sentimento

- [ReLi](#)
- [comentários sobre hotéis publicados no TripAdvisor](#)
- [comentários sobre produtos publicados no Buscapé](#)
- [TweetSentBR](#)
- [UTLCorpus](#)



# Obrigada pela atenção

---