



# PROCESSAMENTO DE LINGUAGEM NATURAL

## Similaridade de Textos



# Similaridade de Textos

Informa o quão  
"próximos" são dois  
textos

# Onde é usada?



**Motor de busca**



**Assuntos  
burocráticos**



**Atendimento ao  
consumidor**

# Onde é usada?



**Motores de  
busca**



**Assuntos  
burocráticos**

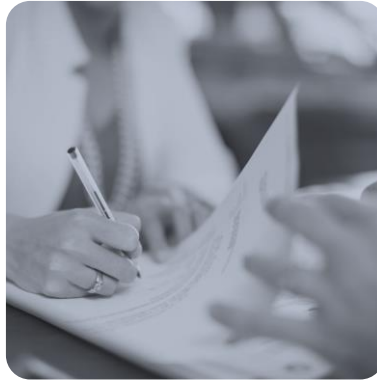


**Atendimento ao  
consumidor**

# Onde é usada?



**Motores de  
busca**



**Assuntos  
burocráticos**



**Atendimento ao  
consumidor**

# Granularidade



Caracteres



Palavras



Frases



Documentos

# Granularidade



Caracteres



Palavras



Frases



Documentos

# Granularidade



Caracteres



Palavras



Frases



Documentos



# Granularidade



Caracteres



Palavras



Frases



Documentos

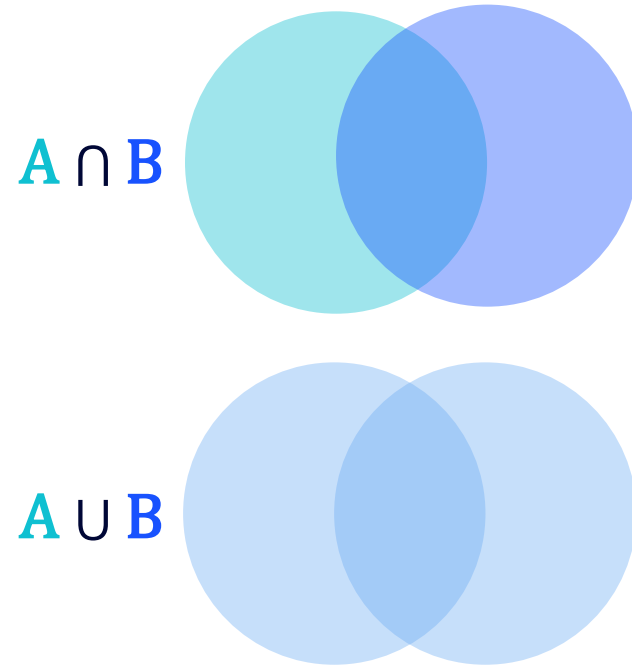


01. Similaridade de Jaccard
02. Distância Euclidiana
03. Similaridade de Cosseno

# Métodos

# Similaridade de Jaccard

$$J(A, B) = \frac{A \cap B}{A \cup B}$$

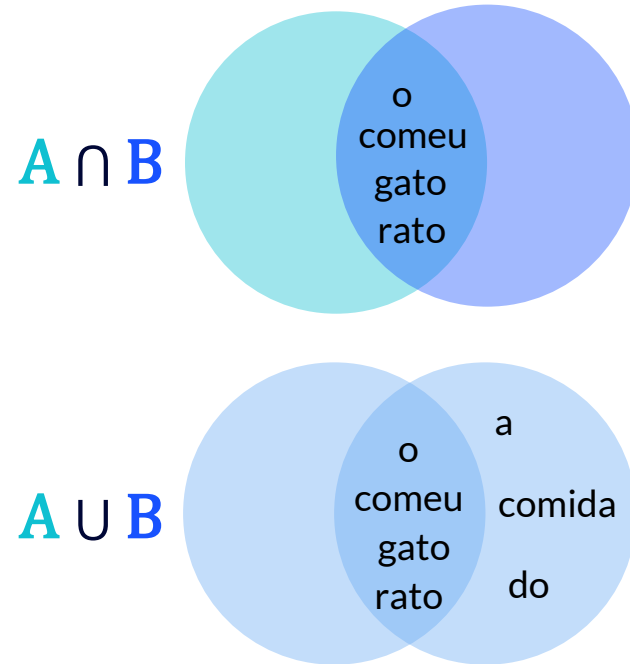


# Similaridade de Jaccard

A: O gato comeu o rato

B: O rato comeu a comida do gato

$$J(A, B) = \frac{4}{7} = 0,57$$

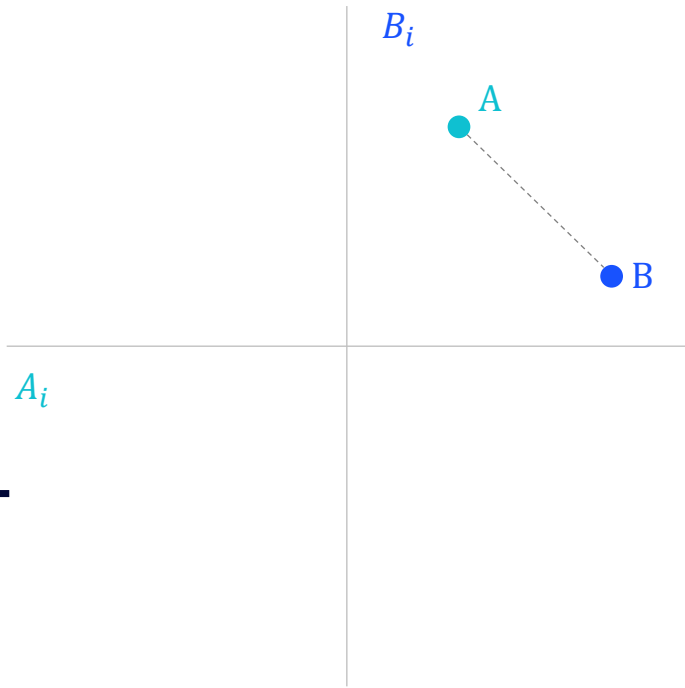


01. Similaridade de Jaccard
02. Distância Euclidiana
03. Similaridade de Cosseno

# Métodos

# Distância Euclidiana

$$d(A,B) = \sqrt{(A_i - B_i)^2}$$

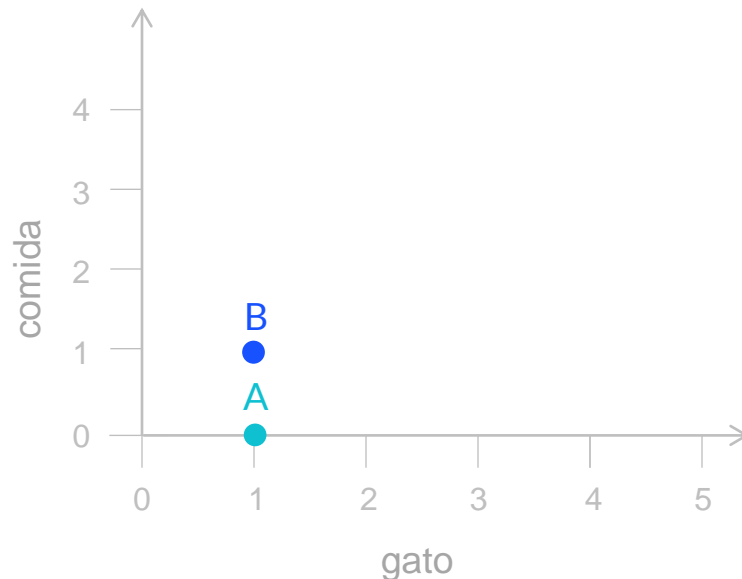


# Distância Euclidiana

A: O gato comeu o rato

B: O rato comeu a comida do gato

	o	gato	comeu	rato	a	comida	do
A	2	1	1	1	0	0	0
B	1	1	1	1	1	1	1



# Distância Euclidiana

	o	gato	comeu	rato	a	comida	do
A	2	1	1	1	0	0	0
B	1	1	1	1	1	1	1

$$d(A,B) = \sqrt{(2-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + (0-1)^2 + (0-1)^2 + (0-1)^2}$$

$$d(A,B) = 2.0$$

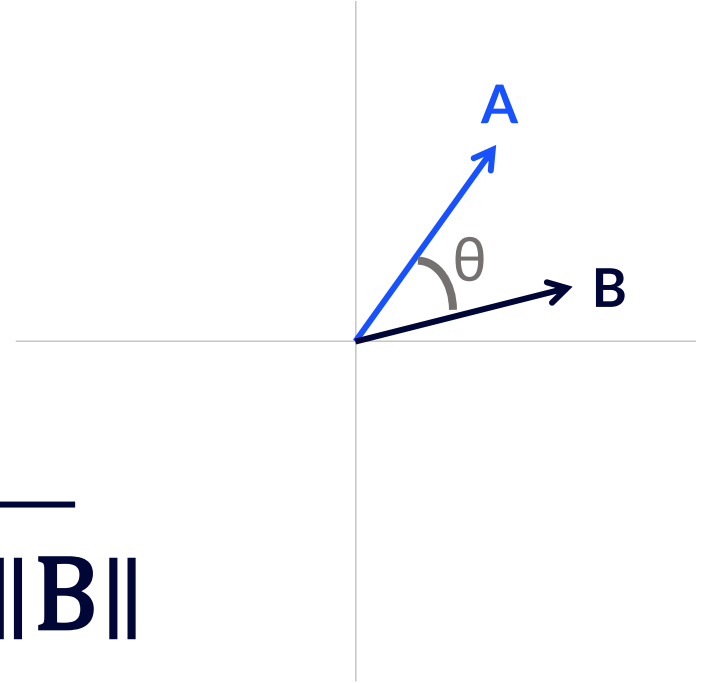


01. Similaridade de Jaccard
02. Distância Euclidiana
03. Similaridade de Cosseno

# Métodos

# Similaridade de Cosseno

$$\text{cosseno}(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \times \|\mathbf{B}\|}$$



# Similaridade de Cosseno

	o	gato	comeu	rato	a	comida	do
A	2	1	1	1	0	0	0
B	1	1	1	1	1	1	1

01.  $A \cdot B = (2 \times 1) + (1 \times 1) + (1 \times 1) + (1 \times 1) + (0 \times 1) + (0 \times 1) + (0 \times 1) = 5$

# Similaridade de Cosseno

	o	gato	comeu	rato	a	comida	do
A	2	1	1	1	0	0	0
B	1	1	1	1	1	1	1

01.  $A \cdot B = (2 \times 1) + (1 \times 1) + (1 \times 1) + (1 \times 1) + (0 \times 1) + (0 \times 1) + (0 \times 1) = 5$

02.  $\|A\| = \sqrt{2^2 + 1^2 + 1^2 + 1^2 + 1^2 + 0^2 + 0^2} = \sqrt{6}$

# Similaridade de Cosseno

	o	gato	comeu	rato	a	comida	do
A	2	1	1	1	0	0	0
B	1	1	1	1	1	1	1

01.  $A \cdot B = (2 \times 1) + (1 \times 1) + (1 \times 1) + (1 \times 1) + (0 \times 1) + (0 \times 1) + (0 \times 1) = 5$

02.  $\|A\| = \sqrt{2^2 + 1^2 + 1^2 + 1^2 + 1^2 + 0^2 + 0^2} = \sqrt{6}$

03.  $\|B\| = \sqrt{1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2} = \sqrt{7}$

# Similaridade de Cosseno

	o	gato	comeu	rato	a	comida	do
A	2	1	1	1	0	0	0
B	1	1	1	1	1	1	1

01.  $A \cdot B = (2 \times 1) + (1 \times 1) + (1 \times 1) + (1 \times 1) + (0 \times 1) + (0 \times 1) + (0 \times 1) = 5$

02.  $\|A\| = \sqrt{2^2 + 1^2 + 1^2 + 1^2 + 1^2 + 0^2 + 0^2} = \sqrt{6}$

03.  $\|B\| = \sqrt{1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2} = \sqrt{7}$

04.  $\text{cosseno}(\theta) = \frac{5}{\sqrt{6} \times \sqrt{7}} = 0.71$



**PUC Minas**  
**Virtual**