

Análise de texto de fontes desestruturadas e Web

Aula 12

Maciel Calebe Vidal

Aula passada

	Secao	Titulo
0	politica	Novas parcelas do auxilio emergencial ganham f...
1	economia	Café arábica recua de máxima de 4 anos na ICE;...
2	economia	Milho e soja tocam máximas de vários anos em C...
3	tecnologia	Brasil termina abril com 82.000 mortes por cov...
4	tecnologia	Vai comprar como?
5	politica	'Há uma grande operação abafa em curso', diz e...
6	politica	Conheça os projetos que propõem prorrogar o au...
7	tecnologia	A reestruturação do SAS
8	politica	TSE mantém protocolos sanitários no segundo turno
9	tecnologia	Laboratório Moderna, pioneiro da vacina antico...

Text Vectorization - Aula passada

Converter texto em um conjunto de números (vetores)

Bag of Words

Notícia 1

CSN estende venda de
suas ações da Usiminas

Notícia 2

Google estende ações
para proibição de
anúncios políticos



Palavra	Notícia 1	Notícia 2
CSN		
Ações		
Anúncios		
Usiminas		
proibição		
estende		
Google		
políticos		
venda		

Resultado - Aula passada

Notícia	governo	dólar	STF	COVID	...	juros	Categoria
1	1	1	1	1		0	política
2	0	0	0	1		0	tecnologia
3	1	1	0	1		1	economia
4	1	1	0	0		1	economia
5	0	0	1	1		0	política
6	0	1	1	1		0	política
7	0	1	1	1		1	economia
...
n	1	0	0	0		1	política

Na aula de hoje

	Titulo
0	UPL tem novo CEO para operação no Brasil
1	Brasil atinge a marca de 450 mil mortes por co...
2	Papo responsável
3	Expectativas de alta do PIB estão indo em dire...
4	Para atrair investimento, Bolsonaro é garoto-p...
5	Lava atinge aeroporto na República Democrática...
6	Diretora do Fed alerta para riscos de stableco...
7	IPC-Fipe sobe 0,33% na 3ª quadrissemana de mai...

Interpretabilidade

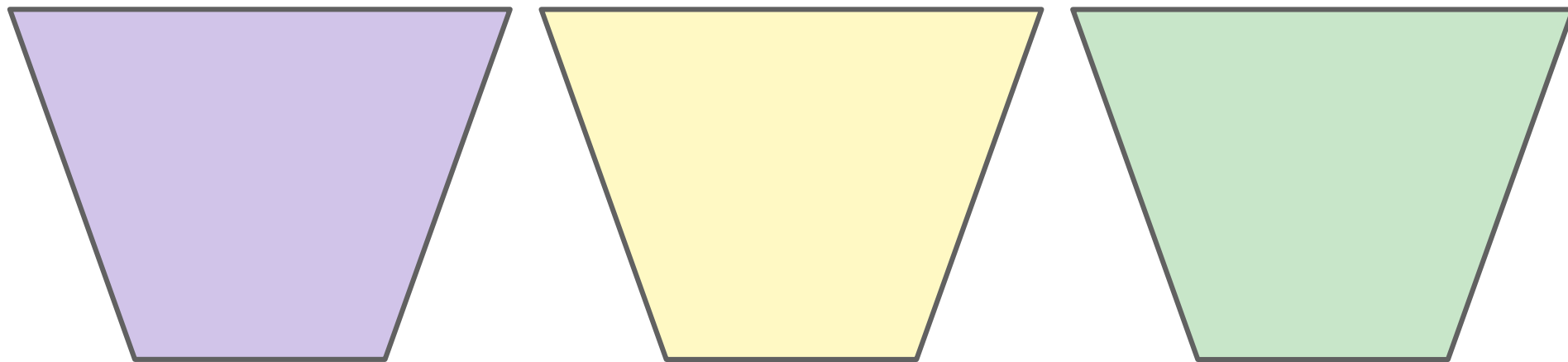
Até agora, os britânicos podiam ingressar na Espanha apenas em casos de extrema necessidade. A partir de 24 de maio, não será mais preciso apresentar um PCR negativo na chegada, como é feito atualmente, conforme indicado pelo governo em um site de promoção do turismo.

Análise de Tópicos

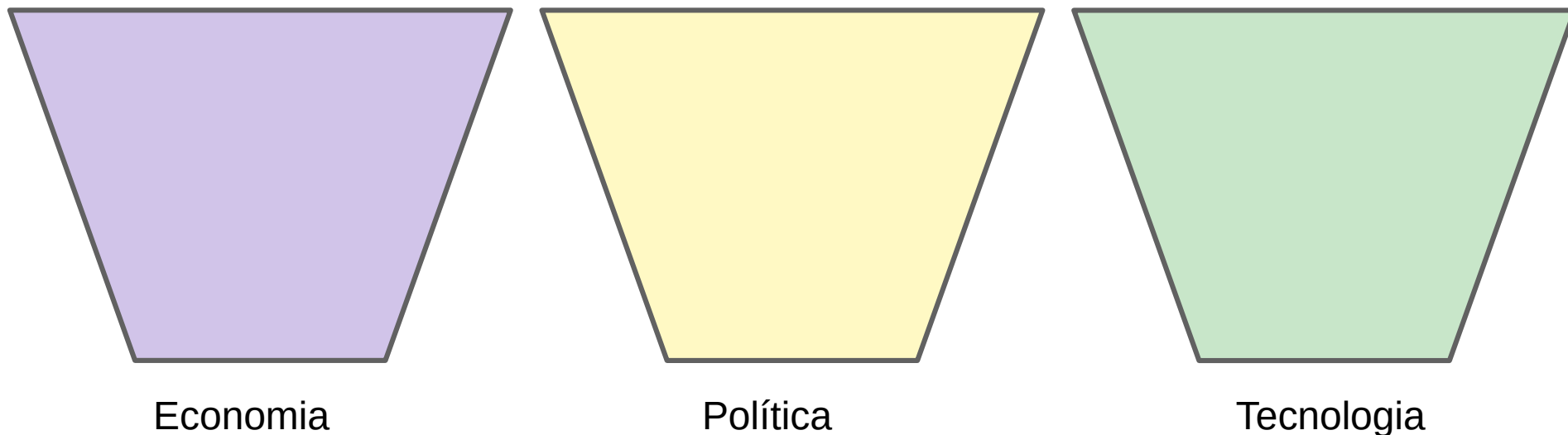
- Grande quantidade de dados
- Extrair informações relevantes
- Em forma de tópicos
- Modelagem de tópicos

	Título
0	UPL tem novo CEO para operação no Brasil
1	Brasil atinge a marca de 450 mil mortes por co...
2	Papo responsável
3	Expectativas de alta do PIB estão indo em dire...
4	Para atrair investimento, Bolsonaro é garoto-p...
5	Lava atinge aeroporto na República Democrática...
6	Diretora do Fed alerta para riscos de stableco...
7	IPC-Fipe sobe 0,33% na 3ª quadrissemana de mai...

Latent Dirichlet Allocation (LDA)



Latent Dirichlet Allocation (LDA)



Análise de Tópicos

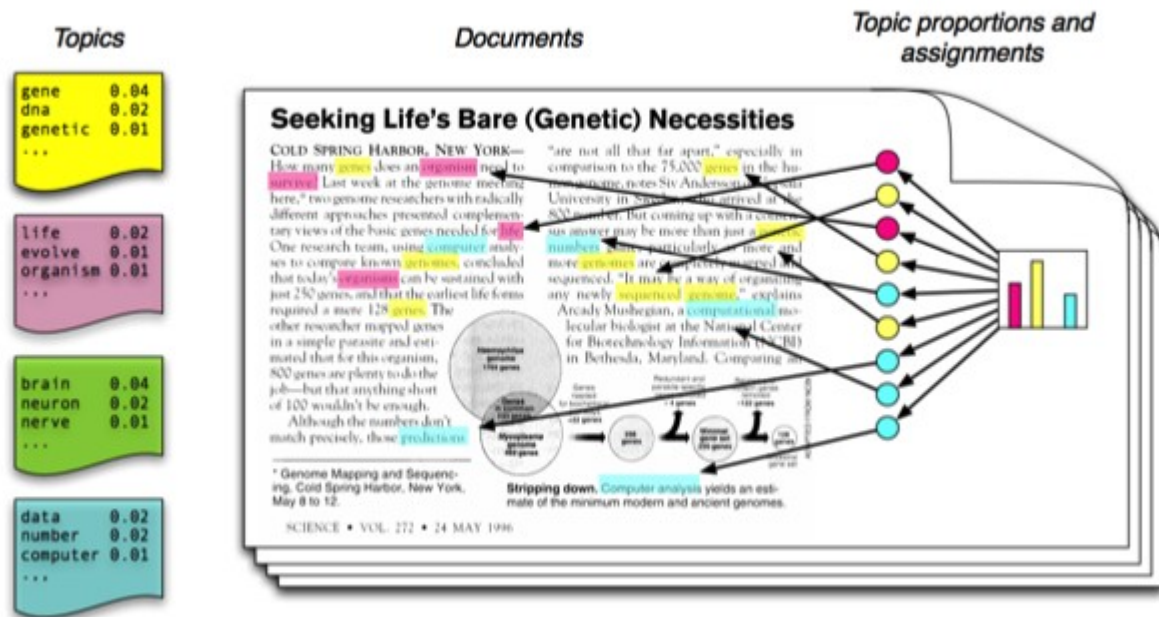


Figure source: Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77-84.

Latent Dirichlet allocation (LDA)

- Palavras mais associadas com cada tópico
- Com quais tópicos um documento está associado

Até agora, os britânicos podiam ingressar na Espanha apenas em casos de extrema necessidade. A partir de 24 de maio, não será mais preciso apresentar um PCR negativo na chegada, como é feito atualmente, conforme indicado pelo governo em um site de promoção do turismo.



Dúvidas?

Obrigado pela participação!