

**Códex, Regex, LLMs:  
texto e pesquisa nas  
Humanidades Digitais**

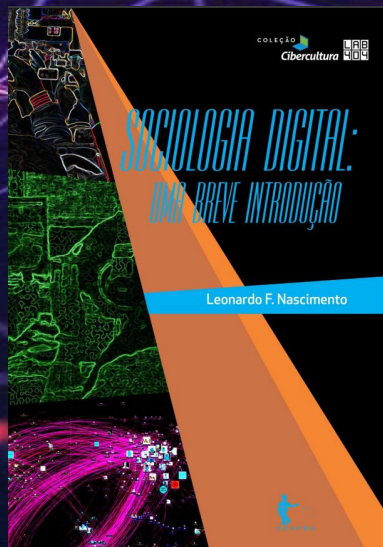


**Leonardo Nascimento**

Computational Social Scientist  
Federal University of Bahia

**15% Programador,**  
**20% Químico,**  
**25% Psicólogo,**  
**40% Cientista Social**

Clique nas imagens para acessar



Públicos refratados:  
grupos de  
extrema-direita  
brasileiros na  
plataforma Telegram



Surge em 2018 junto ao Programa de Pós-graduação em Ciências Sociais (PPGCS) e o Instituto de Ciência, Tecnologia e Inovação (ICTI/UFBA)

Reunir pesquisadores interessados em métodos digitais/computacionais e ciências sociais

# Pesquisadores



**Eric Brasil**

Digital Historian  
Unilab



**Letícia Cesarino**

Digital anthropologist  
Federal University of Santa Catarina



**Rosana Moore**

Pós Doc LABHDUFBA



**Paulo Fonseca**

Digital sociologist  
Federal University of Bahia



**Pedro Moraes**

Data engineer  
Ibotirama Sistemas



**Tarssio Barreto**

Data scientist  
BIT::Analytics

# Equipe

**Anna Carollyne dos Santos Vieira (UFBA) - Bacharelado Interdisciplinar em Humanidades**

**Beatriz Leal Fraga (UFBA) - Bacharelado em Ciências Sociais**

**Daniel de Sena Bastos (UFBA) - Bacharelado Interdisciplinar em CTI**

**Emily Caroline de Vasconcelos (USP) - Bacharelado em Ciências Sociais**

**Ingrid da Silva Barreiros (UFBA) - Bacharelado Interdisciplinar em CTI**

**Jéssica Santana de Oliveira - Bacharelado Interdisciplinar em CTI**

**Juciane Pereira de Jesus(UFBA) - Mestrado em ciências sociais**

**Maria Luiza Scheren (UFSC) - Bacharel em Antropologia**

**Louise Lima Karczeski (UFSC) - Mestre em Antropologia Social**

**Matheus da Costa Martins - Bacharelado em Ciências Sociais**

**Pedro Mores - Engenharia de Dados**

**Rosana Moore - Doutora em Sociologia - Bolsista Pós-doc CNPQ**

**Sofia Schurig - Comunicação Social (FACOM/UFBA)**

**Tarssio Barreto - Cientista de Dados**

**Thamirys Albuquerque Cunha (UFBA) - Bacharelado em Ciência, Tecnologia e Inovação**

# Argumento central

**Os modos de busca e recuperação de informação em textos (indexação) é um dos aspectos, senão o mais importante, da pesquisa das ciências humanas e sociais.**

# Mote

**“a revolução das modalidades de produção e de transmissão dos textos é também uma mutação epistemológica fundamental”**

**(Chartier, 2002, p. 108).**

# O que estes objetos têm em comum?



Pedra



Argila



Papiro



Papel



Cilindro de  
cera



Fita  
Cassete

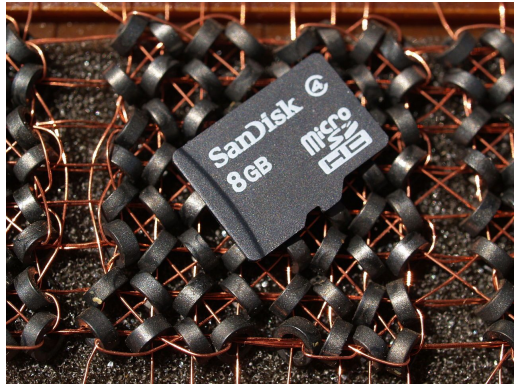


Disco de  
Vinil

<https://en.wikipedia.org/wiki/Writing>

**(Resposta: eles são analógicos)**

# O que estes objetos têm em comum?



8GB (front) vs 8B (back)



Floppy disks (8", 5 1/4", 3 1/2")



Compact disk

**(Resposta: eles são digitais)**

# Mutações epistemológicas (“pré-strings”)

1. Advento tecnológico do Códex (séc I e II d.C) (Chartier, 2002, p. 106)
2. Commonplace Books (sec XVI) (Blair, 2011)
3. Card index (sec XV)(Briggs and Burke, 2010, p. 15)

# ...nova materialidade digital dos bits

(Brasil & Nascimento, 2020)



## FOLHA DE S. PAULO

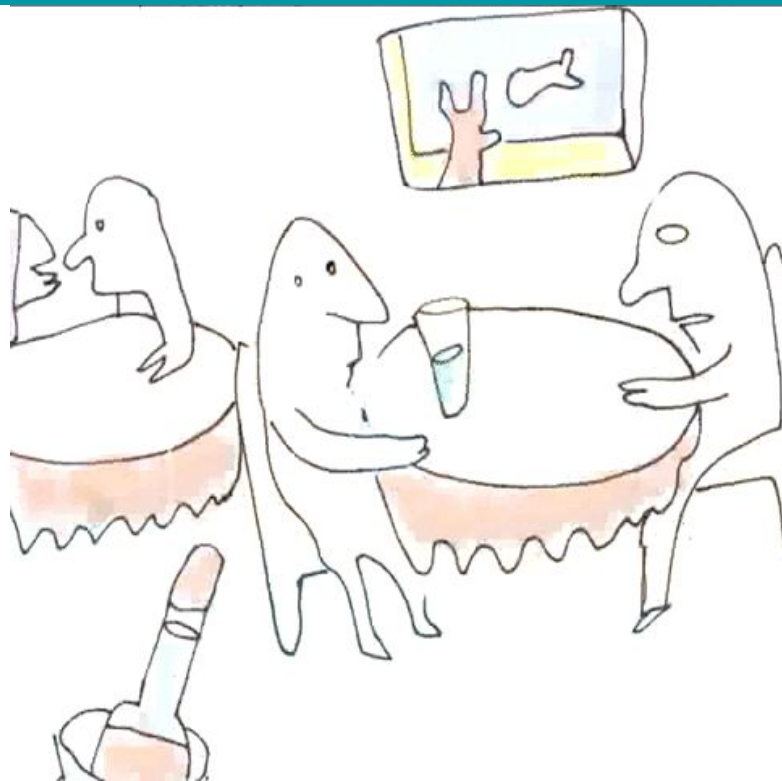


**Folha de S. Paulo** is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s with the release of Letraset sheets containing Lorem Ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem Ipsum

# Mutações epistemológicas (“pós-strings”)

## 4. Expressões regulares (séc XX)

# "Digitalização do eu na vida cotidiana" (Nascimento, 2020)



# Algoritmização de processos sociais



# Traços digitais

(Howison et al. 2011, p. 769)



# Dataficação (Cukier & Mayer-Schoenberger, 2013)



# Era do “Big Data”

The **co-evolution** of  
**storage** capacity,  
**transmission** capacity, and  
**processing** capacity

[Visualcapitalist.com](https://www.visualcapitalist.com) (2021)



# Mutações epistemológicas (“pós-strings”)

## 5. Grandes modelos de linguagem (LLMs) (sec XXI)

# Aumento exponencial do uso do Telegram no Brasil



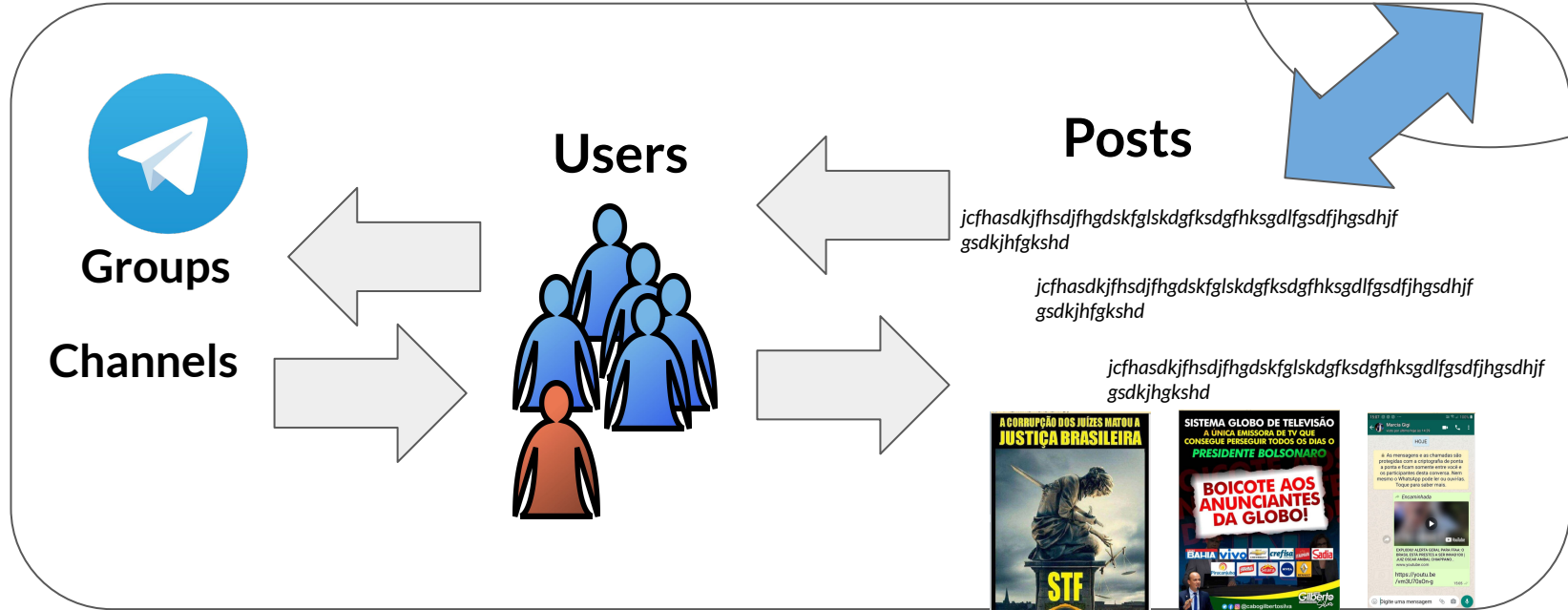
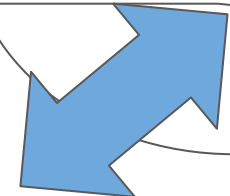
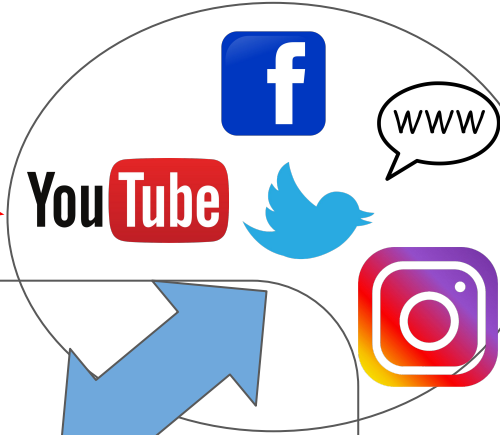
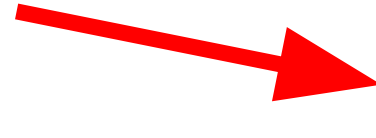
# Vigilância-como-método

**“using computational methods to scrape and collect  
large datasets of posts, comments, and profiles  
(in real-time)”**

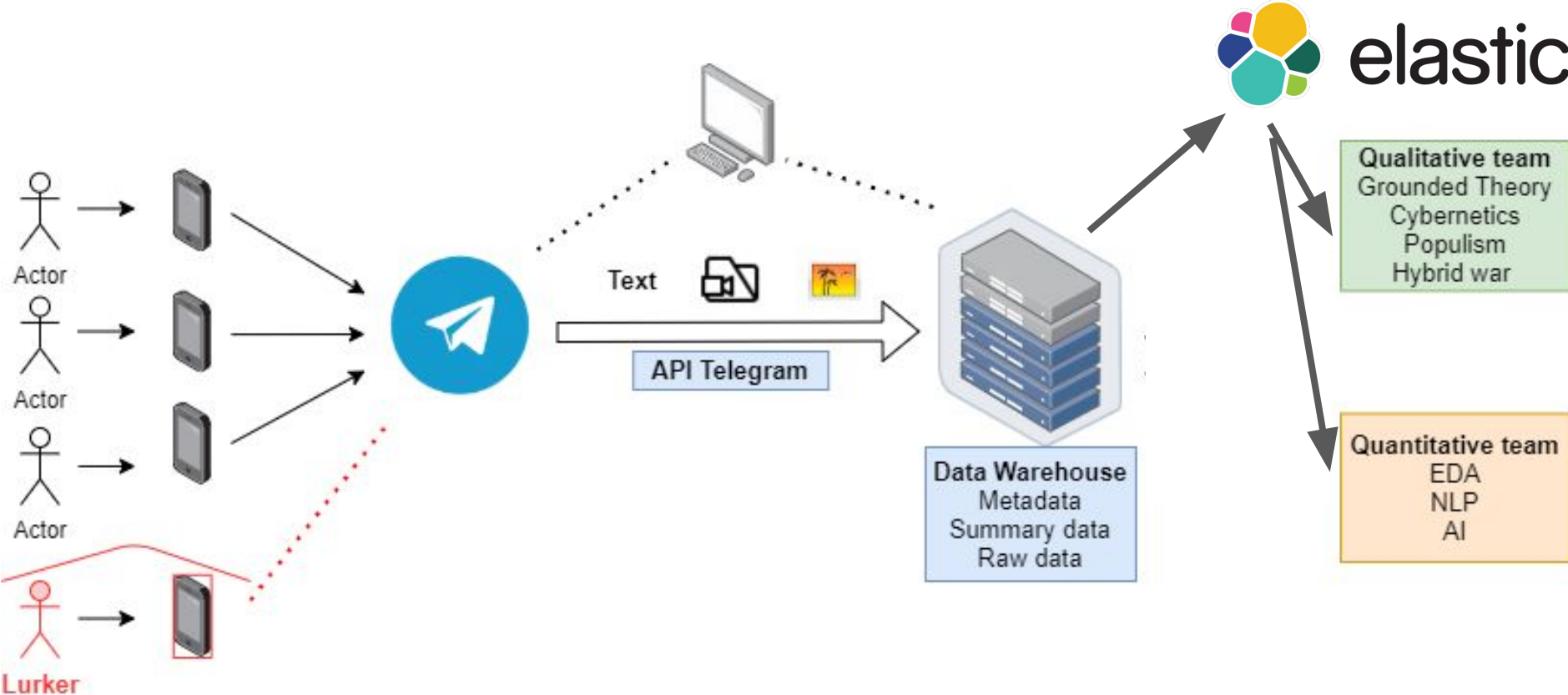
(Topinka, Finlayson & Osborne-Carey, 2021, p. 385)

# Gramática do Telegram

Multi-platform ecosystem



# Extract, Transform and Load





# elastic

- a. Search engine (based on Lucene)
- b. Data visualization (Kibana)

# Relatórios parciais



1º Relatório  
Janeiro a Julho de 2022



2º Relatório  
1 de Agosto a 14 de Setembro de 2022



3º Relatório  
30 de setembro a 15 de novembro de 2022



4º Relatório - Ataques do  
08/01/2023

# Artigos



MAY 8, 2023

## Digital Anthropology Meets Data Science

Analyzing online ecosystems in real time, teams of anthropologists and data scientists can begin to understand rapid social changes as they happen.

↳ artigo

Públicos refratados: grupos de extrema-direita brasileiros na plataforma Telegram

Leonardo F. Nascimento

Leticia Cesarino

Paulo Fonseca

Tarssio Barreto

Vitor Mussa

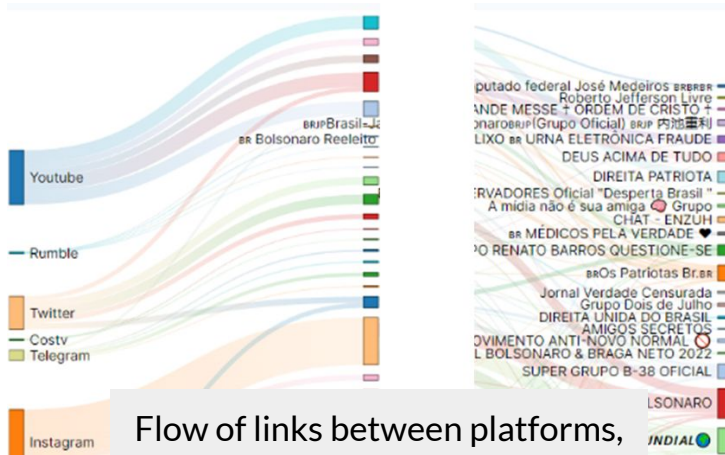
**Far-right publics on brazilian Telegram a mixed-methods approach to digital anthropology**

Corporate author : [UNESCO](#) [66594], [LiiV Center for Innovating Digital Anthropology \(USA\)](#) [16]

Person as author : [Cesarino, Leticia](#) [author] [1], [Nascimento, Leonardo](#) [author] [1], [Fonseca, Paulo](#) [author] [1]

Document code : SHS/MOST/DA/2023/PI/8

# Examples of Analytic Tools



Flow of links between platforms, users and Telegram chats

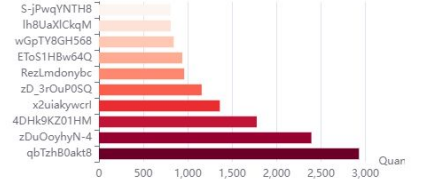
## Ecosistema multiplataforma Word2Vec

Selecione um período

2022-08-05 to 2022-10-23

Selecione a plataforma

- Youtube
- Twitter
- Facebook
- Instagram
- Tiktok
- Kwai

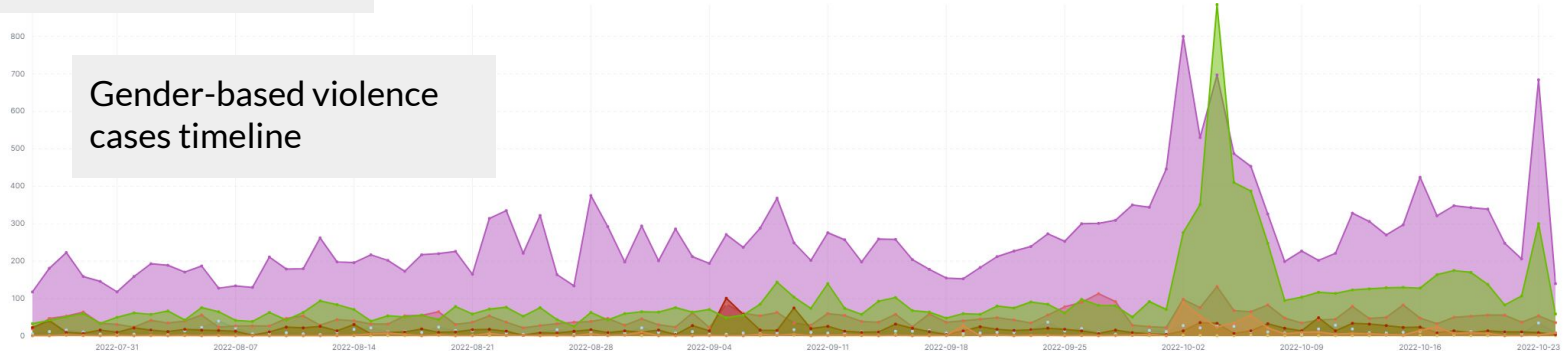


Most shared links on Telegram

10 resultados por página

Pesquisar

id	Link	Quantidade
All	All	All
qbTzhB0akt8	<a href="https://youtu.be/qbTzhB0akt8">https://youtu.be/qbTzhB0akt8</a>	2928
zDuOoyhyN-4	<a href="https://youtu.be/zDuOoyhyN-4">https://youtu.be/zDuOoyhyN-4</a>	2392



# Telegram: entre repertórios e experimentações

Burger King botando criança pra ensinar adulto a dar a bunda virar viado e virar trans.

<https://www.youtube.com/watch?v=301GMPHt7M>

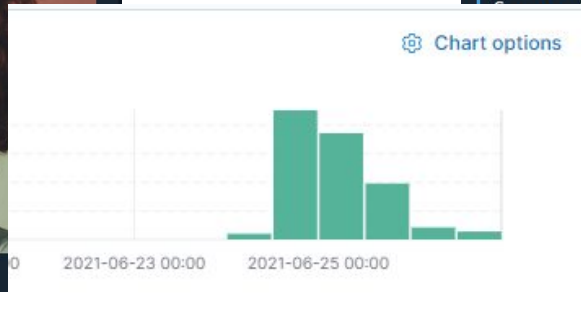
YouTube

Burger King | Como explicar?

Sempre tem um adulto que, quando vê algo relacionado a LGBTQAI+, pensa: "como eu vou explicar isso pra uma criança?". Mas, ao contrário do que eles pensam, as crianças acham esses ...



## Pânico Moral



🟡🔴 \*Vamos bater nos 200mil deslikes 🙄?\*

O Burger King está com uma publicidade onde crianças dizem o que é LGBTQAI+

Acesse agora mesmo o link abaixo e dê o seu DESLIKE.

Faça isso agora pelas nossas crianças! Pelos nossos filhos e netos!

<https://www.youtube.com/watch?v=301GMPHt7M>

Repasse para todos os seus grupos de WhatsApp.

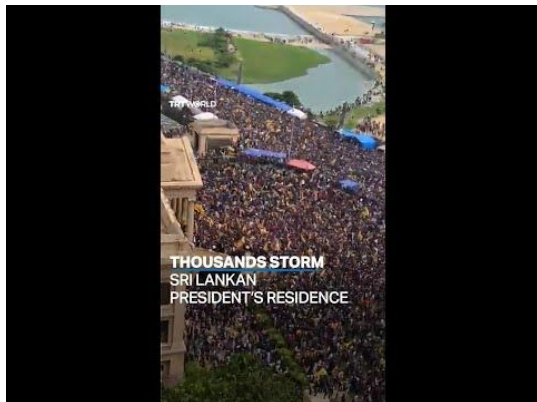
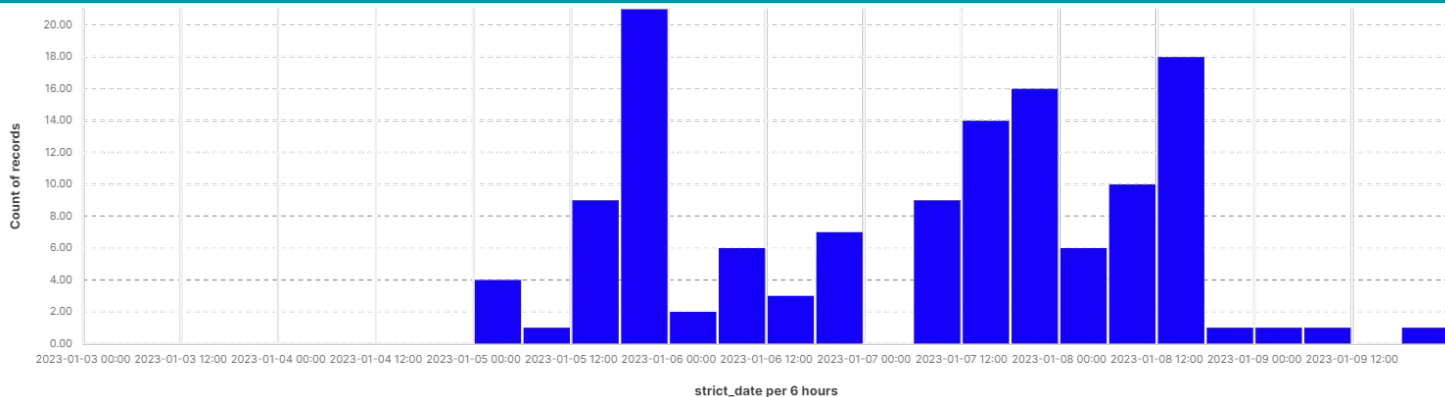
YouTube

Burger King | Como explicar?

um adulto que, quando vê algo relacionado a LGBTQAI+, pensa: "como eu vou explicar isso pra uma criança?". Mas, ao contrário do que eles pensam, as crianças acham esses ...



# Festa da Selma



Vídeo mais compartilhado de 01 a 08/01/2023





**Perguntas?**



<https://github.com/LABHDUFBA>



<https://www.youtube.com/c/LABHDUFBA>



[@labhdufba](https://www.instagram.com/labhdufba)



[@labhdufb](https://twitter.com/labhdufb)

# Referências Bibliográficas

Blair, A., 2011. Too Much to Know: Managing Scholarly Information Before the Modern Age. Yale University Press.

Briggs, A., Burke, P., 2010. Social History of the Media: From Gutenberg to the Internet. Polity.

Chartier, R., 2002. Os desafios da escrita. UNESP.