

Universidade do Minho

Prémio Arquivo.pt 2021

## Major Minors





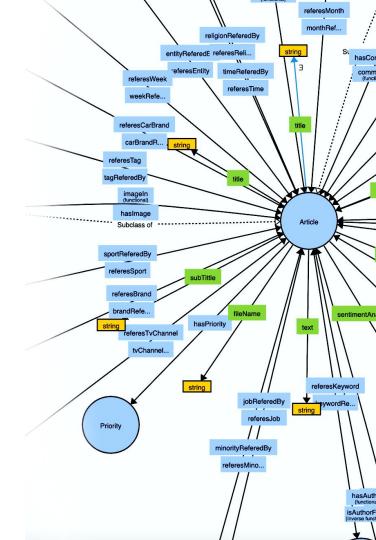
Paulo Martins Leandro Costa



## O que é?

- Knowledge Graph de recortes de imprensa relacionados com questões minoritárias.
- Uma ferramenta de investigação para estudo da representação de minorias pelos meios de comunicação social.
- Corpus ontológico de artigos, comentários e imagens cruzados com milhares de entidades.

Todos os dados foram extraídos a partir do **Arquivo.pt** abrangendo as duas primeiras décadas do séc. XXI.



## Major Minors





















Berço do Projeto: Departamento de Informática da Univ. do Minho

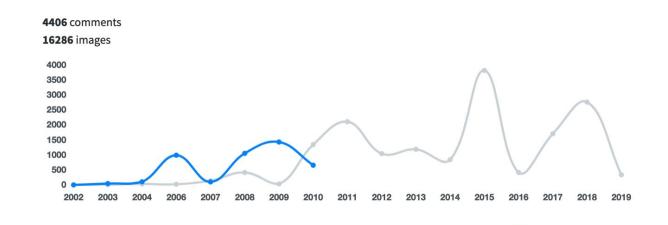








■ Comments ■ Images





## → ARTIGOS

De que forma evoluiu a representação de minorias no contexto jornalístico português?

Que acontecimentos e entidades se relacionaram com temáticas minoritárias?

Qual a evolução linguística (erros ortográficos mais comuns, estrangeirismos introduzidos, etc.) nos últimos 20 anos?





### → IMAGENS

Quais as tendências e influência social das fotos ilustrativas?

Que estereótipos e cargas transmitem?

Tem existido variabilidade ao longo das últimas duas décadas na carga positiva/negativa associada à escolha das fotos para representar determinados grupos minoritários?





## → COMENTÁRIOS

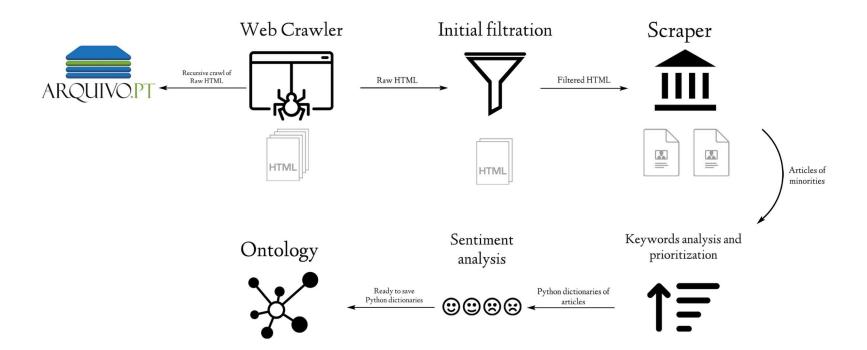
Qual a percepção do público perante diferentes questões minoritárias e evolução ao longos dos últimos 20 anos?

Quais as palavras mais empregues para cada minoria e a carga é negativa ou positiva?

Análise linguística do português na ótica do utilizador digital nos últimos 20 anos.



## Etapas de desenvolvimento



# Ontologias & Knowledge Graphs

«Information is the oil of the 21st century, and analytics is the combustion engine»
(Peter Sondergaard, Gartner Research)

«Hiding within those mounds of **data** is **knowledge** that could change the life of a patient, or change the world.»

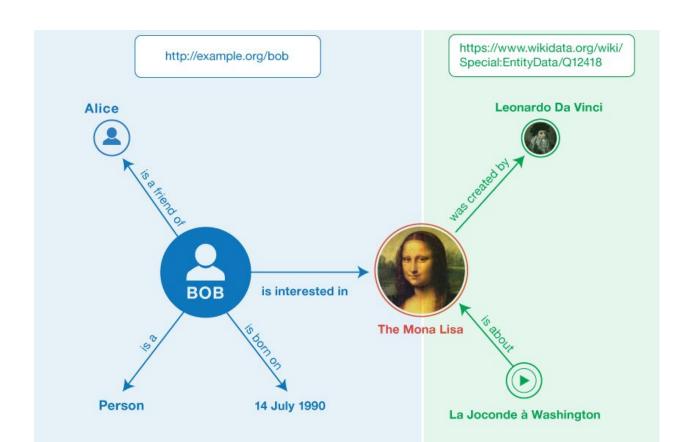
(Atul Butte, Stanford)



BIG DATA



## Ontologias & Knowledge Graphs





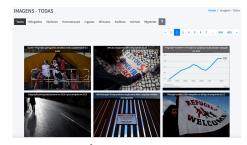
Type	No. of References
Public Figures	32.648
Political Parties	8.525
Minorities	11.530
Entities	11.473
Brands	46.629
Religions	260
Sports	4.333
Ethnicities	1.345
Car brands	969
Animals	6.199
TV Channels	3.799
Continents	7.119
Countries	28.794
Cities	28.933
Other Places	1.067
Newspapers defined tags	10.186
Weekdays	20.575
Months	24.113
Minorities related keywords	36.897







Artigos http://minors.ilch.uminho.pt/articles



**Imagens** http://spargl.ilch.uminho.pt/images



Comentários http://minors.ilch.uminho.pt/comments

## **Galerias**

1º nível de acesso: não técnico. Fácil de extrair dados mas manipulação limitada.





Visual search interface http://minors.ilch.uminho.pt/search



GraphDB Visual http://spargl.ilch.uminho.pt

## **Interfaces Reativas**

2° nível de acesso: intermédio. Manipulação visual dos dados até três camadas.

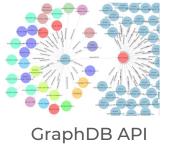




SPARQL API http://minors.ilch.uminho.pt/sparql



WebVOWL http://minors.ilch.uminho.pt/ontology



http://sparql.ilch.uminho.pt/sparql

## SPARQL / APIs

3° nível de acesso: técnico. Liberdade total de manipulação dos dados.





Website
http://minors.ilch.uminho.pt



Crawler
https://tinyurl.com/3vawvncm



Scraper https://tinyurl.com/zk68yy6v



Ontology Builder <a href="https://tinyurl.com/yuyab2we">https://tinyurl.com/yuyab2we</a>



Datasets
<a href="https://tinyurl.com/25dhihdv">https://tinyurl.com/25dhihdv</a>

## **Ferramentas**

Ferramentas de expansão open-source.



























#### **Publicações:**

■ Paulo Martins, Leandro Costa, & José Ramalho (2021). Major Minors - Ontological representation of minorities by newspapers. In 10th Symposium on Languages, Applications and Technologies (SLATE 2021). School of Technology, Polytechnic Institute of Cávado and Ave, Portugal. (in press)

#### **Comunicações:**

■ TechLING 2021 (Univ. de Vigo) - VI Congresso internacional de língua, linguística e tecnologia.

#### **Projetos:**

- Alimentar a base de dados com novos jornais, anos e entidades.
- Replicar o conceito em novos países, fontes, idiomas e temas.

