

Bulk Access to Web Archived data using APIs

vasco.rato@fccn.pt

Arquivo.pt preserves national and international historical web content

GALILEO Versão 2.0

Janeiro 1993

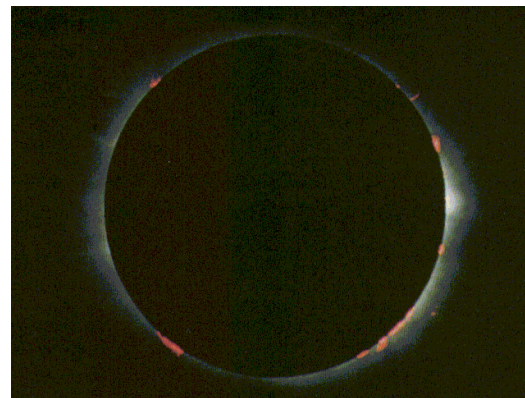
Manual de utilização e sugestões de exploração	AUTORES Programação: João Veloso Manual: Elisa Prata Pina e M. Augusta Patrício Orientação: Carlos Fiolhais
--	--

Departamento de Física da Universidade de Coimbra

ÍNDICE

- [1. Apresentação sumária do programa](#)
- [2. Características do equipamento](#)
- [3. Configuração mínima do computador](#)
- [4. Ficheiros do programa](#)

nautilus.fis.uc.pt - 1993



spacelink.nasa.gov - 1992

How to **search** and **access**
the preserved information?



Search **texts** from the past in any language

[Menu](#)[Opções](#)

ARQUIVO.PT

[Pesquisar](#)

1991
6 Ago

2025
5 Jun

[Páginas](#)[Imagens](#)[Narrativa](#)

[Pesquisa avançada](#)

Cerca de 415 resultados desde 1991 até 2025

resaw.eu/wp-content/uploads/2014/06/Resaw-conference-programme-v.3.pdf
[PDF] [Microsoft Word - Resaw conference programme v.3.docx](#)
3 Junho 2017
and Perspectives Programe **RESAW conference**, 8 – 10 June 2015, Aarhus University **Conference** venue: Incuba Science Park, Abogade 15, 820 ... is always available at the website: <http://resaw.eu/events/international> ...

resaw.eu/events
[Events | RESAW](#)
3 Junho 2017
Events Second **RESAW Conference** London, June 14-15 June 2017 The second **RESAW conference** will be held at ... Perspectives, international **conference**, Aarhus, 8-10 June 2015 **RESAW** Seminar, London, 3-4 December 2014 ... Events | **RESAW RESAW** A Research infrastructure for the Study of Archived Web materials Menu Skip to ...

resaw.eu/forum [page](#)

Search **images** from the past

[Menu](#)

ARQUIVO.PT

[Opções](#)

Pesquisar

2021
6 Ago

2025
5 Jun

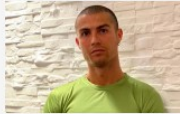
Páginas

Imagens

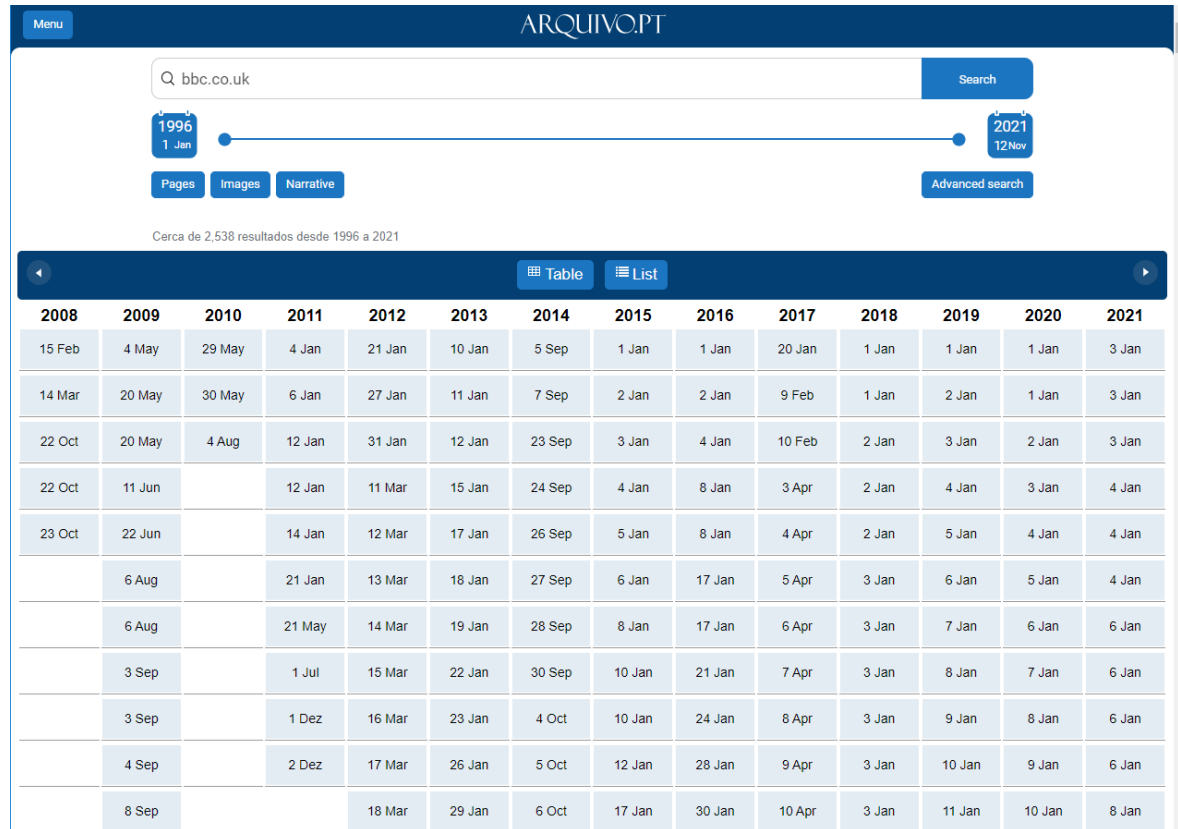
Narrativa

Pesquisa avançada

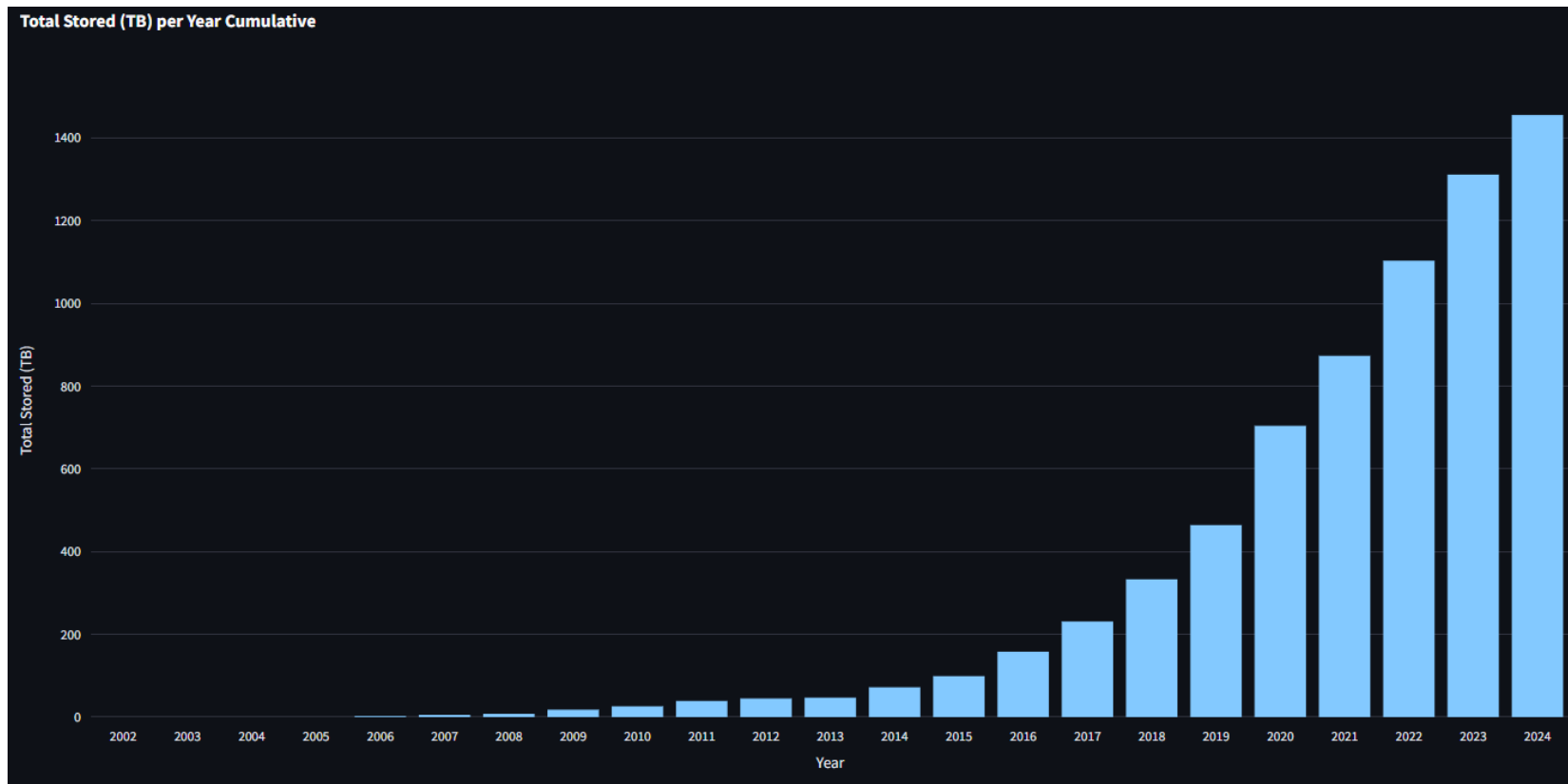
Cerca de 1.871.523 resultados desde 1991 até 2025

 → https://www.transfermark... 29 Fevereiro 2021	 → https://www.transfermark... 27 Outubro 2021	 → https://www.transfermark... 15 Setembro 2021	 → https://www.transfermark... 14 Setembro 2019	 → https://www.transfermark... 7 Janeiro 2021	 → https://www.90min.com/p... 4 Janeiro 2021	 → http://teknomatika.blogspot... 11 Maio 2020
 → https://www.realmadrid.c... 8 Outubro 2019	 → https://www.brigitte.de/ak... 1 Março 2021	 → https://www.stern.de/spo... 28 Fevereiro 2021	 → https://greatmentordolnet... 5 Julho 2020	 → https://sportdaten.spiegel... 17 Junho 2021	 → https://www.90min.com/it... 31 Março 2021	 → https://www.abola.pt/80i... 24 Dezembro 2020
 → https://www.abola.pt/80/s... 17 Abril 2021	 → https://www.abola.pt/80/n... 22 Março 2021	 → https://www.denik.cz/tipil... 8 Março 2022	 → http://www.abola.pt/jogod... 2 Outubro 2017	 → https://www.livefutbol.com/ 27 Março 2019	 → http://www.novagente.pt/ 25 Janeiro 2021	 → http://www.noticiasdeala... 22 Junho 2022

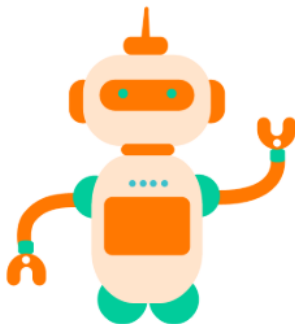
Search the **history** of a web address



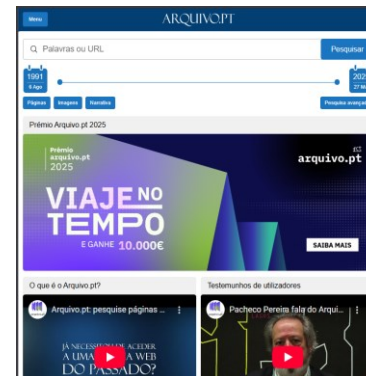
Too much data to search manually...



Solution: API!



Website



API

```
1 // 20250327112712
2 // https://arquivo.pt/textsearch?
  q=fccn%20collection:AMP2&from=19970806&to=20140107&maxItems=16
3
4 {
5   "serviceName": "Arquivo.pt - the Portuguese web-archive",
6   "linkToService": "https://arquivo.pt",
7   "estimated_nr_results": 0,
8   "request_parameters": {
9     "offset": 0,
10    "dedupValue": 2,
11    "from": "19970806000000",
12    "to": "20140107000000",
13    "dedupField": "title",
14    "q": "fccn collection:AMP2",
15    "maxItems": 16
16  },
17  "response_items": [
18  ]
19 }
```

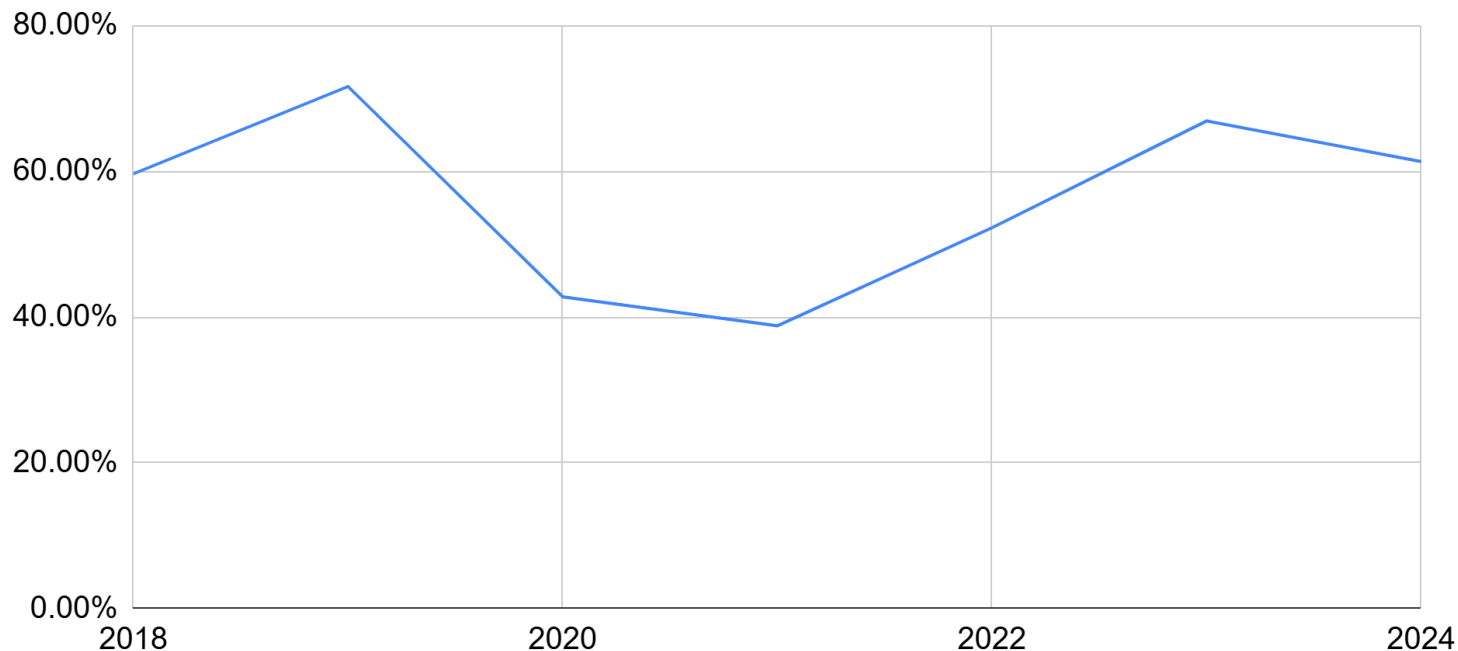

Arquivo.pt **Annual** Awards



Support of the President of the Republic of Portugal

➤ arquivo.pt/awards

Requests made to Arquivo.pt using APIs



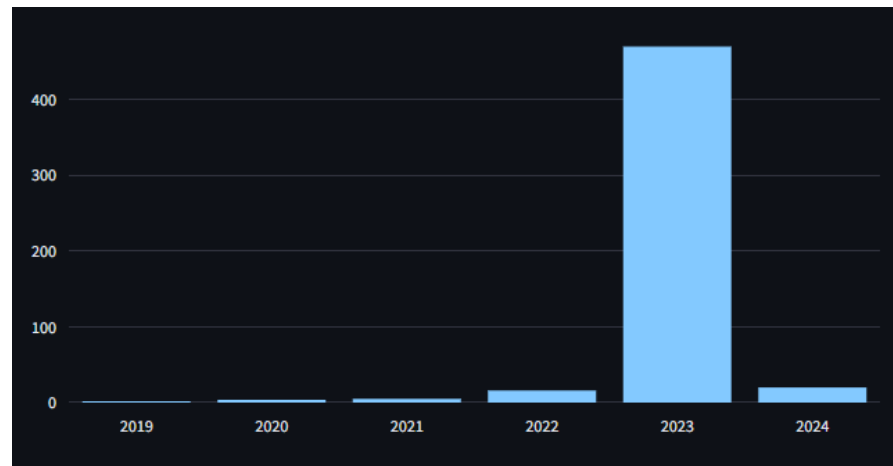
Arquivo.pt nowadays has 4 different APIs:

- [Arquivo.pt API \(text search\)](#)
- [Image Search API](#)
- [CDX-server API \(URL search\)](#)
- [Memento API \(URL search\)](#)

Bulk download of CDXJ

Index of /cdxj

 [ICO]	Name	Last modified	Size	Description
	[PARENTDIR] Parent Directory	-	-	-
	[] AWP1.cdxj	2023-02-03 13:01	18G	
	[] AWP2.cdxj	2023-02-03 13:04	20G	
	[] AWP3.cdxj	2023-02-03 13:07	21G	
	[] AWP4.cdxj	2023-02-03 13:25	30G	
	[] AWP5.cdxj	2023-02-03 14:08	49G	
	[] AWP6.cdxj	2023-02-03 14:44	52G	
	[] AWP7.cdxj	2023-02-03 14:23	39G	
	[] AWP8.cdxj	2022-12-04 02:12	36G	
	[] AWP9.cdxj	2023-02-03 14:27	40G	
	[] AWP10.cdxj	2023-02-03 13:36	34G	
	[] AWP11.cdxj	2023-02-03 13:29	32G	
	[] AWP12.cdxj	2023-02-03 14:00	45G	
	[] AWP15.cdxj	2023-02-03 14:38	65G	
	[] AWP16.cdxj	2023-02-03 15:59	104G	



arquivo.pt/cdxj

Glória –LLM for European Portuguese!

Model Checkpoint:



Source code:



A Large Language Model
for European Portuguese

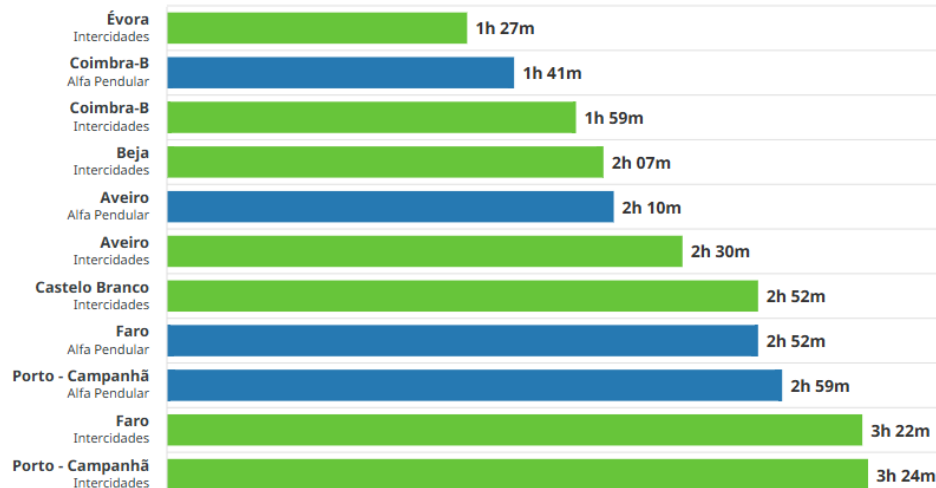


2023 Award winner: “Viajar no tempo sobre carris”

Viajar no tempo **sobre carris** em...



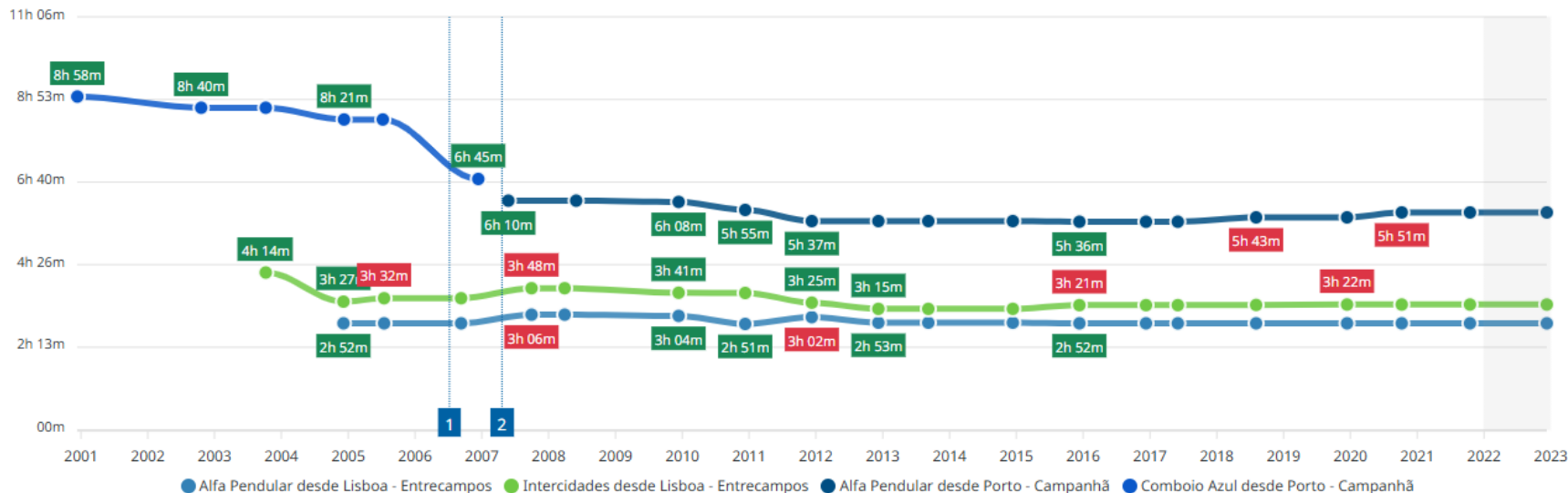
...desde **Lisboa**



viajarsobrecarris.anteropires.com

2023 Award winner: “Viajar no tempo sobre carris”

Tempo médio de viagens



1 09/07/2006 Comboio Azul passou a seguir por Lisboa e pela Linha do Sul em vez da Linha de Vendas Novas e da Linha do Alentejo

2 22/04/2007 Novo serviço Alfa Pendular entre Porto e Faro



viajarsobrecarris.anteropires.com

2023 Award 2nd prize: Existo.pt

E

Artistas Artes Noticias Contactos

É difícil não encontrar algumas destas mulheres! Estes nomes e estes rostos representam percursos de vida extraordinários, nem todas se mantêm entre nós, mas a marca da sua passagem vai-se manter muito para além da sua vida. Não têm idade, porque a idade é apenas uma forma de medir a sua evolução e a transformação que provocaram. Como uma equipa, estão representadas em todas as artes e em todas elas é visível a sua marca.

Quanto destas vidas está visível para o mundo?

Em cada artista vais encontrar algumas das marcas que deixaram como equipa ou a simplesmente ser... existir.

Subjects +

All
Música
Cinema
Teatro
Dança
Literatura
Artes
Plásticas
Arquitectura
Design

ABCDEF GHIJ KLMNOPQRST UVXYZ


Rossana Torres


Catarina Alves Costa


Crista Alfaiate


Surma


Maria Alzira Seixo


Alexandra Moura


Dora Rodrigues


Sofia Dinger









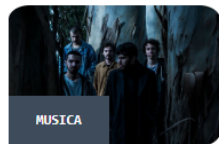
existo.pt

2023 Award 2nd prize: Existo.pt

E

[Artistas](#) [Artes](#) [Noticias](#) [Contactos](#)


Surma

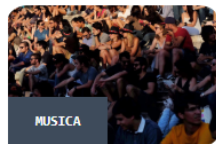
[Voltar](#)


sons-sopradados-de-leiria-first-breath-after-coma-e-twin-transistors-em-preescuta

2016

Nascida em 2012, em Leiria, a Omnichord Records tem-se revelado uma

artisticamente editada

[Read Article](#)


depois-do-calor-e-antes-dos-national-houve-kurt-vile-no-sbsr

2016

Mandou o calor nas primeiras horas do Super Bock Super Rock (SBSR), no Parque das Nações, em Lisboa. Em

[Read Article](#)


festival-nos-alive-em-alges-esta-esgotado

2018

O festival de música Nos Alive, marcado para Julho no Passeio Marítimo de Algés, está esgotado.

[Read Article](#)


dbandada-com-bonga-club-kitten-cinema-e-concertos-num

2016

Os concertos espalham-se por ruas, praças, jardins, bares e outros locais da Baixa da cidade, festa em

[Read Article](#)


preparemonos-que-ha-milhoes-de-festa-mesmo-a-chegar

2016

O Milhões de Festa em Barcelos, isto é. Cenário de festa, como o nome indica, festa

[Read Article](#)


musica-portuguesa-estara-em-destaque-na-austria

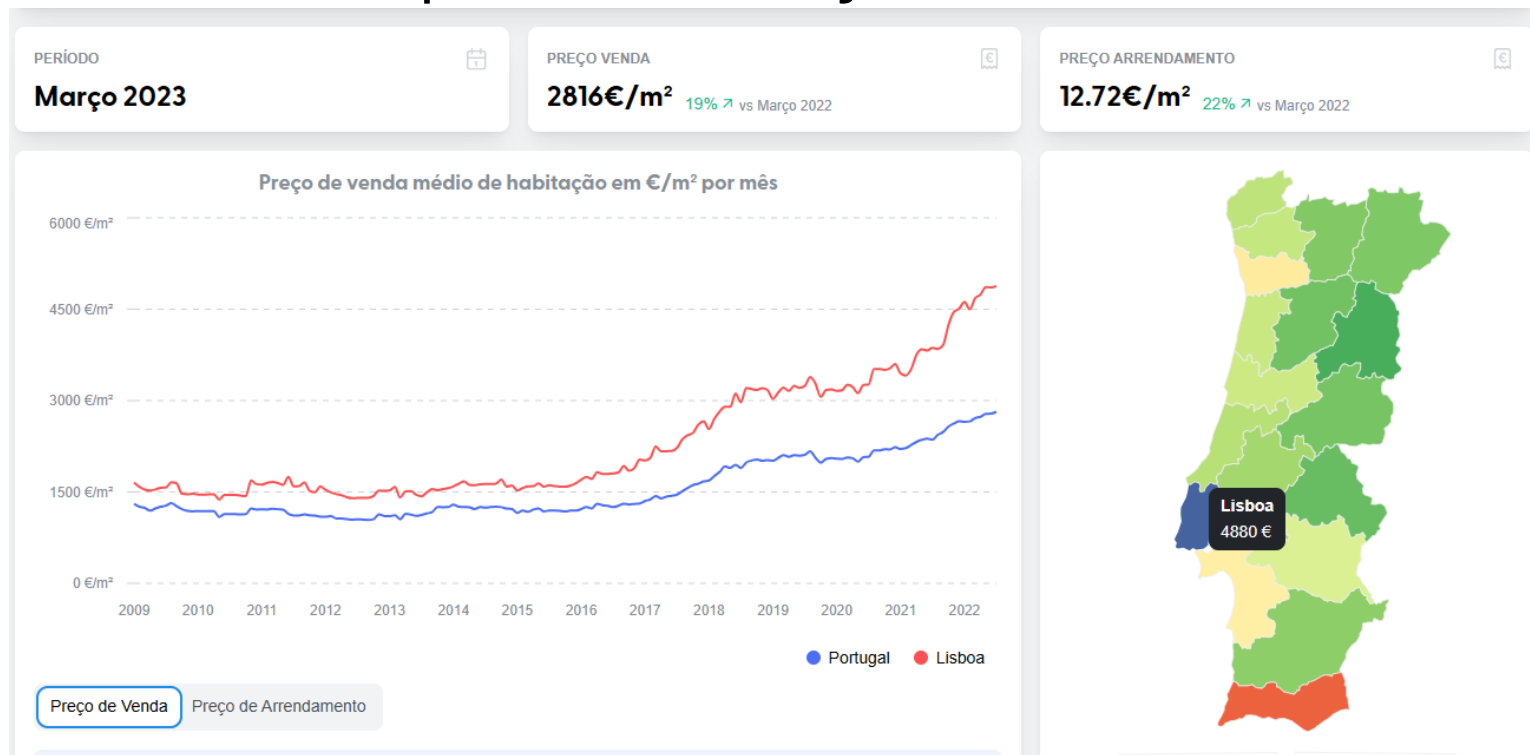
2018

O festival Waves Vienna, que se realiza de 27 a 29 de Setembro na capital austríaca, terá Portugal como um

[Read Article](#)

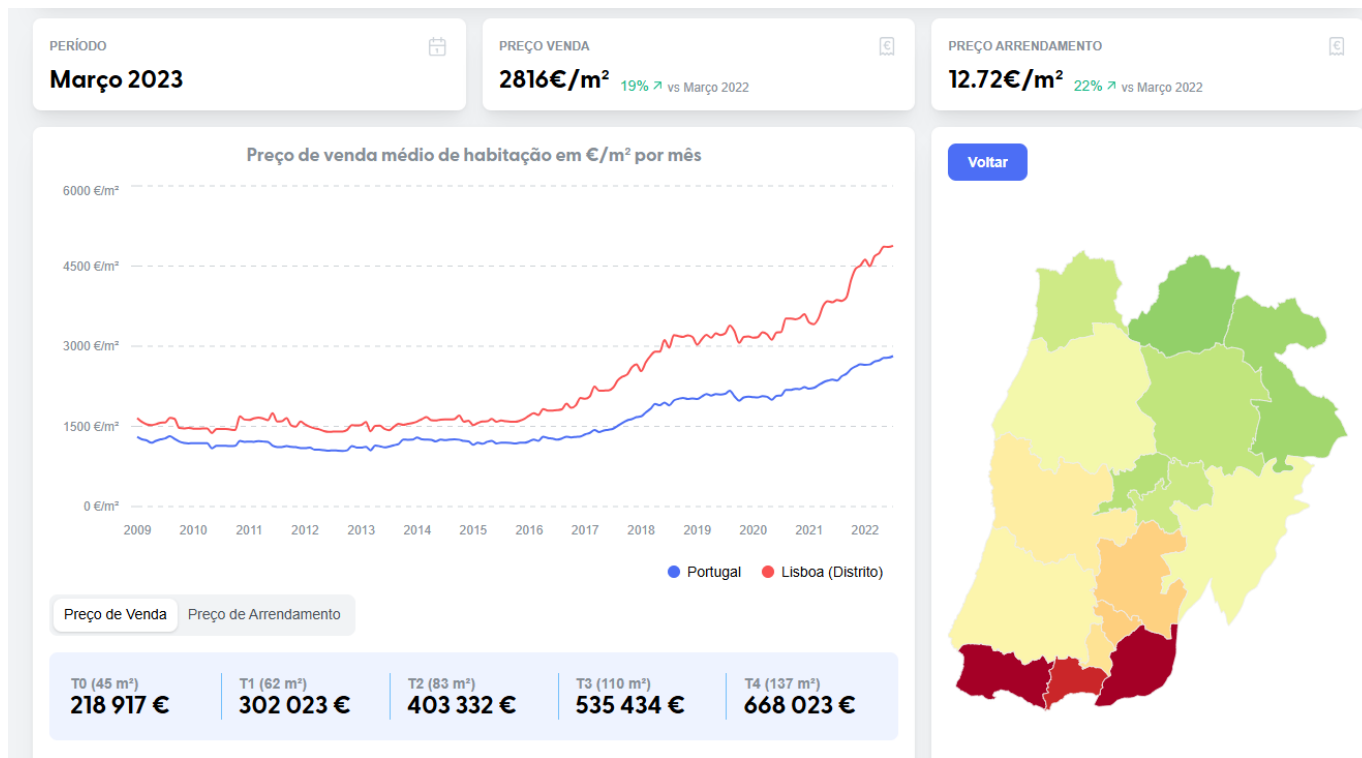
existo.pt

2024 Award 2nd prize: Habitação.NET



habitacao.net

2024 Award 2nd prize: Habitação.NET



habitacao.net

Questions?
arquivo.pt/api

contacto@arquivo.pt