

# Arquivo.pt Query Logs

Pedro Gomes, Arquivo.pt, [pedro.gomes@fccn.pt](mailto:pedro.gomes@fccn.pt)



# Arquivo.pt preserves national and international historical web content

***GALILEO*** Versão 2.0

Janeiro 1993

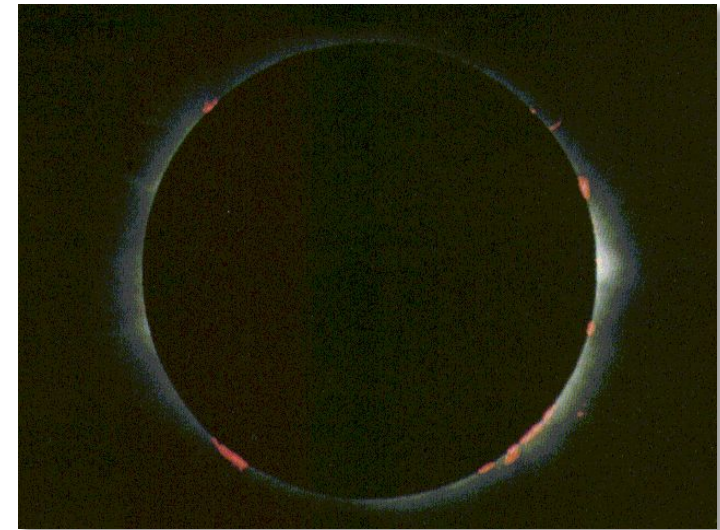
Manual de utilização e sugestões de exploração	<b><i>AUTORES</i></b> <b>Programação:</b> João Veloso <b>Manual:</b> Elisa Prata Pina e M. Augusta Patrício <b>Orientação:</b> Carlos Fiolhais
--	---

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](http://nautilus.fis.uc.pt) - 1993  
(oldest page)



[spacelink.nasa.gov](http://spacelink.nasa.gov) - 1992  
(oldest image)

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



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

The screenshot displays the ARQUIVO.PT search interface. At the top, there is a dark blue header with 'Menu' on the left, 'ARQUIVO.PT' in the center, and 'Options' on the right. Below the header is a search bar containing the text 'Ambasada RP' and a blue 'Search' button. A date range slider is positioned below the search bar, with '1996 1 Jan' on the left and '2021 12 Nov' on the right. Below the slider are three buttons: 'Pages', 'Images', and 'Narrative', and an 'Advanced search' button on the right. The main content area shows search results. The first result is for 'Ambasada RP w Berlinie' with a date of '30 March 2019'. It includes a link to 'berlin.msz.gov.pl/pl' and a snippet of text: 'Ambasada RP w Berlinie Masz wyłączoną obsługę języka JavaScript Polski Deutsch RSS Wersja tekstowa Być wiernym Ojczyźnie mej, Rzeczypospolitej Polskiej Aktualności Ambasada Współpraca dwustronna ... się zwrócić? Zasady realizacji projektów polonijnych Praktyki Ambasada Ambasada Ambasada RP w ...'. The second result is for 'Ambasada RP w Dublinie' with a date of '10 June 2017'. It includes a link to 'dublin.msz.gov.pl/pl' and a snippet of text: 'Ambasada RP w Dublinie Masz wyłączoną obsługę języka JavaScript English Polski RSS Wersja tekstowa Polsce – służyć, Europę – tworzyć, Świat – rozumieć Aktualności Ambasada Współpraca dwustronna ... WYDZIAŁEM KONSULARNYM Ambasada Ambasada RP w Dublinie 5 Ailesbury Road Ballsbridge D04 W221 Dublin ...'.

# Search the **history** of a web address

Menu ARQUIVO.PT

Q bbc.co.uk Search

1996 1 Jan 2021 12 Nov

Pages Images Narrative Advanced search

Cerca de 2,538 resultados desde 1996 a 2021

Table List

2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
15 Feb	4 May	29 May	4 Jan	21 Jan	10 Jan	5 Sep	1 Jan	1 Jan	20 Jan	1 Jan	1 Jan	1 Jan	3 Jan
14 Mar	20 May	30 May	6 Jan	27 Jan	11 Jan	7 Sep	2 Jan	2 Jan	9 Feb	1 Jan	2 Jan	1 Jan	3 Jan
22 Oct	20 May	4 Aug	12 Jan	31 Jan	12 Jan	23 Sep	3 Jan	4 Jan	10 Feb	2 Jan	3 Jan	2 Jan	3 Jan
22 Oct	11 Jun		12 Jan	11 Mar	15 Jan	24 Sep	4 Jan	8 Jan	3 Apr	2 Jan	4 Jan	3 Jan	4 Jan
23 Oct	22 Jun		14 Jan	12 Mar	17 Jan	26 Sep	5 Jan	8 Jan	4 Apr	2 Jan	5 Jan	4 Jan	4 Jan
	6 Aug		21 Jan	13 Mar	18 Jan	27 Sep	6 Jan	17 Jan	5 Apr	3 Jan	6 Jan	5 Jan	4 Jan
	6 Aug		21 May	14 Mar	19 Jan	28 Sep	8 Jan	17 Jan	6 Apr	3 Jan	7 Jan	6 Jan	6 Jan
	3 Sep		1 Jul	15 Mar	22 Jan	30 Sep	10 Jan	21 Jan	7 Apr	3 Jan	8 Jan	7 Jan	6 Jan
	3 Sep		1 Dez	16 Mar	23 Jan	4 Oct	10 Jan	24 Jan	8 Apr	3 Jan	9 Jan	8 Jan	6 Jan
	4 Sep		2 Dez	17 Mar	26 Jan	5 Oct	12 Jan	28 Jan	9 Apr	3 Jan	10 Jan	9 Jan	6 Jan
	8 Sep			18 Mar	29 Jan	6 Oct	17 Jan	30 Jan	10 Apr	3 Jan	11 Jan	10 Jan	8 Jan

# Browse [bbc.co.uk](http://bbc.co.uk) 10 years ago

The screenshot shows the BBC website's archive interface. At the top, it says 'ARQUIVO.PT' and 'bbc.co.uk 1 July at 07:43, 2011'. The main navigation bar includes 'BBC Mobile', 'News', 'Sport', 'Weather', 'Travel', 'TV', 'Radio', 'More', and a search box. A left sidebar contains a 'Table' view and a date selector for the year 2011, with 'July' selected. The main content area is divided into several sections:

- Top News story:** A large article titled "'Doubts' over Strauss-Kahn case" with a photo of Dominique Strauss-Kahn. The text reads: "The sex assault case against Dominique Strauss-Kahn appears to be in trouble amid reported concerns about the alleged victim's credibility."
- Spotlight:** A section titled "One in a Billion" featuring a photo of an elderly woman. Below it are three sub-articles: "China's haves and have-nots", "Building dreams: Dai Zhikang", "From hawk to billionaire: Zhou Xiaoguang", and "China's top art collector: Liu Yiqian".
- News:** A section with three smaller articles:
  - "Moroccans vote on king's reforms" with a photo of a crowd.
  - "US in CIA detainee death inquiry" with a photo of a man speaking.
  - "Poland to take over EU presidency" with a photo of a crowd.
- Most Popular in News:** A list of popular stories with a 'Watched/Listened' indicator:
  - "Hu warns communists on corruption" with a photo of a crowd and a red flag with a yellow hammer and sickle.
  - "Chavez tells of cancerous tumour" with a photo of Hugo Chavez.
  - "Crowds welcome royals in Canada" with a photo of a crowd.
  - "Japan manufacturers 'pessimistic'" with a photo of a crowd.

# Search **images** from the past

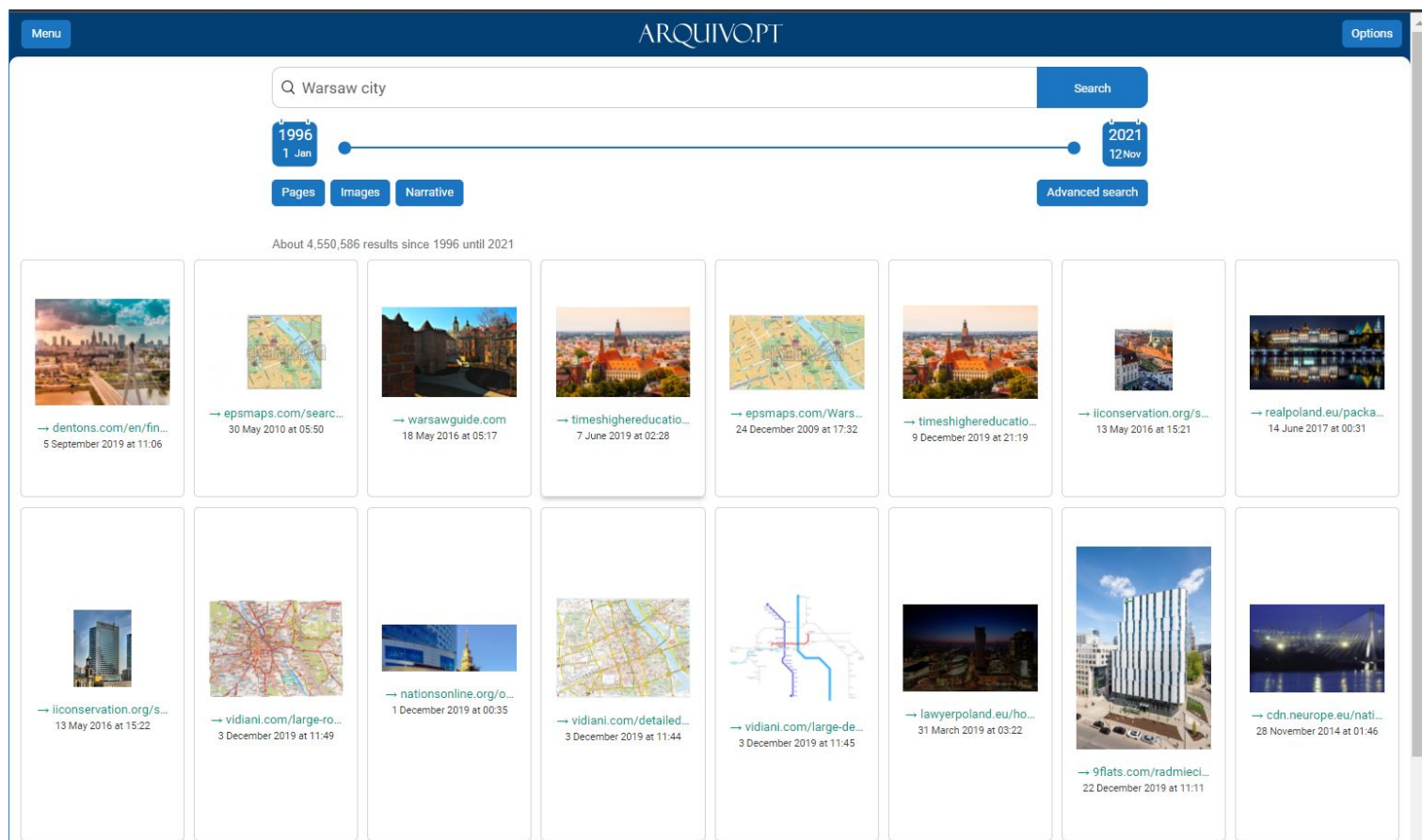
Menu ARQUIVO.PT Options

Q Warsaw city Search

1996 1 Jan 2021 12 Nov


Pages Images Narrative Advanced search

About 4,550,586 results since 1996 until 2021



Source	Capture Date
dentons.com/en/fin...	5 September 2019 at 11:06
epsmaps.com/searc...	30 May 2010 at 05:50
warsawguide.com	18 May 2016 at 05:17
timeshighereducatio...	7 June 2019 at 02:28
epsmaps.com/Wars...	24 December 2009 at 17:32
timeshighereducatio...	9 December 2019 at 21:19
iiconservation.org/s...	13 May 2016 at 15:21
realpoland.eu/packa...	14 June 2017 at 00:31
iiconservation.org/s...	13 May 2016 at 15:22
vidiani.com/large-ro...	3 December 2019 at 11:49
nationsonline.org/o...	1 December 2019 at 00:35
vidiani.com/detailed...	3 December 2019 at 11:44
vidiani.com/large-de...	3 December 2019 at 11:45
lawyerpoland.eu/ho...	31 March 2019 at 03:22
9flats.com/radmieci...	22 December 2019 at 11:11
cdn.neurope.eu/nati...	28 November 2014 at 01:46

ARQUIVO.PT



Visit Details

**Image**

**Title:** Warsaw city centre

**Alt text:** Warsaw city centre

**Caption:** Poland's education system has seen some extraordinary changes in recent years. Following a huge drive in the nation's Katowice =6 1001+ Wroclaw University of Science and Technology Wroclaw Read mor...

**URL:** timeshighereducation.com/sites/default/files/styles/the\_breaking\_news\_image\_style/public/warsaw\_city\_centre.jpg?tok=NYFWGvNj

**Resolution:** jpeg 620 x 413

**Capture date:** 7 June 2019 at 02:28

**Page**

**Title:** Best universities in Poland | Times Higher Education (THE)

**URL:** timeshighereducation.com/student/best-universities/best-universities-poland

**Capture date:** 7 June 2019 at 02:28



VS





# Analyzing user behavior

Analyzing user behavior is an important research topic to understand users' information needs and enhance the quality of search results.

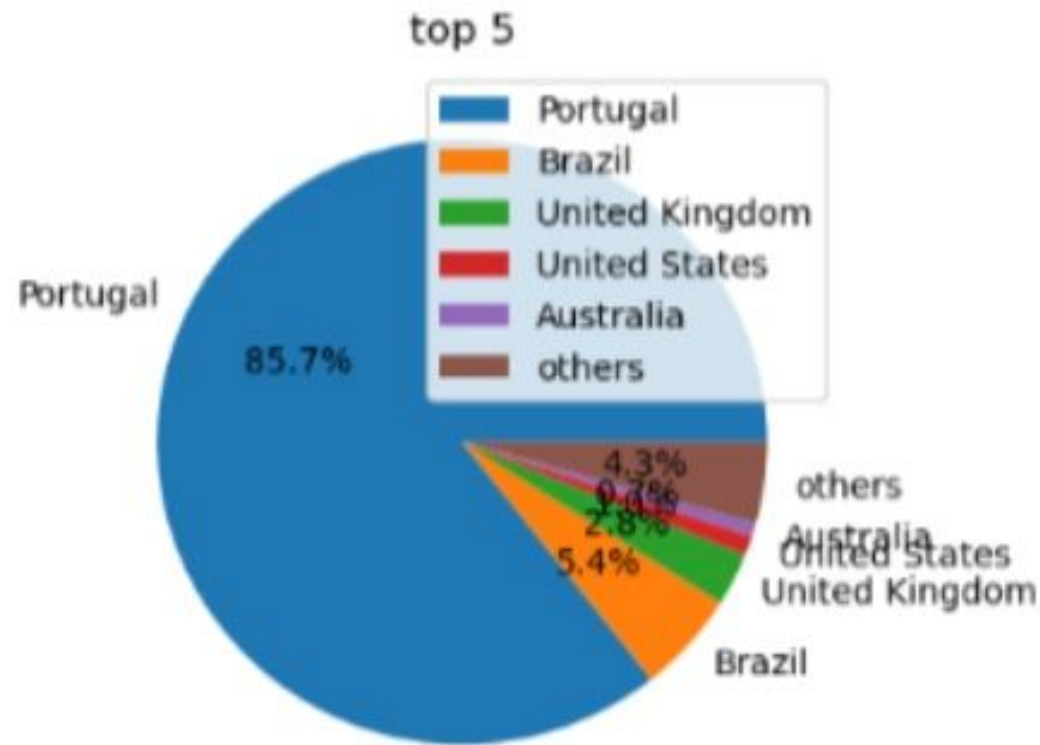


# Query Log Dataset

- 3 months from June to September 2021

	A	B	C	D	E	F	G	H	I	J	K
1	REQUEST	TIMESTAMP	YEAR	TYPE_SEARCH	QUERY	PAGE	POSITION	COUNTRY	CITY	TYPE_DEVICE	BROWSER_FAMILY
2	GET /textsearch?q=NDA%20CREW%:	1625359337	2021	textsearch	NDA CREW DHEO	0	1	Portugal	Lisbon	Browser	Chrome
3	GET /textsearch?q=FYND%20GRAFF	1625359561	2021	textsearch	FYND GRAFFITI	0		Portugal	Lisbon	Browser	Chrome
4	GET /imagesearch?q=FYND%20GRAI	1625359565	2021	imagesearch	FYND GRAFFITI	0		Portugal	Lisbon	Browser	Chrome
5	GET /textsearch?q=FYND%20GRAFF	1625359571	2021	textsearch	FYND GRAFFITI	0	1	Portugal	Lisbon	Browser	Chrome
6	GET /textsearch?q=FYND%20GRAFF	1625359607	2021	textsearch	FYND GRAFFITI	0	4	Portugal	Lisbon	Browser	Chrome
7	GET /textsearch?q=escola%20secund	1625364834	2021	textsearch	escola secundaria cascais	0	1	Portugal	Lisbon	Browser	Chrome
8	GET /textsearch?q=escola%20secund	1625365794	2021	textsearch	escola secundaria cascais	0		Portugal	Lisbon	Browser	Chrome
9	GET /textsearch?q=escola%20secund	1625365813	2021	textsearch	escola secundaria inês castro	0		Portugal	Lisbon	Browser	Chrome
10	GET /textsearch?q=escola%20secund	1625365828	2021	textsearch	escola secundaria inês castro	1		Portugal	Lisbon	Browser	Chrome
11	GET /textsearch?q=esic&from=199601	1625366051	2021	textsearch	esic	0	1	Portugal	Lisbon	Browser	Chrome
12	GET /textsearch?q=d.%20joao%20ii&f	1625366200	2021	textsearch	d. joao ii	0	1	Portugal	Lisbon	Browser	Chrome
13	GET /textsearch?q=d%20joao%20ii&fr	1625366280	2021	textsearch	d joao ii	0	1	Portugal	Lisbon	Browser	Chrome
14	GET /textsearch?q=d%20joao%20ii&fr	1625366322	2021	textsearch	d joao ii	0		Portugal	Lisbon	Browser	Chrome
15	GET /textsearch?q=d%20joao%20ii%2	1625366331	2021	textsearch	d joao ii setubal	0		Portugal	Lisbon	Browser	Chrome
16	GET /textsearch?q=d%20joao%20ii%2	1625366340	2021	textsearch	d joao ii setubal 1983	0		Portugal	Lisbon	Browser	Chrome
17	GET /textsearch?q=d%20jo%C3%A3o	1625366345	2021	textsearch	d joão ii setúbal 1983	0		Portugal	Lisbon	Browser	Chrome
18	GET /textsearch?q=escola%20cavaqu	1625366811	2021	textsearch	escola cavaquinhas	0		Portugal	Lisbon	Browser	Chrome
19	GET /textsearch?q=escola%20dr.%20j	1625366838	2021	textsearch	escola dr. josé afonso	0		Portugal	Lisbon	Browser	Chrome
20	GET /textsearch?q=escola%20bocage	1625368558	2021	textsearch	escola bocage	0	8	Portugal	Lisbon	Browser	Chrome

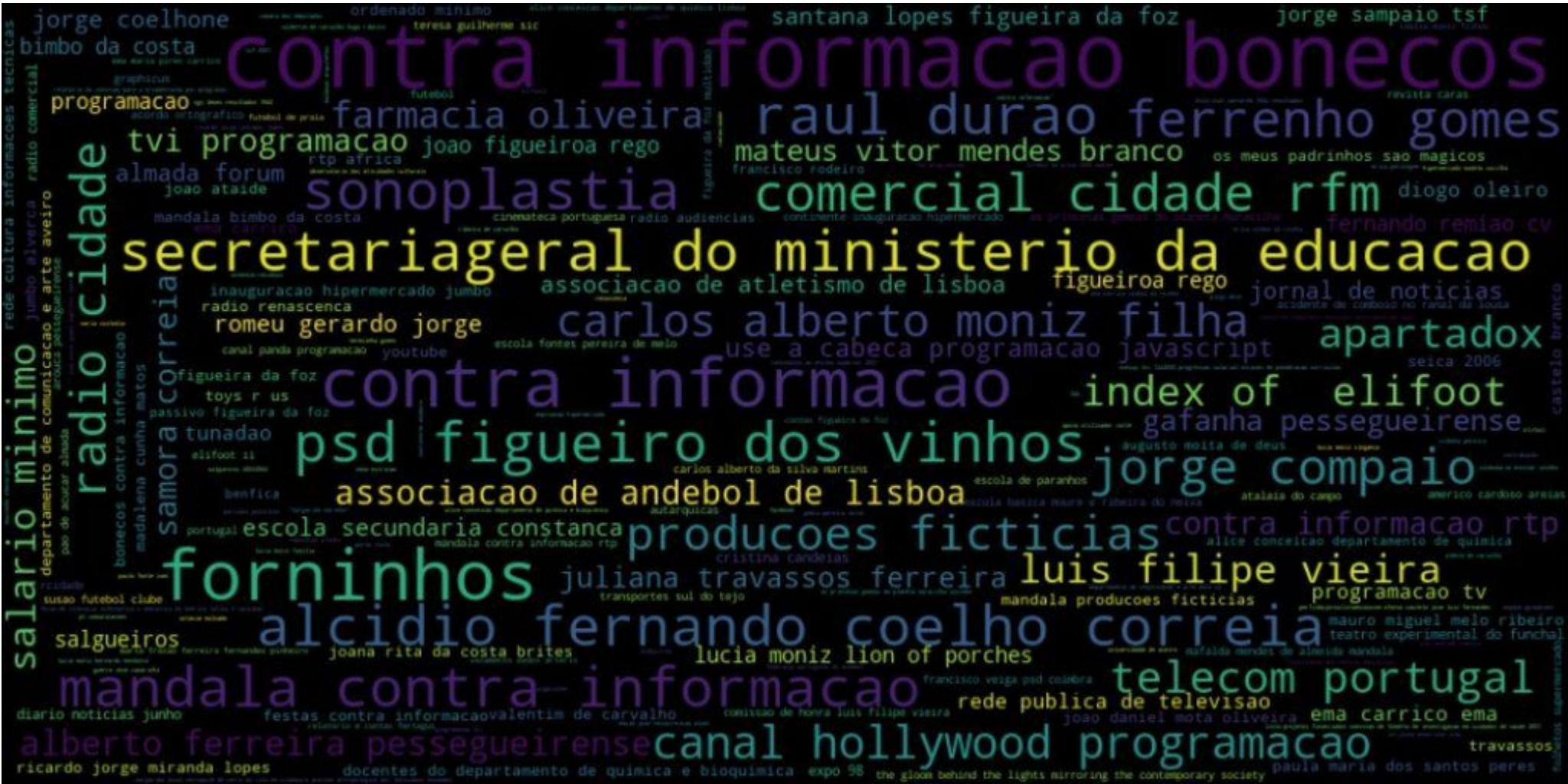
# 85.7% of the queries came from Portugal



# Only 37% of the queries are Portuguese

Portuguese	37.06%
Spanish	11.74%
Italian	7.68%
English	7.59%
Catalan	3.99%
Romanian	2.85%

# Word cloud of the most frequent query terms submitted to Arquivo.pt



# User tend to use longer/multiple queries

- Users of Arquivo.pt tend to use longer queries.
- Users tend to issue multiple queries within a session rather than a single query.



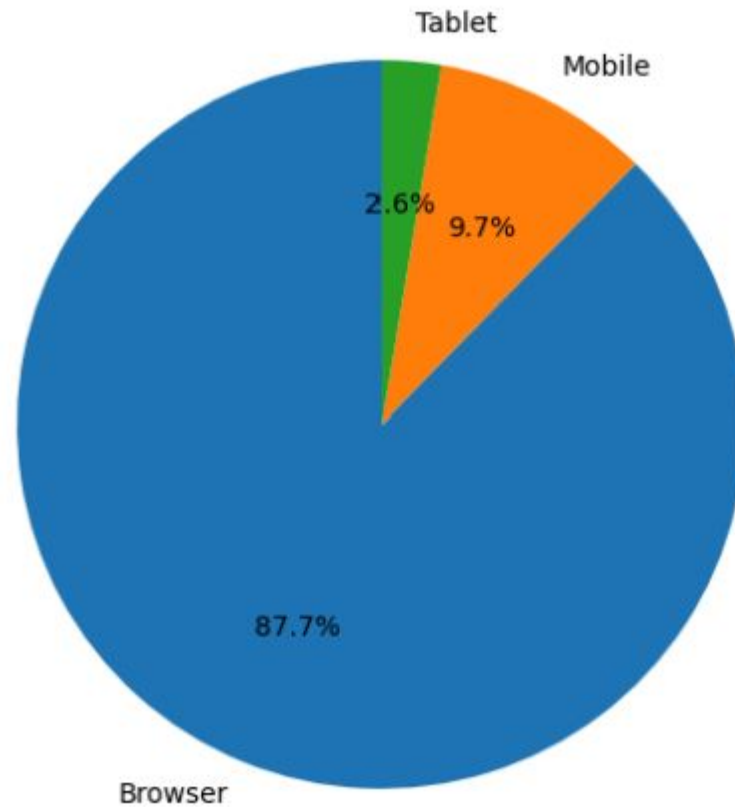
# Users tend to refine more the time span

- Users refined the time span of the search (using the datepickers) in about 50% of queries.
- Only a small percentage of users included specific years in their query terms (4%).



# Type of devices

87.7% of queries submitted to Arquivo.pt were made using Desktop Browsers.





# Thank You

[arquivo.pt/datasets/querylogs](https://arquivo.pt/datasets/querylogs)

[contacto@arquivo.pt](mailto:contacto@arquivo.pt)

