

Developing computational skills for digital collections, Thursday, 5th of May, 2022



Timeline summarization for large-scale past-web events with
Python: the case of Arquivo.pt

Daniel Gomes and Ricardo Campos
daniel.gomes@fccn.pt, ricardo.campos@ipt.pt



80%

of the web pages disappear
or change their content
after only 1 year.

What is Arquivo.pt?



Free online service to research the Past Web

Preserves **publicly accessible** information related with:

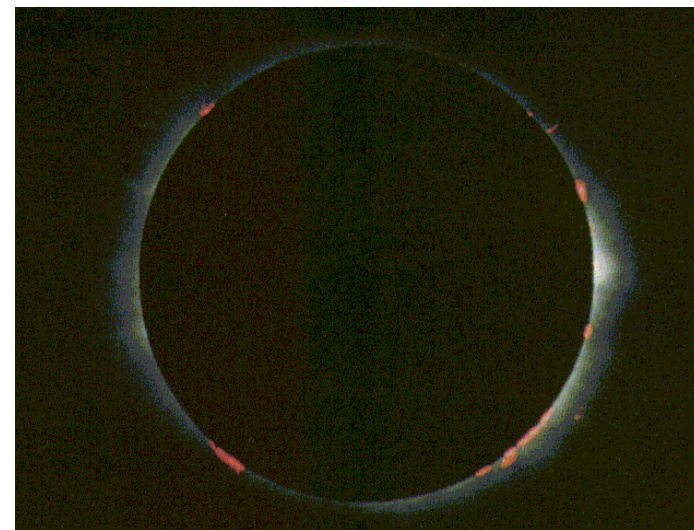
- Portugal
- **Research** and **Education** (international)

Preserves historical information published online
since the 1990s

Arquivo.pt preserves national and international historical web content

<i>GALILEO</i> Versão 2.0	
Janeiro 1993	
Manual de utilização e sugestões de exploração	<i>AUTORES</i> Programação: João Veloso Manual: Elisa Prata Pina e M. Augusta Patrício Orientação: Carlos Fiolhais
Departamento de Física da Universidade de Coimbra	
<i>ÍNDICE</i>	
1. Apresentação sumária do programa 2. Características do equipamento 3. Configuração mínima do computador 4. Ficheiros do programa	





















nautilus.fis.uc.pt - 1993
(oldest page)



spacelink.nasa.gov – 1992
(oldest image)

Arquivo.pt is used word-wide



Country	Users	% Users
1.  Portugal	46,891	 46.56%
2.  United States	26,373	 26.19%
3.  Brazil	2,266	 2.25%
4.  Russia	2,234	 2.22%
5.  United Kingdom	2,231	 2.22%
6.  Japan	2,172	 2.16%
7.  Canada	1,237	 1.23%
8.  Mozambique	1,213	 1.20%
9.  India	902	 0.90%
10.  Germany	894	 0.89%

53% of users are international

User Interfaces and documentation also in English

Combined with Google Translate enables **cross-lingual access** to historical information from the Past Web

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



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

[Menu](#) ARQUIVO.PT [Options](#)

[Search](#)

1996
1 Jan

2021
12 Nov

[Pages](#) [Images](#) [Narrative](#) [Advanced search](#)

About 23,322 results since 1996 until 2021

[→ berlin.msz.gov.pl/pl](#)

Ambasada RP w Berlinie
30 March 2019
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 ...

[→ dublin.msz.gov.pl/pl](#)

Ambasada RP w Dublinie
10 June 2017
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 10 years ago

Menu

Table

2008

February

15 February at 02:28

March

October

2009

2010

2011

January

May

July

1 July at 07:43

December

2012

2013

2014

2015

2016

ARQUIVO.PT

bbc.co.uk 1 July at 07:43, 2011

BBC Mobile


News Sport Weather Travel TV Radio More

Search the BBC


Top News story

'Doubts' over Strauss-Kahn case

The sex assault case against Dominique Strauss-Kahn appears to be in trouble amid reported concerns about the alleged victim's credibility.




News



Hu warns communists on corruption

President Hu Jintao warns China's communists that corruption could cost them the support of the people, as the party marks its 90th anniversary.



Chavez tells of cancerous tumour

Venezuelan President Hugo Chavez says he has had a second operation to remove a cancerous tumour, in his first televised speech since flying to Cuba for treatment.

Moroccans vote on king's reforms

US in CIA detainee death inquiry


Poland to take over EU presidency

Crowds welcome royals in Canada

Japan manufacturers 'pessimistic'

Spotlight

One in a Billion



China's haves and have-nots

It is officially a Communist country - but economic data points to a stark gap between rich and poor.

Building dreams: Dai Zhikang

From hawk to billionaire: Zhou Xiaoguang

China's top art collector: Liu Yiqian

Most Popular in News

Shared Read Watched/Listened

Meet the world's most powerful computer

Monaco Prince Albert to wed Olympic star

China celebrates longest sea bridge


Search **images** from the past

Menu
ARQUIVO.PT
Options


1996
1 Jan

2021
12 Nov


About 4,550,586 results since 1996 until 2021




→ [dentons.com/en/fin...](#)
5 September 2019 at 11:06




→ [epsmaps.com/search...](#)
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




→ [ilconservation.org/s...](#)
13 May 2016 at 15:21




→ [realpoland.eu/packa...](#)
14 June 2017 at 00:31




→ [ilconservation.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



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+ Wrocław University of Science and Technology Wrocław 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

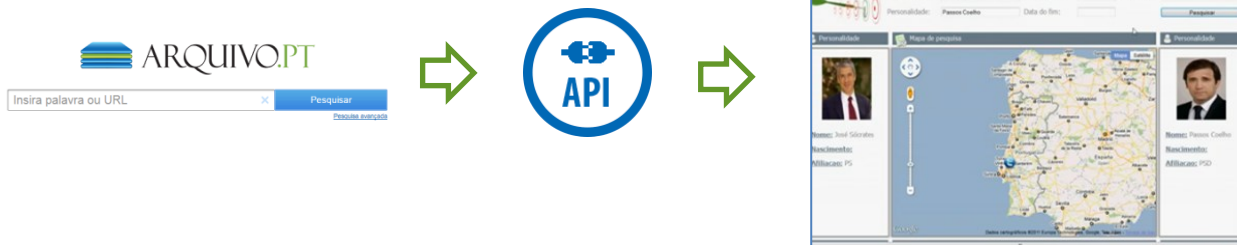
Title:
[Best universities in Poland | Times Higher Education \(THE\)](#)

URL: [timeshighereducation.com/student/best-universities/best-universities-poland](#)

Capture date: 7 June 2019 at 02:28

Application Programming Interfaces to develop new applications

Your new app



Arquivo.pt APIs

APIs

Daniel Gomes edited this page 11 days ago · 33 revisions

APIs specific to Arquivo.pt that enable the full exploration of its functions

- [Arquivo.pt API \(Full-text & URL search\)](#)
- [Image Search API v1.1 \(beta version\)](#)

APIs based on international standards to enable interoperability among web archives and code reuse

- [CDX-server API \(URL search\)](#)
- [Memento API \(URL search\)](#)

arquivo.pt/api

APIs requests and responses are detailed on GitHub

Arquivo.pt API

Daniel Gomes edited this page on 17 Dec 2021 · 11 revisions

The Arquivo.pt API allows full-text search and access preserved web content and related metadata. It is also possible to search by URL, accessing all versions of preserved web content.

API returns a **JSON** object.

EndPoint: <https://arquivo.pt/textsearch>

Note: a text search query returns a **maximum of 2000 response results**

Request Parameters

Search for terms

Parameter Name	Description	Examples
q	Query search terms. The query can contain advanced search operators. Advanced search operators can be: <ul style="list-style-type: none">" ": search for items containing expression composed by search terms (e.g. phrase or named entity).- : excludes items that contains the given terms.	<code>q=James Davis</code> <code>q="Antonio Costa"</code> <code>q=Albert -Einstein</code>
from	Set an initial date for the time span of the search. Format: YYYYMMDDHHMMSS, also accepts a shorter date fotmat, e.g. (YYYY). Default: 1996	<code>from=19960101000000</code>
to	Set a end date for the time span of the search. Format: YYYYMMDDHHMMSS, also accepts a shorter date format, for example (YYYY). Default: Current Year-1	<code>to=20151022163016</code>

Response fields

Type of Search	Field Name	Description	Examples
Full-text, URL, Metadata	serviceName	Service Name.	"serviceName": "Arquivo.pt - the Portuguese web-archive"
Full-text, URL, Metadata	linkToService	URL for the service, in the case arquivo.pt	"linkToService": " https://arquivo.pt "
Full-text, URL, Metadata	request_parameters	field with all request parameters.	"request_parameters": { "q": "Albert Einstein", "offset": 10, "limit": 5, "itemsPerSite": 2, "prettyPrint": "true" }
Full-text, URL	next_page	URL for the next N items. N = offset + limit. (offset=0 and limit=50). Default=50	"next_page": " http://arquivo.pt/textsearch?q=hello%20world&offset=50 "
Full-text, URL	previous_page	URL for previous N items. N = offset - limit. (offset=0 and limit=50). Default=0.	"previous_page": " http://arquivo.pt/textsearch?q=hello%20world&offset=0 "

Usage examples

Examples Requests

Example Request - Full-text Search

Full-text Search for the terms Albert and Einstein, with a limit of 5 results, with the offset to 10. Finally, returns the results with formatting.

<https://arquivo.pt/textsearch?q=Albert%20Einstein&maxItems=5&prettyPrint=true>

Example Response - Full-text Search

```
{
  "serviceName": "Arquivo.pt - the Portuguese web-archive",
  "linkToService": "https://arquivo.pt",
  "next_page": "https://arquivo.pt/textsearch?q=Albert%20Einstein&maxItems=5&prettyPrint=true&offset=5",
  "previous_page": "https://arquivo.pt/textsearch?q=Albert%20Einstein&maxItems=5&prettyPrint=true&offset=0",
  "request_parameters": {
    "q": "Albert Einstein",
    "maxItems": "5",
    "prettyPrint": "true"
  },
  "response_items": [
    {
      "title": "Albert Einstein ? Wikipédia, a enciclopédia livre",
      "originalURL": "http://pt.wikipedia.org/wiki/Albert_Einstein",
      "linkToArchive": "https://arquivo.pt/wayback/20150408223215/http://pt.wikipedia.org/wiki/Albert_Einstein",
      "tstamp": "20150408223215",
      "contentLength": "513829",
      "digest": "3bf5a50a6ec8adddc5225569d360123b",
      "mimeType": "text/html",
      "linkToScreenshot": "https://arquivo.pt/screenshot?url=https%3A%2F%2Farquivo.pt%2FnoFrame%2Fpreplay%2F20150408223215",
      "date": "1428532335",
      "encoding": "UTF-8",
      "linkToNoFrame": "https://arquivo.pt/noFrame/replay/20150408223215/http://pt.wikipedia.org/wiki/Albert_Einstein",
      "linkToOriginalFile": "https://arquivo.pt/noFrame/replay/20150408223215id_/http://pt.wikipedia.org/wiki/Albert_Einstein",
      "snippet": "<em>Albert</em> <em>Einstein</em> &#8211; Wikipédia, a enciclopédia livre <em>Albert</em>",
      "collection": "AWP17",
      "linkToExtractedText": "https://arquivo.pt/textextracted?m=http%3A%2F%2Fpt.wikipedia.org%2Fwiki%2FAlbert_Einstein",
      "linkToMetadata": "https://arquivo.pt/textsearch?metadata=http%3A%2F%2Fpt.wikipedia.org%2Fwiki%2FAlbert_Einstein"
    }
  ]
}
```

Example Request - URL Search

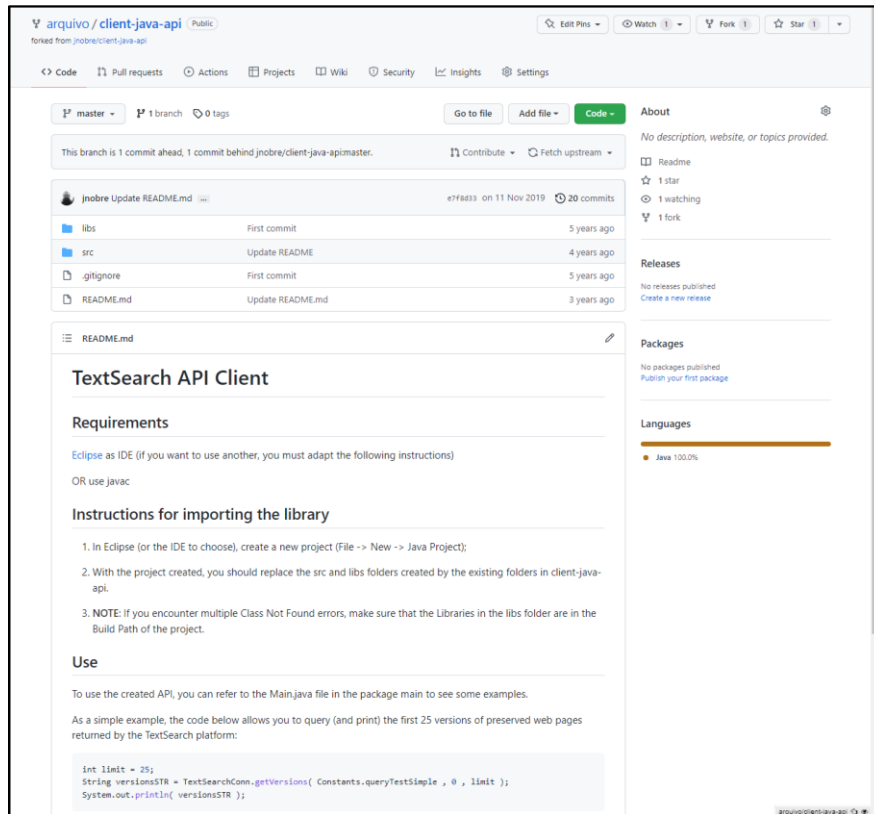
URL Search for the *expresso.pt* versions, with a limit of 5 results. Finally, returns the results with formatting.

<https://arquivo.pt/textsearch?prettyPrint=true&versionHistory=expresso.pt&maxItems=5>

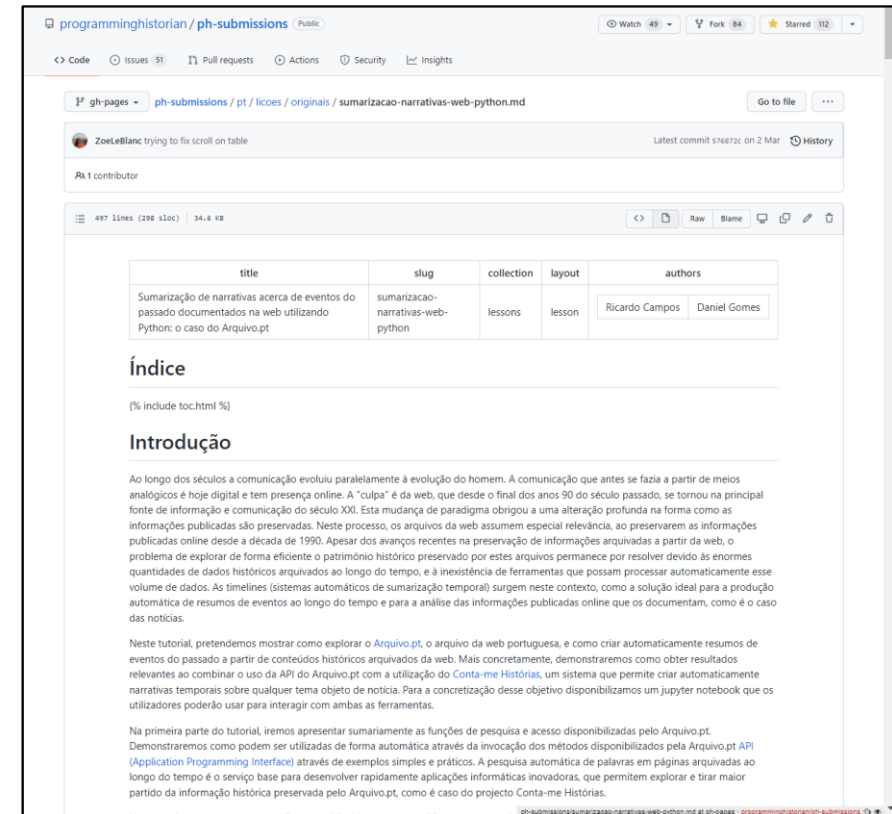
Example Response - URL Search

```
{
  "serviceName": "Arquivo.pt - the Portuguese web-archive",
  "linkToService": "https://arquivo.pt",
  "next_page": "https://arquivo.pt/textsearch?prettyPrint=true&versionHistory=expresso.pt&maxItems=5&offset=5",
  "previous_page": "https://arquivo.pt/textsearch?prettyPrint=true&versionHistory=expresso.pt&maxItems=5&offset=0",
  "request_parameters": {
    "maxItems": "5",
    "from": "19960101000000",
    "to": "20181231235959",
    "prettyPrint": "true"
  },
  "response_items": [
    {
      "title": "302 Found",
      "originalURL": "http://www.expresso.pt/",
      "linkToArchive": "https://arquivo.pt/wayback/20160719170305/http://www.expresso.pt/",
      "tstamp": "20160719170305",
      "contentLength": "318",
      "digest": "POPLDMMERZ2PUX2BRPCFNS2BB6ELQHMY",
      "mimeType": "text/html",
      "linkToScreenshot": "https://arquivo.pt/screenshot?url=https%3A%2F%2Farquivo.pt%2FnoFrame%2Fpreplay%2F20160719170305",
      "date": "1468947785",
      "encoding": "iso-8859-1",
      "linkToNoFrame": "https://arquivo.pt/noFrame/replay/20160719170305/http://www.expresso.pt/",
      "linkToOriginalFile": "https://arquivo.pt/noFrame/replay/20160719170305id_/http://www.expresso.pt/",
      "status": "302",
      "collection": "FAMP2620160719",
      "linkToExtractedText": "https://arquivo.pt/textextracted?m=http%3A%2F%2Fwww.expresso.pt%2F%2F20160719170305",
      "linkToMetadata": "https://arquivo.pt/textsearch?metadata=http%3A%2F%2Fwww.expresso.pt%2F%2F20160719170305"
    }
  ]
}
```


API clients can be written in any programming language

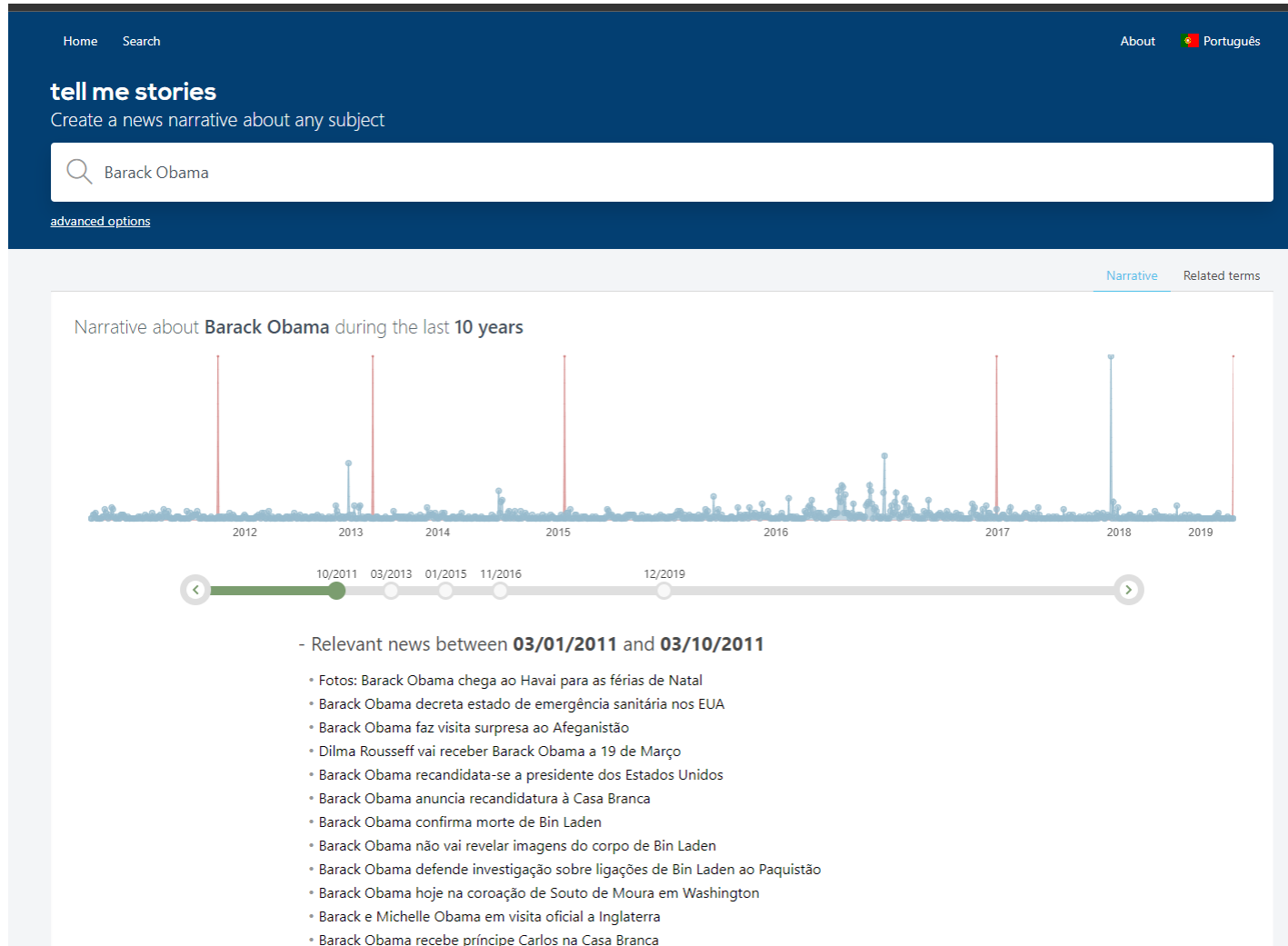


Example of Java client at
github.com/arquivo



Example of Python client at
github.com/programminghistorian

Arquivo.pt API in action: Timeline summarization for large-scale past-web events



TellMeStories
(ContaMeHistorias.pt)

Automatically generates
narratives about any subject
based on online news from the
Past.

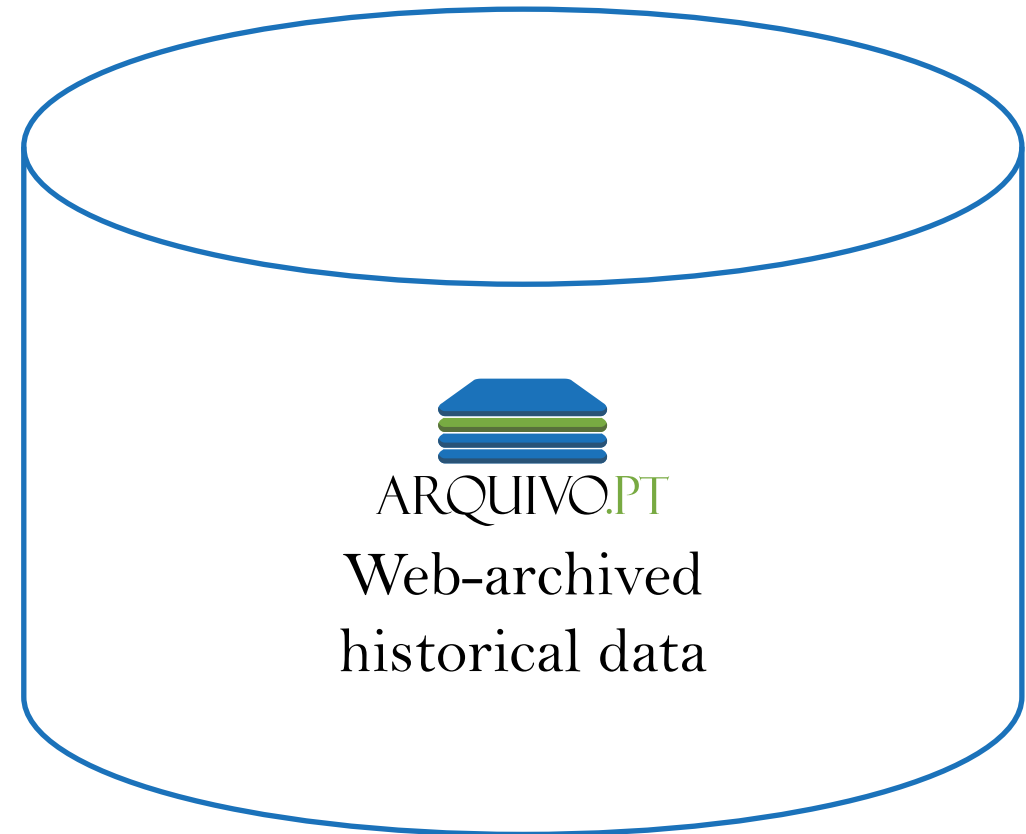
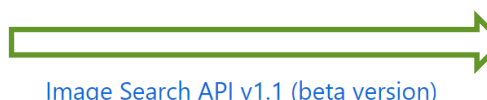
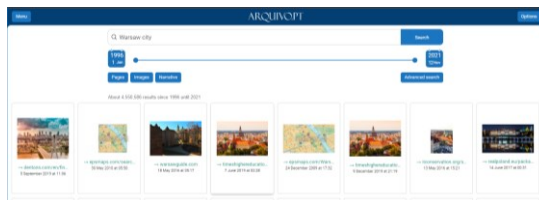
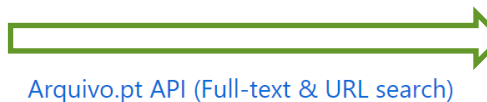
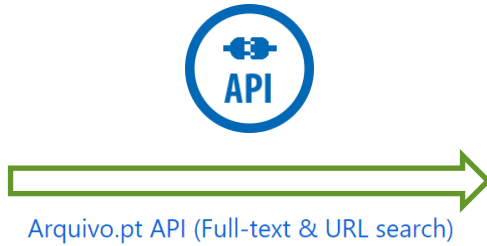
ContaMeHistorias.pt won the Arquivo.pt **Annual Awards** in 2018



Support of the President of the Republic of Portugal

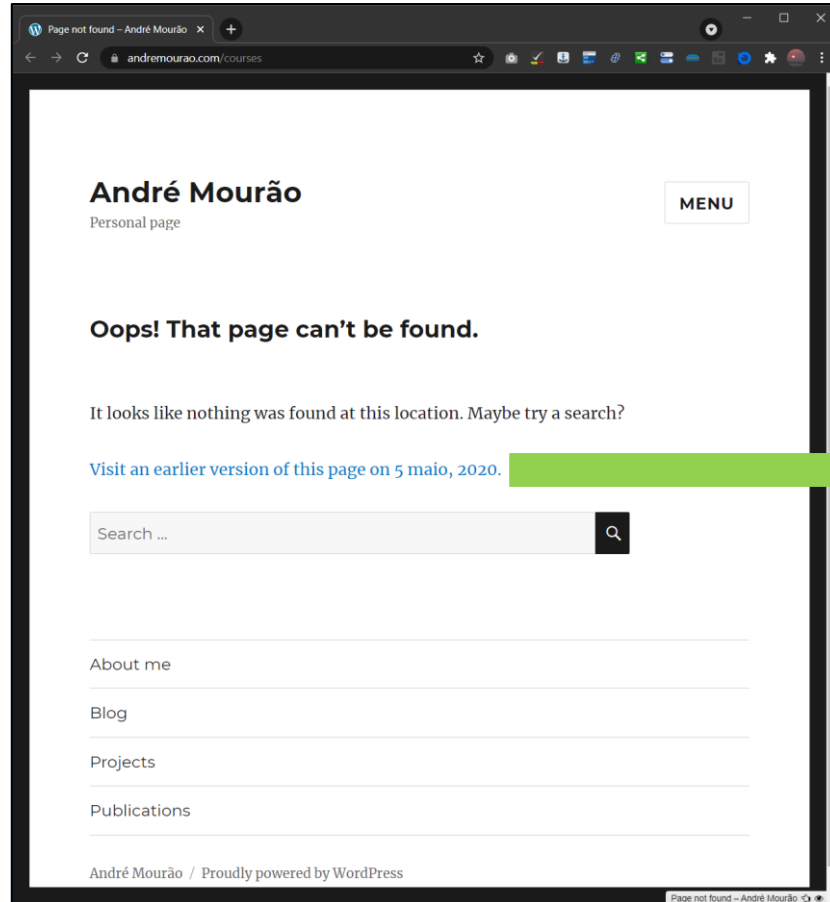
➤ arquivo.pt/awards

Arquivo.pt APIs: “we eat our own dog food!”



arquivo.pt/arquivo404

fix the **broken links** in your website!



Page available at Arquivo.pt

Page not found at live website

Training about web archiving

Attend Training courses about web preservation and research

New ways of searching the past (module A)

- Any Internet user

Publish preservable information on the web (B)

- Web authors

Automatic processing of information preserved from the Web (C)

- Developers, Researchers

Do-it-yourself web archiving (D)

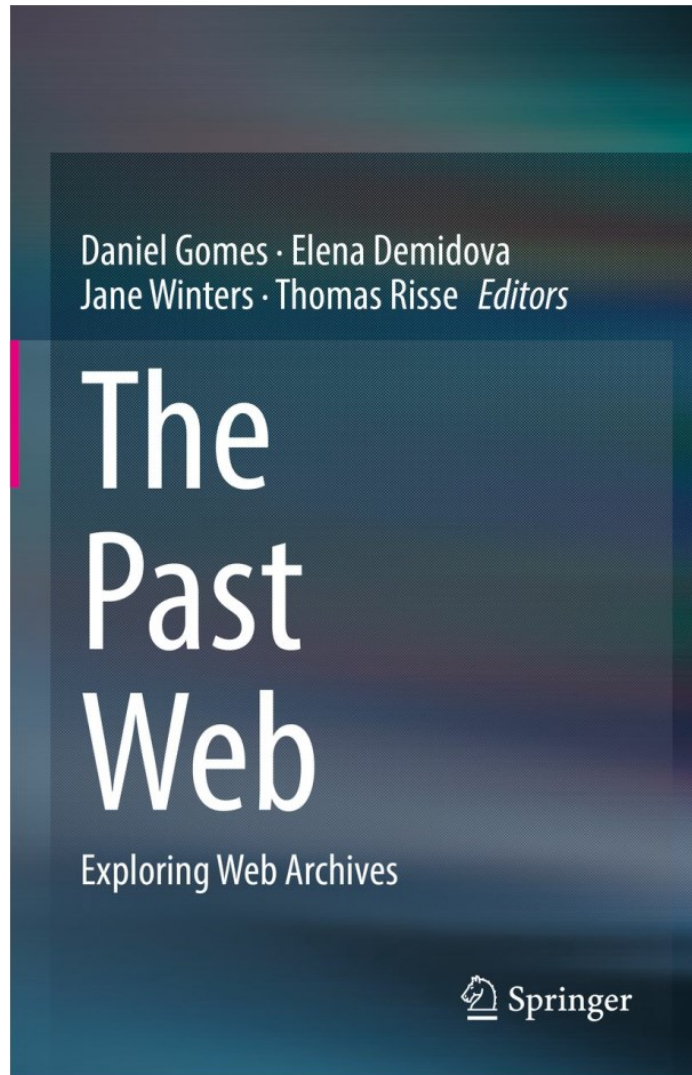
- Researchers, Librarians, Archivists

Videos in English also available!



arquivo.pt/training

Read this book 😊



Preprint version available in open access at:

arquivo.pt/book

Web Archives as Infrastructures to Develop Innovative Services

Front Matter

Pages 215-216

PDF [↓](#)

The Need for Research Infrastructures for the Study of Web Archives

Niels Brügger

Pages 217-224

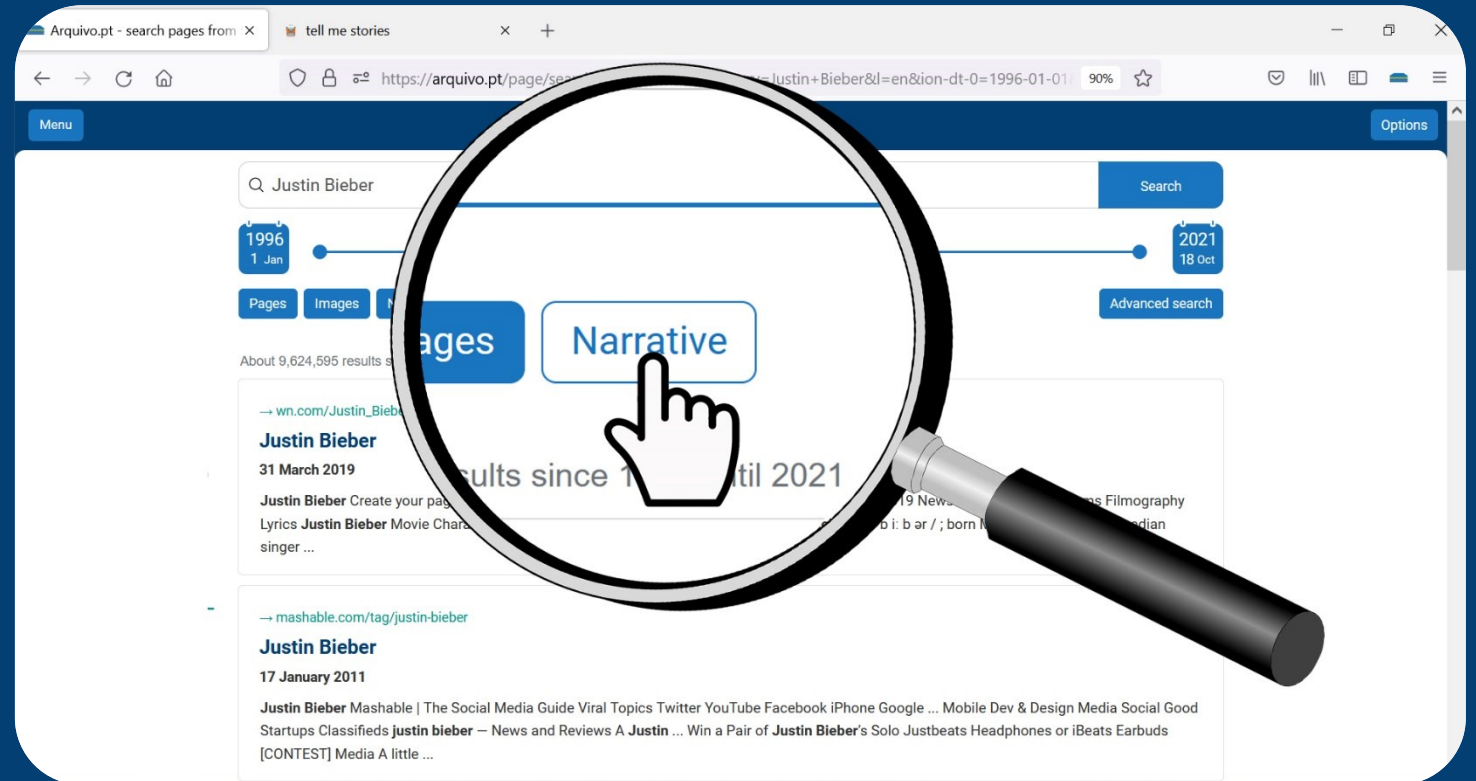
Automatic Generation of Timelines for Past-Web Events

Ricardo Campos, Arian Pasquali, Adam Jatowt, Vítor Mangaravite, Alípio Mário Jorge

Pages 225-242

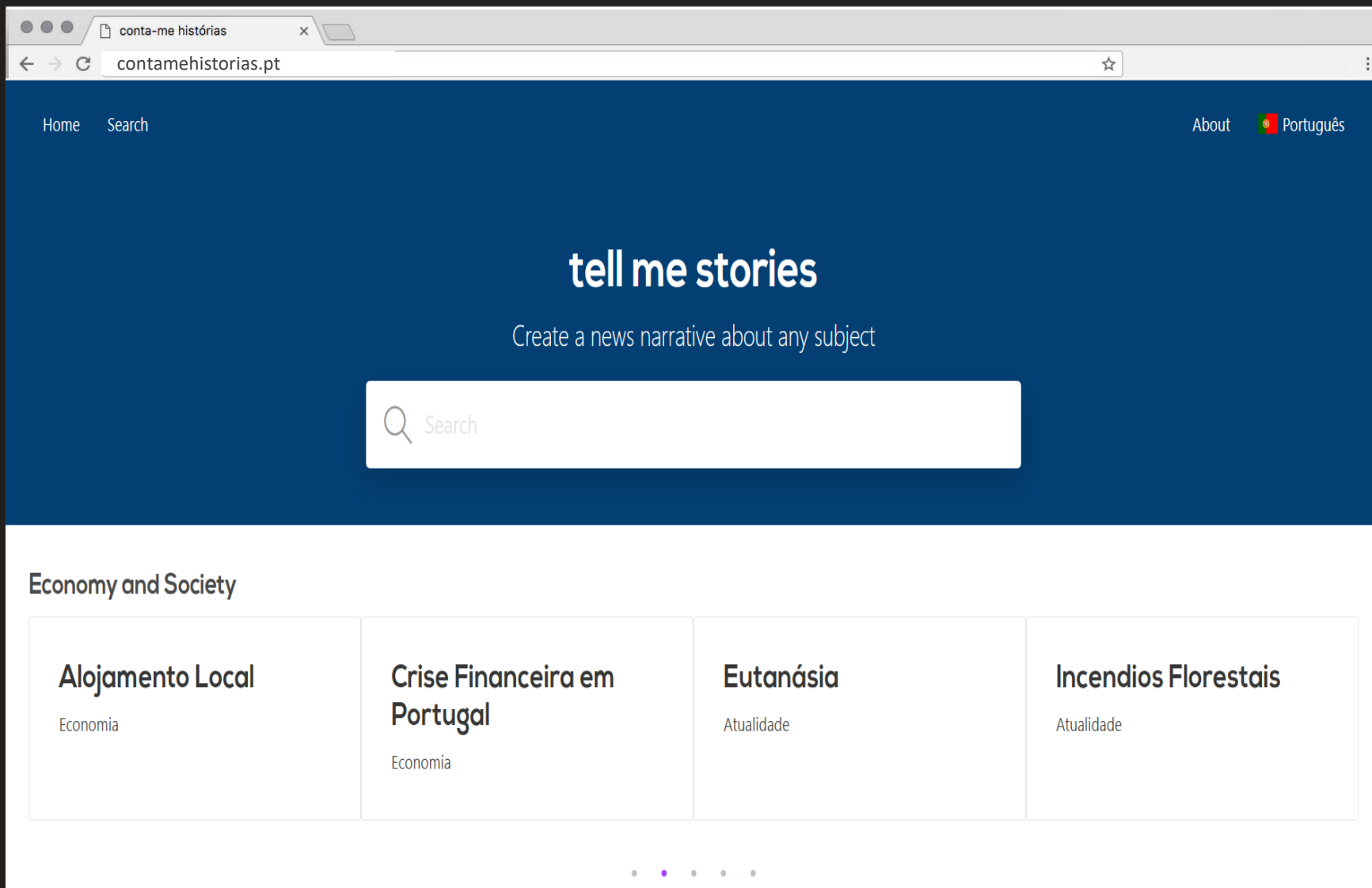
Tell me Stories

Arquivo.pt Awards 2018



<http://contamehistorias.pt>







conta-me histórias

contamehistorias.pt

- Relevant news between **06/01/2011** and **23/02/2016**

2012-02-03

- Donald Trump inaugura o Hotel Trump SoHo New York rodeado pelos filhos
- Donald Trump vai apoiar Mitt Romney nas primárias
- Donald Trump é o 12º candidato republicano à Presidência dos EUA
- Donald Trump anuncia candidatura à presidência dos EUA
- Donald Trump é o novo candidato na corrida presidencial í Casa Branca
- Donald Trump é candidato à presidência dos Estados Unidos
- Donald Trump e Jeb Bush anunciam candidaturas à Casa Branca
- Costa Rica e México sem 'Miss Universo' por causa de Donald Trump
- Donald trump diz que john mccain não pode ser considerado herói de guerra
- Donald Trump é o candidato favorito dos republicanos à Casa Branca
- Campanha de donald trump para presidenciais nos eua sofre primeira baixa
- Donald Trump diz que rival republicano Jeb Bush deve falar inglês em vez de espanhol
- Ivanka Trump, a filha de Donald Trump quer deixar a sua marca
- Donald Trump quer proibir a entrada de muçulmanos nos Estados Unidos
- Donald Trump quer barrar "completamente" a entrada de muçulmanos nos Estados Unidos
- Donald Trump quer proibir entrada de muçulmanos nos EUA
- Obama graceja sobre possível discurso do Estado da União de Donald Trump
- Ted Cruz vence Donald Trump no caucus do Iowa
- Donald Trump e Bernie Sanders favoritos nas primárias de New Hampshire
- Hillary Clinton e Donald Trump vencem, Jeb Bush sai de cena

Horizontal

Vertical

News sources : Lux lol, Tvi24 lol, Dn, Abola, Jn, Ionline, Sol Sapo, Tsf, Noticias Sapo, Publico, Expresso Sapo, Sapo, Economico Sapo, Sabado, Jornaldenegocios, Aeiou, Dinheirovivo, Rtp, Dnoticias, News Google, Cmjornal, Sicnoticias Sapo, Abola:80, Sol Sapo:80,

10925

PAGES FOUND

3222

CONSIDERED NEWS ARTICLES

24

NEWS SOURCES

conta-me histórias

Arquivo.pt: pesquise páginas

arquivo.pt

Menu

ARQUIVO.PT

Opções

dn.pt/inicio/globo/interior.aspx?content_id=2281455&seccao=EUA%20e%20Am%F9ricas

3 Fevereiro às 23:02, 2012

Tabela

Cerca de 3 resultados desde 1996 a 2021

Tabela

Lista

2012

3 versões

Fevereiro

3 versões

3 Fevereiro às 23:02


5 Fevereiro às 05:12

5 Fevereiro às 22:08

Estados Unidos

Donald Trump vai apoiar Mitt Romney nas primárias

por DN.pt



Trump fala com jornalistas em Las Vegas Fotografia © REUTERS/Rick Wilking

Corriam rumores sobre um eventual apoio a Newt Gingrich, mas o empresário deverá anunciar nesta quinta-feira o seu desejo de ver Mitt Romney vencer a nomeação republicana

De acordo a Associated Press (AP), o multimilionário Donald Trump irá, esta noite,

ESPECIAIS

Recuar

Avançar

FERRAMENTAS

Enviar por Email

Partilhar

Imprimir

Comentar este Artigo

Aumentar Texto

Diminuir Texto

PARTILHAR NOTÍCIA

Share

Tweet


Share

Recomendar

12 pessoas recomendam

amigos recomendam.

PUBLICIDADE



Oscar Mascarenhas

Provedor do Leitor

provedordoleitor@dn.pt

CLIQUE AQUI

COMPRE

O JORNAL DE HOJE

TAGS

Partido Republicano, EUA,

primárias, Globo, EUA e Américas

FEIRA DO LIVRO

VEJA AQUI

conta-me histórias

contamehistorias.pt

- Relevant news between **19/01/2017** and **07/05/2017**

- Donald Trump torna-se o 45.º Presidente dos Estados Unidos
- Theresa May quer discutir comércio e NATO com Donald Trump
- Rússia discorda de Donald Trump e diz que Irão não é um Estado terrorista
- Madonna, Scarlett Johanson e Michael Moore unidos contra Trump
- Parlamento unido na crítica a Donald Trump e em defesa do Acordo de Paris sobre clima
- Viola Davis defende Meryl Streep de Donald Trump
- Britânicos protestam contra visita de Donald Trump ao Reino Unido
- Donald Trump quer que Putin devolva Crimeia à Ucrânia
- Donald Trump volta a defender supremacia nuclear dos EUA e diz que quer expandir arsenal
- Hillary Clinton goza com Donald Trump
- Donald Trump vai faltar ao jantar de correspondentes da Casa Branca
- Putin ridiculariza acusações de espionagem russa sobre Donald Trump
- Secretário da Segurança Interna dos EUA desmente Donald Trump
- Donald Trump quer assinar nova ordem sobre imigração
- Donald Trump acusa Barack Obama de o ter colocado sob escuta durante a campanha
- Presidente Peña Nieto cancela encontro com Donald Trump
- Donald Trump recebe Angela Merkel na Casa Branca
- Donald Trump afasta Steve Bannon do Conselho de Segurança Nacional
- Donald Trump ameaça Coreia do Norte
- Papa Francisco vai receber Donald Trump no Vaticano

Horizontal

Vertical

News sources : Lux lol, Tvi24 lol, Dn, Abola, Jn, Ionline, Sol Sapo, Tsf, Noticias Sapo, Publico, Expresso Sapo, Sapo, Economico Sapo, Sabado, Jornaldenegocios, Aeiou, Dinheirovivo, Rtp, Dnoticias, News Google, Cmjornal, Sicnoticias Sapo, Abola:80, Sol Sapo:80,

10925

PAGES FOUND

3222

CONSIDERED NEWS ARTICLES

24

NEWS SOURCES

conta-me histórias

contamehistorias.pt

HomeSearch

AboutPortuguês

tell me stories

Create a news narrative about any subject

donald trump

[advanced options](#)

NarrativeRelated terms

Terms related with Donald Trump during the last 10 years

Acordo de Paris

Trump Tower

EUA

Nova Iorque

Estados Unidos

Hillary Clinton

Vaticano

Obama

Vladimir Putin

NATO

Reino Unido

Donald Trump

Trump

Putin

Jeb Bush

China

Coreia do Norte

Estado

Papa Francisco

Russia

Mexico

Ted Cruz

Barack Obama

Bernie Sanders

Casa Branca

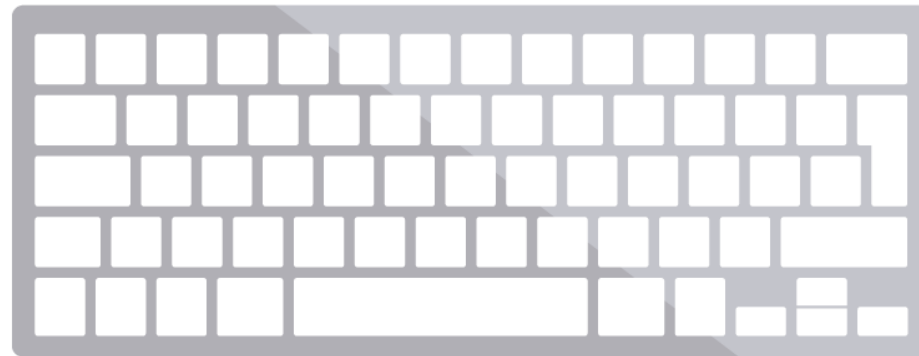
Hillary

Ucrania

Kim Jong-un

Melania

<http://yake.inesctec.pt>

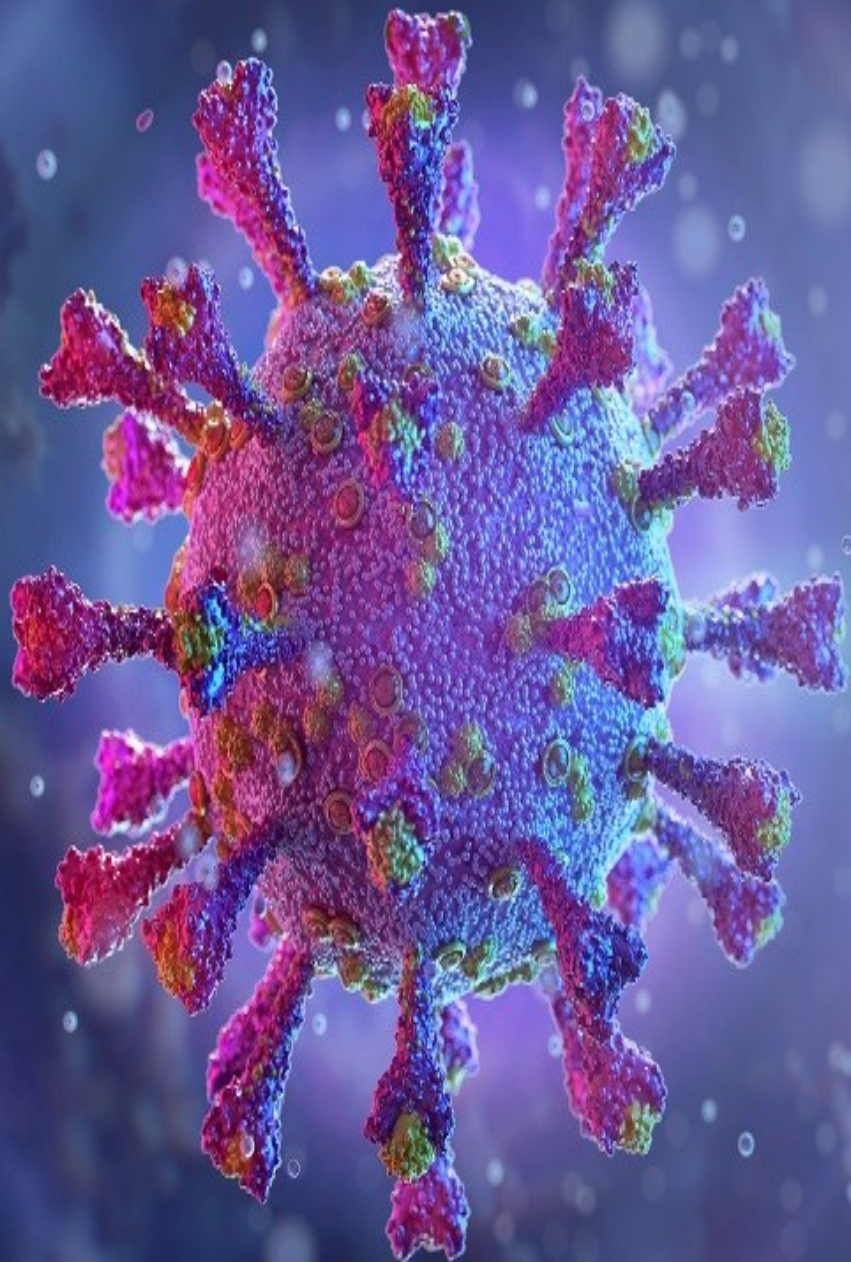


Time spent to run YAKE algorithm 0.47 ms.

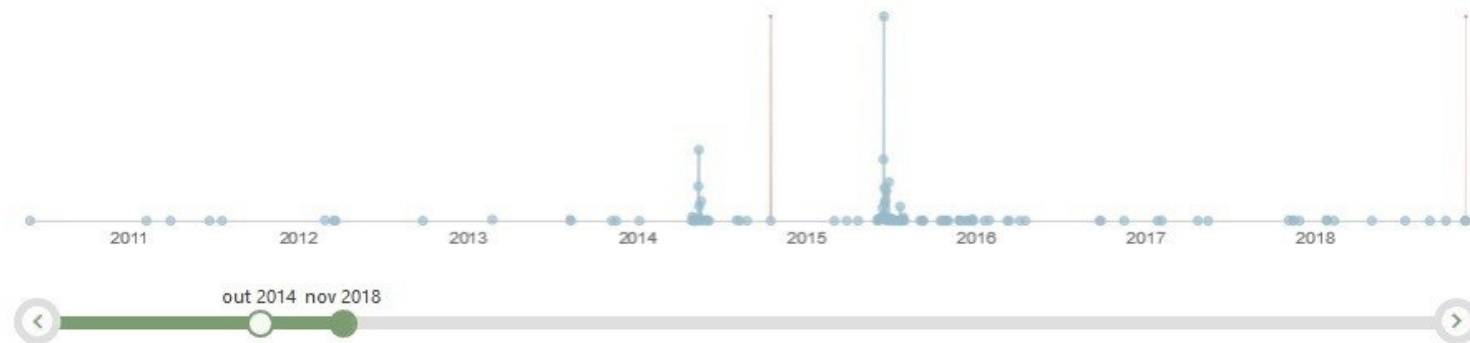
Annotated text

The top 20 keywords in terms of relevance are annotated in the text.

brain aging **gene discovered**. Summary: Researchers have identified a genetic variant that can accelerate normal **brain aging** in older people by up to 12 years. Source: **columbia university medical center**. Genetic variant accelerates normal **brain aging** in older people by up to 12 years. **columbia university medical center** (CUMC) researchers have discovered a common genetic variant that greatly impacts normal **brain aging**, starting at around age 65, and may modify the risk for neurodegenerative diseases. The findings could point toward a novel biomarker for the evaluation of anti-aging interventions and highlight potential new targets for the prevention or treatment of age-associated **brain** disorders such as Alzheimer's **disease**. The study was published online today in the journal Cell Systems. "If you look at a group of seniors, some will look older than their peers and some will look younger," said the study's co-leader Asa Abeliovich, PhD, professor of pathology and neurology in the Taub Institute for Alzheimer's **disease** and the **aging brain** at CUMC. "The same differences in **aging** can be seen in the frontal cortex, the **brain** region responsible for higher mental processes. Our findings show that many of these differences are tied to variants of a **gene** called TMEM106B. People who have two 'bad' copies of this **gene** have a frontal cortex that, by various biological measures, appears 12 years older than those who have two normal copies." Studies have identified individual genes that increase one's risk for various neurodegenerative disorders, such as apolipoprotein E (APOE) for Alzheimer's **disease**. "But those genes explain only a small part of these diseases," said study co-leader Herve Rhinn, PhD, assistant professor of pathology and cell biology in the Taub Institute. "By far, the major risk factor for neurodegenerative **disease** is **aging**. Something changes in the **brain** as you age that makes you more susceptible to **brain disease**. That got us thinking, 'What, on a genetic level, is driving healthy **brain** aging?'" In the current study, Drs. Abeliovich and Rhinn analyzed genetic data from autopsied human **brain** samples taken from 1,904 people without neurodegenerative **disease**. First, the researchers looked at the subjects' transcriptomes (the initial products of **gene** expression), compiling an average picture of the **brain** biology of people at a given age. Next, each person's transcriptome was compared to the average transcriptome of people at the same age, looking specifically at about 100 genes whose expression was found to increase or decrease with **aging**. From this comparison, the researchers derived a measure that they call differential **aging**: the difference between an individual's apparent (biological) age and his or her true (chronological) age. "This told us whether an individual's frontal cortex looked older or younger than expected," said Dr. Abeliovich. The researchers then searched the genome of each individual, looking for genetic variants that were associated with an increase in differential age. "One variant stood out: TMEM106B," said Dr. Rhinn. "It's very common. About one-third of people have two copies and another third have one copy." "TMEM106B begins to exert its effect once people reach age 65," said Dr. Abeliovich. "Until then, everybody's in the same boat, and then there's some yet-to-be-defined stress that kicks in. If you have two good copies of the **gene**, you respond well to that stress. If you have two bad copies, your **brain** ages quickly." The researchers found a second variant—inside the progranulin gene—that contributes to **brain aging**, though less so than TMEM106B. Progranulin and TMEM106B are located on different chromosomes but are involved in the same signaling pathway. Both have also been associated with a rare neurodegenerative **disease** called frontotemporal dementia. The study did not address what role the two genetic variants might have in neurodegenerative **disease**. "We were studying healthy individuals, so it is not about **disease**, per se," said Dr. Abeliovich. "But of course, it's in healthy tissue that you start to get **disease**. It appears that if you have these genetic variants, **brain aging** accelerates and that increases vulnerability to **brain disease**. And vice versa: if you have **brain disease**, the **disease** accelerates **brain aging**. It's a vicious cycle." About this genetics and **aging** research article Funding: The study was supported by grants from the National Institute of **aging** (AG042317), the National Institute of Neurological Disorders and Stroke, and the Michael J. Fox Foundation for Parkinson's Research. Dr. Abeliovich is a co-founder of and consultant for Alektor. Dr. Rhinn is a consultant for Alektor. The researchers declare no other financial conflicts of interest. Source: Karin Eskenazi – **columbia university medical center** Image Source:



Narrativa sobre Síndrome Respiratória Do Médio Oriente durante os últimos 10 anos

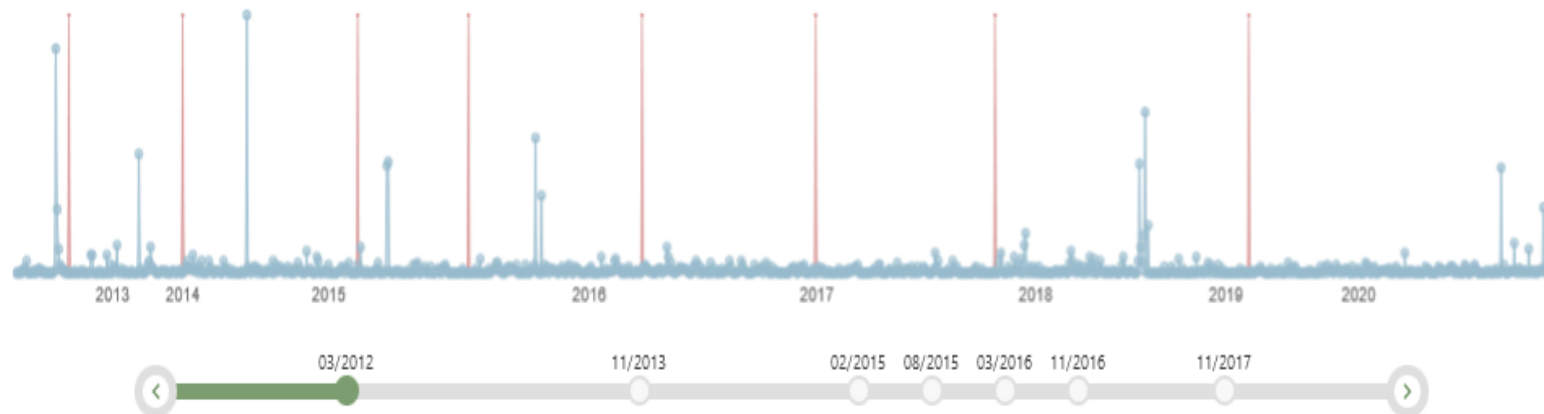


- Notícias relevantes entre **01, março 2015** e **22, novembro 2018**

Identificados anticorpos contra coronavírus da Síndrome Respiratória do Médio Oriente.
EUA confirmam primeiro caso de contaminação por MERS.
Confirmado terceiro caso de coronavírus nos EUA.
Cientistas anunciam composto que combate coronavírus como o da MERS.
DGS renova recomendações sobre coronavírus do Médio Oriente.
Sobe para cinco número de mortes devido ao coronavírus na coreia do sul.
Coreia do sul regista sexta vítima mortal devido ao coronavírus mers.
Coronavírus mers causa sétima vítima mortal na coreia do sul.
Hong kong emite alerta vermelho para viagens à coreia do sul devido a surto de coronavírus.
Coronavírus mers já matou nove pessoas na coreia do sul.
Oms convoca reunião de comissão de emergência sobre síndrome respiratória.
Oms discute na terça-feira situação do coronavírus mers na coreia do sul.
Oms aconselha coreia do sul a proibir viagens de potenciais infectados com coronavírus mers.
Equipa da OMS vai visitar Coreia do Sul para analisar surto de MERS.
Presidente da Coreia do Sul cancela visita aos EUA devido ao surto do novo coronavírus.
Coreia do sul regista mais sete casos de coronavírus mers.
Síndrome respiratória do Médio Oriente é a nova preocupação da OMS.
Medo da síndrome respiratória do médio oriente faz disparar compras online.
Síndrome Respiratória propaga-se na Coreia do Sul.
China e EUA anunciam testes bem sucedidos na luta contra a Síndrome Respiratória do Médio Oriente.
OMS recusa declarar emergência sanitária mundial devido ao vírus MERS.
Coreia do sul regista três novos casos de síndrome respiratória do médio oriente.
Turismo da coreia do sul afetado por síndrome respiratória do médio oriente.
homem morre com síndrome do Médio Oriente.



Narrativa sobre Putin durante os últimos 10 anos



- Notícias relevantes entre **03/01/2012** e **24/03/2012**

- Mikhail Gorbatchov aconselha Vladimir Putin a abandonar a corrida à Presidência russa
- Tensão entre Putin e Medvedev face à intervenção na Líbia
- Partido de Vladimir Putin vence eleições parlamentares
- Marcha da oposição autorizada em Moscovo, Putin responde com manifestação
- Serviços secretos impediram atentado contra Putin
- Putin evoca vitória militar sobre Napoleão para apelar a 130 mil apoiantes
- Putin confirma Medvedev no cargo de primeiro-ministro e promete diálogo com oposição
- Putin deverá ser escolhido para liderar Rússia nos próximos seis anos
- Putin volta a ser Presidente
- Putin vence eleições presidenciais russas: Primeiro-ministro ganhou acto eleitoral com mais d
- Czar Vladimir Putin regressa ao Kremlin para terceiro mandato
- Vladimir Putin vence presidenciais à primeira volta
- Vladimir Putin como Mr. Burns dos Simpsons [com vídeo]
- Dezenas de milhares de manifestantes contra Putin em Moscovo e São Petersburgo
- Vladimir Putin participa em encontro de motards na Ucrânia
- Vídeo: Putin vence eleições presidenciais russas
- Medvedev apoia candidatura de Putin a presidente
- Super Putin, novo herói de Banda Desenhada
- Eduardo dos Santos saúda eleição de Vladimir Putin
- Vídeo: 20 mil em manifestação contra Putin

Tell me Stories Python Package

The screenshot shows the GitHub repository page for 'LIAAD / TemporalSummarizationFramework'. The browser address bar displays 'https://github.com/LIAAD/TemporalSummarizationFramework'. The repository page includes a search bar, navigation links for Pull requests, Issues, Marketplace, and Explore, and a header with the repository name and icons for Unwatch (7), Star (5), and Fork (2). Below the header, there are tabs for Code, Issues, Pull requests, Actions, Security, Insights, and Settings. The main content area shows the repository structure with a table of files and folders, including 'contamehistorias', 'LICENSE.txt', 'MANIFEST.in', 'README.md', 'requirements.txt', 'setup.cfg', 'setup.py', and 'test_conteme.py'. The right sidebar contains sections for 'About' (Temporal summarization framework), 'Releases' (No releases published), and 'Packages' (No packages published).

Search or jump to... Pull requests Issues Marketplace Explore

LIAAD / TemporalSummarizationFramework

Unwatch 7 Star 5 Fork 2

<> Code Issues Pull requests Actions Security Insights Settings

master 2 branches 0 tags

Go to file Add file Code

rncampos Update README.md 3c1a36b on 20 Sep 2019 10 commits

contamehistorias	fix bing	2 years ago
LICENSE.txt	first commit	2 years ago
MANIFEST.in	first commit	2 years ago
README.md	Update README.md	2 years ago
requirements.txt	first commit	2 years ago
setup.cfg	first commit	2 years ago
setup.py	first commit	2 years ago
test_conteme.py	fix test example	2 years ago

README.md

About

Temporal summarization framework

nlplab.inesctec.pt/tellmestories

Readme

View license

Releases

No releases published

[Create a new release](#)

Packages

No packages published

[Publish your first package](#)

<http://github.com/LIAAD/TemporalSummarizationFramework>

DEMO



MONTHLY



Code

Memory: 776.6 MB

```
In [2]: from contamehistorias.datasources.webarchive import ArquivoPT
        from datetime import datetime

        # Specify website and time frame to restrict your query
        domains = [ 'http://publico.pt/', 'http://www.dn.pt/', 'http://dnoticias.pt/', 'http://www.rtp.pt/', 'http://www.sapo.pt/' ]

        params = { 'domains':domains, 'from':datetime(year=2011, month=1, day=1), 'to':datetime(year=2021, month=1, day=1) }

        query = 'putin'

        apt = ArquivoPT()
        search_result = apt.getResult(query=query, **params)
```

7.5.2 Percorrer os resultados obtidos do Arquivo.pt

O objeto `search_result` devolve o número total de resultados obtidos a partir da chamada à API do Arquivo.pt. O número total de resultados excede facilmente as 10,000 entradas. Um tal volume de dados é praticamente impossível de processar por qualquer utilizador que a partir dele queira retirar conhecimento em tempo útil.

```
In [3]: len(search_result)
```

```
Out[3]: 10834
```

Para lá do número total de resultados o objeto `search_result` reúne informação extremamente útil para o passo seguinte do algoritmo, i.e., a seleção das notícias mais relevantes ao longo do tempo. Em concreto, este objeto permite ter acesso à:

- `datetime`: data de coleta do recurso;

Team



Adam Jatowt



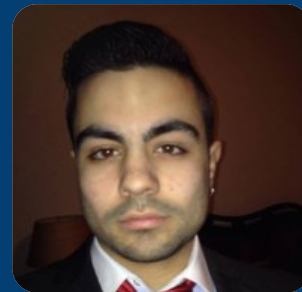
Alexandre Ribeiro



Alípio Jorge



Arian Pasquali



João Campos



Simão Samouco



Vítor Mangaravite

Acknowledgements

