

APIs do Arquivo.pt

Como processar informação em larga escala?

NovaSearch Reading Group

19 de Outubro de 2020

André Mourão

Engenheiro I&D

andre.mourao@fccn.pt

Quem sou eu?

- Doutorado em Informática pela FCT/UNL
 - Pesquisa Multimédia em Larga Escala

Quem sou eu?

- Doutorado em Informática pela ~~FCT/UNL~~ FCT NOVA
 - Pesquisa Multimédia em Larga Escala

Quem sou eu?

- Doutorado em Informática pela ~~FCT/UNL~~ ~~FCT NOVA~~ NOVA SST
 - Pesquisa Multimédia em Larga Escala

Quem sou eu?

- Doutorado em Informática pela ~~FCT/UNL~~ ~~FCT NOVA~~ NOVA SST
 - Pesquisa Multimédia em Larga Escala
- Engenheiro de I&D no Arquivo.pt
 - Responsável pela pesquisa de imagens, texto e APIs

Quem sou eu?

- Doutorado em Informática pela ~~FCT/UNL~~ ~~FCT NOVA~~ NOVA SST
 - Pesquisa Multimédia em Larga Escala
- Engenheiro de I&D no Arquivo.pt
 - Responsável pela pesquisa de imagens, texto e APIs
- revisionista.pt, effortly.run, freelancer....

O que é o Arquivo.pt?



MILHÕES DE PÁGINAS
ARQUIVADAS

Os dados do Arquivo.pt

- Os dados do Arquivo.pt podem ser explorados de várias maneiras:
 - Consulta de histórico de páginas arquivadas
 - Pesquisa de texto e metadados
 - Pesquisa de imagens
- Um factor diferenciador do Arquivo.pt é possibilitar a exploração do **componente temporal** dos dados recolhidos
 - A mesma página é recolhida várias vezes em alturas diferentes

Consulta de histórico de páginas: www.fct.unl.pt

ARQUIVO.PT

Q fct.unl.pt Pesquisar

1996 1 Jan 2020 15 Out

Páginas Imagens Pesquisa avançada

Cerca de 428 resultados desde 1996 a 2020

Tabela Lista

1996	1 versão
1997	3 versões
1998	7 versões
1999	7 versões
2000	18 versões
2001	13 versões
2002	8 versões
2003	12 versões

ARQUIVO.PT

Menu

Q fct.unl.pt Pesquisar

1996 1 Jan 2020 15 Out

Páginas Imagens Pesquisa avançada

Cerca de 428 resultados desde 1996 a 2020

Tabela Lista

1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
13 Out	26 Fev	12 Jan	25 Jan	29 Fev	2 Fev	28 Mar	4 Fev	24 Mar	4 Jan	1 Jan	4 Jan	3 Jan	21 Mai	31 Mai	20 Jan	22 Jan	6 Nov	30 Set	23 Abr	9 Jan	1 Fev	7 Abr	29 Jan	
	26 Jul	14 Mai	25 Jan	2 Mar	5 Fev	1 Jun	14 Fev	11 Mai	9 Jan	8 Jan	9 Jan	8 Jan	23 Jun	5 Jun	22 Jan	23 Jan	8 Nov	1 Out	24 Abr	11 Jan	22 Fev	15 Abr	4 Fev	
	10 Dez	12 Jul	8 Fev	8 Abr	1 Mar	3 Jun	27 Mar	6 Jun	12 Jan	9 Jan	14 Jan	12 Jan	24 Set	9 Jun	23 Jan		9 Nov	2 Out	26 Abr	12 Jan	22 Fev	15 Abr	5 Fev	
		1 Dez	19 Fev	8 Abr	1 Mar	6 Jun	5 Abr	14 Jun	13 Jan	12 Jan	19 Jan	16 Jan	25 Set	4 Ago	29 Jan			28 Nov	15 Set	13 Jan	23 Fev	16 Abr	14 Fev	
		6 Dez	27 Abr	10 Mai	2 Mar	2 Ago	22 Abr	15 Jun	14 Jan	14 Jan	24 Jan	20 Jan	26 Set	5 Ago	29 Jan				15 Set	14 Jan	6 Jun	20 Abr	17 Fev	
		12 Dez	29 Abr	10 Mai	9 Mar	29 Set	31 Mai	23 Jun	19 Jan	14 Jan	2 Fev	24 Jan	17 Dez	6 Ago	19 Mai				24 Nov	15 Jan	13 Jun	15 Jun	1 Mar	
		12 Dez	12 Out	11 Mai	1 Abr	22 Nov	19 Jun	1 Jul	22 Jan	14 Jan	22 Fev	29 Jan	18 Dez		20 Mai				26 Nov	17 Jan	6 Ago	15 Jun	22 Mar	

https://arquivo.pt/wayback/*/http://www.fct.unl.pt

Consulta de histórico de páginas: www.fct.unl.pt

Menu

ARQUIVO.PT

fct.unl.pt 13 Outubro às 20:31, 1996

Opções

Tabela

1996

Outubro

13 Outubro às 20:31

1997

1998

1999

2000

2001

2002

2003

2004

2005

 **Universidade Nova de Lisboa**

Faculdade de Ciências e Tecnologia



[Air photos of the campus](#)

Index

General Information	Departments & Sections	Portuguese References	Curiosities
-------------------------------------	--	---------------------------------------	-----------------------------

 [Esta página em português](#)

<https://arquivo.pt/wayback/19961013203140/http://www.fct.unl.pt/>

1996 10 13 20 31 40
ano mês dia hora min seg

Pesquisa de texto



The screenshot shows the ARQUIVO.PT search interface. At the top, there's a dark blue header with 'ARQUIVO.PT' in the center, a 'Menu' button on the left, and an 'Opções' button on the right. Below the header, a search bar contains the text 'departamento de informática fct unl' with a 'Pesquisar' button to its right. Under the search bar, there's a timeline slider from 1996 to 2020, with '1 Jan' and '15 Out' markers. Below the slider are buttons for 'Páginas', 'Imagens', and 'Pesquisa avançada'. The search results section shows 'Cerca de 536.260 resultados desde 1996 a 2020'. The first result is for 'di.fct.unl.pt' with the title 'Departamento de Informática - FCT/UNL' and date '25 Setembro, 2009'. The snippet describes the department's focus on research and teaching. The second result is also for 'di.fct.unl.pt' with the title 'DI – Departamento de Informática - FCT/UNL' and date '18 Dezembro, 2009'. Its snippet mentions English contacts, internal logs, and recruitment efforts.

Menu

ARQUIVO.PT

Opções

Q departamento de informática fct unl

Pesquisar

1996 1 Jan

2020 15 Out

Páginas Imagens Pesquisa avançada

Cerca de 536.260 resultados desde 1996 a 2020

→ di.fct.unl.pt

Departamento de Informática - FCT/UNL

25 Setembro, 2009

Departamento de Informática - FCT/UNL Início Sobre o DI Ensino Investigação Pessoas Serviços ... University. Read more... [arquivo de destaques] Bem-vindo ao Departamento de Informática da FCT/UNL . O DI ... Departamento de Informática da FCT/UNL encontram-se estruturadas e enquadradas em três secções de ensino ...

↳ di.fct.unl.pt

DI – Departamento de Informática - FCT/UNL

18 Dezembro, 2009

DI – Departamento de Informática - FCT/UNL English Contactos WIKI Vida Interna Log in Acesso restrito Departamento de Informática Pesquisar Pesquisa Avançada... Sobre o DI Ensino Investigação Pessoas ... Luís Atribuição de Bolsas de Estudo O Departamento de Informática pretende recrutar alunos, em ...

Ver mais resultados de www.di.fct.unl.pt

<https://arquivo.pt/page/search?query=departamento+de+informática+fct+unl>










Pesquisa de imagens

Menu ARQUIVO.PT Opções

Q campus fct unl Pesquisar

Páginas Imagens Pesquisa avançada

Cerca de 39.348 resultados desde 1996 a 2020

 <p>Fig. 3 - Pistola de abastecimento (modelo Australiano)</p> <p>→ insinuacoes.blogspot.com/... 11 Março às 19:29, 2008</p>	 <p>→ computerworld.com.pt 1 Março às 16:07, 2011</p>	 <p>→ computerworld.com.pt/201... 25 Março às 03:46, 2011</p>	 <p>→ mar.ist.utl.pt/captar/en/ove... 17 Agosto às 09:52, 2016</p>	<p>→ noitedosinvestigadores.org/... 18 Julho às 21:41, 2018</p>	 <p>→ dcsa.fct.unl.pt/index.php?o... 10 Junho às 12:57, 2007</p>
					

<https://arquivo.pt/image/search?query=campus+fct+unl>

Pesquisa avançada de imagens ou texto

- Pesquisa por frase
- Filtragem por intervalo temporal
- Site/Domínio (ex. dentro de fct.unl.pt)
- Formato
 - pdf, doc, pptx,
 - png, jpeg, gif....

Palavras ▼

Tenham estas palavras:

ex.: expo 98

Tenham esta expressão ou frase:

ex.: pavilhão da utopia

Não tenham estas palavras:

ex.: concerto

Data ▼

2008
1 Jan

2017
31 Dez

Formato ▼

Ver as páginas com o formato:

Selecione ▼

Sítio Web ▼

Com este endereço:

Casos de uso

Gestão: documentação/auditoria de projeto concluído



“Estava tudo no site”

Linguateca
[Estrutura](#)
[Equipa](#)
Apresentação
Acesso a recursos

- AC/DC
- CETENPublico
- CETENFolha
- CHAVE
- COMPARA
- Corpografo
- CotTrad
- Esfinge
- Floresta Sintática
- GeoNetPT
- PAPEL
- PANTERA
- PoNTE
- ReEscreve
- REPENTINO
- Repositório
- WebSpell
- WPT

Avaliação conjunta
[Catálogo de recursos](#)
[Catálogo de ferramentas](#)
[Catálogo de atores](#)
[Catálogo de publicações](#)
[Informação interessante](#)
[Perguntas já respondidas](#)
[Gramateca](#)
[English](#)
[Norsk](#)

Repositório de recursos, publicações e programas

[Linguateca](#)

Neste espaço providenciamos uma "estante electrónica" para todos quantos quiserem ter os seus trabalhos relacionados com o processamento computacional da língua portuguesa acessíveis directamente da rede.

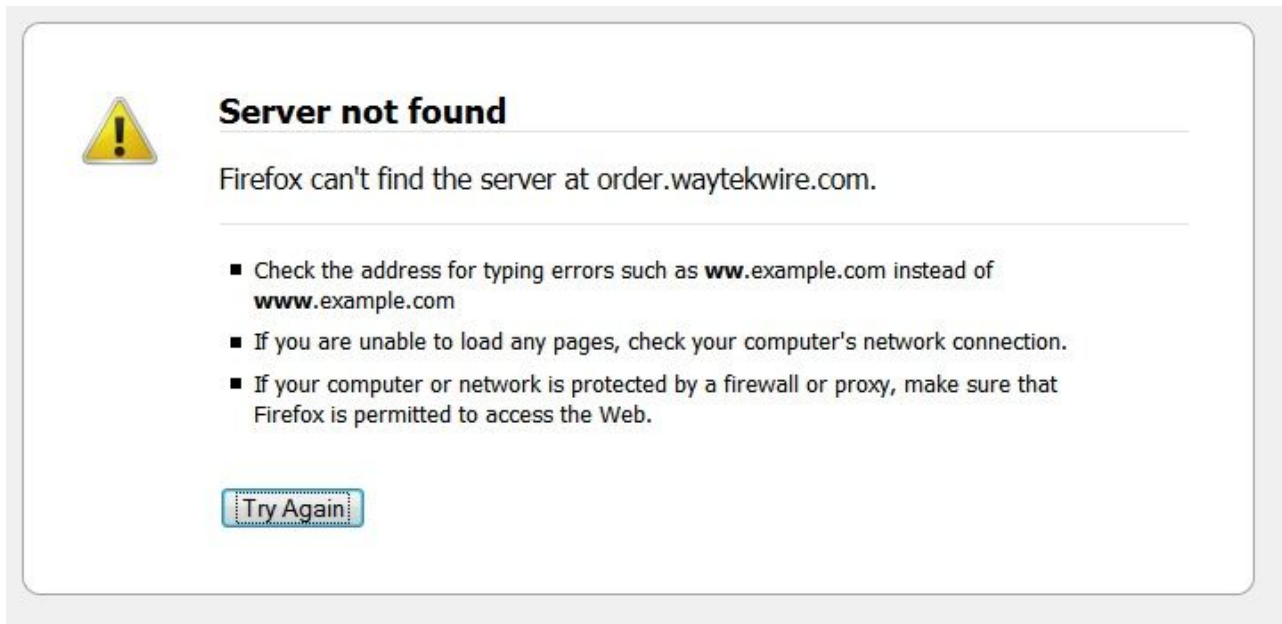
Basta enviarem-nos a versão electrónica, indicando a referência que queiram que apareça no nosso [catálogo de publicações](#), ou, no caso de programas ou outros recursos, enviarem-nos o respectivo programa e documentação básica.

De momento, temos os seguinte recursos importantes (por ordem de disponibilização), em que indicamos também, com gratidão, o responsável pela sua cedência, e a data a partir da qual ficaram disponíveis no Repositório

- [Regras de expressões bilingues inglês-português](#), Logos Corporation (Anabela Barreiro, Janeiro de 2001)
- [Base de Dados para Sintetizador de Fala em Português Europeu, FEUP/IPB-BD](#) (João Paulo Teixeira, Abril de 2001)
- [Dicionário Português-Latino de Pedro José da Fonseca](#) (Digitalizado por Ricardo Ueda Karpischek, Julho de 2003)
- [Coleção FolhaRIcol](#) (Vera Lúcia Strube de Lima, Agosto de 2003)
- [Corpus TeMário](#) (Lucia H. Machado Rino, Outubro de 2003)
- [Coleção WBR-99](#) (Berthier Ribeiro-Neto, Nivio Ziviani, Dezembro de 2003)
- [Port4NooI: recursos linguísticos para o português no formato NooI](#) (Anabela Barreiro, Novembro de 2007)
- [Léxico-gramática dos adjectivos intransitivos humanos do português europeu](#) (Paula Carvalho, Dezembro de 2008)
- [Erros ortográficos em português de Portugal e sua correcção](#) (José Carlos Medeiros, Agosto de 2009)
- [Memórias de tradução de português de Portugal para inglês britânico](#) (Março de 2010)
- [Correcção ortográfica de português](#) (Abril de 2010)
- [Ref.i: como de resenhas de livros](#) (Cláudia Freitas, Maio de 2013)

Portfólio/CV

- “O meu melhor trabalho foi um site que já não existe.”
 - 60% dos sites desaparecem após apenas 2 anos.



Documentação de tecnologia “legacy”



NOKIA
CONNECTING PEOPLE

Mapa Site | Pesquisa Avançada

Pesquisa Rápida 

Nokia 3210

Divertido por fora, sério por dentro

O novo Nokia 3210. Divertido por fora, sério por dentro. Com tampas Xpress-on pode trocá-las por ambos os lados.

O que há de novo
[Componha](#) o seu toque de telefone pessoal. [Envie uma mensagem](#) que vale mil palavras, contendo uma imagem. Navegue sem fronteiras nas redes GSM 900 e GSM 1800.

Divertido por fora
Exprima-se livremente com as tampas Xpress-on™ [que pode trocar](#), de frente e de trás. Jogue três jogos. Dê um gosto especial às suas mensagens enviando imagens - é como mandar um postal electrónico.



Se não sabe o modelo do seu telemóvel, pode-o encontrar no manual de instruções com a ajuda do tipo técnico do seu telemóvel, siga os seguintes passos:

- Retire a bateria do telemóvel.
- O tipo técnico do telemóvel está na etiqueta branca, dentro do telemóvel.

Modelo	Tipo Técnico	Língua
Nokia 3210	nse-8	Inglês
Nokia 3210	nse-8	Português
Nokia 5110	nse-1	Inglês
Nokia 5110	nse-1	Português
Nokia 6090	nme-3	Inglês
Nokia 6090	nme-3	Português
Nokia 6110	nse-3	Inglês
Nokia 6110	nse-3	Português
Nokia 6150	nsm-1	Inglês
Nokia 6150	nsm-1	Português
Nokia 7110	nse-5	Inglês
Nokia 7110	nse-5	Português
Nokia 8210	nsm-3	Inglês
Nokia 8210	nsm-3	Português

NOKIA
Connecting People

Mapa Site | Pesquisa Avançada

Pesquisa Rápida 

Manuais

Telemóveis recentes:
[Escolha um](#) 

Nokia PC Suite

- Informação Geral
- Download
- Apoio
 - Manual do Utilizador
 - Estudo de Iniciação
 - Perguntas mais Frequentes
 - PC Suite Cleaner
- Funcionalidades suportadas pelo seu telemóvel
- Aplicações adicionais

Software Avançado de PC

Controladores de cabo

Nokia Lifeblog

Catalogs

Manual do utilizador



O manual do utilizador contém informação para o ajudar a instalar o PC Suite e a familiarizar-se com o mesmo.

Em caso de problemas no download de ficheiros, clique com o botão direito e abra em uma nova janela ou desactive o bloqueio de pop-ups.

Download para Telemóvel Nokia em Português

Manual do Utilizador para o Nokia PC Suite 6.82 em Português.  [Faça o download agora](#)

Nokia_PC_Suite_682_UG_por.pdf
15-Nov-06 | Version 1.0 | 378 KB | Portuguese (Portugal)

“Legacy”: ninguém gosta, mas temos de lidar com ela.

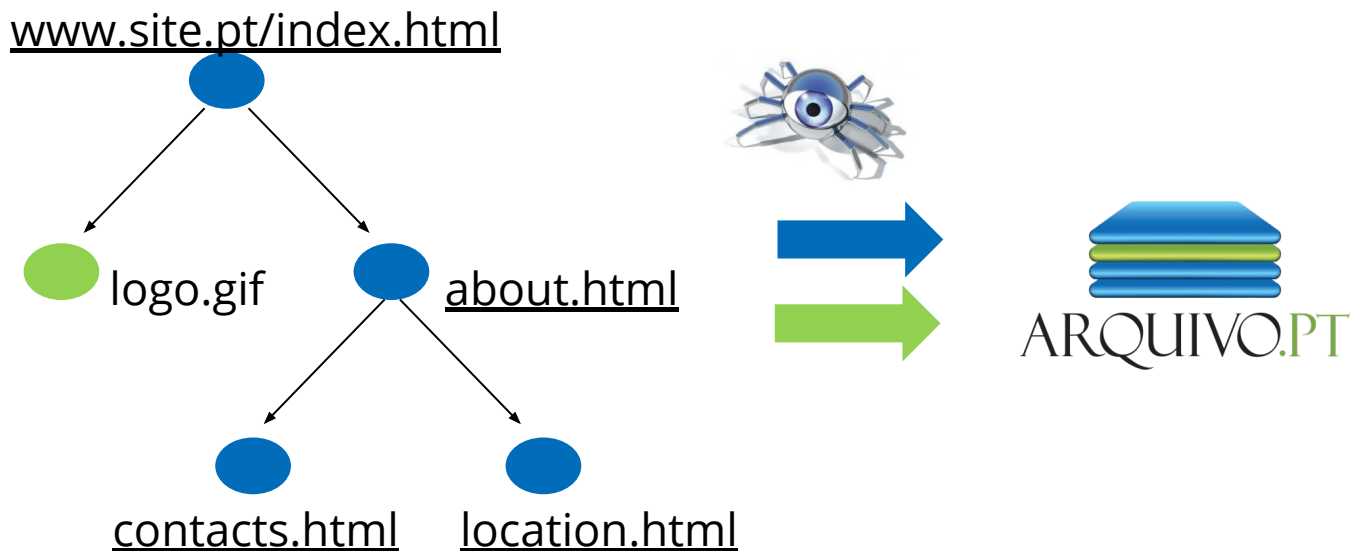
Como funciona o Arquivo.pt?

Funcionamento do Arquivo.pt é maioritariamente automatizado



Informação é recolhida automaticamente da Web

- Conjunto de endereços iniciais (e.g. páginas de entrada)
- Recolha automática e iterativa de ficheiros hiperligados



As APIs do Arquivo.pt

O Arquivo.pt nos tempos do Big Data

- Eu quero processar:
 - muitos dados
 - de um período de tempo específico
 - de um site específico
- O Arquivo.pt possui **+10.000 milhões de páginas** e **+31 milhões de imagens** guardadas e disponíveis para pesquisa
- Será que existe uma **alternativa** a fazer o processo de consulta e recolha manual?

Escala dos dados do Arquivo

- 10,318,854,181 de ficheiros recolhidos
 - ~31,000,000 imagens indexadas
 - 600-700,000,000 no início do próximo ano
- 662 TB de dados (comprimidos) capturados
- Processados em 90 servidores
 - 30 no Hadoop

As APIs do Arquivo

- Estes processos podem ser mapeados diretamente para as nossas APIs:
 - Consulta de histórico de páginas arquivadas
 - Pesquisa de texto e metadados
 - Pesquisa de imagens

As APIs do Arquivo

- Estes processos podem ser mapeados diretamente para as nossas APIs:
 - Consulta de histórico de páginas arquivadas -> **Text Search API**/Memento/CDX Server
 - Pesquisa de texto e metadados -> **Text Search API**
 - Pesquisa de imagens -> **Image Search API**

<https://arquivo.pt/api>

API para consulta de histórico de uma página

```
serviceName: "Arquivo.pt - the Portuguese web-archive"
linkToService: "https://arquivo.pt"
▼ next_page: "https://arquivo.pt/nutchwax/textsearch?versionHistory=www.igot.ulisboa.pt&offset=50"
▼ previous_page: "https://arquivo.pt/nutchwax/textsearch?versionHistory=www.igot.ulisboa.pt&offset=0"
  estimated_nr_results: "119"
▼ request_parameters:
  from: "19960101000000"
  to: "20201231235959"
▼ response_items:
▼ 0:
  ▼ title: "IGOT – Instituto de Geografia e Ordenamento do Território"
  originalURL: "http://www.igot.ulisboa.pt/"
  ▼ linkToArchive: "https://arquivo.pt/wayback/20190608075150/http://www.igot.ulisboa.pt/"
    timestamp: "20190608075150"
    contentLength: "33492"
    digest: "CSD6QKQYKCGE6NXUQA20FNPXURQM306B"
    mimeType: "text/html"
  ▼ linkToScreenshot: "https://arquivo.pt/screenshot?url=https%3A%2F%2Farquivo.pt%2FnoFrame%2Fpreplay%2F20190608075150%2Fhttp%3A%2F%2Fwww.igot.ulisboa.pt%2F"
    date: "1559980310"
    encoding: "UTF-8"
  ▼ linkToNoFrame: "https://arquivo.pt/noFrame/replay/20190608075150/http://www.igot.ulisboa.pt/"
    status: "200"
    collection: "EAWP2320190608"
  ▼ linkToExtractedText: "https://arquivo.pt/textextracted?m=http%3A%2F%2Fwww.igot.ulisboa.pt%2F%2F20190608075150"
  ▼ linkToMetadata: "https://arquivo.pt/textsearch?metadata=http%3A%2F%2Fwww.igot.ulisboa.pt%2F%2F20190608075150"
  ▼ linkToOriginalFile: "https://arquivo.pt/noFrame/replay/20190608075150id_/http://www.igot.ulisboa.pt/"
▼ 1:
  ▼ title: "IGOT – Instituto de Geografia e Ordenamento do Território"
  originalURL: "http://www.igot.ulisboa.pt/"
  ▼ linkToArchive: "https://arquivo.pt/wayback/20190523044005/http://www.igot.ulisboa.pt/"
    timestamp: "20190523044005"
    contentLength: "33600"
```

[https://arquivo.pt/textsearch?
versionHistory=www.igot.ulisboa.pt](https://arquivo.pt/textsearch?versionHistory=www.igot.ulisboa.pt)

API para consulta de histórico de uma página

[https://arquivo.pt/textsearch?
versionHistory=www.igot.ulisboa.pt&
maxItems=50&
offset=50&
from=20010224174130&
to=2019](https://arquivo.pt/textsearch?versionHistory=www.igot.ulisboa.pt&maxItems=50&offset=50&from=20010224174130&to=2019)

- **textsearch**
 - Serviço requisitado
- **versionHistory=www.igot.ulisboa.pt**
 - URL a requisitar histórico
- **offset=50**
 - Começar no resultado 50
- **maxItems=50**
 - Número de resultados pedidos
- **from=20010224174130**
 - resultados a partir de 24/02/2001 às 17:41:30
- **to=2019**
 - resultados até 2019

```
serviceName: "Arquivo.pt - the Portuguese web-archive"
linkToService: "https://arquivo.pt"
▼ next_page: "https://arquivo.pt/nutchwax/textsearch?versionHistory=www.igot.ulisboa.pt&maxItems=50&offset=100&from=20010224174130&to=2019"
▼ previous_page: "https://arquivo.pt/nutchwax/textsearch?versionHistory=www.igot.ulisboa.pt&maxItems=50&offset=0&from=20010224174130&to=2019"
estimated_nr_results: "116"
▼ request_parameters:
  offset: "50"
  maxItems: "50"
  from: "20010224174130"
  to: "201900000000000"
▼ response_items:
  ▼ 0:
    title: "http://www.igot.ulisboa.pt/"
    originalURL: "http://www.igot.ulisboa.pt/"
    ▼ linkToArchive: "https://arquivo.pt/wayback/20181111005825/http://www.igot.ulisboa.pt/"
    tstamp: "20181111005825"
    contentLength: "34732"
    digest: "3VCAJPV3FRCET73XMZUFVRLY6TD03CWR"
    mimeType: "text/html"
    ▼ linkToScreenshot: "https://arquivo.pt/screenshot/?url=https%3A%2F%2Farquivo.pt%2FnoFrame%2Fpreplay%2F20181111005825%2Fhttp%3A%2F%2Fwww.igot.ulisboa.pt%2F"
    date: "1541897905"
    encoding: ""
    ▼ linkToNoFrame: "https://arquivo.pt/noFrame/replay/20181111005825/http://www.igot.ulisboa.pt/"
    status: "200"
    collection: ""
    linkToExtractedText: ""
    ▼ linkToMetadata: "https://arquivo.pt/textsearch?metadata=http%3A%2F%2Fwww.igot.ulisboa.pt%2F%2F20181111005825"
    ▼ linkToOriginalFile: "https://arquivo.pt/noFrame/replay/20181111005825id_/http://www.igot.ulisboa.pt/"
  ▼ 1:
    title: "http://www.igot.ulisboa.pt/"
    originalURL: "http://www.igot.ulisboa.pt/"
    ▼ linkToArchive: "https://arquivo.pt/wayback/20181110010634/http://www.igot.ulisboa.pt/"
    tstamp: "20181110010634"
    contentLength: "34732"
```

Campos dos resultados

<https://arquivo.pt/textsearch?versionHistory=www.igot.ulisboa.pt&maxItems=50&offset=50&from=20010224174130&to=2019>

- **linkToExtractedText**
 - todo o texto extraído da página
- **linkToScreenshot**
 - imagem da página
- **linkToOriginalFile**
 - link para ficheiro original

```

serviceName: "Arquivo.pt - the Portuguese web-archive"
linkToService: "https://arquivo.pt"
next_page: "https://arquivo.pt/nutchwax/textsearch?versionHistory=www.igot.ulisboa.pt&maxItems=50&offset=100&from=20010224174130&to=2019"
previous_page: "https://arquivo.pt/nutchwax/textsearch?versionHistory=www.igot.ulisboa.pt&maxItems=50&offset=0&from=20010224174130&to=2019"
estimated_nr_results: "116"
request_parameters:
  offset: "50"
  maxItems: "50"
  from: "20010224174130"
  to: "20190000000000"
response_items:
  0:
    title: "http://www.igot.ulisboa.pt/"
    originalURL: "http://www.igot.ulisboa.pt/"
    linkToArchive: "https://arquivo.pt/wayback/20181111005825/http://www.igot.ulisboa.pt/"
    tstamp: "20181111005825"
    contentLength: "34732"
    digest: "3VCAJPV3FCET73XMZUFVRLY6TD03CWR"
    mimeType: "text/html"
    linkToScreenshot: "https://arquivo.pt/screenshot/?url=https%3A%2F%2Farquivo.pt%2FnoFrame%2Fpreplay%2F20181111005825%2Fhttp%3A%2F%2Fwww.igot.ulisboa.pt%2F"
    date: "1541897905"
    encoding: ""
    linkToNoFrame: "https://arquivo.pt/noFrame/replay/20181111005825/http://www.igot.ulisboa.pt/"
    status: "200"
    collection: ""
    linkToExtractedText: ""
    linkToMetadata: "https://arquivo.pt/textsearch?metadata=http%3A%2F%2Fwww.igot.ulisboa.pt%2F%2F20181111005825"
    linkToOriginalFile: "https://arquivo.pt/noFrame/replay/20181111005825id_/http://www.igot.ulisboa.pt/"
  1:
    title: "http://www.igot.ulisboa.pt/"
    originalURL: "http://www.igot.ulisboa.pt/"
    linkToArchive: "https://arquivo.pt/wayback/20181110010634/http://www.igot.ulisboa.pt/"
    tstamp: "20181110010634"
    contentLength: "34732"
  
```

Consulta de texto extraído de uma página

IGOT – Instituto de Geografia e Ordenamento do Território Ensino Licenciaturas Licenciatura em Geografia Licenciatura em Planeamento e Gestão do Território Licenciatura em Estudos Europeus Mestrados Geografia Física e Ordenamento do Território Ordenamento do Território e Urbanismo Políticas Europeias: Desenvolvimento e Coesão Socioterritorial Geografia Humana: Globalização, Sociedade e Território SIG e Modelação Territorial Aplicados ao Ordenamento Ensino de Geografia Turismo e Comunicação Doutoramentos Doutoramento em Geografia Doutoramento em Migrações Doutoramento em Território, Risco e Políticas Públicas Doutoramento em Turismo Doutoramento em Estudos de Desenvolvimento Mobilidade Nacional e Internacional Provas Académicas Centro de Formação Projeto Nós Propomos! Cidadania e Inovação na Educação Geográfica Investigação Candidaturas Condições de candidatura 2018/2019 Estudantes Internacionais – 2018/2019 Licenciaturas – 2018/2019 Concursos Especiais de Acesso 2018/2019 Regimes de Reingresso e Mudança de par estabelecimento/curso 2018/2019 Mestrados – 2018/2019 Doutoramentos – 2018/2019 Unidades Curriculares Isoladas 2018/2019 Informações e Serviços Centro de Documentação do IGOT Biblioteca Mapoteca Fototeca Serviços Online Avisos e prazos de atos académicos Formulários e Regulamentos Informações, Calendários e Horários – 2018/2019 Informações, Calendários e Horários – 2017/2018 Serviços de Apoio – Unidade de Gestão de Recursos Humanos O IGOT Corpo Docente Estatutos Estrutura Órgãos Associação de Estudantes Aluguer de espaços Loading... Faculdade TOP100 Mundial Rankings QS e Shanghai: IGOT é N.º 1 em Portugal e TOP100 do Mundo IGOT no TOP 100 Mundial Faculdade de Geografia e Ordenamento n.º 1 em Portugal Os cursos do IGOT são aliciantes Afirmam talentos, impulsionam o sucesso e promovem a inserção no mercado de trabalho Focado no sucesso dos estudantes O IGOT dispõe de um edifício totalmente equipado para o ensino de Geografia e Planeamento do Território No IGOT, as ferramentas para o sucesso são grátis! ARCGIS, SPSS e 5 licenças oficiais e completas do Office 365 igot 2017-12-13T15:09:36+00:00 Eventos Dia Aberto da Geografia 4 Janeiro, 2019 Inscrições abertas para o Projeto Nós propomos 2018/19 29 Abril, 2019 Ver todos os eventos Provas académicas Mestrado em Gestão do Território e Urbanismo. Discussão pública do Relatório de Estágio de Bernardo Veloso Barros “ Coesão Territorial em Portugal - O contributo dos equipamentos educativos no período 2007-2013 ” 9 Novembro, 2018, 11:00, Sala de Conferências (IGOT) Mestrado em Turismo e Comunicação. Discussão pública do Relatório de Estágio de Solange Amorim Fernandes “ Parques de Sintra – Monte da Lua: a influência da gestão patrimonial na experiência turística ” 12 Novembro, 2018, 10:30, Sala de Conferências (IGOT) Mestrado em SIG e Modelação Territorial Aplicados ao Ordenamento. Discussão pública da Dissertação de Pedro António Faria Freitas “ Avaliação de novas plataformas de deteção remota para monitorização de lagoas termocársicas (subártico canadiano) ” 12 Novembro, 2018, 15:00, Sala de Conferências (IGOT) Mestrado em Geografia Física e Ordenamento Território. Discussão pública da Dissertação de Cláudia Estêvão dos Reis “ A eficiência térmica no potencial de arrefecimento de dois espaços verdes em Lisboa ” 13 Novembro, 2018, 10:30, Sala de Conferências (IGOT) Mestrado em Geografia Física e Ordenamento do Território. Discussão pública da Dissertação de Márcia Alexandra Neves Matias “ O balanço radiativo de um conjunto de edifícios em Telheiras com recurso a termografia infravermelha ” 20 Novembro, 2018, 15:00, Sala de Conferências (IGOT) Mestrado em Geografia Física e Ordenamento do Território. Discussão pública da Dissertação de Raquel Alexandra dos Santos Fernandes “ Avaliação da posibilidade e vulnerabilidade associadas a condições biometeorológicas adversas

<https://arquivo.pt/textextracted?>

<m=http://www.igot.ulisboa.pt/20181109014435>

API de Pesquisa de texto

[https://arquivo.pt/textsearch?
q=igot&
maxItems=50&
offset=50&
from=20100402000000&
to=2019](https://arquivo.pt/textsearch?q=igot&maxItems=50&offset=50&from=20100402000000&to=2019)

- **textsearch**
 - Serviço requisitado
- **q=igot**
 - Texto a pesquisar
- **offset=50**
 - Começar no resultado 50
- **maxItems=50**
 - Número de resultados pedidos
- **from=20100402000000**
 - de 02/04/2010 às 00:00:00
- **to=2019**
 - até 2019

```
serviceName: "Arquivo.pt - the Portuguese web-archive"
linkToService: "https://arquivo.pt"
▼ next_page: "https://arquivo.pt/nutchwax/textsearch?q=igot&maxItems=50&offset=100&from=20100402000000&to=2019"
▼ previous_page: "https://arquivo.pt/nutchwax/textsearch?q=igot&maxItems=50&offset=0&from=20100402000000&to=2019"
estimated_nr_results: "177029"
▼ request_parameters:
  q: "igot"
  offset: "50"
  maxItems: "50"
  from: "20100402000000"
  to: "20190000000000"
▼ response_items:
  ▼ 0:
    title: "ULisboa - VII CONFERÊNCIA ANUAL DO IGOT"
    originalURL: "http://live.fccn.pt/ulisboa/igot/"
    ▼ linkToArchive: "https://arquivo.pt/wayback/20151127113743/http://live.fccn.pt/ulisboa/igot/"
    timestamp: "20151127113743"
    contentLength: "10922"
    digest: "1c55ae2094b682e821a0a294cdaa0f9a"
    mimeType: "text/html"
    ▼ linkToScreenshot: "https://arquivo.pt/screenshot?url=https%3A%2F%2Farquivo.pt%2FnoFrame%2Fpreplay%2F20151127113743%2Fhttp%3A%2F%2Flive.fccn.pt%2Fulisboa%2Figot%2F"
    date: "1448624263"
    encoding: "iso-8859-1"
    ▼ linkToNoFrame: "https://arquivo.pt/noFrame/replay/20151127113743/http://live.fccn.pt/ulisboa/igot/"
    ▼ snippet: "ULisboa - VII CONFERÊNCIA ANUAL DO <em>IGOT</em> VII CONFERÊNCIA ANUAL DO <em>IGOT</em> Data Emiss&atilde;o: 21 de maio de 2015 Hor&aacute;cuta;rio: 09h00 &#150; 19h00 As cidades constituem umas das constru&ccedil;es mais impressionantes da sociedade com express&atilde;o generalizada a todos os espa&ccedil;os do globo, incluindo mesmo os mais remotos, que<span class="\"ellipsis\"> ... </span>"
    "AMP19"
    ▼ linkToExtractedText: "https://arquivo.pt/textextracted?m=http%3A%2F%2Flive.fccn.pt%2Fulisboa%2Figot%2F2F20151127113743"
    ▼ linkToMetadata: "https://arquivo.pt/textsearch?metadata=http%3A%2F%2Flive.fccn.pt%2Fulisboa%2Figot%2F2F20151127113743"
    ▼ linkToOriginalFile: "https://arquivo.pt/wayback/noFrame/replay/20151127113743id_/http://live.fccn.pt/ulisboa/igot/"
  ▼ 1:
    title: "Conferência no I.G.O.T."
    ▼ originalURL: "http://www.fa.ulisboa.pt/index.php?view=article&catid=164%3Apresidente-depachos&id=3176%3Aconferencia-no-i-format-pdf&option=com_content&Itemid=272&lang=pt"
    ▼ linkToArchive: "https://arquivo.pt/wayback/20150913080020/http://www.fa.ulisboa.pt/index.php?view=article&catid=164%3Apresidente-depachos&id=3176%3Aconferencia-no-igot&format-pdf&option=com_content&Itemid=272&lang=pt"
    timestamp: "20150913080020"
```

Campos dos resultados

[https://arquivo.pt/textsearch?
q=igot&
maxItems=50&
offset=50&
from=20100402000000&
to=2019](https://arquivo.pt/textsearch?q=igot&maxItems=50&offset=50&from=20100402000000&to=2019)

- **linkToExtractedText**
 - todo o texto extraído da página
- **linkToScreenshot**
 - imagem da página
- **linkToOriginalFile**
 - link para ficheiro original

(semelhante à consulta de histórico)

```
serviceName: "Arquivo.pt - the Portuguese web-archive"
linkToService: "https://arquivo.pt"
▼ next_page: "https://arquivo.pt/nutchwax/textsearch?q=igot&maxItems=50&offset=100&from=20100402000000&to=2019"
▼ previous_page: "https://arquivo.pt/nutchwax/textsearch?q=igot&maxItems=50&offset=0&from=20100402000000&to=2019"
estimated_nr_results: "177029"
▼ request_parameters:
  q: "igot"
  offset: "50"
  maxItems: "50"
  from: "20100402000000"
  to: "20190000000000"
▼ response_items:
  ▼ 0:
    title: "ULisboa - VII CONFERÊNCIA ANUAL DO IGOT"
    originalURL: "http://live.fccn.pt/ulisboa/igot/"
    ▼ linkToArchive: "https://arquivo.pt/wayback/20151127113743/http://live.fccn.pt/ulisboa/igot/"
    timestamp: "20151127113743"
    contentLength: "10922"
    digest: "1c55ae2094b682e821a0a294cdaa0f9a"
    mimeType: "text/html"
    ▼ linkToScreenshot: "https://arquivo.pt/screenshot?url=https%3A%2F%2Farquivo.pt%2FnoFrame%2Fpreplay%2F20151127113743%2Fhttp%3A%2F%2Flive.fccn.pt%2Fulisboa%2Figot%2F"
    date: "1448624263"
    encoding: "iso-8859-1"
    ▼ linkToNoFrame: "https://arquivo.pt/noFrame/replay/20151127113743/http://live.fccn.pt/ulisboa/igot/"
    ▼ snippet: "ULisboa - VII CONFERÊNCIA ANUAL DO <em>IGOT</em> VII CONFERÊNCIA ANUAL DO <em>IGOT</em> Data Emiss&atilde;o: 21 de maio de 2015 Hor&aacute;cuta;rio: 09h00 &#150; 19h00 As cidades constituem umas das constru&ccedil;es mais impressionantes da sociedade com express&atilde;o generalizada a todos os espa&ccedil;os do globo, incluindo mesmo os mais remotos, que<span class="\"ellipsis\" > ... </span>"
    collection: "AMP19"
    ▼ linkToExtractedText: "https://arquivo.pt/textextracted?m=http%3A%2F%2Flive.fccn.pt%2Fulisboa%2Figot%2F%2F20151127113743"
    ▼ linkToMetadata: "https://arquivo.pt/textsearch?metadata=http%3A%2F%2Flive.fccn.pt%2Fulisboa%2Figot%2F%2F20151127113743"
    ▼ linkToOriginalFile: "https://arquivo.pt/wayback/noFrame/replay/20151127113743id_/http://live.fccn.pt/ulisboa/igot/"
  ▼ 1:
    title: "Conferência no I.G.O.T."
    ▼ originalURL: "http://www.fa.ulisboa.pt/index.php?view=article&catid=164%3Apresidente-depachos&id=3176%3Aconferencia-no-i-format=pdf&option=com_content&Itemid=272&lang=pt"
    ▼ linkToArchive: "https://arquivo.pt/wayback/20150913080020/http://www.fa.ulisboa.pt/index.php?view=article&catid=164%3Apresidente-depachos&id=3176%3Aconferencia-no-igot&format=pdf&option=com_content&Itemid=272&lang=pt"
    timestamp: "20150913080020"
```

API de Pesquisa de imagens

[https://arquivo.pt/imagesssearch?
q=igot&
size=md&type=jpg&
from=20150101235959&to=20190101235959&
offset=5&maxItems=100&
more=imgThumbnailBase64](https://arquivo.pt/imagesssearch?q=igot&size=md&type=jpg&from=20150101235959&to=20190101235959&offset=5&maxItems=100&more=imgThumbnailBase64)

- **imagesssearch**
 - Serviço requisitado
- **q=igot**
 - Texto a pesquisar
- **from=20150101235959 &to=20190101235959**
 - mostrar resultados de 2015 a 2010
- **offset=5 &maxItems=100**
 - Começar no resultado 5, devolver 100 resultados
- **type=jpg &size=md**
 - Pedir imagens em JPG de tamanho médio
- **more=imgThumbnailBase64.....**
 - campos de saída requisitados

```

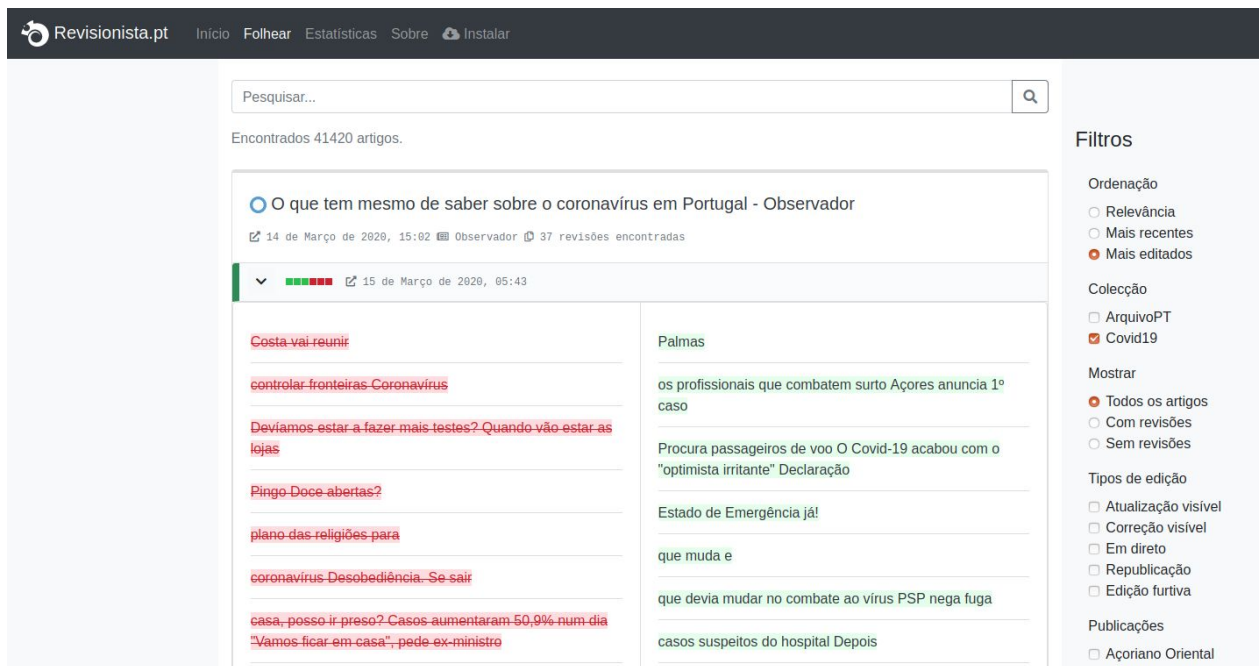
serviceName: "Arquivo.pt - image search service."
linkToService: "https://arquivo.pt/images.jsp"
▶ linkToMoreFields: "https://arquivo.pt/image...pageHost,pageImages,safe"
nextPage: "https://arquivo.pt/image...humbnailBase64&offset=49"
▶ previousPage: "https://arquivo.pt/image...ThumbnailBase64&offset=0"
totalItems: 49
numberOfResponseItems: 44
offset: 5
▼ responseItems:
  ▼ 0:
    ▶ imgAlt: "Prof. Gonalo Vieira int...ic research in Portugal"
    imgMimeType: "image/jpeg"
    imgTstamp: 20180719185629
    pageTstamp: 20181014010543
    ▶ pageTitle: "Prof. Oliver Bakewell at...T on April, 21st - IGOT"
    ▶ pageURL: "http://www.igot.ulisboa...t-on-april-21st/?lang=en"
    imgHeight: 371
    collection: "AWP28"
    ▶ imgSrc: "http://www.igot.ulisboa...2017/04/GVieira_IGOT.jpg"
    ▶ imgThumbnailBase64: "/9j/4AAQSkZJRgABAQAAQAB...r1Fj6Mm3GS/EcbJWR//2Q=="
    imgWidth: 579
    ▶ imgLinkToArchive: "https://arquivo.pt/wayba...2017/04/GVieira_IGOT.jpg"
    ▶ pageLinkToArchive: "https://arquivo.pt/wayba...t-on-april-21st/?lang=en"
  ▼ 1:
    imgTstamp: 20160214200636
    imgWidth: 448
    ▶ imgThumbnailBase64: "/9j/4AAQSkZJRgABAQAAQAB...gAooooAKKKKACiigD/2Q=="
    pageTstamp: 20160212072542
    ▶ pageTitle: "Pavilhão do Conhecimento...Quando o ao nos falha"
    ▶ pageURL: "http://centros.cvtv.pt/n...as/index.asp?id_obj=3195"
    collection: "AWP28"
  
```


Exemplos de utilização



- Conta-me Histórias: <http://contamehistorias.pt/arquivopt/>
 - APIs Usadas: Pesquisa de texto filtrada por domínios de sites de notícias

Exemplos de utilização

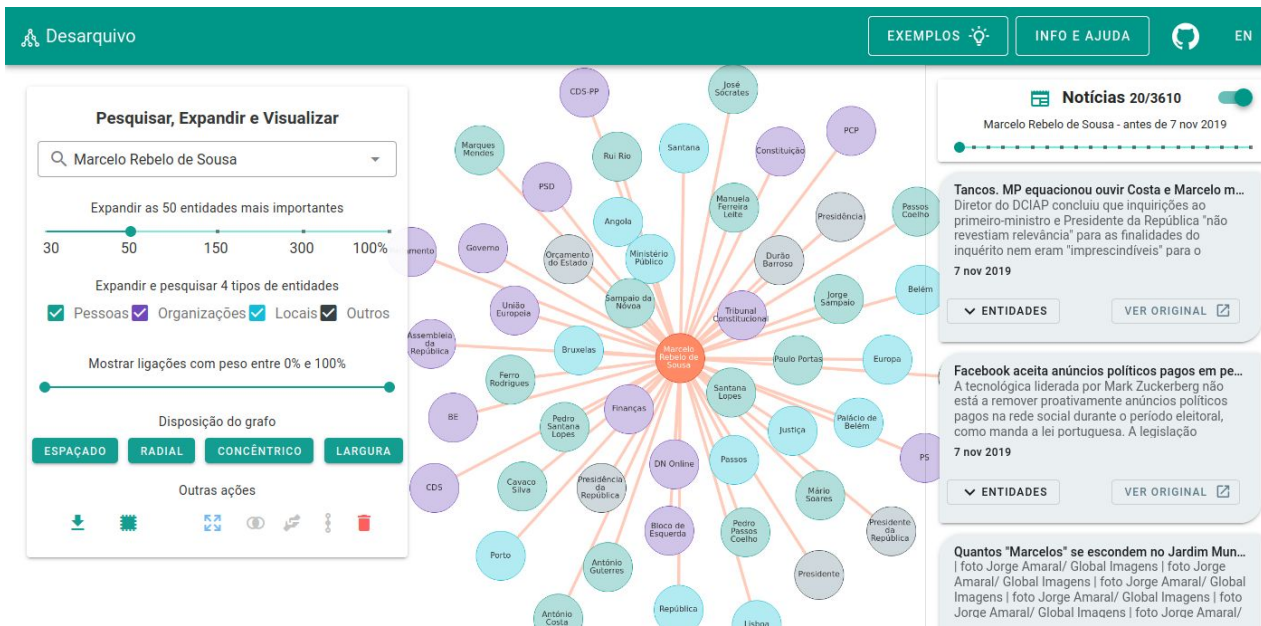


The screenshot shows the Revisionista.pt website. The header includes the logo and navigation links: Início, Folhear, Estatísticas, Sobre, and Instalar. A search bar at the top contains the text "Pesquisar...". Below the search bar, it states "Encontrados 41420 artigos." The main content area displays a list of articles. The first article is titled "O que tem mesmo de saber sobre o coronavírus em Portugal - Observador" and is dated "14 de Março de 2020, 15:02". Below this, there is a section for "15 de Março de 2020, 05:43" which lists several headlines in red text, such as "Costa vai reunir", "controlar fronteiras Coronavirus", and "Devíamos estar a fazer mais testes? Quando vão estar as lojas". To the right of these headlines, there are more articles listed in green text, including "Palmas", "os profissionais que combatem surto Açores anuncia 1º caso", and "Procura passageiros de voo O Covid-19 acabou com o 'optimista irritante' Declaração". On the right side of the page, there is a "Filtros" (Filters) section with options for "Ordenação" (Ordering), "Colecção" (Collection), "Mostrar" (Show), "Tipos de edição" (Types of editing), and "Publicações" (Publications).

- Revisionista.PT: <https://revisionista.pt>

- APIs Usadas: Consulta de histórico de versões arquivadas

Exemplos de utilização



- Desarquivo: <https://msramalho.github.io/desarquivo/>
 - APIs Usadas: Pesquisa de texto filtrada e mostra de páginas capturadas

Exemplos de utilização

*(insira a sua
ideia aqui)*

Prémio Arquivo.pt
2020

Concorra até 4 de maio

VIAJE NO TEMPO

e ganhe 10.000€



1º Classificado **10.000€**
2º Classificado **3.000€**
3º Classificado **2.000€**

Crie um trabalho individual ou em grupo
que use o **Arquivo.pt**

arquivo.pt/premio2020

Demonstração

<https://github.com/arquivo/demo-extrair-entidades/>

Non-GPE locations, mountain ranges, bodies of water:

IGOT; Instituto de Geografia e Ordenamento do Território Ensino Licenciaturas; Geografia Licenciatura; Planeamento; Gestão do Território Licenciatura em Estudos Europeus; Ordenamento Ensino de Geografia Turismo e Comunicação Doutoramentos Doutoramento; Geografia Doutoramento; Migrações Doutoramento; Território; Risco; Políticas Públicas Doutoramento em Turismo; ...

Companies, agencies, institutions, etc.:

Internacional Provas Académicas; Estudantes Internacionais; Associação de Estudantes Aluguer; Faculdade TOP100 Mundial Rankings QS; Novembro; Dissertação; IGOT; Notícias; Prof; Unidade de Investigação Científica em Geografia do país; Centro de Estudos Geográficos; CEG; Fundação para a Ciência e Tecnologia; Serviços Financeiros

Named person or family.:

Bernardo Veloso Barros; Pedro António Faria Freitas; Alexandra Neves Matias; Raquel Alexandra; Carlos José Simões Alves; Isabel André; Mário Vale; Professor Jorge Gaspar; Excelente; Website

Mais informações

- APIs Disponíveis em <https://arquivo.pt/api>
- **Código Aberto** e **Acesso Aberto** para todos!
 - Licenças e copyrights dos dados continuam a pertencer aos criadores originais
- Incentivamos o uso das nossas APIs para **investigação colaborativa** e **interdisciplinar**
- Em caso de **dúvidas** ou **problemas**:
 - Enviar mensagem para o formulário de contacto do Arquivo
 - <https://sobre.arquivo.pt/pt/contacto/>
 - Abrir *issue* no GitHub
 - <https://github.com/arquivo/pwa-technologies/issues>

Questões?

andre.mourao@fccn.pt

Inscrevam-se na lista de email para não perder
as novidades mais importantes!

<https://arquivo.pt/inscrever>