

WORKSHOP

O Arquivo.pt e a preservação da memória digital da Internet

19 de abril | 09h00 - 13h00

ARDITI, piso -1, Ursa Menor 2



API Arquivo.pt

Acesso e processamento
automático da informação
preservada da Web

O que é uma API?

“Application Programming Interface”

Acesso automático

Fácil integração

Elimina a necessidade de conhecer o código núcleo

Os dados do Arquivo.pt

- Os dados do Arquivo.pt podem ser explorados de várias maneiras:
 - Pesquisa de texto e metadados
 - Pesquisa de imagens
 - Consulta de histórico de páginas arquivadas
- 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

Pesquisa de texto

Termos de pesquisa: Arditi



The screenshot shows the ARQUIVO.PT search interface. At the top, the search bar contains the term 'Arditi' and a 'Pesquisar' button. Below the search bar, a date range filter is set from 1991 (6 Ago) to 2017 (19 Abr). There are buttons for 'Páginas', 'Imagens', 'Narrativa', and 'Pesquisa avançada'. The search results section displays two entries from www.arditi.pt:

- ARDITI**
21 Fevereiro 2017
ARDITI Apoios Bolsas FDCTI-RAM Concursos Bolsas FDCTI-RAM Contratação de Doutorados Projetos ... Documentos Ligações Institucionais Oportunidades de financiamento ARDITI Natureza Estrutura Associados ... disponíveis em <http://oom.arditi.pt/>. O pedido de esclarecimento de dúvidas deverá ser feito através do e ...
- ARDITI**
23 Abril 2015

Pesquisa de texto - Pesquisa avançada

Termos de pesquisa: Arditi

Data entre 1 de Janeiro de 2000 e 1 de Janeiro de 2020

Formato: pdf

Menu **ARQUIVO.PT**

Tenham estas **palavras**:

ex. expo 98

Tenham esta **expressão ou frase**:

ex. pavilhão da utopia

Não tenham estas palavras:

ex. concerto

Data:

2000 1 Jan ————— 2020 1 Jan

Formato:

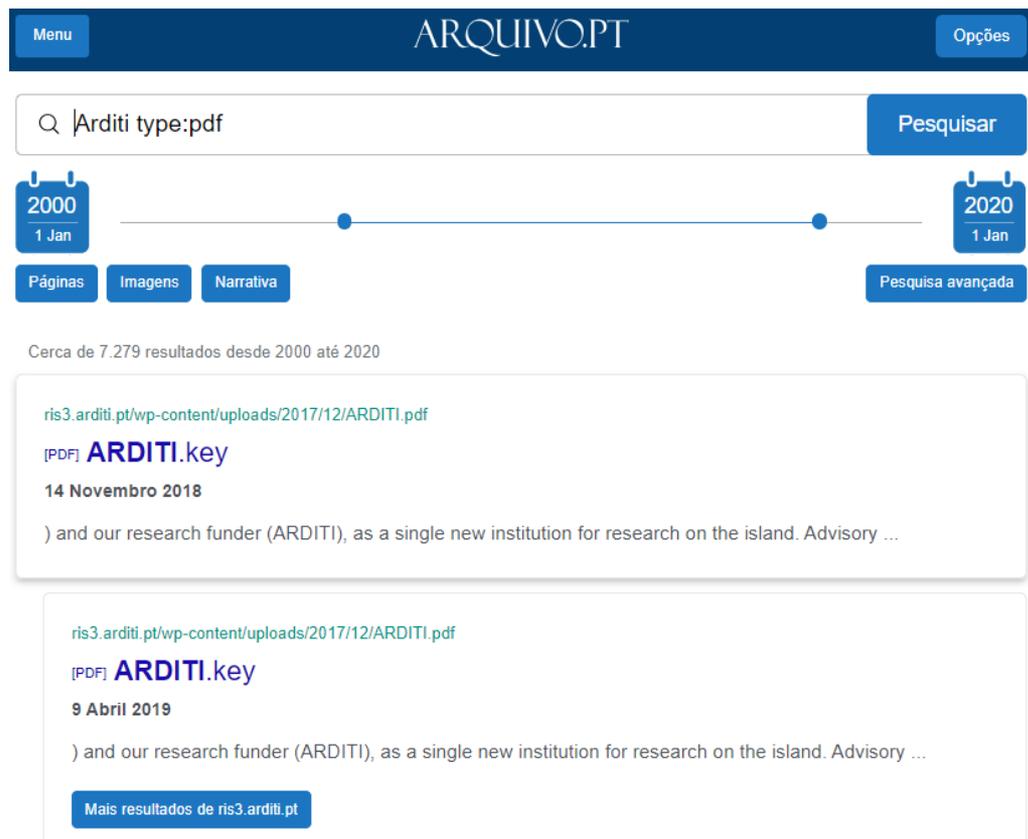
<input type="checkbox"/> Todos os formatos	<input type="checkbox"/> Microsoft Excel (.xls)
<input type="checkbox"/> Microsoft PowerPoint (.ppt)	<input checked="" type="checkbox"/> Adobe PDF (.pdf)
<input type="checkbox"/> Microsoft Word (.doc)	<input type="checkbox"/> Adobe PostScript (.ps)
<input type="checkbox"/> HTML (.htm, .html)	<input type="checkbox"/> Rich Text Format (.rtf)

Pesquisa de texto - Pesquisa avançada

Termos de pesquisa: Arditi

Data entre 1 de Janeiro de 2000 e 1 de Janeiro de 2020

Formato: pdf



The screenshot shows 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 'Opções' on the right. Below the header is a search bar containing the text 'Arditi type:pdf' and a 'Pesquisar' button. Underneath the search bar is a date range selector showing '2000 1 Jan' and '2020 1 Jan' with a horizontal timeline between them. Below the date range are three filter buttons: 'Páginas', 'Imagens', and 'Narrativa', and a 'Pesquisa avançada' button. The search results section shows 'Cerca de 7.279 resultados desde 2000 até 2020'. Two results are visible, both from 'ris3.arditi.pt/wp-content/uploads/2017/12/ARDITI.pdf'. The first result is dated '14 Novembro 2018' and the second is dated '9 Abril 2019'. Both results include a '[PDF] ARDITI.key' link and a snippet of text: ') and our research funder (ARDITI), as a single new institution for research on the island. Advisory ...'. At the bottom of the second result is a button that says 'Mais resultados de ris3.arditi.pt'.

Pesquisa de texto - API

```
{
  "serviceName": "Arquivo.pt - the Portuguese web-archive",
  "linkToService": "https://arquivo.pt",
  "next_page": "https://arquivo.pt/textsearch?q=arditi&offset=50",
  "estimated_nr_results": 676019,
  "request_parameters": {
    "offset": 0,
    "dedupValue": 2,
    "dedupField": "site",
    "q": "arditi",
    "maxItems": 50
  },
  "response_items": [
    {
      "title": "ARDITI - Arditi",
      "originalURL": "http://www.arditi.tk/",
      "linkToArchive": "https://arquivo.pt/wayback/20180407222823/http://www.arditi.tk/",
      "tstamp": "20180407222823",
      "contentLength": 79649,
      "digest": "f8a9f5dcaad17675727b73e4116b07c7",
      "mimeType": "text/html",
      "encoding": "UTF-8",
      "date": "1523140103",
      "linkToScreenshot": "https://arquivo.pt/screenshot?url=https%3A%2F%2Fwww.arditi.tk%2F",
      "linkToNoFrame": "https://arquivo.pt/noFrame/replay/20180407222823/http://www.arditi.tk/",
      "linkToExtractedText": "https://arquivo.pt/textextracted?m=http%3A%2F%2Fwww.arditi.tk%2F2F20180407222823",
      "linkToMetadata": "https://arquivo.pt/textsearch?metadata=http%3A%2F%2Fwww.arditi.tk%2F2F20180407222823",
      "linkToOriginalFile": "https://arquivo.pt/noFrame/replay/20180407222823/id/http://www.arditi.tk/",
      "snippet": "<em>ARDITI</em> &#8211; <em>Arditi</em> <em>ARDITI</em> &#8211; <em>Arditi</em> Skip to content DigipakCD and limited edition LP out now on<span class=\"ellipsis\"> ... </span> list to receive updates on <em>ARDITI</em> and related projects. Spearheaded by H. M&ouml;ller and M. Bj&ouml;rkm&ouml;n, the <em>ARDITI</em> torch was set alight in 1997, inspired by the Italian Futurist Movement of the early 20 th<span class=\"ellipsis\"> ... </span>",
      "fileName": "IAH-20180407213215-05217-p82.arquivo.pt",
      "collection": "AMP26",
      "offset": 74982195
    }
  ]
}
```

<https://arquivo.pt/textsearch?q=arditi>

Pesquisa de texto - API

[https://arquivo.pt/textsearch
?q=arditi
&from=2000
&to=20200101
&type=pdf](https://arquivo.pt/textsearch?q=arditi&from=2000&to=20200101&type=pdf)

- **textsearch**
 - Serviço requisitado
- **q=arditi**
 - Termos a pesquisar
- **from=2000**
 - resultados a partir do ano 2000
- **to=20200101**
 - resultados até 1 de Janeiro de 2019
- **type=pdf**
 - Ficheiros do formato pdf

```
{
  "serviceName": "Arquivo.pt - the Portuguese web-archive",
  "linkToService": "https://arquivo.pt",
  "next_page": "https://arquivo.pt/textsearch?q=arditi&from=2000&to=20200101&type=pdf&offset=50",
  "estimated_nr_results": 7279,
  "request_parameters": {
    "offset": 0,
    "dedupValue": 2,
    "from": "2000000000000",
    "to": "20200101000000",
    "type": [
      "pdf"
    ],
    "dedupField": "site",
    "q": "arditi",
    "maxItems": 50
  },
  "response_items": [
    {
      "title": "ARDITI.key",
      "originalURL": "https://ris3.arditi.pt/wp-content/uploads/2017/12/ARDITI.pdf",
      "linkToArchive": "https://arquivo.pt/wayback/20181114050602/https://ris3.arditi.pt/wp-content/",
      "tstamp": "20181114050602",
      "contentLength": 5892724,
      "digest": "473e35ebae4e056cbf9fbec99c27e265",
      "mimeType": "application/pdf",
      "date": "1542171962",
      "linkToScreenshot": "https://arquivo.pt/screenshot?url=https%3A%2F%2Farquivo.pt%2FnoFrame%2Fcontent%2Fuploads%2F2017%2F12%2FARDITI.pdf",
      "linkToNoFrame": "https://arquivo.pt/noFrame/replay/20181114050602/https://ris3.arditi.pt/wp-c
```

Pesquisa de texto - API

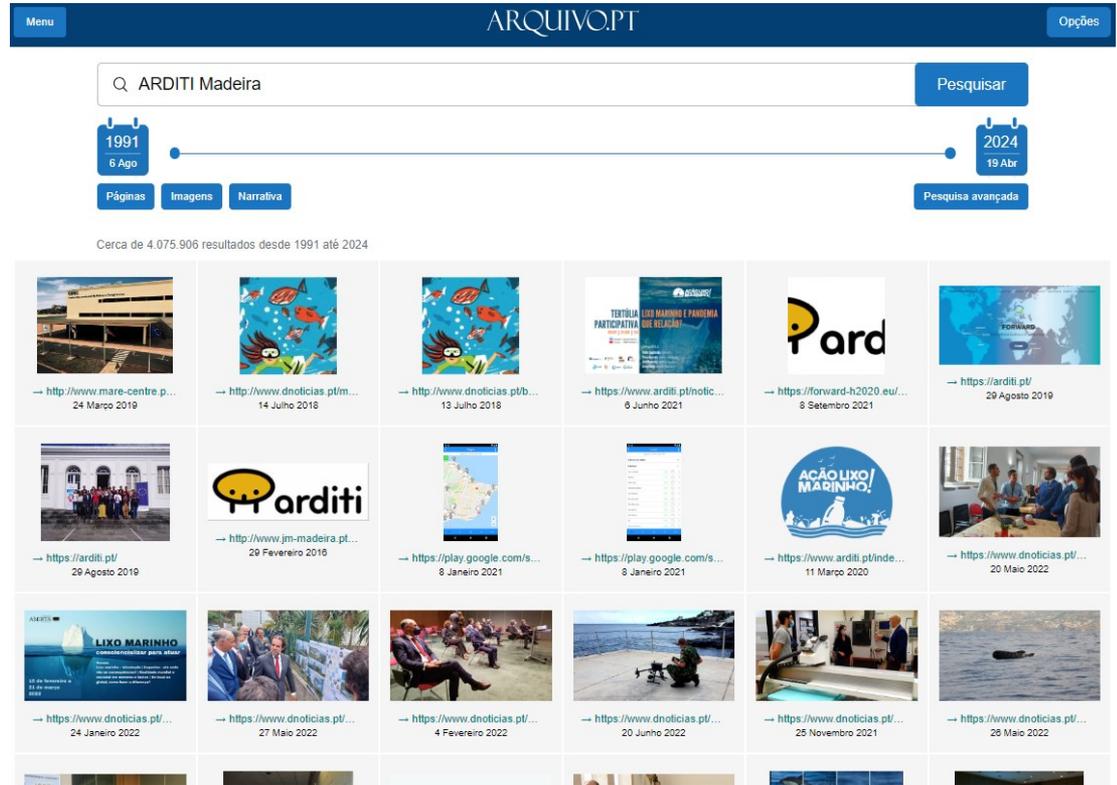
<https://arquivo.pt/textsearch?q=arditi&from=2000&to=20200101&type=pdf>

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

```
--  
"response_items": [  
  {  
    "title": "ARDITI.key",  
    "originalURL": "https://ris3.arditi.pt/wp-content/uploads/2017/12/ARDITI.",  
    "linkToArchive": "https://arquivo.pt/wayback/20181114050602/https://ris3.",  
    "tstamp": "20181114050602",  
    "contentLength": 5892724,  
    "digest": "473e35ebae4e056cbf9fbec99c27e265",  
    "mimeType": "application/pdf",  
    "date": "1542171962",  
    "linkToScreenshot": "https://arquivo.pt/screenshot?url=https%3A%2F%2Farqu",  
    "linkToOriginalFile": "https://arquivo.pt/wayback/20181114050602/https://ris3.",  
    "linkToNoFrame": "https://arquivo.pt/noFrame/replay/20181114050602/https://ris3.",  
    "linkToExtractedText": "https://arquivo.pt/textextracted?m=https%3A%2F%2Farqu",  
    "linkToMetadata": "https://arquivo.pt/textsearch?metadata=https%3A%2F%2Farqu",  
    "linkToOriginalFile": "https://arquivo.pt/noFrame/replay/20181114050602id",  
    "snippet": ") and our \nresearch funder (<em>ARDITI</em>), as a single ne",  
    "fileName": "WEB-20181114050600010-p81.arquivo.pt",  
    "collection": "AWP28",  
    "offset": 12575849  
  },  
]
```

Pesquisa de Imagens

Termo de pesquisa: ARDITI
Madeira



The screenshot shows the ARQUIVO.PT search interface. At the top, there is a navigation bar with 'Menu', 'ARQUIVO.PT', and 'Opções'. Below this is a search bar containing the text 'ARDITI Madeira' and a 'Pesquisar' button. A timeline slider is positioned below the search bar, with markers for '1991 6 Ago' and '2024 19 Abr'. Below the timeline are buttons for 'Páginas', 'Imagens', and 'Narrativa', and a 'Pesquisa avançada' button. The search results are displayed in a grid of 18 items, each with a thumbnail image, a URL, and a date. The results include various images related to ARDITI Madeira, such as buildings, marine life, logos, and people.

Menu ARQUIVO.PT Opções

Q ARDITI Madeira Pesquisar

1991 6 Ago 2024 19 Abr

Páginas Imagens Narrativa Pesquisa avançada

Cerca de 4.075.906 resultados desde 1991 até 2024

 → http://www.mare-centre.p... 24 Março 2019	 → http://www.dnocias.pt/m... 14 Julho 2018	 → http://www.dnocias.pt/m... 13 Julho 2018	 → https://www.arditi.pt/notic... 6 Junho 2021	 → https://forward-h2020.eu/... 8 Setembro 2021	 → https://arditi.pt/ 29 Agosto 2019
 → https://arditi.pt/ 29 Agosto 2019	 → http://www.jm-madeira.pt... 29 Fevereiro 2016	 → https://play.google.com/s... 8 Janeiro 2021	 → https://play.google.com/s... 6 Janeiro 2021	 → https://www.arditi.pt/inde... 11 Março 2020	 → https://www.dnocias.pt/... 20 Maio 2022
 → https://www.dnocias.pt/... 24 Janeiro 2022	 → https://www.dnocias.pt/... 27 Maio 2022	 → https://www.dnocias.pt/... 4 Fevereiro 2022	 → https://www.dnocias.pt/... 20 Junho 2022	 → https://www.dnocias.pt/... 25 Novembro 2021	 → https://www.dnocias.pt/... 26 Maio 2022

Pesquisa de Imagens - Pesquisa avançada

Termo de pesquisa: ARDITI
Madeira

Data entre 6 de Agosto de
2013 e a data de hoje

Imagens no formato PNG

Imagens provenientes do
website arditi.pt

Menu **ARQUIVO.PT**

Não relacionadas com estas palavras:

ex. concerto

Data:

2013 6 Ago ● ● 2024 19 Abr

Imagens com o formato:

<input type="checkbox"/> Todos os formatos	<input type="checkbox"/> Graphics Interchange Format (.gif)
<input type="checkbox"/> Joint Photographic Experts Group (.jpg)	<input type="checkbox"/> Bitmap Image File (.bmp)
<input checked="" type="checkbox"/> Portable Network Graphics (.png)	<input type="checkbox"/> WEBP (.webp)

Imagens:

Mostrar todas ▼

Pesquisa segura:

Excluir impróprias ▼

Sítios Web, separados por vírgulas:

ex. www.fcn.pt, www.fct.pt

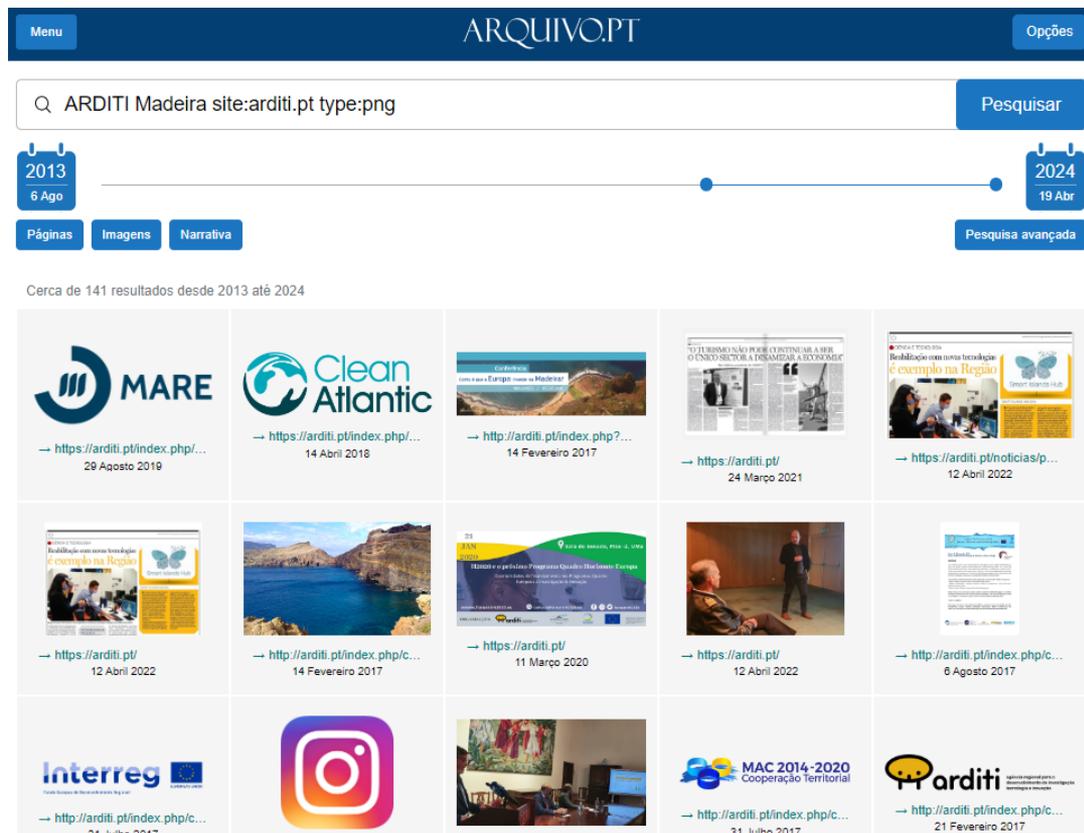
Pesquisa de Imagens - Pesquisa avançada

Termo de pesquisa: ARDITI Madeira

Data entre 6 de Agosto de 2013 e a data de hoje

Imagens no formato PNG

Imagens provenientes do website arditi.pt



Menu ARQUIVO.PT Opções

Q ARDITI Madeira site:arditi.pt type:png Pesquisar

2013 6 Ago 2024 19 Abr

Páginas Imagens Narrativa Pesquisa avançada

Cerca de 141 resultados desde 2013 até 2024

 → https://arditi.pt/index.php/... 29 Agosto 2019	 → https://arditi.pt/index.php/... 14 Abril 2018	 → http://arditi.pt/index.php?... 14 Fevereiro 2017	 → https://arditi.pt/ 24 Março 2021	 → https://arditi.pt/noticias/p... 12 Abril 2022
 → https://arditi.pt/ 12 Abril 2022	 → http://arditi.pt/index.php/c... 14 Fevereiro 2017	 → https://arditi.pt/ 11 Março 2020	 → https://arditi.pt/ 12 Abril 2022	 → http://arditi.pt/index.php/c... 6 Agosto 2017
 → http://arditi.pt/index.php/c... 24 Julho 2017			 → https://arditi.pt/index.php/c... 31 Julho 2017	 → http://arditi.pt/index.php/c... 21 Fevereiro 2017

Pesquisa de Imagens - API

```
{
  "serviceName": "Arquivo.pt - image search service.",
  "linkToService": "https://arquivo.pt/images.jsp",
  "linkToMoreFields": "https://arquivo.pt/imagesearch?q=ARDITI%20Madeira&more=pageHost,matchingImages,safe",
  "nextPage": "https://arquivo.pt/imagesearch?q=ARDITI%20Madeira&offset=50",
  "previousPage": "https://arquivo.pt/imagesearch?q=ARDITI%20Madeira&offset=0",
  "totalItems": 4075906,
  "numberOfResponseItems": 50,
  "offset": 0,
  "responseItems": [
    {
      "imgDigest": "1f086292517b214156990080fca8e8047910ee2fa81d29150b80f61834cbe463",
      "imgSrc": "http://www.mare-centre.pt/sites/marecentre/files/galeria/arditi1.jpg",
      "pageTitle": "Infraestruturas | MARE",
      "imgHeight": 658,
      "imgWidth": 921,
      "imgMimeType": "image/jpeg",
      "pageURL": "http://www.mare-centre.pt/pt/mare/infraestruturas",
      "imgTstamp": "20190324124110",
      "pageTstamp": "20190324085317",
      "imgCaption": [
        "ARDITI (Madeira)",
        "ARDITI (Madeira)"
      ],
      "imgAlt": [
        "ARDITI (Madeira)"
      ],
      "collection": [
        "AWP36",
        "AWP29",
        "EAWP28",
        "EAWP36",
        "AWP39",
      ]
    }
  ]
}
```

arquivo.pt/imagesearch?q=ARDITI%20Madeira

Pesquisa de Imagens - API

<https://arquivo.pt/imagesearch?q=ARDITI%20Madeira&type=png&site=arditi.pt>

- **imagesearch**
 - Serviço requisitado
- **q=ARDITI Madeira**
 - Termos a pesquisar
- **type=PNG**
 - Imagens no formato PNG
- **site=arditi.pt**
 - Resultados do site arditi.pt

```
{
  "serviceName": "Arquivo.pt - image search service.",
  "linkToService": "https://arquivo.pt/images.jsp",
  "linkToMoreFields": "https://arquivo.pt/imagesearch?q=C%C3%A2mara%20Municipal%20de%20Lisboa&format=PNG&from=20010224174130&to=2019&more=pageHost,matchingImages,safe",
  "nextPage": "https://arquivo.pt/imagesearch?q=C%C3%A2mara%20Municipal%20de%20Lisboa&format=PNG&from=20010224174130&to=2019&offset=50",
  "previousPage": "https://arquivo.pt/imagesearch?q=C%C3%A2mara%20Municipal%20de%20Lisboa&format=PNG&from=20010224174130&to=2019&offset=0",
  "totalItems": 11784234,
  "numberOfResponseItems": 50,
  "offset": 0,
  "responseItems": [
    {
      "imgDigest": "de1807dc12bf9d6dbc8b00dbacb46741452d150b8e632c5b9947351ac7d12820",
      "imgSrc": "http://www.voarte.com/contents/festparceirost3/14116527312272.png",
      "pageTitle": "Vo'Arte - InShadow - Parceiros",
      "imgHeight": 80,
      "imgWidth": 150,
      "imgMimeType": "image/png",
      "pageURL": "http://www.voarte.com/pt/festvoarte/inshadow/edicao/parceiros",
      "imgTstamp": "20150408151323",
      "pageTstamp": "20150408151020",
      "imgCaption": [
        "Ap\u00f3io Institucional\nParceiros Escolas",
        "Institutional support\nSchools Partners"
      ],
      "imgAlt": [
        "Camara Municipal de Lisboa",
        "C\u00e2mara Municipal de Lisboa"
      ],
      "imgTitle": [
        "Camara Municipal de Lisboa",
        "C\u00e2mara Municipal de Lisboa"
      ],
      "collection": [
        "AMP17",
        "AMP20",
        "AMP23",
        "AMP25",
        "AMP26",
        "AMP27",
        "AMP28",
        "EAMP28"
      ],
      "imgLinkToArchive": "https://arquivo.pt/wayback/20150408151323im_/http://www.voarte.com/contents/festparceirost3/14116527312272.png",
      "pageLinkToArchive": "https://arquivo.pt/wayback/20150408151020/http://www.voarte.com/pt/festvoarte/inshadow/edicao/parceiros"
    }
  ]
}
```

Pesquisa de Imagens - API

<https://arquivo.pt/imagessearch?q=ARDITI%20Madeira&type=png&site=arditi.pt>

- **pageLinkToArchive**
 - Link para a página arquivada que contém a imagem
- **imgLinkToArchive**
 - Link para a imagem arquivada

```
{
  "serviceName": "Arquivo.pt - image search service.",
  "linkToService": "https://arquivo.pt/images.jsp",
  "linkToMoreFields": "https://arquivo.pt/imagessearch?q=C%C3%A2mara%20Municipal%20de%20Lisboa&format=PNG&from=20010224174130&to=2019&more=pageHost,matchingImages,safe",
  "nextPage": "https://arquivo.pt/imagessearch?q=C%C3%A2mara%20Municipal%20de%20Lisboa&format=PNG&from=20010224174130&to=2019&offset=50",
  "previousPage": "https://arquivo.pt/imagessearch?q=C%C3%A2mara%20Municipal%20de%20Lisboa&format=PNG&from=20010224174130&to=2019&offset=0",
  "totalItems": 11784234,
  "numberOfResponseItems": 50,
  "offset": 0,
  "responseItems": [
    {
      "imgDigest": "de1807dc12bf9d6dbc8b00dbacb46741452d150b8e632c5b9947351ac7d12820",
      "imgSrc": "http://www.voarte.com/contents/festparceirost3/14116527312272.png",
      "pageTitle": "Vo'Arte - InShadow - Parceiros",
      "imgHeight": 80,
      "imgWidth": 150,
      "imgMimeType": "image/png",
      "pageURL": "http://www.voarte.com/pt/festvoarte/inshadow/edicao/parceiros",
      "imgTstamp": "20150408151323",
      "pageTstamp": "20150408151020",
      "imgCaption": [
        "Apoio Institucional\nParceiros Escolas",
        "Institutional support\nSchools Partners"
      ],
      "imgAlt": [
        "Camara Municipal de Lisboa",
        "Câmara Municipal de Lisboa"
      ],
      "imgTitle": [
        "Camara Municipal de Lisboa",
        "Câmara Municipal de Lisboa"
      ],
      "collection": [
        "AMP17",
        "AMP20",
        "AMP23",
        "AMP25",
        "AMP26",
        "AMP27",
        "AMP28",
        "EAMP28"
      ],
      "imgLinkToArchive": "https://arquivo.pt/wayback/20150408151323im_/http://www.voarte.com/contents/festparceirost3/14116527312272.png",
      "pageLinkToArchive": "https://arquivo.pt/wayback/20150408151020/http://www.voarte.com/pt/festvoarte/inshadow/edicao/parceiros"
    }
  ]
}
```

Consulta de histórico de páginas arquivadas

www.arditi.pt

1991 6 Ago — 2024 19 Abr

Páginas | Imagens | Narrativa

Pesquisa avançada

Cerca de 174 resultados desde 1991 até 2024

Tabela | Lista

2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
		9 Nov	30 Set	11 Abr	5 Fev	14 Fev	9 Abr	19 Mar	23 Jan	2 Jan	15 Jan	1 Jan	
				23 Abr	10 Fev	17 Fev	13 Abr	2 Abr	5 Fev	11 Jan	16 Jan	18 Jan	
				11 Set	23 Mai	21 Fev	16 Abr	9 Abr	3 Mar	4 Mar	17 Jan	18 Jan	
				15 Set	30 Mai	9 Jun	17 Abr	22 Abr	11 Mar	10 Mar	19 Jan	31 Jan	
				16 Nov	6 Jun	31 Jul	16 Jul	29 Ago	23 Mar	24 Mar	1 Fev	31 Jan	
				23 Nov	31 Out	6 Ago	16 Jul	30 Ago	26 Mar	6 Jun	10 Fev	15 Fev	
					6 Nov	7 Dez	20 Jul	2 Set	10 Jun	27 Jun	11 Fev	15 Fev	
						12 Dez	26 Jul	4 Set	19 Jun	21 Ago	11 Abr		
						14 Dez	26 Jul	4 Set	19 Jun	21 Ago	21 Abr		
						17 Dez	10 Out	5 Dez	6 Set	22 Ago	21 Abr		
							11 Out	7 Dez	6 Dez	9 Set	23 Abr		

www.arditi.pt

1991 6 Ago — 2024 19 Abr

Páginas | Imagens | Narrativa

Pesquisar

Pesquisa avançada

Cerca de 174 resultados desde 1991 até 2024

Tabela | Lista

2013 1 versão

2014 1 versão

2015 6 versões

Abril 2 versões

11 Abril 17h24, 2015

23 Abril 02h24, 2015

Setembro 2 versões

Novembro 2 versões

2016 7 versões

2017 10 versões

Consulta de histórico de páginas arquivadas

ARQUIVO.PT

www.arditi.pt

1991 6 Ago — 2024 19 Abr

Páginas | Imagens | Narrativa

Pesquisa avançada

Cerca de 174 resultados desde 1991 até 2024

Tabela | Lista

11	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
		9 Nov	30 Set	11 Abr	31 Ev	14 Fev	9 Abr	19 Mar	23 Jan	2 Jan	15 Jan	1 Jan	
				23 Abr	10 Fev	17 Fev	13 Abr	2 Abr	5 Fev	11 Jan	16 Jan	18 Jan	
				11 Set	23 Mai	21 Fev	16 Abr	9 Abr	3 Mar	4 Mar	17 Jan	18 Jan	
				15 Set	30 Mai	9 Jun	17 Abr	22 Abr	11 Mar	10 Mar	19 Jan	31 Jan	
				16 Nov	6 Jun	31 Jul	16 Jul	29 Ago	23 Mar	24 Mar	1 Fev	31 Jan	
				23 Nov	31 Out	6 Ago	16 Jul	30 Ago	26 Mar	6 Jun	10 Fev	15 Fev	
					6 Nov	7 Dez	20 Jul	2 Set	10 Jun	27 Jun	11 Fev	15 Fev	
						12 Dez	26 Jul	4 Set	19 Jun	21 Ago	11 Abr		
						14 Dez	26 Jul	4 Set	19 Jun	21 Ago	21 Abr		
						17 Dez	10 Out	5 Dez	6 Set	22 Ago	21 Abr		
							11 Out	7 Dez	6 Dez	9 Set	23 Abr		

ARQUIVO.PT



agência regional para o desenvolvimento da investigação tecnologia e inovação

Início | ARDITI | Biblioteca On-line | Bolsas | Ligações | Contactos

Bolsas de Curta Duração

Bolsas Individuais de Formação

Bolsas Professores/Investigadores

Candidaturas

- OOM
- Qualidade Agro-Alimentar
- TICS
- Estágios para Doutores Fev 2014
- Doenças Emergentes
- Turismo
- Estágios para Doutores Abr 2014
- Estágios PIDTI – Maio 2014
- MHCI 2014

Dia de Portas Abertas – Banco de Germoplasma ISOPlexis

08 abril 2015



O Banco de Germoplasma ISOPlexis, unidade de investigação da Universidade da Madeira promove um **Dia de Portas Abertas a 18 de Maio de 2015**. Este evento insere-se nas comemorações do Dia Internacional do Fascínio das Plantas (*Fascination of Plants Day*), organizado pela European Plant Science Organisation (EPSO), e ainda do Dia Internacional da Biodiversidade, no âmbito do evento “Pé n’A Terra 2015” promovido pela Biodiversity4All. Inserido nestas comemorações, instituições por todo o país promovem diversas atividades durante o mês de Maio. O BG

Consulta de histórico de páginas arquivadas - API



```
{
  "serviceName": "Arquivo.pt - the Portuguese web-archive",
  "linkToService": "https://arquivo.pt",
  "next_page": "https://arquivo.pt/textsearch?versionHistory=arditi.pt&offset=50",
  "estimated_nr_results": 208,
  "response_items": [
    {
      "title": "https://arditi.pt/",
      "originalURL": "https://arditi.pt/",
      "linkToArchive": "https://arquivo.pt/wayback/20230215184001/https://arditi.pt/",
      "tstamp": "20230215184001",
      "contentLength": 570,
      "digest": "3I42H3S6NMFQ2MSVX7XZKYAYSCX5QBYJ",
      "mimeType": "text/html",
      "linkToScreenshot": "https://arquivo.pt/screenshot?url=https%3A%2F%2Farquivo.pt%2FnoFrame%2Freply%2F20230215184001%2Fhttps%3A%2F%2Farditi.pt%2F",
      "linkToNoFrame": "https://arquivo.pt/noFrame/replay/20230215184001/https://arditi.pt/",
      "linkToMetadata": "https://arquivo.pt/textsearch?metadata=https%3A%2F%2Farditi.pt%2F%2F20230215184001",
      "linkToOriginalFile": "https://arquivo.pt/noFrame/replay/20230215184001id_/https://arditi.pt/",
      "fileName": "WEB-20230215183955832-p100.arquivo.pt.warc.gz",
      "collection": "EAWP42",
      "offset": 22310291,
      "statusCode": 301
    },
    {
      "title": "http://arditi.pt/",
      "originalURL": "http://arditi.pt/",

```

arquivo.pt/textsearch?versionHistory=arditi.pt

Outra forma de consultar o histórico - CDX API



```
pt,arditi/ 20131109035146 {"url": "http://www.arditi.pt/", "mime": "text/html", "status": "200", "digest": "J57CKAX0FV5JZJHMDUBFZPQHEE4WFYB3", "length": "7295", "offset": "1"}
20131109035029-23305-p13.arquivo.pt.arc.gz, "collection": "AWP15", "source": "$root:AWP1-20.cdxj", "source-coll": "$root"}
pt,arditi/ 20140930000155 {"url": "http://www.arditi.pt/", "mime": "text/html", "status": "200", "digest": "PLTYT4YHVWEEWECW5N5NQXASK32VML", "length": "8262", "offset": "2"}
20140929235929-21419-p13.arquivo.pt.arc.gz, "collection": "AWP16", "source": "$root:AWP1-20.cdxj", "source-coll": "$root"}
pt,arditi/ 20150411172413 {"url": "http://arditi.pt/", "mime": "text/html", "status": "200", "digest": "Z5KSZ4VYHE04XU6RIICFJJIQGGREMPIKK", "length": "9254", "offset": "93476"}
20150411170624-02534-p13.arquivo.pt.arc.gz, "collection": "AWP17", "source": "$root:AWP1-20.cdxj", "source-coll": "$root"}
pt,arditi/ 20150423022416 {"url": "http://www.arditi.pt/", "mime": "text/html", "status": "200", "digest": "5KFI6T24I4ZIIY25NWR74R4DBLMEIJP5X", "length": "9365", "offset": "5"}
20150423021914-31264-p13.arquivo.pt.arc.gz, "collection": "AWP17", "source": "$root:AWP1-20.cdxj", "source-coll": "$root"}
pt,arditi/ 20150911191226 {"url": "http://www.arditi.pt/", "mime": "text/html", "status": "200", "digest": "XUXYOTIVAXQIPFRW56CKQTBZVMPMKJKQ", "length": "9240", "offset": "5"}
20150911190856-19476-p81.arquivo.pt.arc.gz, "collection": "AWP18", "source": "$root:AWP1-20.cdxj", "source-coll": "$root"}
pt,arditi/ 20150915062906 {"url": "http://arditi@arditi.pt/", "mime": "text/html", "status": "200", "digest": "QQIZSPNKXSGCQYKCB4FPJE425F3NIUXY", "length": "9250", "offset": "20150915050839-28399-p81.arquivo.pt.arc.gz, "collection": "AWP18", "source": "$root:AWP1-20.cdxj", "source-coll": "$root"}
pt,arditi/ 20151116205054 {"url": "http://arditi@arditi.pt/", "mime": "text/html", "status": "200", "digest": "DRLB2E642YVARFGRCY7LUUZUEJ66RP4", "length": "9258", "offset": "20151116203522-01990-p81.arquivo.pt.arc.gz, "collection": "AWP19", "source": "$root:AWP1-20.cdxj", "source-coll": "$root"}
pt,arditi/ 20151123214504 {"url": "http://www.arditi.pt/", "mime": "text/html", "status": "200", "digest": "A4NNUHYPEUT2ZOSUF2RVUX3XIBSBENGI", "length": "8771", "offset": "6"}
20151123211529-29132-p81.arquivo.pt.arc.gz, "collection": "AWP19", "source": "$root:AWP1-20.cdxj", "source-coll": "$root"}
pt,arditi/ 20160205161645 {"url": "http://arditi@arditi.pt/", "mime": "text/html", "status": "200", "digest": "5YKRYJSIDFOD4FBT3KSORZIMCGIL6IBA", "length": "9826", "offset": "20160205155819-02134-p81.arquivo.pt.arc.gz, "collection": "AWP20", "source": "$root:AWP1-20.cdxj", "source-coll": "$root"}
pt,arditi/ 20160210135332 {"url": "http://www.arditi.pt/", "mime": "text/html", "status": "200", "digest": "NFH2J40G5WIKLGP1HOAFA5GMMPRQ7LBQ", "length": "9779", "offset": "1"}
20160210135216-36162-p81.arquivo.pt.arc.gz, "collection": "AWP20", "source": "$root:AWP1-20.cdxj", "source-coll": "$root"}
pt,arditi/ 20160523082440 {"url": "http://www.arditi.pt/", "mime": "text/html", "status": "200", "digest": "5Y34FEHJWKOYNUY22PHK78YLIESFHODP", "length": "9168", "offset": "6"}
20160523072201-24095-p82.arquivo.pt.arc.gz, "collection": "EAWP11", "source": "$root:EAWP11.cdxj", "source-coll": "$root"}
pt,arditi/ 20160530184117 {"url": "http://arditi@arditi.pt/", "mime": "text/html", "status": "200", "digest": "I4RGSKMZ5UNHCR53YIRG3DNRB4T4V73P6", "length": "9412", "offset": "20160530175459-01303-p81.arquivo.pt.arc.gz, "collection": "AWP21", "source": "$root:AWP21-29.cdxj", "source-coll": "$root"}
pt,arditi/ 20160606020231 {"url": "http://www.arditi.pt/", "mime": "text/html", "status": "200", "digest": "6YURMNT4VHIPISWVFNRSQSYGYYPJ65TB5", "length": "9697", "offset": "9"}
20160606013652-25407-p81.arquivo.pt.arc.gz, "collection": "AWP21", "source": "$root:AWP21-29.cdxj", "source-coll": "$root"}
pt,arditi/ 20161031184940 {"url": "http://arditi@arditi.pt/", "mime": "text/html", "status": "200", "digest": "2V6SP5B4DP65QMIMG4NS3TKGUQJYKUH", "length": "9007", "offset": "20161031184155-01196-p81.arquivo.pt.arc.gz, "collection": "AWP22", "source": "$root:AWP21-29.cdxj", "source-coll": "$root"}
```

arquivo.pt/wayback/cdx?url=arditi.pt

Consulta em massa – download em Bulk dos CDX



Index of /cdxj

 [ICO]	<u>Name</u>	<u>Last modified</u>	<u>Size</u>	<u>Description</u>
 [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

Colecções do Arquivo.pt

List of the collections preserved by Arquivo.pt (publicly available)

Collection ID	Description	Collection Start date	Collection End date	Total files (Number of URLs collected)	Total seeds (Preserved sites)	Total volume of WARCS files (TB)	Collection in Production? (URL, Page and Image indexed)
AWP1	1st complete crawl of the Portuguese web, mainly from the .PT domain, in 2008.	2008-02-12	2008-03-06	56,046,288	154,787	1.60	TRUE
AWP2	2nd complete crawl of the Portuguese web, mainly from the .PT domain, in 2008.	2008-03-11	2008-05-30	48,718,404	-	1.60	TRUE
AWP3	3rd complete crawl of the Portuguese web, mainly from the .PT domain, in 2008.	2008-10-21	2008-12-10	51,863,006	193,294	2.00	TRUE
AWP4	4th complete crawl of the Portuguese web, mainly from the .PT domain, in 2009.	2009-05-01	2009-05-31	68,776,707	366,880	2.50	TRUE
AWP5	5th complete crawl of the Portuguese web, mainly from the .PT domain, in 2009.	2009-10-01	2009-10-31	119,135,566	373,323	3.80	TRUE
AWP6	6th complete crawl of the Portuguese web, mainly from the .PT domain, in 2009.	2009-12-01	2009-12-31	118,810,364	340,018	3.50	TRUE
AWP7	7th complete crawl of the Portuguese web, mainly from the .PT domain, in 2010.	2010-05-01	2010-05-31	87,988,812	389,957	2.90	TRUE
AWP8	Incremental crawl of the Portuguese web, mainly from the .PT domain, in 2010. The AWP8 crawl is incremental because it was performed using DeDuplicator (http://landsbokasafn.github.io/DeDuplicator/) taking the content of AWP7 as baseline. Thus, the files that remained unchanged from the AWP7 complete crawl were not archived (duplicated) on the AWP8 incremental crawl.	2010-08-01	2010-08-31	75,771,317	411,562	1.90	TRUE
AWP9	Incremental crawl of the Portuguese web, mainly from the .PT domain, in 2011. The AWP9 crawl is incremental because it was performed using DeDuplicator (http://landsbokasafn.github.io/DeDuplicator/) taking the content of AWP7 as baseline. Thus, the files that remained unchanged from the AWP7 complete crawl were not archived (duplicated) on the AWP9 incremental	2011-01-20	2011-03-22	81,114,575	473,588	2.10	TRUE

arquivo.pt/collections

Um exemplo de utilização: Modelo LLM



Glória, um modelo para a língua portuguesa

Um dos projetos que utilizou o Arquivo.pt para obter grande quantidade de texto denomina-se Glória e é um modelo de linguagem em larga escala (*LLM, Large Language Model*) focado na língua portuguesa europeia.

“Apesar da abundância de LLMs para muitas línguas de recursos elevados, a disponibilidade de tais modelos continua a ser limitada para o português europeu”, como explicam os autores do Glória, Ricardo Lopes, João Magalhães, David Semedo, investigadores da [Faculdade de Ciência e Tecnologia da Universidade Nova de Lisboa](#), no seu artigo *Glória – A Generative and Open Large Language Model for Portuguese*.

Memento API – Standard internacional

Arquivo404

Página não encontrada

Veja uma versão arquivada desta página de 2024-1-19 em arquivo.pt

VOLTAR AO INÍCIO



Questões?

arquivo.pt/api

contacto@arquivo.pt