

# APIs do Arquivo.pt

Como processar informação em larga escala?

Café com o Arquivo.pt

08 de Maio de 2020

André Mourão

Engenheiro I&D - Arquivo.pt

[andre.mourao@fccn.pt](mailto:andre.mourao@fccn.pt)

# 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: uc.pt

ARQUIVOPT

uc.pt

1996 1 Jan 2020 20 Abr

Páginas Imagens Pesquisa avançada

Tabela Lista

1996

Outubro

1997

Junho

Outubro

10 Outubro às 07:56

1998

Dezembro

12 Dezembro às 03:36

1999

2000

2001

2002

2003

2004

ARQUIVOPT

uc.pt

1996 1 Jan 2020 20 Abr

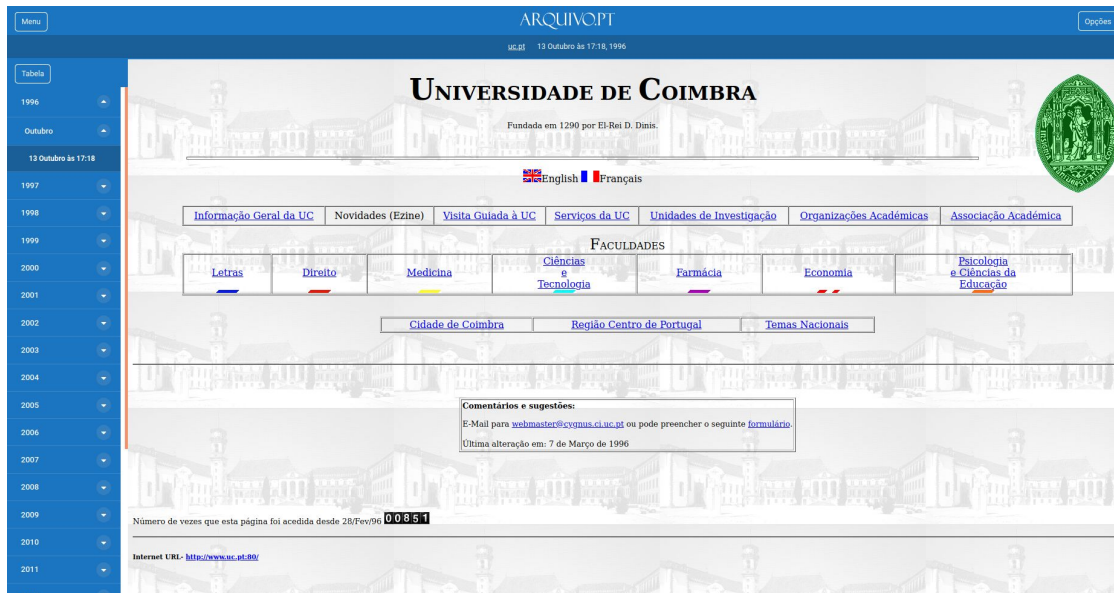
Páginas Imagens Pesquisa avançada

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 | 18 Jun | 12 Dez |        |        |        |        |        |        |        |        |        |        |        |        |        |        |       |        |        |        |        |        |        |      |
|        | 10 Out |        | 8 Fev  | 8 Abr  | 2 Mar  | 29 Mai | 16 Fev | 5 Abr  | 4 Jan  | 1 Jan  | 5 Jan  | 8 Jan  | 20 Mai | 2 Jun  | 3 Jan  | 23 Jan | 7 Nov | 29 Set | 22 Abr | 9 Jan  | 22 Fev | 26 Mar | 14 Mar |      |
|        |        |        | 20 Abr | 11 Mai | 2 Mar  | 31 Mai | 21 Mar | 9 Jun  | 15 Jan | 1 Jan  | 15 Jan | 14 Jan | 25 Jun | 9 Jun  | 8 Jan  |        | 7 Nov | 9 Out  | 24 Abr | 11 Fev | 23 Fev | 28 Mar | 22 Mai |      |
|        |        |        | 27 Abr | 11 Mai | 2 Mar  | 2 Ago  | 19 Abr | 11 Jun | 17 Jan | 1 Jan  | 20 Jan | 20 Jan | 25 Jun | 11 Jun | 21 Jan |        | 8 Nov | 23 Out | 6 Mai  | 11 Fev | 6 Jun  | 29 Mar | 6 Jun  |      |
|        |        |        |        | 11 Mai | 2 Abr  | 28 Set | 11 Jun | 12 Jun | 18 Jan | 1 Jan  | 25 Jan | 27 Jan | 25 Set | 3 Ago  | 21 Jan |        |       | 26 Nov | 15 Set | 18 Fev | 7 Ago  | 30 Mar | 7 Jun  |      |
|        |        |        |        | 17 Out | 28 Abr | 30 Set | 12 Jun | 18 Jun | 19 Jan | 1 Jan  | 2 Fev  | 5 Fev  | 26 Set | 4 Ago  | 22 Jan |        |       |        |        | 14 Mar | 7 Ago  | 31 Mar | 7 Jun  |      |
|        |        |        |        |        | 9 Mai  | 22 Nov | 26 Jul | 19 Jun | 20 Jan | 1 Jan  | 2 Fev  | 15 Fev | 26 Set | 4 Ago  | 24 Jan |        |       |        | 25 Nov | 15 Mai | 7 Ago  | 1 Abr  | 19 Jun |      |
|        |        |        |        |        | 11 Jul |        | 23 Set | 22 Jun | 23 Jan | 2 Jan  | 10 Fev | 15 Fev | 26 Set | 5 Ago  | 19 Mai |        |       |        | 4 Dez  | 20 Mai | 13 Dez | 2 Abr  | 20 Jun |      |
|        |        |        |        |        |        |        | 21 Out | 23 Jun | 5 Fev  | 2 Jan  | 17 Fev | 18 Fev | 6 Out  |        | 20 Mai |        |       |        |        | 6 Jun  | 13 Dez | 3 Abr  | 31 Out |      |
|        |        |        |        |        |        |        | 26 Out | 23 Jun | 6 Fev  | 3 Jan  | 25 Fev | 14 Mar | 17 Dez |        | 22 Mai |        |       |        |        | 11 Jun | 14 Dez | 4 Abr  | 31 Out |      |
|        |        |        |        |        |        |        | 19 Nov | 25 Jun | 9 Fev  | 3 Jan  | 6 Mar  | 14 Mar | 17 Dez |        | 22 Mai |        |       |        |        | 17 Ago |        |        |        |      |
|        |        |        |        |        |        |        | 27 Nov | 26 Jun | 10 Fev | 3 Jan  | 14 Mar | 15 Mar | 18 Dez |        | 3 Jul  |        |       |        |        | 17 Ago |        |        |        |      |
|        |        |        |        |        |        |        |        | 27 Jun | 11 Fev | 3 Jan  | 23 Mar | 21 Mar | 18 Dez |        | 3 Jul  |        |       |        |        | 6 Nov  |        |        |        |      |
|        |        |        |        |        |        |        |        | 28 Jun | 15 Fev | 4 Jan  | 29 Mar | 21 Out | 19 Dez |        |        |        |       |        |        | 7 Nov  |        |        |        |      |
|        |        |        |        |        |        |        |        | 30 Jun | 22 Fev | 11 Jan | 6 Abr  | 21 Out |        |        |        |        |       |        |        |        |        |        |        |      |
|        |        |        |        |        |        |        |        | 1 Jul  | 25 Fev | 13 Jan | 27 Abr | 22 Out |        |        |        |        |       |        |        |        |        |        |        |      |
|        |        |        |        |        |        |        |        | 1 Jul  | 28 Fev | 14 Jan | 30 Abr | 22 Out |        |        |        |        |       |        |        |        |        |        |        |      |
|        |        |        |        |        |        |        |        | 3 Jul  | 3 Mar  | 15 Jan | 5 Mai  | 22 Out |        |        |        |        |       |        |        |        |        |        |        |      |
|        |        |        |        |        |        |        |        | 4 Jul  | 3 Mar  | 27 Jan | 28 Mai | 22 Out |        |        |        |        |       |        |        |        |        |        |        |      |

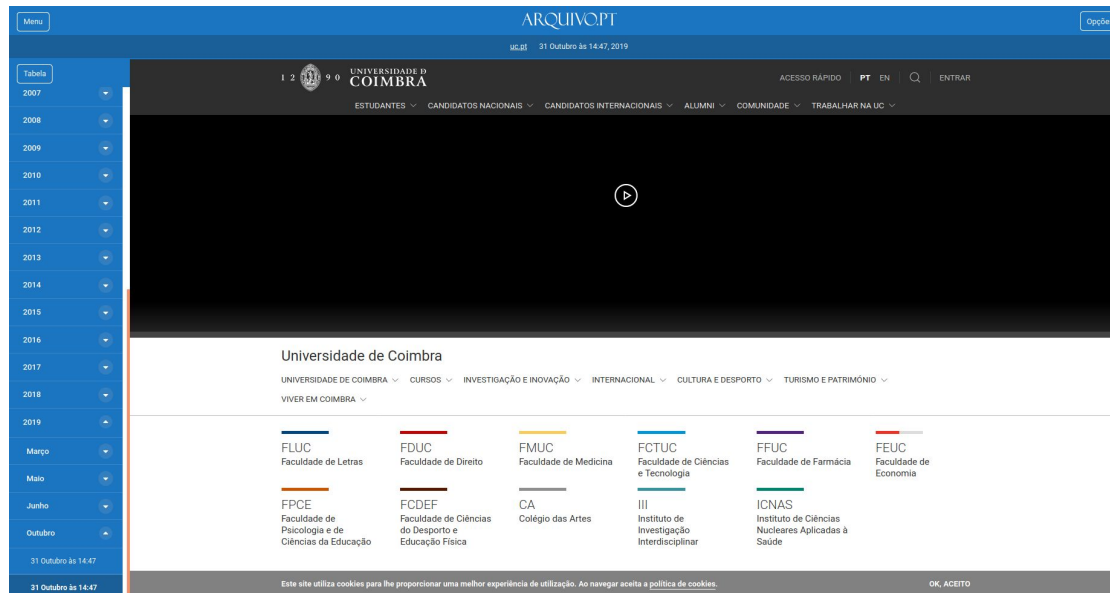
[https://arquivo.pt/wayback/\\*/http://www.uc.pt/](https://arquivo.pt/wayback/*/http://www.uc.pt/)

# Consulta de histórico de páginas: uc.pt



<https://arquivo.pt/wayback/19961013171816/http://www.uc.pt/>

# Consulta de histórico de páginas: uc.pt



<https://arquivo.pt/wayback/20191031144724/http://www.uc.pt/>

2019 10 21 14 47 24  
ano mês dia hora min seg

# Consulta de histórico de páginas: uc.pt

Menu

ARQUIVOPT

1 Janeiro às 04:41, 2006

Opções

Tabela

1996

1997

1998

1999

2000

2001

2002

2003

2004

2005

2006

1 Janeiro

1 Janeiro às 04:41

1 Janeiro às 06:57

1 Janeiro às 08:04

1 Janeiro às 09:59

1 Janeiro às 12:58

1 Janeiro às 15:49

1 Janeiro às 22:31

Universidade de Coimbra

Informação

Pesquisa

Serviços

Ensino e Investigação

Organizações

Letras

Direito

Medicina

Ciências

Farmácia

Economia

Psicologia

PRÊMIO "FOREIGN DIRECT INVESTMENT"

AR - UNIVERSIDADE DE COIMBRA - 2005

60 234 651

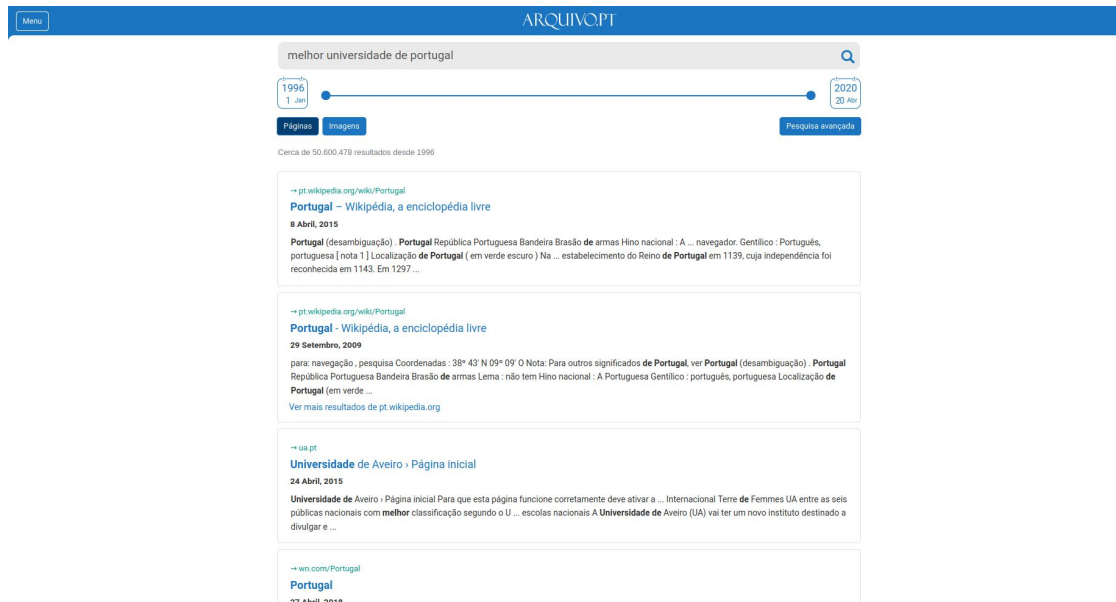
acessados em 2004

<https://arquivo.pt/wayback/2006/http://www.uc.pt/>





# Pesquisa de texto: melhor universidade de Portugal



The screenshot shows the ARQUIVO.PT search interface. At the top, there's a blue header with 'Menu' and 'ARQUIVO.PT'. Below it, a search bar contains the text 'melhor universidade de portugal'. To the right of the search bar is a magnifying glass icon. Below the search bar, there's a timeline slider from 1996 to 2020. Below the slider are buttons for 'Páginas', 'Imagens', and 'Pesquisa avançada'. Below these buttons, it says 'Cerca de 50.600.478 resultados desde 1996'. The search results are listed below, with the first result being 'Portugal - Wikipédia, a enciclopédia livre' dated 8 Abril, 2015. The second result is 'Portugal - Wikipédia, a enciclopédia livre' dated 29 Setembro, 2009. The third result is 'Universidade de Aveiro - Página inicial' dated 24 Abril, 2015. The fourth result is 'Portugal' dated 07 Abril, 2018.

[https://arquivo.pt/search.jsp?  
query=melhor+universidade+de+portugal](https://arquivo.pt/search.jsp?query=melhor+universidade+de+portugal)

# Pesquisa avançada de texto

- Eu quero páginas sobre a:
  - **Universidade de Coimbra**
  - capturadas durante **1996-1997**
  - parte do site **www.uc.pt**

<https://arquivo.pt/advanced.jsp?query=universidade+de+coimbra&site=www.uc.pt&dateStart=01/01/1996&dateEnd=31/12/1997>

Palavras

Tenham estas palavras:

universidade de coimbra

ex.: expo 98

Tenham esta expressão ou frase:

ex.: pavilhão da utopia

Não tenham estas palavras:

ex.: concerto

Data

1996  
1 Jan

1997  
31 Dez

Formato

Ver as páginas com o formato:

Selecione

Sítio Web

Com este endereço:

uc.pt

ex.: www.arquivo.pt

# Pesquisa de imagens

- Eu quero imagens da:
  - **Universidade de Coimbra**
  - no site da **uc.pt** (diferente de **www.uc.pt**)
  - capturadas entre **1996-1997**

[https://arquivo.pt/advancedImages.jsp?  
query=universidade+de+coimbra&  
site=uc.pt&  
dateStart=01/01/1996&  
dateEnd=31/12/1997](https://arquivo.pt/advancedImages.jsp?query=universidade+de+coimbra&site=uc.pt&dateStart=01/01/1996&dateEnd=31/12/1997)

**Palavras**

Tenham estas palavras:

  
ex.: expo 98

Tenham esta expressão ou frase:

  
ex.: pavilhão da utopia

Não tenham estas palavras:

  
ex.: concerto

---

**Data**

1996  
1 Jan

1997  
20 Abr

---

**Imagens**

Formato:

  
Selecione

Tamanho:

  
mostrar todas

Pesquisa segura:

  
Excluir impróprias

---

**Sítio Web**

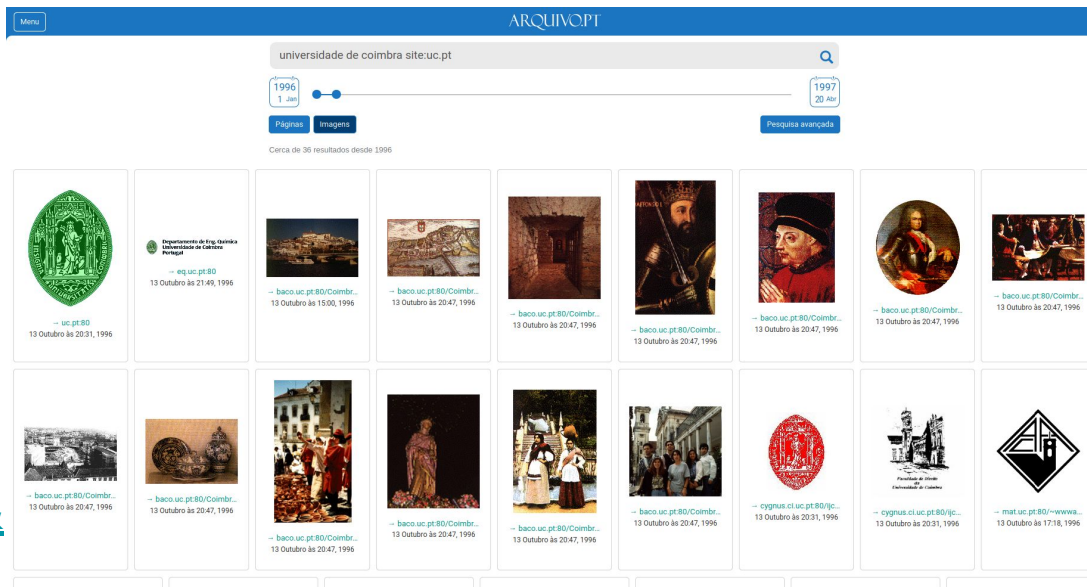
Com este endereço:

  
ex.: www.arquivo.pt

# Pesquisa avançada de imagens

- Eu quero imagens da:
  - **Universidade de Coimbra**
  - no site da **uc.pt** (diferente de **www.uc.pt**)
  - capturadas entre **1996-1997**

<https://arquivo.pt/images.jsp?query=universidade+de+coimbra&site=uc.pt&dateStart=01/01/1996&dateEnd=31/12/1997>



# Questões?

# 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 **+8.000 milhões de páginas** e **+22 milhões de imagens** guardadas e disponíveis para pesquisa
- Será que existe uma **alternativa** a fazer o processo de consulta e recolha manual?

# O que é uma API? (Application Programming Interface)

- Interface normalizada para aceder a serviços disponibilizados por uma plataforma;
- No caso do Arquivo.pt, as APIs permitem processar páginas e imagens automaticamente;
- Pedidos são feitos por URL, e.g. <https://arquivo.pt/textsearch?versionHistory=uc.pt>:
  - <https://arquivo.pt/textsearch>: serviço
  - [versionHistory=uc.pt](https://arquivo.pt/textsearch?versionHistory=uc.pt): parâmetros do pedido
- Resultados são disponibilizados num formato comum (JSON), e podem ser testados no Firefox, Chrome, Safari...

# 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=uc.pt&offset=50"
▼ previous_page: "https://arquivo.pt/nutchwax/textsearch?versionHistory=uc.pt&offset=0"
estimated_nr_results: "869"
▼ request_parameters:
  from: "19960101000000"
  to: "20201231235959"
▼ response_items:
  ► 0: {...}
  ► 1: {...}
  ▼ 2:
    title: "Universidade de Coimbra"
    originalURL: "https://www.uc.pt/"
    ▼ linkToArchive: "https://arquivo.pt/wayback/20190620043525/https://www.uc.pt/"
      tstamp: "20190620043525"
      contentLength: "9737"
      digest: "BNU455XH3H5UZZMINT2LRQP2057TD7F"
      mimeType: "text/html"
    ▼ linkToScreenshot: "https://arquivo.pt/screenshot?url=https%3A%2F%2Farquivo.pt%2FnoFrame%2Freplay%2F20190620043525id%2Fhttps%3A%2F%2Fwww.uc.pt%2F"
      date: "1561005325"
      encoding: "UTF-8"
    ▼ linkToNoFrame: "https://arquivo.pt/noFrame/replay/20190620043525/https://www.uc.pt/"
      status: "200"
      collection: "EAMP2320190620"
    ▼ linkToExtractedText: "https://arquivo.pt/textextracted?m=https%3A%2F%2Fwww.uc.pt%2F%2F20190620043525"
    ▼ linkToMetadata: "https://arquivo.pt/textsearch?metadata=https%3A%2F%2Fwww.uc.pt%2F%2F20190620043525"
    ▼ linkToOriginalFile: "https://arquivo.pt/noFrame/replay/20190620043525id /https://www.uc.pt/"

```

[https://arquivo.pt/textsearch?  
versionHistory=uc.pt](https://arquivo.pt/textsearch?versionHistory=uc.pt)

# API para consulta de histórico de uma página

[https://arquivo.pt/textsearch?  
versionHistory=www.uc.pt&  
maxItems=50&  
offset=0&  
from=20010224174130&  
to=2019](https://arquivo.pt/textsearch?versionHistory=www.uc.pt&maxItems=50&offset=0&from=20010224174130&to=2019)

- **textsearch**
  - Serviço requisitado
- **versionHistory=www.uc.pt**
  - URL a requisitar histórico
- **offset=0**
  - Começar no primeiro resultado
- **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=uc.pt&offset=50"
▼ previous_page:  "https://arquivo.pt/nutchwax/textsearch?versionHistory=uc.pt&offset=0"
estimated_nr_results: "869"
▼ request_parameters:
  from:           "19960101000000"
  to:             "20201231235959"
▼ response_items:
  ► 0:            {...}
  ► 1:            {...}
  ▼ 2:
    title:        "Universidade de Coimbra"
    originalURL:   "https://www.uc.pt/"
    ▼ linkToArchive: "https://arquivo.pt/wayback/20190620043525/https://www.uc.pt/"
    tstamp:       "20190620043525"
    contentLength: "9737"
    digest:        "BNU455XHV3H5UZZMINT2LRQP2057D7F"
    mimeType:      "text/html"
    ▼ linkToScreenshot: "https://arquivo.pt/screenshot?url=https%3A%2F%2Farquivo.pt%2FnoFrame%2Fplay%2F20190620043525%2Fhttps%3A%2F%2Fwww.uc.pt%2F"
    date:          "1561005325"
    encoding:      "UTF-8"
    ▼ linkToNoFrame:   "https://arquivo.pt/noFrame/replay/20190620043525/https://www.uc.pt/"
    status:         "200"
    collection:     "EAWP2320190620"
    ▼ linkToExtractedText: "https://arquivo.pt/textextracted?m=https%3A%2F%2Fwww.uc.pt%2F%2F20190620043525"
    ▼ linkToMetadata:    "https://arquivo.pt/textsearch?metadata=https%3A%2F%2Fwww.uc.pt%2F%2F20190620043525"
    ▼ linkToOriginalFile: "https://arquivo.pt/noFrame/replay/20190620043525id /https://www.uc.pt/"

```

- **linkToExtractedText**

- **linkToScreenshot**

- **linkToOriginalFile**

- ```

serviceName: "Arquivo.pt - the Portuguese web archive"
linkToService: "https://arquivo.pt"
▼ next_page: "https://arquivo.pt/nutchwax/textsearch?versionHistory=uc.pt&offset=50"
▼ previous_page: "https://arquivo.pt/nutchwax/textsearch?versionHistory=uc.pt&offset=0"
estimated_nr_results: "869"
▼ request_parameters:
  from: "19960101000000"
  to: "20201231235959"
▼ response_items:
  ► 0: {...}
  ► 1: {...}
  ▼ 2:
    title: "Universidade de Coimbra"
    originalURL: "https://www.uc.pt/"
    ▼ linkToArchive: "https://arquivo.pt/wayback/20190620043525/https://www.uc.pt/"
    tstamp: "20190620043525"
    contentLength: "9737"
    digest: "BNU455XH3H5UZZMINT2LRQP2057TD7F"
    mimeType: "text/html"
    ▼ linkToScreenshot: "https://arquivo.pt/screenshot?url=https%3A%2F%2Farquivo.pt%2FnoFrame%2Freplay%2F20190620043525%2Fhttps%3A%2F%2Fwww.uc.pt%2F%2F1561005325"
    date: "1561005325"
    encoding: "UTF-8"
    ▼ linkToNoFrame: "https://arquivo.pt/noFrame/replay/20190620043525/https://www.uc.pt/"
    status: "200"
    collection: "EAWP2320190620"
    ▼ linkToExtractedText: "https://arquivo.pt/textextracted?m=https%3A%2F%2Fwww.uc.pt%2F%2F20190620043525"
    ▼ linkToMetadata: "https://arquivo.pt/textsearch/metadata=https%3A%2F%2Fwww.uc.pt%2F%2F20190620043525"
    ▼ linkToOriginalFile: "https://arquivo.pt/noFrame/replay/20190620043525id /https://www.uc.pt/"

```

# Consulta de texto extraído de uma página

Universidade de Coimbra Este site utiliza cookies para lhe proporcionar uma melhor experiência de utilização. Ao navegar aceita a política de cookies . OK, ACEITO Acesso rápido PT EN Entrar Universidade de Coimbra Sobre nós Órgãos de governo Extensão Cultural e Apoio à Formação Unidades de Ensino e Investigação Património Mundial Organizacional Cursos Licenciaturas Mestrados integrados Mestrados Doutoramentos Ensino a distância Outros cursos Cursos preparatórios Aprender português Investigação e Inovação Unidades de I&D Ligação a empresas Instituto de investigação interdisciplinar IPN – Incubadora de empresas Biocant – Parque Tecnológico Congressos e conferências Projetos financiados UC Digitalis Estudo Geral Imprensa da Universidade Centro de Documentação 25 de Abril Energia para a Sustentabilidade Internacional Linhas Estratégicas Candidatos Internacionais Erasmus e outros programas de mobilidade Serviços de Apoio Cultura e Desporto Agenda UC EUG 2018 Desporto Universitário Estádio Universitário Teatro Académico de Gil Vicente Rua Larga - Revista da UC Bibliotecas Semana Cultural da UC AnoZero Associação Académica de Coimbra Turismo e Património UC, Alta e Sofia, Património Mundial Turismo Museu da Ciência Jardim Botânico Arquivo da UC Viver em Coimbra A Cidade Alojamento e Serviços de Apoio Academia e Lazer Custo de Vida Transportes Estudantes Matrículas Serviços de apoio Certidões e diplomas Provedor do Estudante Para além das aulas Contactos Pré-universitários Candidaturas e matrículas Por quê escolher a UC Cursos Faculdades Um dia na UC: programas para escolas Universidade de Verão Contactos Estudantes internacionais Brasil China Outros países Oportunidades de estudo Estudar e viver em Coimbra Informações práticas Contactos Alumni Rede UC Inscreva-se na Rede UC Global Alumni Contactos Comunidade Empresas Serviços à comunidade Emprego público Bibliotecas Energia para a Sustentabilidade Produção científica Trabalhar na UC Docentes Investigadores Dirigentes Não Docentes ESTUDAR NA UC Mais de 300 opções por onde escolher. Universidade de Coimbra Universidade de Coimbra Sobre nós Órgãos de governo Extensão Cultural e Apoio à Formação Unidades de Ensino e Investigação Património Mundial Organizacional Cursos Licenciaturas Mestrados integrados Mestrados Doutoramentos Ensino a distância Outros cursos Cursos preparatórios Aprender português Investigação e Inovação Unidades de I&D Ligação a empresas Instituto de investigação interdisciplinar IPN – Incubadora de empresas Biocant – Parque Tecnológico Congressos e conferências Projetos financiados UC Digitalis Estudo Geral Imprensa da Universidade Centro de Documentação 25 de Abril Energia para a Sustentabilidade Internacional Linhas Estratégicas Candidatos Internacionais Erasmus e outros programas de mobilidade Serviços de Apoio Cultura e Desporto Agenda UC EUG 2018 Desporto Universitário Estádio Universitário Teatro Académico de Gil Vicente Rua Larga - Revista da UC Bibliotecas Semana Cultural da UC AnoZero Associação Académica de Coimbra Turismo e Património UC, Alta e Sofia, Património Mundial Turismo Museu da Ciência Jardim Botânico Arquivo da UC Viver em Coimbra A Cidade Alojamento e Serviços de Apoio Academia e Lazer Custo de Vida Transportes FLUC Faculdade de Letras

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

[m=https://www.uc.pt/20181227215357](https://www.uc.pt/20181227215357)



# API de Pesquisa de texto

[https://arquivo.pt/textsearch?  
q=universidade%20de%20coimbra&  
maxItems=50&  
offset=50&  
from=20100402000000&  
to=2019](https://arquivo.pt/textsearch?q=universidade%20de%20coimbra&maxItems=50&offset=50&from=20100402000000&to=2019)

- **textsearch**
  - Serviço requisitado
- **q=universidade de coimbra**
  - 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?versionHistory=uc.pt&offset=50"
▼ previous_page: "https://arquivo.pt/nutchwax/textsearch?versionHistory=uc.pt&offset=0"
  estimated_nr_results: "869"
▼ request_parameters:
  from: "19960101000000"
  to: "20201231235959"
▼ response_items:
  ► 0: {}
  ► 1: {}
  ▼ 2:
    title: "Universidade de Coimbra"
    originalURL: "https://www.uc.pt/"
    ▼ linkToArchive: "https://arquivo.pt/wayback/20190620043525/https://www.uc.pt/"
      tstamp: "20190620043525"
      contentLength: "9737"
      digest: "BNU455XHV3H5UZZMINT2LRQP2057TD7F"
      mimeType: "text/html"
    ▼ linkToScreenshot: "https://arquivo.pt/screenshot?url=https%3A%2F%2Fwww.uc.pt%2F%2FnoFrame%2Fplay%2F20190620043525%2Fhttps%3A%2F%2Fwww.uc.pt%2F%2F%2FnoFrame/replay/20190620043525/https://www.uc.pt/"
      date: "1561005325"
      encoding: "UTF-8"
    ▼ linkToNoFrame: "https://arquivo.pt/noFrame/replay/20190620043525/https://www.uc.pt/"
      status: "200"
      collection: "EAWP2320190620"
    ▼ linkToExtractedText: "https://arquivo.pt/textextracted?m=https%3A%2F%2Fwww.uc.pt%2F%2F20190620043525"
    ▼ linkToMetadata: "https://arquivo.pt/textsearch?metadata=https%3A%2F%2Fwww.uc.pt%2F%2F20190620043525"
    ▼ linkToOriginalFile: "https://arquivo.pt/noFrame/replay/20190620043525id_/https://www.uc.pt/"

```

# Campos dos resultados

<https://arquivo.pt/textsearch?q=universidade%20de%20coimbra&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?versionHistory=uc.pt&offset=50"
▼ previous_page: "https://arquivo.pt/nutchwax/textsearch?versionHistory=uc.pt&offset=0"
estimated_nr_results: "869"
▼ request_parameters:
  from: "19960101000000"
  to: "20201231235959"
▼ response_items:
  ► 0: {}
  ► 1: {}
  ▼ 2:
    title: "Universidade de Coimbra"
    originalURL: "https://www.uc.pt/"
    ▼ linkToArchive: "https://arquivo.pt/wayback/20190620043525/https://www.uc.pt/"
      tstamp: "20190620043525"
      contentLength: "9737"
      digest: "BNU455XHV3H5UZZMINT2LRQP2057TD7F"
      mimeType: "text/html"
      ▼ linkToScreenshot: "https://arquivo.pt/screenshot?url=https%3A%2F%2Fwww.uc.pt%2F%2FnoFrame%2Fplay%2F20190620043525%2Fhttps%3A%2F%2Fwww.uc.pt%2F%2F%2FnoFrame/replay/20190620043525/https://www.uc.pt/"
        date: "1561005325"
        encoding: "UTF-8"
        ▼ linkToNoFrame: "https://arquivo.pt/noFrame/replay/20190620043525/https://www.uc.pt/"
          status: "200"
          collection: "EAWP2320190620"
          ▼ linkToExtractedText: "https://arquivo.pt/textextracted?m=https%3A%2F%2Fwww.uc.pt%2F%2F20190620043525"
          ▼ linkToMetadata: "https://arquivo.pt/textsearch?metadata=https%3A%2F%2Fwww.uc.pt%2F%2F20190620043525"
          ▼ linkToOriginalFile: "https://arquivo.pt/noFrame/replay/20190620043525id_/https://www.uc.pt/"

```



# API de Pesquisa de imagens

[https://arquivo.pt/imagesearch?  
q=Universidade%20de%20Coimbra&  
size=lg&type=png&  
from=20050101000000&to=20100101235959&  
offset=5&maxItems=100&  
more=imgThumbnailBase64](https://arquivo.pt/imagesearch?q=Universidade%20de%20Coimbra&size=lg&type=png&from=20050101000000&to=20100101235959&offset=5&maxItems=100&more=imgThumbnailBase64)

- **imagesearch**
  - Serviço requisitado
- **q=Universidade de Coimbra**
  - Texto a pesquisar
- **from=20050101000000 &to=20100101235959**
  - mostrar resultados de 2005 a 2010
- **offset=5 &maxItems=100**
  - Começar no resultado 5, devolver 100 resultados
- **type=png &size=lg**
  - Pedir imagens em PNG de tamanho grande
- **more=imgThumbnailBase64.....**
  - campos de saída requisitados

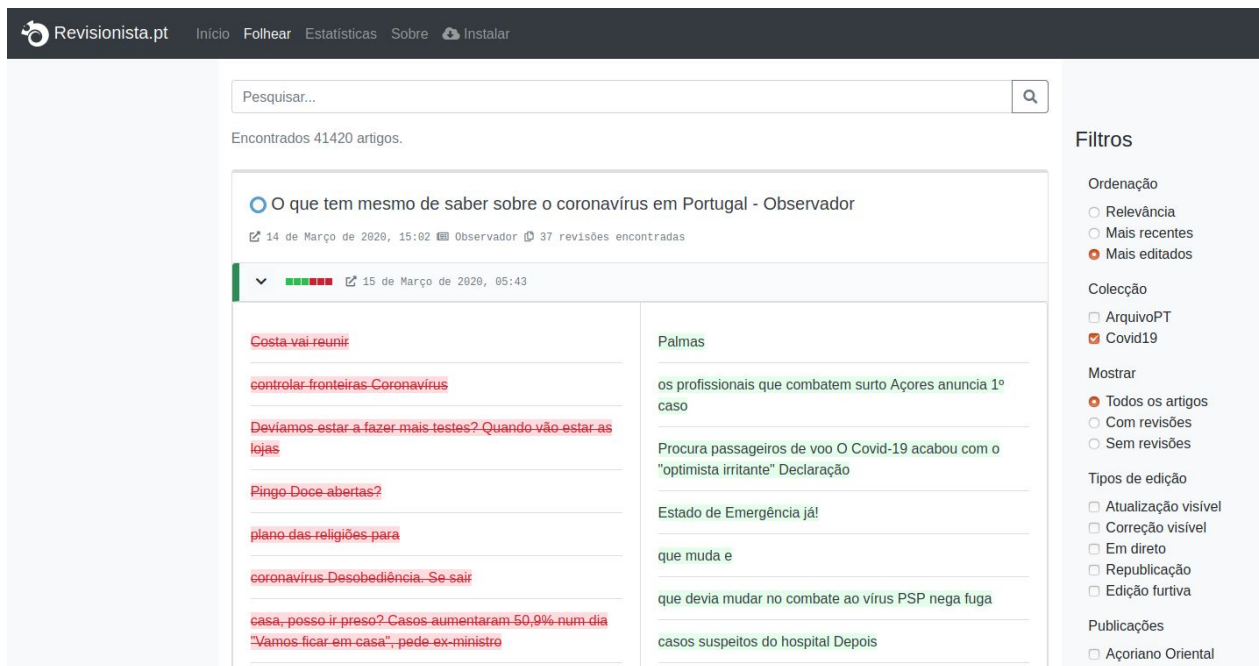
```
serviceName: "Arquivo.pt - image search service."
linkToService: "https://arquivo.pt/images.jsp"
▶ linkToMoreFields: "https://arquivo.pt/image...pageHost,pageImages,safe"
▶ nextPage: "https://arquivo.pt/image...png&collection=&offset=8"
▶ previousPage: "https://arquivo.pt/image...png&collection=&offset=0"
totalItems: 8
numberOfResponseItems: 3
offset: 5
▼ responseItems:
  ▼ 0:
    pageHost: "www.reitoria.utl.pt"
    imgTstamp: 20091218192049
    imgHeight: 1325
    imgWidth: 1603
    imgMimeType: "image/png"
    safe: 0.994
    pageProtocol: "http"
    ▶ imgDigest: "834616d09a238916826b9052...f3bf7b4760410af186d4eb3"
    ▶ imgThumbnailBase64: "iVBORw0KGgoAAAANSUUEUgAA...a3QnwkAAAAASUVORK5CYII="
    ▶ imgSrcURLDigest: "dd9f8f3640222571cedb6701...1193f75f7c24aac742739f7"
    pageTstamp: 20110523223951
    ▶ pageURL: "http://www.reitoria.utl....php?area=243&id=letonia"
    collection: "AWP10"
    ▶ imgSrc: "http://www.reitoria.utl...dB/world_map_turquia.png"
    ▶ pageTitle: "Universidade Técnica de ...: Partner Institutions"
    pageImages: 44
    ▶ imgLinkToArchive: "https://arquivo.pt/wayba...dB/world_map_turquia.png"
    ▶ pageLinkToArchive: "https://arquivo.pt/wayba...php?area=243&id=letonia"
  ▶ 1: {...}
```

# 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

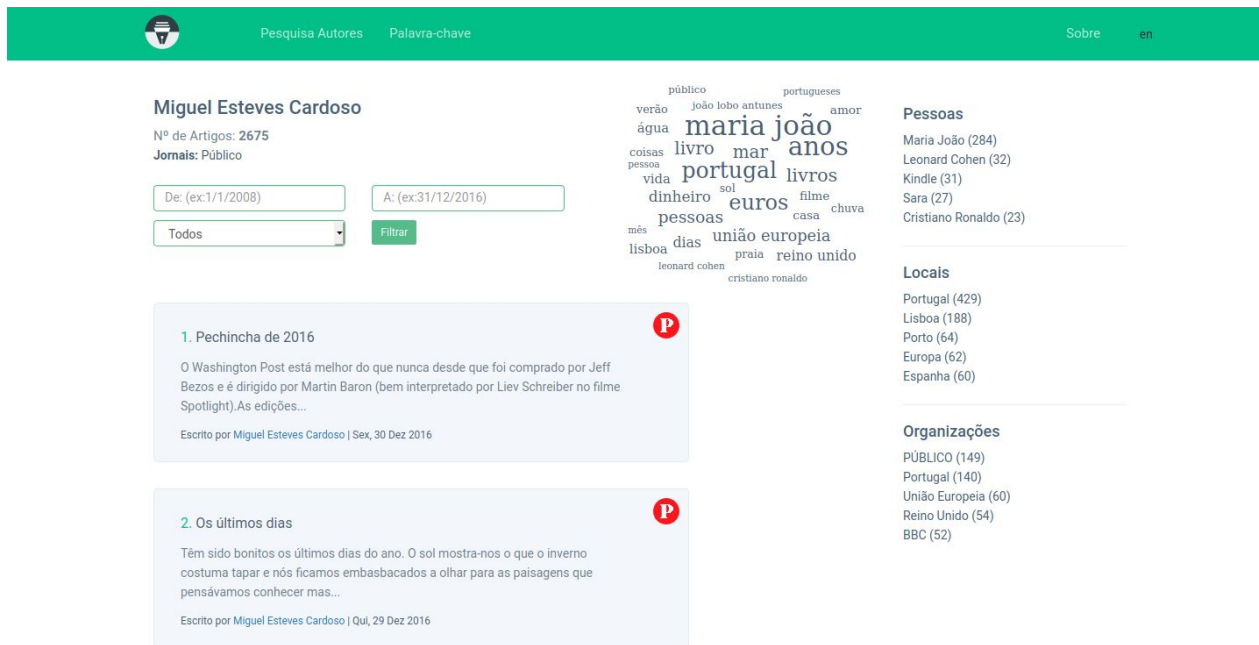


The screenshot shows the Revisionista.pt website. The header includes the logo and navigation links: Início, Folhear, Estatísticas, Sobre, and Instalar. A search bar at the top contains the text "Pesquisar...". Below the search bar, it states "Encontrados 41420 artigos." The main content area displays 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



**Miguel Esteves Cardoso**  
Nº de Artigos: 2675  
Jornais: Público

De: (ex:1/1/2008) A: (ex:31/12/2016)  
Todos **Filtrar**

**1. Pechincha de 2016**  
O Washington Post está melhor do que nunca desde que foi comprado por Jeff Bezos e é dirigido por Martin Baron (bem interpretado por Liev Schreiber no filme Spotlight). As edições...  
Escrito por Miguel Esteves Cardoso | Sex, 30 Dez 2016

**2. Os últimos dias**  
Têm sido bonitos os últimos dias do ano. O sol mostra-nos o que o inverno costuma tapar e nós ficamos embasbacados a olhar para as paisagens que pensávamos conhecer mas...  
Escrito por Miguel Esteves Cardoso | Qui, 29 Dez 2016

**Word Cloud:** público, portugueses, maria joão, livro, mar, anos, portugal, livros, dinheiro, sol, filme, euros, pessoas, casa, chuva, dias, união europeia, reino unido, lisboa, praia, cristiano ronaldo, leonard cohen.

**Pessoas**  
Maria João (284)  
Leonard Cohen (32)  
Kindle (31)  
Sara (27)  
Cristiano Ronaldo (23)

**Locais**  
Portugal (429)  
Lisboa (188)  
Porto (64)  
Europa (62)  
Espanha (60)

**Organizações**  
PÚBLICO (149)  
Portugal (140)  
União Europeia (60)  
Reino Unido (54)  
BBC (52)

- Arquivo de opinião: <http://arquivodeopinioao.pt/pt/>
  - APIs Usadas: Pesquisa de texto filtrada e mostra de páginas capturadas

# Exemplos de utilização

*(insira a sua ideia aqui)*

# Demonstração

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

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

Antártida Investigadores da UC; Auditório da Reitoria da Universidade de Coimbra; Biocant; Cidade Alojamento; Coimbra; Estudar; Museu da Ciência Jardim Botânico Arquivo da UC Palácio de São Marcos; Oportunidades; Parque Tecnológico Congressos; Serviços de Apoio Academia...

## **Miscellaneous entities, e.g. events, nationalities, products or works of art:**

A Tua Viagem; ACEITO Acesso; Agenda UC Redes Sociais UC Mapa; Alta e Sofia; Apoio à Formação da Universidade de Coimbra; Aprender português Investigação e Inovação Unidades de I&D.

## **Companies, agencies, institutions, etc.:**

Ciências da Educação FCDEF Faculdade de Ciências do Desporto e Educação Física CA Colégio das Artes III; Comunidade Empresas Serviços; Contactos Alumni Rede UC; Diretores das Unidades de Extensão Cultural; Elite mundial da Ciência Polar; Estudantes internacionais da; FCTUC; Instituto; ...

## **Named person or family.:**

AÇÃO SOCIAL INTERNACIONAL PATRIMÓNIO; Carlos Reis; Eduardo Lourenço 2019 Descobrir a UC CANDIDATOS INVESTIGAÇÃO; Gil Vicente Rua Larga; Lazer Custo de Vida Transportes; ...



# Mais informações

- APIs Disponíveis em <https://arquivo.pt/api>
- **Código Aberto** e **Acesso Aberto** para todos!
- 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](mailto:andre.mourao@fccn.pt)