

Arquivo.pt

Acesso automático ao Arquivo.pt (APIs)

Fernando Melo < fernando.melo@fccn.pt >

API – Application Programming Interface



Acesso automático

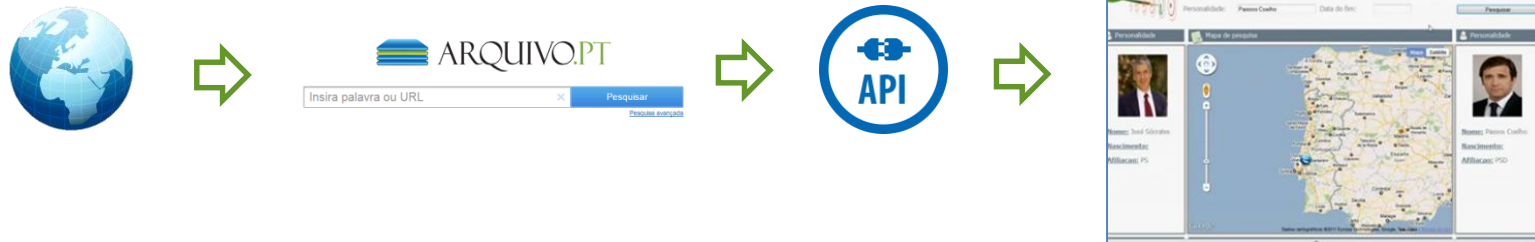
Fácil integração

Elimina necessidade de compreender o código do núcleo

Rápido desenvolvimento de novos serviços e aplicações

APIs: desenvolvimento de novas aplicações Web sobre dados preservados

A sua aplicação



- Pesquisa por texto
- Pesquisa por URL

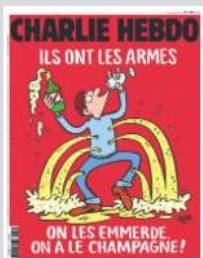
Para que serve?

Pesquisa de Imagens



charlie hebdo

entre: 01/01/2015 e: 31/12/2015 [Pesquisa avançada de imagens](#)



expresso.sapo.pt/
18 Novembro, 2015



jornaldiabo.com/o...
20 Novembro, 2015



www.mulherportugu...
25 Novembro, 2015



www.abola.pt/
31 Dezembro, 2015



jpn.up.pt/tag/media/
1 Dezembro, 2015



jpn.up.pt/media/p...
11 Novembro, 2015



jornais.sapo.pt/i...
1 Outubro, 2015



portpravda.ru/so...
5 Outubro, 2015



jornais.sapo.pt/i...
13 Outubro, 2015



portpravda.ru/bu...
14 Outubro, 2015

www.publico.pt/mu...
4 Dezembro, 2015



jpn.up.pt/author/...
6 Dezembro, 2015



visao.sapo.pt/act...
8 Dezembro, 2015



portpravda.ru/ar...
11 Outubro, 2015

Para que serve?

Agregador informação política



Para que serve?

Eventos desportivos marcantes



Ronaldo Vence Bola de Ouro (2008)



Espanha Campeã Europeia de futebol (2008)



Vasco Ribeiro é campeão mundial de surf de juniores (2014)



Portugal é campeão da Europa de Ténis de Mesa (2014)

Arquivo.pt – acesso automático aos
conteúdos



Sintaxe Wayback – Arquivo.pt

Automático (Web Scrapping)

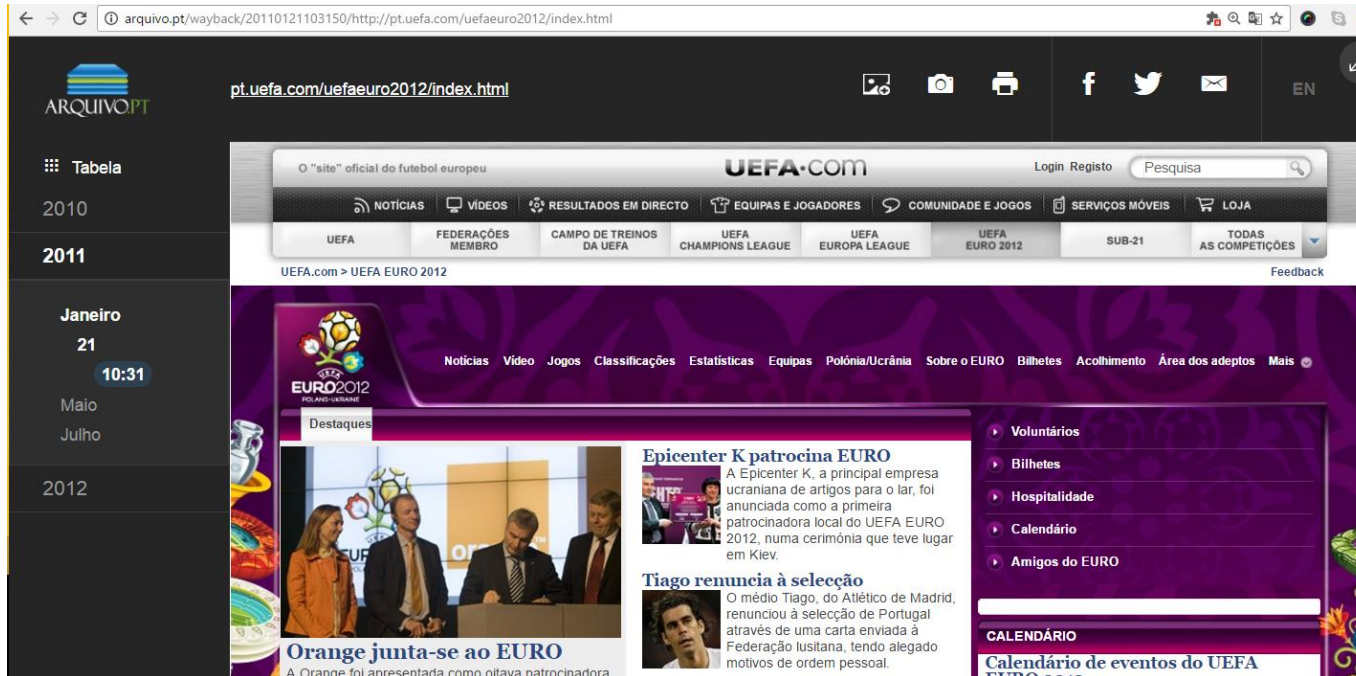
Acesso Manual

Aceder a Página preservada

Listar Páginas preservadas

Sintaxe Wayback

<http://arquivo.pt/wayback/{Timestamp}/{URL}>



{Timestamp} = 20110121103150

{URL} = <http://pt.uefa.com/uefaeuro2012/index.html>

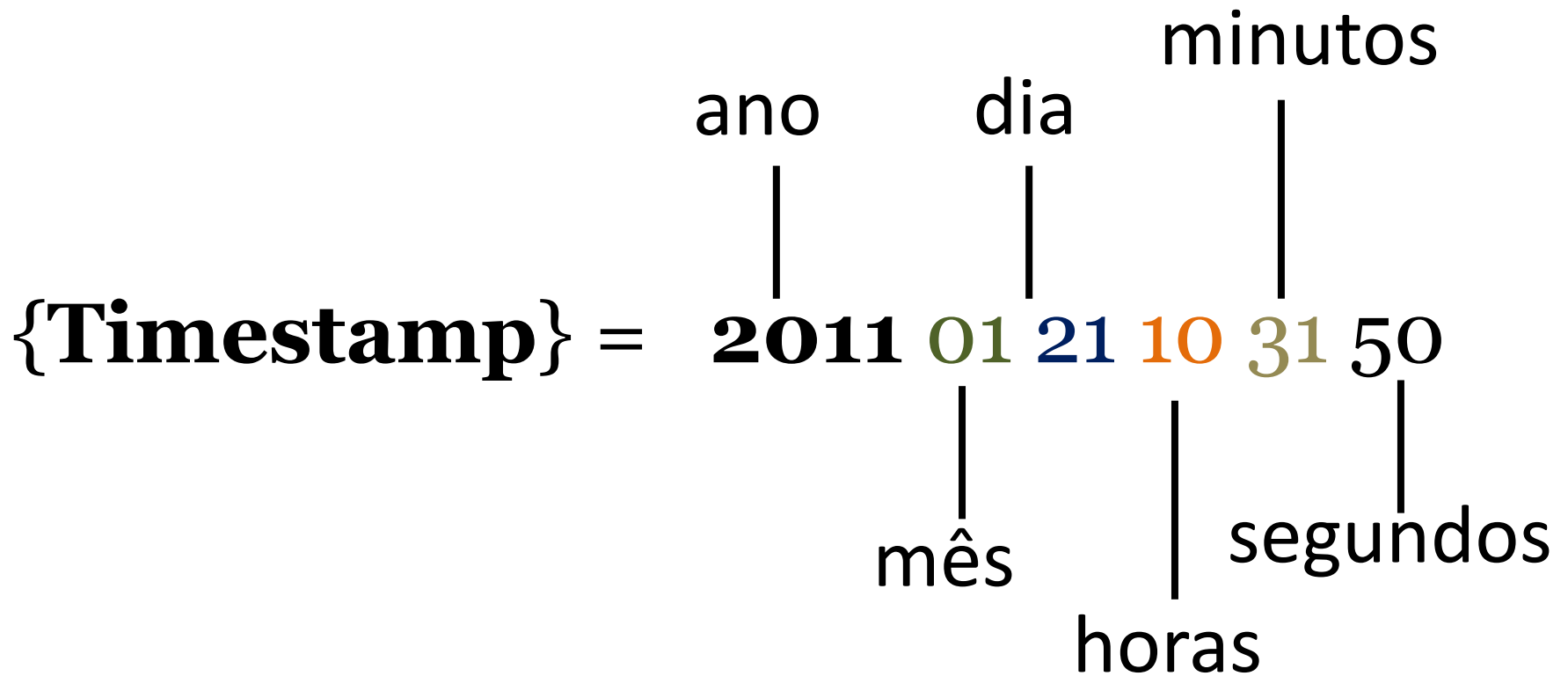
Sintaxe Wayback – **Timestamp**

14 Dígitos

{Timestamp} = **2011 01 21 10 31 50**

Sintaxe Wayback – **Timestamp**

14 Dígitos



Sintaxe Wayback- Closest Date

E se não soubermos o Timestamp exato?

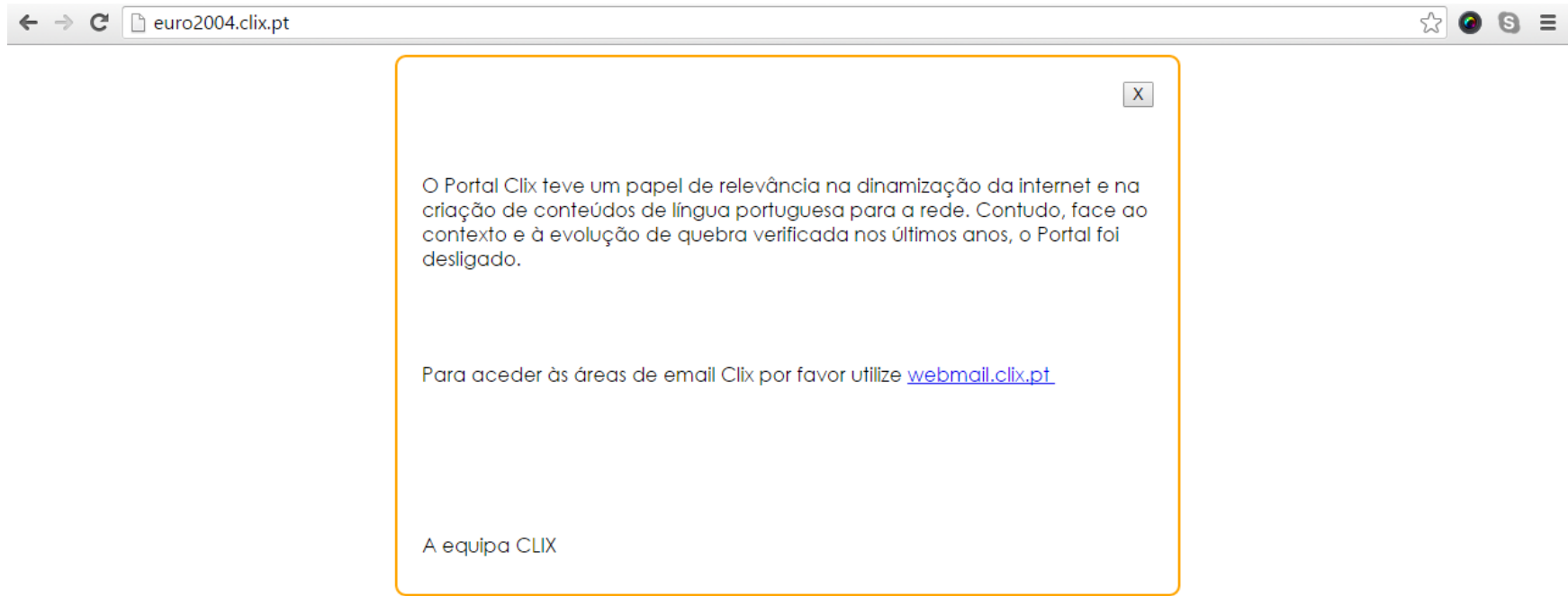
Sintaxe Wayback- Closest Date

<http://arquivo.pt/wayback/2004/http://www.euro2004.clix.pt/>

The screenshot shows the website euro2004.clix.pt. The top navigation bar includes the ARQUIVOPT logo, the site URL, and social media icons for Facebook, Twitter, and email. A sidebar on the left contains a calendar for 2004, with the date 21:27 selected. The main content area features a news article titled "Euro 2004 Jogadores da selecção portuguesa já escolheram números das camisolas". The article includes a photo of players and text stating that the 12 players of the Portuguese national football team have chosen their jersey numbers for the final phase of the Euro 2004 tournament. A secondary article below it is titled "Óbidos: primeiro treino da selecção foi leve". To the right of the main article is a "PUB" section with several short news items, including "Responsáveis desdramatizam insegurança no Euro 2004" and "Cristiano Ronaldo e Jorge Andrade já treinam com a selecção". At the bottom right, there is a list of participating countries: Portugal, Grécia, Espanha, Rússia, França, Inglaterra, Croácia, Suíça, Suécia, Itália, Dinamarca, and Bulaária.

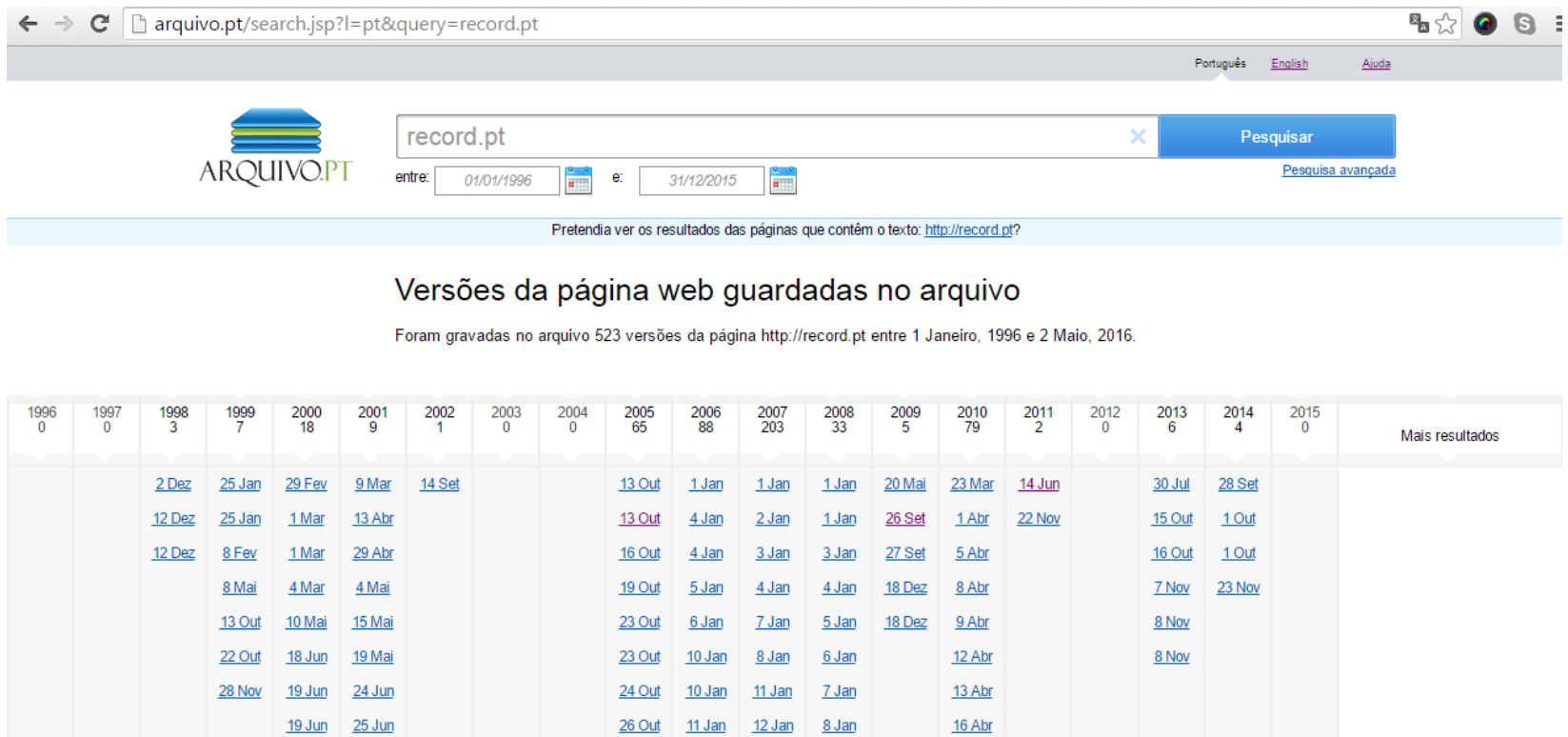
<http://arquivo.pt/wayback/20040525212736/http://euro2004.clix.pt/>

<http://www.euro2004.clix.pt/>




Sintaxe Wayback: E se quisermos todas as Versões?

http://arquivo.pt/wayback/*/record.pt



← → ↻ arquivo.pt/search.jsp?!=pt&query=record.pt

Português English Ajuda

 record.pt

entre: 01/01/1996 e: 31/12/2015

Pretendia ver os resultados das páginas que contém o texto: <http://record.pt>

Versões da página web guardadas no arquivo

Foram gravadas no arquivo 523 versões da página <http://record.pt> entre 1 Janeiro, 1996 e 2 Maio, 2016.

1996 0	1997 0	1998 3	1999 7	2000 18	2001 9	2002 1	2003 0	2004 0	2005 65	2006 88	2007 203	2008 33	2009 5	2010 79	2011 2	2012 0	2013 6	2014 4	2015 0	Mais resultados
		2 Dez	25 Jan	29 Fev	9 Mar	14 Set			13 Out	1 Jan	1 Jan	1 Jan	20 Mai	23 Mar	14 Jun		30 Jul	28 Set		
		12 Dez	25 Jan	1 Mar	13 Abr				13 Out	4 Jan	2 Jan	1 Jan	26 Set	1 Abr	22 Nov		15 Out	1 Out		
		12 Dez	8 Fev	1 Mar	29 Abr				16 Out	4 Jan	3 Jan	3 Jan	27 Set	5 Abr			16 Out	1 Out		
			8 Mai	4 Mar	4 Mai				19 Out	5 Jan	4 Jan	4 Jan	18 Dez	8 Abr			7 Nov	23 Nov		
			13 Out	10 Mai	15 Mai				23 Out	6 Jan	7 Jan	5 Jan	18 Dez	9 Abr			8 Nov			
			22 Out	18 Jun	19 Mai				23 Out	10 Jan	8 Jan	6 Jan		12 Abr			8 Nov			
			28 Nov	19 Jun	24 Jun				24 Out	10 Jan	11 Jan	7 Jan		13 Abr						
				19 Jun	25 Jun				26 Out	11 Jan	12 Jan	8 Jan		16 Abr						

Memento API



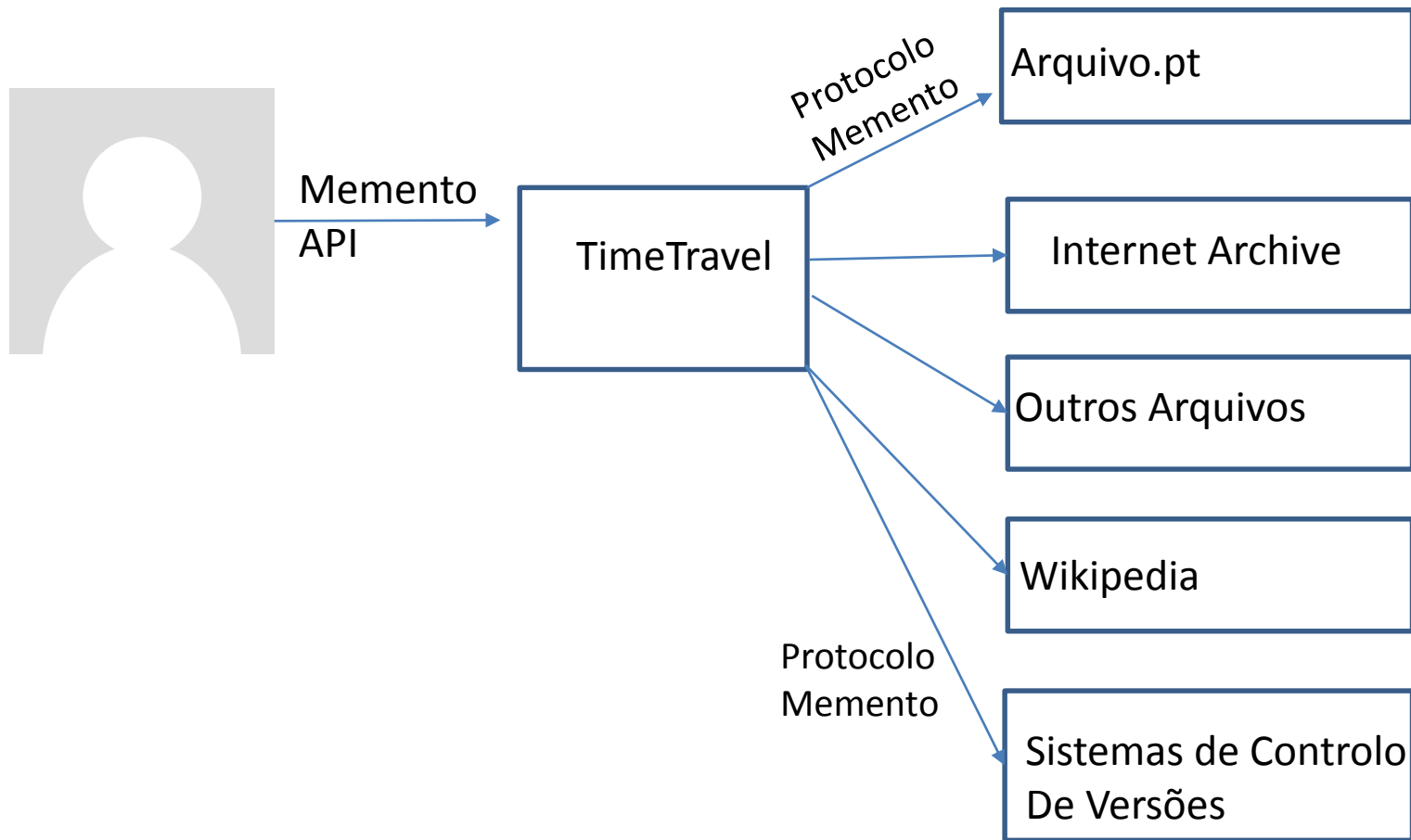
Memento API



<http://timetravel.mementoweb.org/guide/api/>

“Time Travel helps you find and view versions of web pages that existed at some time in the past.

Memento API



Memento API json



<http://timetravel.mementoweb.org/api/json/20130115102033/http://ojogo.pt>

```
{ "original_uri": "http://ojogo.pt",
  "mementos":
    { "last": { "datetime": "2016-04-27T00:21:36Z",
                "uri": ["http://web.archive.org/web/20160427002136/http://ojogo.pt"]
              },
      "next": { "datetime": "2013-01-21T20:31:31Z",
                "uri": ["http://web.archive.org/web/20130121203131/http://ojogo.pt"] },
      "closest": { "datetime": "2013-01-15T16:02:24Z",
                   "uri": ["http://arquivo.pt/wayback/wayback/20130115160224/http://www.ojogo.pt/"] },
      "first": { "datetime": "1998-12-05T18:11:57Z",
                  "uri": ["http://web.archive.org/web/19981205181157/http://ojogo.pt"] },
      "prev": { "datetime": "2013-01-14T22:57:54Z",
                 "uri": ["http://web.archive.org/web/20130114225754/http://ojogo.pt"] },
      "timegate_uri": "http://timetravel.mementoweb.org/timegate/http://ojogo.pt",
      "timemap_uri": { "json_format": "http://timetravel.mementoweb.org/timemap/json/http://ojogo.pt",
                       "link_format": "http://timetravel.mementoweb.org/timemap/link/http://ojogo.pt"
                     }
    }
}
```

Memento TimeMap URL



[http://timetravel.mementoweb.org/timemap/json/
http://ojogo.pt](http://timetravel.mementoweb.org/timemap/json/http://ojogo.pt)

```
{
  "original_uri": "http://ojogo.pt",
  "timegate_uri": "http://timetravel.mementoweb.org/timegate/http://ojogo.pt",
  "timemap_index": [
    {
      "uri": "http://timetravel.mementoweb.org/webcite/timemap/link/http://ojogo.pt",
      "memento_compliant": "no",
      "archive_id": "webcite"
    },
    {
      "uri": "http://timetravel.mementoweb.org/si/timemap/link/http://ojogo.pt",
      "memento_compliant": "no",
      "archive_id": "si"
    },
    {
      "uri": "http://web.archive.org/web/timemap/link/http://ojogo.pt",
      "memento_compliant": "yes",
      "archive_id": "ia"
    }
  ],
  "timemap_uri": {
    "json_format": "http://timetravel.mementoweb.org/timemap/json/http://ojogo.pt",
    "link_format": "http://timetravel.mementoweb.org/timemap/link/http://ojogo.pt"
  }
}
```

Memento TimeMap URL



[http://timetravel.mementoweb.org/timemap/json/
http://ojogo.pt](http://timetravel.mementoweb.org/timemap/json/http://ojogo.pt)

```
<http://wayback.archive-it.org/all/20100605142349/http://www.ojogo.pt/>; rel="first memento"; datetime="Sat, 05 Jun 2010 14:23:49 GMT",  
<http://wayback.archive-it.org/all/20101225093200/http://www.ojogo.pt/>; rel="memento"; datetime="Sat, 25 Dec 2010 09:32:00 GMT",  
<http://wayback.archive-it.org/all/20120512011130/http://www.ojogo.pt/>; rel="memento"; datetime="Sat, 12 May 2012 01:11:30 GMT",  
<http://wayback.archive-it.org/all/20120528141318/http://www.ojogo.pt/>; rel="memento"; datetime="Mon, 28 May 2012 14:13:18 GMT",  
<http://wayback.archive-it.org/all/20120913093151/http://www.ojogo.pt/>; rel="last memento"; datetime="Thu, 13 Sep 2012 09:31:51 GMT"
```

OpenSearch API

OpenSearch - Definition

<http://www.opensearch.org/Home>

“OpenSearch is a collection of simple formats for the sharing of search results.”

OpenSearch – Arquivo.pt Wiki

<https://github.com/arquivo/pwa-technologies/wiki/OpenSearch>

OpenSearch

Fernando-Melo edited this page on Dec 22, 2015 · 10 revisions

Introduction

The Portuguese Web Archive provides an interface for users and tools to easily query the system. The response is a XML-based file (RSS 2.0).

Details

The PWA interface follows the [OpenSearch 1.1 \(Draft 5\) namespace](http://www.opensearch.org/Specifications/OpenSearch/1.1/Draft_5) defined at http://www.opensearch.org/Specifications/OpenSearch/1.1/Draft_5.

It also follows the [OpenSearch Time extension \(Draft 1\) namespace](http://www.opensearch.org/Specifications/OpenSearch/Extensions/Time/1.0/Draft_1) at http://www.opensearch.org/Specifications/OpenSearch/Extensions/Time/1.0/Draft_1 that describes how to set temporal search parameters.

The [OpenSearch Description Document](http://arquivo.pt/opensearch.jsp) at <http://arquivo.pt/opensearch.jsp> describes the public interface and how the search client should make search requests.

See the URL syntax at http://www.opensearch.org/Specifications/OpenSearch/1.1/Draft_5#OpenSearch_URL_template_syntax for more information.

Full-text Search

[OpenSearch Full-text Search API](#)

URL Search


[OpenSearch URL Search API](#)


▼ Pages **19**

- Home
- Compile
- ConfigureSearch
- Install
- L2R4WAIR
- MainFeatures
- milestones
- Open Search**
- Open Search API – Full text Search**
- Open Search API – URL Search**
- Operation
- PwaArchiveAccess
- PwaLucene
- PwaProcessor
- PwaSpellchecker

Show 4 more pages...

Clone this wiki locally

<https://github.com/arquivo/> 

 Clone in Desktop

OpenSearch – Pesquisa por Texto

OpenSearch – Pesquisa por Texto

<http://arquivo.pt/opensearch?query=euro 2004>

start=0

hitsPerSite=2

hitsPerPage=10

OpenSearch – Exemplo Resposta

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<rss xmlns:opensearch="http://a9.com/-/spec/opensearch/1.1/" xmlns:pwa="http://arquivo.pt/opensearchrss/1.0/" xmlns:time
<channel>
<title>PWA Search Engine</title>
<description>PWA search results for query: euro 2004</description>
<link>http://arquivo.pt</link>
<opensearch:totalResults>14631253</opensearch:totalResults>
<opensearch:startIndex>0</opensearch:startIndex>
<opensearch:itemsPerPage>10</opensearch:itemsPerPage>
<opensearch:Query role="request" searchTerms="euro 2004" startPage="1"/>
<item>
<title>Dvd's do Euro 2004 - Aveiro</title>
<source url="http://aveiro-aveiro.olx.pt/dvd-s-do-euro-2004-iid-6414187">Original URL of Dvd's do Euro 2004 - Aveiro</so
<link>http://arquivo.pt/wayback/20080219183920/http://aveiro-aveiro.olx.pt/dvd-s-do-euro-2004-iid-6414187</link>
<pwa:id>37240</pwa:id>
<pwa:index>19</pwa:index>
<pwa:arcname>IAH-20080219065613-10957-T1</pwa:arcname>
<pwa:arcoffset>36702036</pwa:arcoffset>
<pwa:digest>a762b112cd5f0d385c9056a9771972ae</pwa:digest>
<pwa:cache>all</pwa:cache>
<pwa:tstamp>20080219071641000</pwa:tstamp>
<pwa:contentLength>12642</pwa:contentLength>
<pwa:primaryType>text</pwa:primaryType>
<pwa:subType>html</pwa:subType>
</item>
<item>
...
</item>
<item>
...
</item>
...
</channel>
</rss>
```

OpenSearch – Exemplo Resposta

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<rss xmlns:opensearch="http://a9.com/-/spec/opensearch/1.1/" xmlns:pwa="http://arquivo.pt/opensearchrss/1.0/" xmlns:time
<channel>
<title>PWA Search Engine</title>
<description>PWA search results for query: euro 2004</description>
<link>http://archive.pt</link>
<opensearch:totalResults>14631253</opensearch:totalResults>
<opensearch:startIndex>0</opensearch:startIndex>
<opensearch:itemsPerPage>10</opensearch:itemsPerPage>
<opensearch:Query role="request" searchTerms="euro 2004" startPage="1"/>
<item>
<title>Dvd's do Euro 2004 - Aveiro</title>
<source url="http://aveiro-aveiro.olx.pt/dvd-s-do-euro-2004-iid-6414187">Original URL of Dvd's do Euro 2004 - Aveiro</so
<link>http://arquivo.pt/wayback/20080219183920/http://aveiro-aveiro.olx.pt/dvd-s-do-euro-2004-iid-6414187</link>
<pwa:id>37240</pwa:id>
<pwa:index>19</pwa:index>
<pwa:arcname>IAH-20080219065613-10957-T1</pwa:arcname>
<pwa:arcoffset>36702036</pwa:arcoffset>
<pwa:digest>a762b112cd5f0d385c9056a9771972ae</pwa:digest>
<pwa:cache>all</pwa:cache>
<pwa:tstamp>20080219071641000</pwa:tstamp>
<pwa:contentLength>12642</pwa:contentLength>
<pwa:primaryType>text</pwa:primaryType>
<pwa:subType>html</pwa:subType>
</item>
<item>
...
</item>
<item>
...
</item>
...
</channel>
</rss>
```

OpenSearch – Exemplo Resultado

```
<item>
<title>abola.pt</title>
<source url="http://abola.pt/">Original URL of abola.pt</source>
<link>http://arquivo.pt/wayback/20091218033620/http://abola.pt/</link>
<pwa:id>176069</pwa:id>
<pwa:index>29</pwa:index>
<pwa:arcname>IAH-20091217170150-00002-awp01.fccn.pt</pwa:arcname>
<pwa:arcoffset>22344</pwa:arcoffset>
<pwa:digest>e09099b7d70bfe6fc36fb720c69ba20d</pwa:digest>
<pwa:tstamp>20091217170338000</pwa:tstamp>
<pwa:contentLength>76975</pwa:contentLength>
<pwa:primaryType>text</pwa:primaryType>
<pwa:subType>html</pwa:subType>
</item>
```

OpenSearch – Pesquisa Por Texto

[http://arquivo.pt/openserach?query=euro 2004
&start=10](http://arquivo.pt/openserach?query=euro 2004&start=10)

start=10 (próximos 10 resultados)

OpenSearch – Pesquisa Por Texto

[http://arquivo.pt/opensearch?query=euro 2004
&start=0&hitsPerPage=2000](http://arquivo.pt/opensearch?query=euro 2004&start=0&hitsPerPage=2000)

OpenSearch – Pesquisa por Texto

<http://arquivo.pt/opensearch?query=“euro 2004”>

Procura resultados que contenham a expressão:
euro 2004

OpenSearch – Pesquisa por Texto

<http://arquivo.pt/opensearch?query=euro 2004+-moeda>

Procura resultados que contenham as palavras ***euro*** , ***2004*** e que **não conttenham** a palavra ***moeda***

OpenSearch – Pesquisa por Texto

[http://arquivo.pt/openserach?query=euro 2004+
type=application/pdf](http://arquivo.pt/openserach?query=euro 2004+type=application/pdf)

Procura resultados que contenham as palavras ***euro , 2004 no formato .pdf***

OpenSearch – Pesquisa Por Texto

[http://arquivo.pt/openserach?query=euro2004+
type=image/png&hitsPerPage=2000&start=0](http://arquivo.pt/openserach?query=euro2004+type=image/png&hitsPerPage=2000&start=0)

Procura imagens que tenham nome do ficheiro
“euro2004.png”

OpenSearch – Pesquisa por URL

OpenSearch – Exemplo Pesquisa URL

[http://arquivo.pt/openserch?query=
date:19960101000000-20150101000000+
exacturlexpand:http://abola.pt/
&hitsPerSite=10000&waybackQuery=true
&start=0&hitsPerPage=10000](http://arquivo.pt/openserch?query=<u>date:19960101000000-20150101000000+</u><u>exacturlexpand:http://abola.pt/</u><u>&hitsPerSite=10000&waybackQuery=true</u><u>&start=0&hitsPerPage=10000</u>)

OpenSearch – Arquivo.pt Pesquisa URL

<http://arquivo.pt/openserach?query=>

date: Search versions of a URL between 2 dates.

ou

closestdate: For each collection, lists the version of a URL whose date is closest to the requested date.

OpenSearch – Arquivo.pt Pesquisa URL

<http://arquivo.pt/openserach?query=>

date:19960101000000-20151022163016

ou

closestdate:20140101010101

OpenSearch – Arquivo.pt URL-Search

exacturl: <http://abola.pt>

<http://abola.pt/>

<http://www.abola.pt/>

<http://www.abola.pt>

ou

exacturlexpand: <http://abola.pt>

<http://abola.pt/>

<http://www.abola.pt/>

<http://www.abola.pt>

/index.html , /index.htm ,

/index.shtml , /default.html, /default.htm,

/index.php, /index.asp,

/index.aspx and /index.jsp

OpenSearch – Arquivo.pt URL-Search

hitsPerSite=10000

waybackQuery=true (obrigatório)

OpenSearch – Exemplo URL-Search

[http://arquivo.pt/openserach?query=closestdate:19960101000000+exacturlexpand:http://abola.pt/&hitsPerSite=10000&waybackQuery=true&start=0&hitsPerPage=10000](http://arquivo.pt/openserach?query=<u>closestdate:19960101000000+</u><u>exacturlexpand:http://abola.pt/</u><u>&hitsPerSite=10000&waybackQuery=true</u><u>&start=0&hitsPerPage=10000</u>)

Obrigado

<http://arquivo.pt/formacao>

Fernando Melo < fernando.melo@fccn.pt >