

Acesso automático ao Arquivo.pt (APIs)

Fernando Melo <fernando.melo@fccn.pt>

API – Application Programming Interface



Rápido desenvolvimento de novas aplicações

Acesso automático

Fácil integração

Elimina necessidade de compreender código núcleo.

Para que servem as APIs do Arquivo.pt?

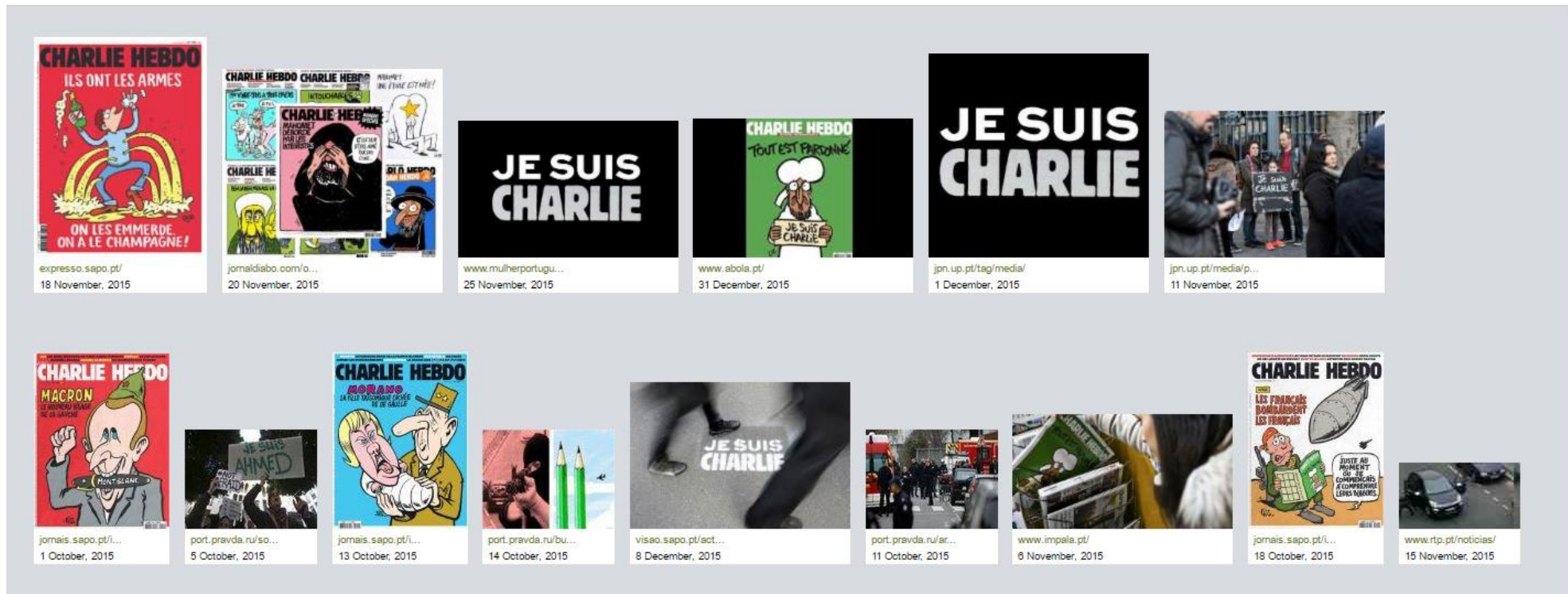
Para que servem as APIs do Arquivo.pt?

Pesquisa de Imagens



Charlie Hebdo Search

between: 01/01/2015 and: 31/12/2015 [Advanced image search](#)



Para que servem as APIs do Arquivo.pt?

Agregador de informação sobre políticos

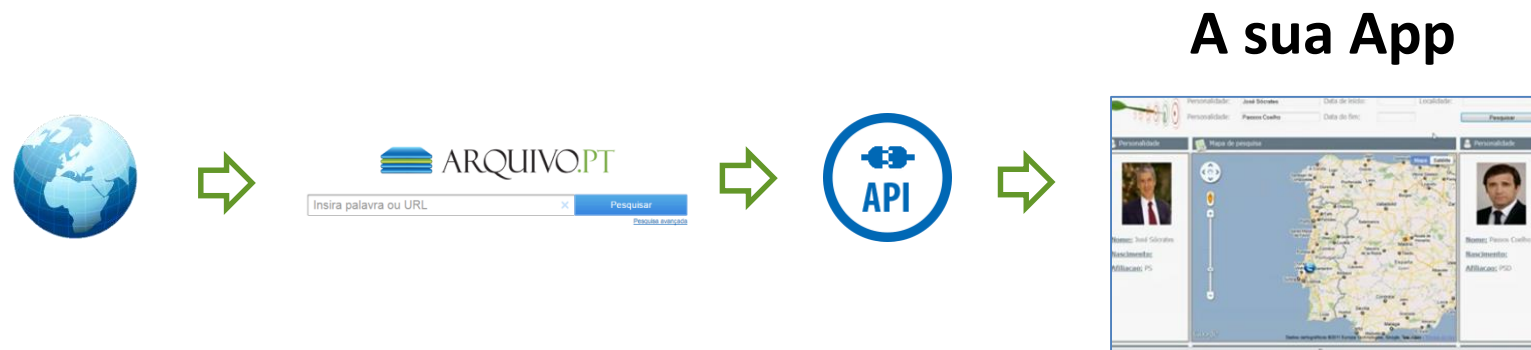


Para que servem as APIs do Arquivo.pt?

Prémio Arquivo.pt 2018



Interfaces acesso automático ao Arquivo.pt



Pesquisa por URL: CDX-server API

Pesquisa por texto: Opensearch API

Sintaxe *Wayback* para acesso a
conteúdos arquivados

Sintaxe Wayback

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



{Timestamp} = 20110121103150

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

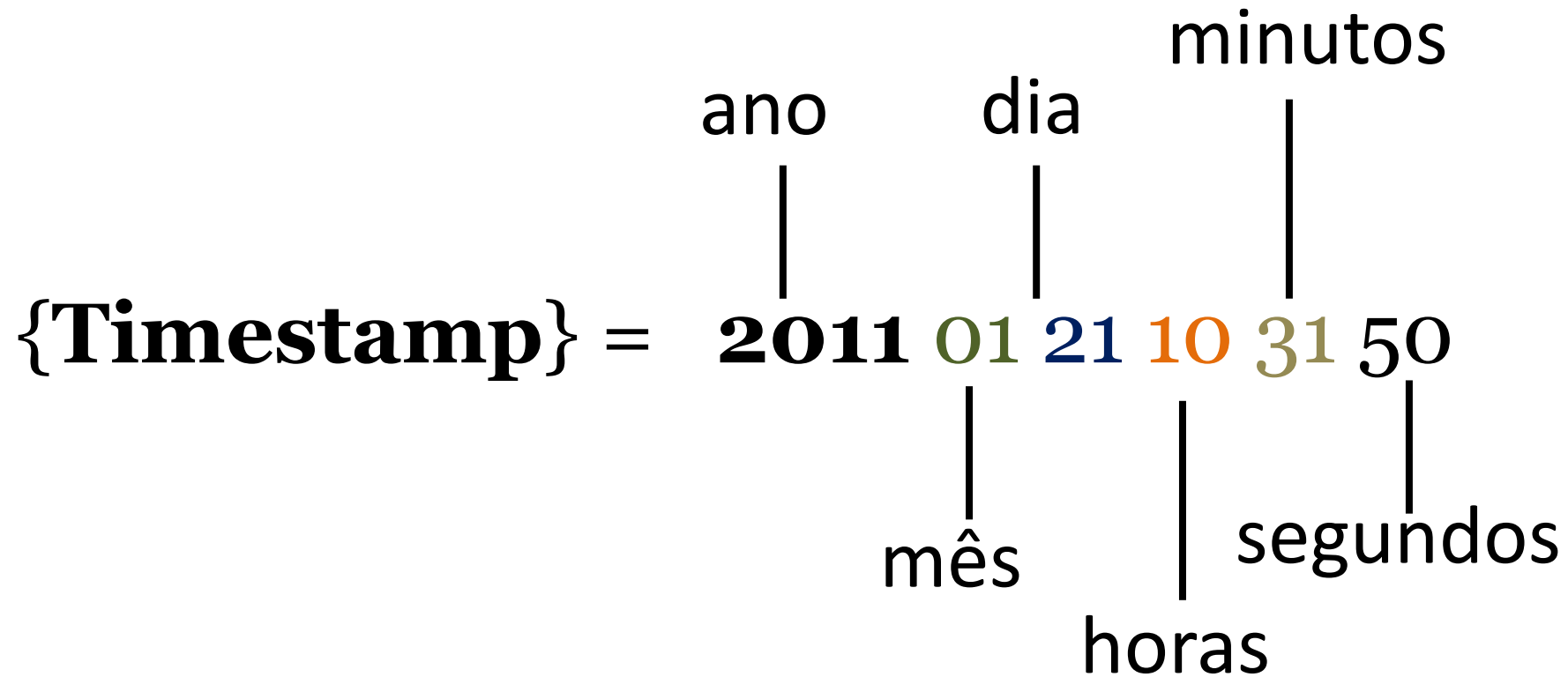
Sintaxe Wayback – **Timestamp**

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/20040000000000/http://www.euro2004.clix.pt/>

← → ↻ ☆ 🌐 S ☰

1ª página
Em Directo
Notícias
Calendário
Estádios
Seleções
Equipamentos
Arbitragem
Galeria
Histórico
Curiosidades
Opinião

Fórum
Chat

Qual é a selecção que vai ganhar o Euro2004?
 Portugal
 Espanha

Euro 2004
Jogadores da selecção portuguesa já escolheram números das camisolas
Os presumíveis titulares da selecção portuguesa de futebol vão envergar as camisolas de 1 a 11 na fase final do Euro 2004, que se disputa em Portugal de 12 de Junho a 4 de Julho, segundo a escolha hoje divulgada.



Óbidos: primeiro treino da selecção foi leve
Os 12 jogadores da selecção portuguesa de futebol já concentrados com vista ao Euro 2004 cumpriram hoje de manhã o primeiro treino do estágio de Óbidos, onde a comitiva se encontra desde ontem.

Segundo o preparador físico e o médico da selecção nacional
Euro 2004: Tiago estará a cem por cento
O estágio em Óbidos servirá para recuperar "física e mentalmente" os jogadores. O preparador físico e o médico da selecção nacional, respectivamente Darlan Schneider e Henrique Jones, dissiparam ontem todas as dúvidas em relação à condição física do médio Tiago. O jogador do Benfica é o jogador que apresenta um quadro clínico mais problemático no início do estágio de preparação para o Euro 2004, uma vez que sofre de uma pubalgia. Mas os clínicos responsáveis pelos 23 convocados portugueses garantiram que o jovem internacional português estará plenamente preparado para dar o seu melhor durante a competição.

Fyssas é um dos 23 eleitos de Otto Rehhagel
Convocatória da Grécia para o Euro 2004 sem surpresas
O seleccionador da Grécia, o alemão Otto Rehhagel, divulgou hoje o lote de 23 futebolistas para o Euro 2004 confinado a oito "estrangeiros", entre os quais o defesa esquerdo Fyssas, do Benfica, e 15 jogadores dos três maiores clubes de Atenas.

PUB

Responsáveis desdramatizam insegurança no Euro 2004 2004/05/25
O primeiro-ministro Durão Barroso garantiu ontem...

Cristiano Ronaldo e Jorge Andrade já treinam com a selecção 2004/05/25
O treino matinal de hoje da selecção portuguesa de...

Linha telefónica dá informações sobre medicamentos a turistas 2004/05/25
A Associação Nacional de Farmácias (ANF) lança...

Euro conta com meios do INEM para as áreas da saúde 2004/05/25
O ministério da Saúde já reservou o seu lugar nas...

Treino luso rendeu oito golos 2004/05/25
Um jogo a meio-campo e um total de oito golos...

Primeiro apoio popular à Selecção 2004/05/24
Eram apenas 12 os jogadores que Luiz Felipe...



Portugal
Grécia
Espanha
Rússia

França
Inglaterra
Croácia
Suíça

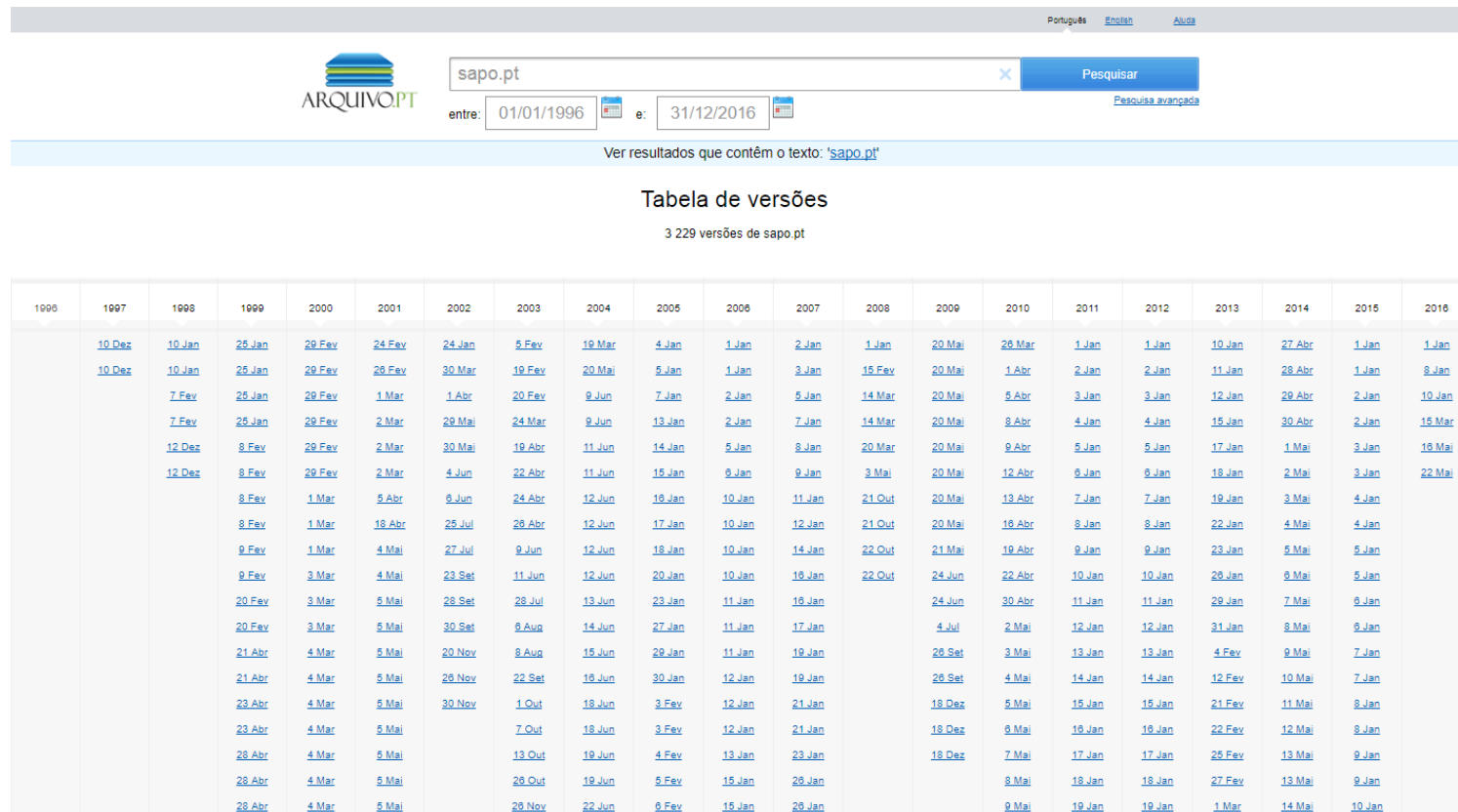
<http://arquivo.pt/wayback/20040525212736/http://euro2004.clix.pt/>

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



Sintaxe Wayback: listar todas as versões

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



Portuguesa Inglês Ajuda

ARQUIVO.PT

sapo.pt

entre: 01/01/1996 e: 31/12/2016

Pesquisar

Pesquisa avançada

Ver resultados que contêm o texto: 'sapo.pt'

Tabela de versões

3 229 versões de sapo.pt

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

Pesquisa por URL

(CDX-server API)

CDX-server API: para que serve?

arquivo.pt/cdxserverapi

Efetuar pesquisas por endereço

Filtrar e ordenar a lista de versões obtida



Pesquisa por URL: exemplo **pesquisa**

<http://arquivo.pt/wayback/-cdx?url=http://sapo.pt>

Listar todas as versões do URL:

<http://sapo.pt>

Pesquisa por URL: exemplo de resposta

```
pt,sapo)/ 20111027150223
{
  "url": "http://www.sapo.pt/",
  "mime": "text/html",
  "status": "200",
  "digest": "UJOQ5C4V7TAI2B47I3GFGZVT3MWZXH7D",
  "length": "22559",
  "offset": "290805",
  "filename": "IAH-20111027150152-00003-p12.arquivo.pt.arc.gz"
}
```

<http://arquivo.pt/wayback/2011027150223/http://sapo.pt>

Pesquisa por URL- exemplo **pesquisa json**

<http://arquivo.pt/wayback/-cdx?output=json>
[&url=http://sapo.pt](http://sapo.pt)

Listar todas as versões do URL

<http://sapo.pt>

output json

Pesquisa por URL - exemplo **resposta json**

```
{  
  "urlkey": "pt,sapo)/",  
  "timestamp": "20151221180211",  
  "status": "200",  
  "url": "http://www.sapo.pt/",  
  "filename": "IAH-20151221180151-00074-p84.arquivo.pt.arc.gz",  
  "length": "37073",  
  "mime": "text/html",  
  "offset": "1682",  
  "digest": "Z6QZU53FB4HMGCLJOHIFVLTX5ZX3GZM2"  
}
```

<http://arquivo.pt/wayback/20151221180211/http://sapo.pt>

Pesquisa por URL: **ordenar** os resultados da **pesquisa**

[http://arquivo.pt/wayback/-cdx?output=json
&url=http://sapo.pt&sort=reverse](http://arquivo.pt/wayback/-cdx?output=json&url=http://sapo.pt&sort=reverse)

Listar todas as versões do URL:

<http://sapo.pt>

Mostrar versões:

mais recentes primeiro (sort=reverse)

Pesquisa por URL: exemplo de pesquisa com definição de **data inicial**

[http://arquivo.pt/wayback/-cdx?output=json
&url=http://sapo.pt
&from=20010224174130](http://arquivo.pt/wayback/-cdx?output=json&url=http://sapo.pt&from=20010224174130)

Listar todas as versões do URL:

<http://sapo.pt>

data igual ou posterior:

24 de fevereiro de 2001, 17h41m30s

Pesquisa por URL dentro de intervalo **temporal**

[http://arquivo.pt/wayback/-cdx?output=json
&url=http://sapo.pt
&from=20010224174130
&to=20150424181416](http://arquivo.pt/wayback/-cdx?output=json&url=http://sapo.pt&from=20010224174130&to=20150424181416)

Listar todas as versões do URL:

<http://sapo.pt>

data entre:

24 de fevereiro de 2001

e:

24 de abril de 2015

Pesquisa por **prefixo** de URL

```
http://arquivo.pt/wayback/-cdx?output=json  
&url=http://noticias.sapo.pt  
&matchType=prefix  
&limit=10000
```

Listar todas as versões de URLs começados pelo **prefixo**:

[http://noticias.sapo.pt](#)

apresentando os primeiros:

10 000 resultados

Pesquisa por URL– exemplo pesquisa por **mimetype**

<http://arquivo.pt/wayback/-cdx?output=json>
[&url=http://sapo.pt](http://arquivo.pt/wayback/-cdx?output=json&url=http://sapo.pt)
[&limit=10000](http://arquivo.pt/wayback/-cdx?output=json&limit=10000)
[&filter=mime:text/html](http://arquivo.pt/wayback/-cdx?output=json&filter=mime:text/html)

Listar todas as versões de:

páginas html

do URL:

<http://sapo.pt>

primeiros **10 000** resultados

Pesquisa por URL – desafio

<http://arquivo.pt/wayback/-cdx?output=json>
[&url=http://sapo.pt](http://sapo.pt)
[**&matchType=prefix**](#)
[&limit=10000](#)
[**&filter=mime:image/jpeg**](#)

Qual a resposta a este pedido?

Pesquisa por URL com **filtro de resposta**

```
http://arquivo.pt/wayback/-cdx?  
output=json  
&url=http://sapo.pt  
&fl=url,timestamp,status
```

Filtrar resposta para mostrar apenas:
URL da página preservada
data (timestamp) em que foi preservada
status code HTTP

Pesquisa por URL: **resposta filtrada**

```
{"url": "http://sapo.pt/", "timestamp": "19980207015433", "status": "200"}  
{"url": "http://sapo.pt/", "timestamp": "19981212020853", "status": "302"}  
{"url": "http://sapo.pt/", "timestamp": "19981212020853", "status": "302"}
```

Pesquisa por texto

(OpenSearch API)

OpenSearch - Definição

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

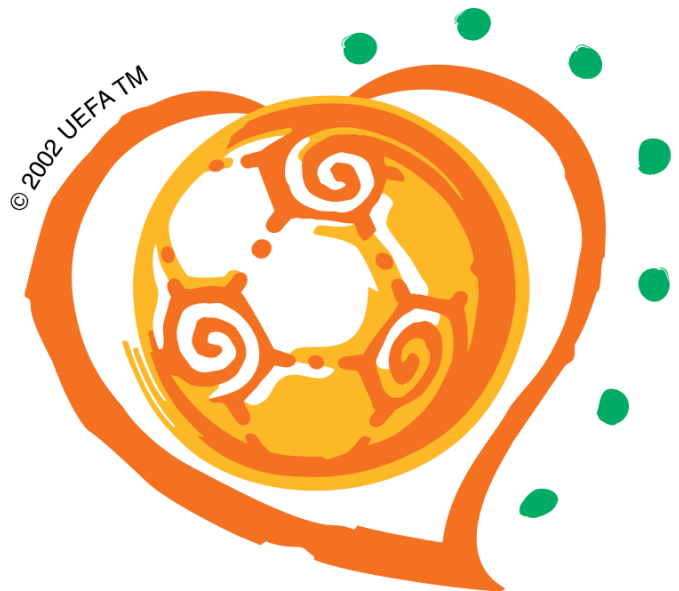
“OpenSearch is a collection of simple formats for the sharing of search results.”

API do Arquivo.pt – baseada no OpenSearch

Suporta pesquisas temporais sobre o Arquivo.pt

Pesquisa por texto: documentação

arquivo.pt/openserchapi



UEFA
Euro 2004
PORTUGAL

Pesquisa por texto – exemplo de **pesquisa**

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

Parâmetros pré-definidos:

start=0 (primeiro resultado)

hitsPerPage=10 (número de resultados)

Pesquisa por texto - pesquisar **próximos 10 resultados**

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

start=10 (próximos 10 resultados)

Pesquisa por texto – **2 000 resultados**

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

Pesquisa por texto – **pesquisar por uma expressão**

[http://arquivo.pt/openserch?query="euro 2004"](http://arquivo.pt/openserch?query=)

Resultados com expressão:
“euro 2004”





Pesquisa por texto: **excluir uma palavra**

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

Pesquisar pelas palavras:

euro 2004

mas sem a palavra:

moeda

Pesquisa por texto - **pesquisa por mimetype PDF**

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

Pesquisa por que contenham as palavras:
euro 2004
do tipo:
PDF

Resposta a pesquisa por texto: resumo dos resultados

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<rss xmlns:opensearch="http://a9.com/-/spec/opensearch/1.1/">
<channel>
  <title>PWA Search Engine</title>
  <description>PWA search results for query: euro 2004</description>
  <link>http://archive.pt</link>
  <opensearch:totalResults>18185160</opensearch:totalResults>
  <opensearch:startIndex>0</opensearch:startIndex>
  <opensearch:itemsPerPage>10</opensearch:itemsPerPage>
  <opensearch:Query role="request" searchTerms="euro 2004" startPage="1"/>
  <item>...</item>
  <item>...</item>
  <item>...</item>
  <item>...</item>
  <item>...</item>
  <item>...</item>
  <item>...</item>
  <item>...</item>
  <item>...</item>
  <item>...</item>
  <item>...</item>
  <item>...</item>
</channel>
</rss>
```

Resposta a pesquisa por texto: *items* dos resultados

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<rss xmlns:opensearch="http://a9.com/-/spec/opensearch/1.1/">
<channel>
  <title>PWA Search Engine</title>
  <description>PWA search results for query: euro 2004</description>
  <link>http://archive.pt</link>
  <opensearch:totalResults>18185160</opensearch:totalResults>
  <opensearch:startIndex>0</opensearch:startIndex>
  <opensearch:itemsPerPage>10</opensearch:itemsPerPage>
  <opensearch:Query role="request" searchTerms="euro 2004" startPage="1"/>
  <item>...</item>
  <item>...</item>
  <item>...</item>
  <item>...</item>
  <item>...</item>
  <item>...</item>
  <item>...</item>
  <item>...</item>
  <item>...</item>
  <item>...</item>
  <item>...</item>
</channel>
</rss>
```

Item de resultado

```
<item>
  <title>Euro 2004 - Estádios do Euro 2004</title>
  <source url="http://helder2004.no.sapo.pt/">
    Original URL of Euro 2004 - Estádios do Euro 2004
  </source>
  <link>http://arquivo.pt/wayback/20131106141720/http://helder2004.no.sapo.pt/</link>
  <pwa:id>2506798</pwa:id>
  <pwa:index>24</pwa:index>
  <pwa:arcname>IAH-20131105224049-01265-p13.arquivo.pt</pwa:arcname>
  <pwa:arcoffset>86869029</pwa:arcoffset>
  <pwa:digest>7bb196359aef86a45767a34b21a78a97</pwa:digest>
  <pwa:tstamp>20131105224456000</pwa:tstamp>
  <pwa:contentLength>1718</pwa:contentLength>
  <pwa:primaryType>text</pwa:primaryType>
  <pwa:subType>html</pwa:subType>
</item>
```

Obrigado

<http://arquivo.pt/api>

Fernando Melo <fernando.melo@fccn.pt>