



Automatic Processing

fernando.melo@fccn.pt

Tutorial outline

Research the Past Web using Web archives

1. Search and access

- The Past Web: examples and use cases
- Public online services

2. Publish and preserve

- Recommendations to publish preservable information
- How to preserve information collected from the Web

3. Automatic processing

- Interoperability protocols
- Application Programming Interfaces (API)

Wayback Syntax

Fast access to Web archived content

Wayback Syntax – it works on other web archives



`http://arquivo.pt/wayback/{Timestamp}/{URL}`



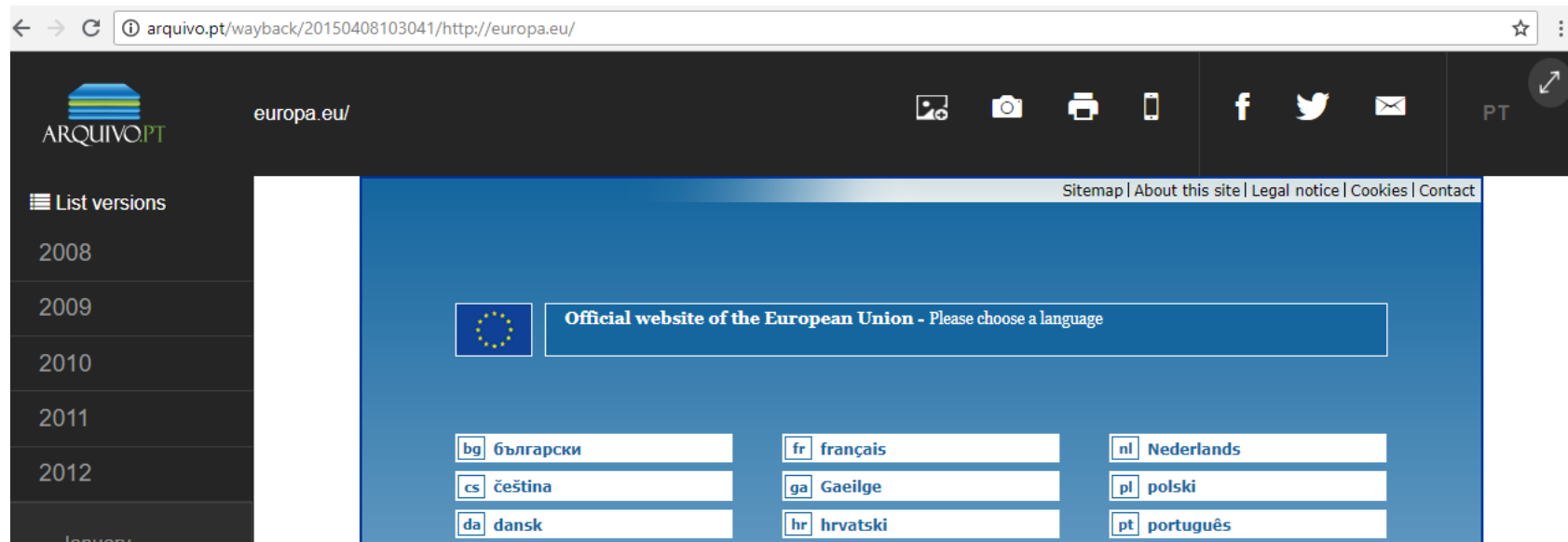
`http://web.archive.org/web/{Timestamp}/{URL}`



`http://webarchive.org.uk/wayback/archive/{Timestamp}/{URL}`

Wayback Syntax

`http://arquivo.pt/wayback/{Timestamp}/{URL}`



{Timestamp} = 20150408103041

{URL} = <http://europa.eu>

Wayback Syntax

```
http://arquivo.pt/wayback/{Timestamp}/{URL}
```

14 Digits

{Timestamp} = 2015 04 08 10 30 41

Wayback Syntax

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

14 Digits

14 Digits

{Timestamp} = 2015 04 08 10 30 41

year day minutes

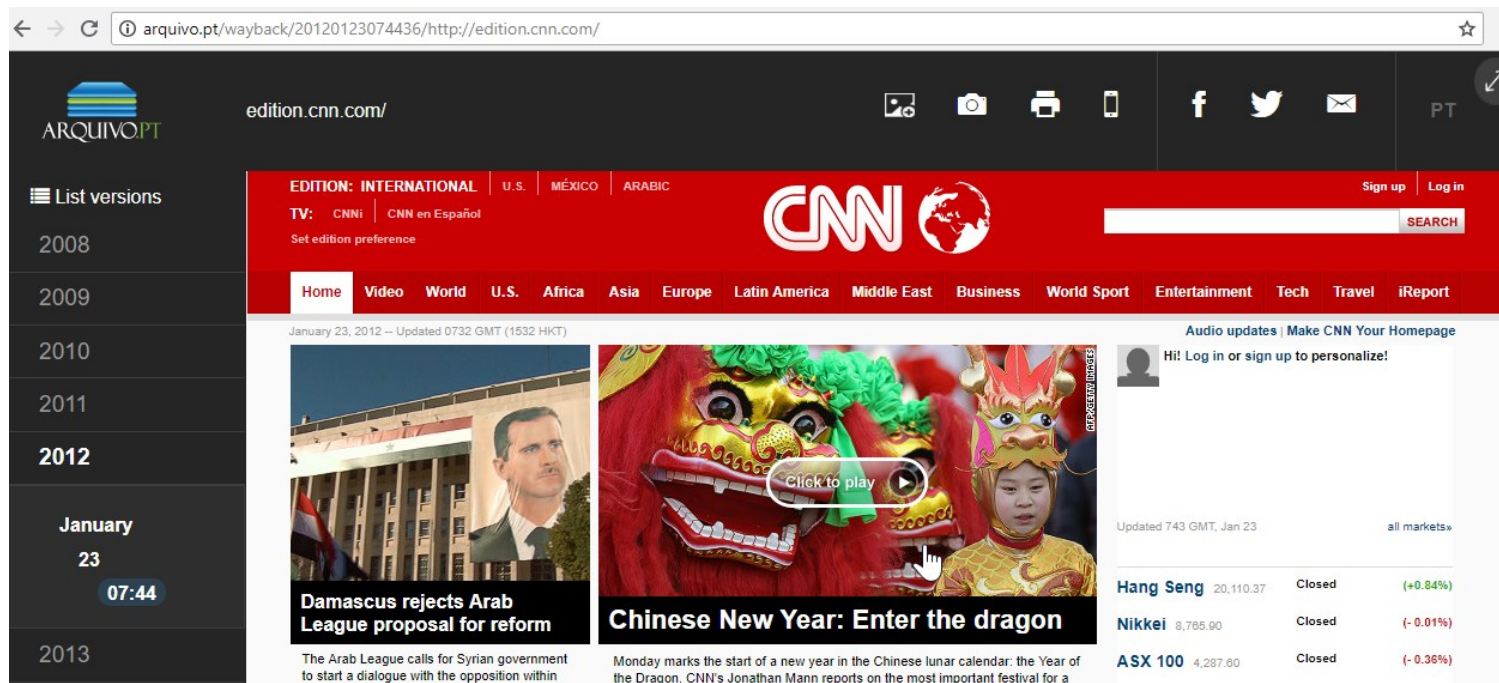
month hours seconds

Wayback Syntax - closest date

What if we don't know the exact timestamp?

Wayback Syntax – closest date

<http://arquivo.pt/wayback/20120000000000/edition.cnn.com>



<http://arquivo.pt/wayback/20120123074436/http://edition.cnn.com/>

Wayback Syntax – list all archived versions

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

[About](#) [Português](#) 



× Search

between:  and:  [Advanced search](#)

See webpages with the text: 'sapo.pt'

Versions list

3 611 versions of sapo.pt

1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
10 Dec	10 Jan	25 Jan	29 Feb	24 Feb	24 Jan	5 Feb	19 Mar	4 Jan	1 Jan	2 Jan	1 Jan	20 May	26 Mar	1 Jan	1 Jan	10 Jan	27 Apr	1 Jan	1 Jan	Available next year
10 Dec	10 Jan	25 Jan	29 Feb	26 Feb	30 Mar	19 Feb	20 May	5 Jan	1 Jan	3 Jan	15 Feb	20 May	1 Apr	2 Jan	2 Jan	11 Jan	28 Apr	1 Jan	1 Jan	
	7 Feb	25 Jan	29 Feb	1 Mar	1 Apr	20 Feb	9 Jun	7 Jan	2 Jan	5 Jan	14 Mar	20 May	5 Apr	3 Jan	3 Jan	12 Jan	29 Apr	2 Jan	2 Jan	
	7 Feb	25 Jan	29 Feb	2 Mar	29 May	24 Mar	9 Jun	13 Jan	2 Jan	7 Jan	14 Mar	20 May	8 Apr	4 Jan	4 Jan	15 Jan	30 Apr	2 Jan	2 Jan	
	12 Dec	8 Feb	29 Feb	2 Mar	30 May	19 Apr	11 Jun	14 Jan	5 Jan	8 Jan	20 Mar	20 May	9 Apr	5 Jan	5 Jan	17 Jan	1 May	3 Jan	3 Jan	
	12 Dec	8 Feb	29 Feb	2 Mar	4 Jun	22 Apr	11 Jun	15 Jan	6 Jan	9 Jan	3 May	20 May	12 Apr	6 Jan	6 Jan	18 Jan	2 May	3 Jan	4 Jan	

Wayback Syntax - challenge

1. Closest version: October 2008 of the URL nytimes.com
2. Number of archived versions of the website theguardian.com

Wayback Syntax – challenge 1

<http://arquivo.pt/wayback/20081000000000/nytimes.com>



www.nytimes.com/

Log In | Register Now

List versions

2008

February

March

October

22

12:52

24

2009

2010

2011

2012

2013

2014

2015

2016

HOME PAGE | MY TIMES | TODAY'S PAPER | VIDEO | MOST POPULAR | TIMES TOPICS

The New York Times

Wednesday, October 22, 2008 Last Update: 9:54 AM ET

Personalize Your Weather

Search

JOB | REAL ESTATE | AUTOS | Blogs | Cartoons / Humor

ALL CLASSIFIEDS | WORLD | U.S.

POLITICS | N.Y./REGION | BUSINESS

TECHNOLOGY | SPORTS | SCIENCE

HEALTH | OPINION | Classifieds

ARTS

BOOKS | MOVIES | MUSIC

TELEVISION | THEATER

STYLE

DINING & WINE | FASHION & STYLE

HOME & GARDEN

WEDDINGS/ CELEBRATIONS

TRAVEL | Corrections | Crossword/ Games

First Look

Learning Network

Multimedia

NYC Guide

Obituaries

Podcasts

Public Editor

Sunday Magazine

T Magazine

Video

Week in Review

Weather

SERVICES

My Alerts

NYT Mobile

NYT Store

McCain Camp Fights to Keep Crucial Blue State in Play

By ELISABETH BUMILLER and JEFF ZELENY

Strategists insist that Pennsylvania's electoral votes are within reach for John McCain's campaign.

• Palin's Hand Seen in Mine Battle

On Health Plans, the Numbers Fly

By KEVIN SACK

Senators Barack Obama and John McCain often cite rough estimates when talking about their health care plans.

Post a Comment | Read (18)

THE 2008 CAMPAIGN



Ozler Muhammad/The New York Times

Wayback Syntax – challenge 2

`http://arquivo.pt/wayback/*/theguardian.com`

 [Advanced search](#)

between: and:

See webpages with the text: '[theguardian.com](#)'

Versions list

122 versions of theguardian.com

1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
															30 Jul	6 Sep	9 Apr	8 Jan	Available next year
															6 Nov	7 Sep	14 Aug	21 Jan	
															8 Nov	23 Nov	13 Nov	28 Jan	
																26 Nov		5 Feb	
																		14 Mar	
																		21 Apr	
																		13 May	
																		14 May	
																		20 Jun	
																		31 Oct	
																		8 Nov	

Wayback Syntax

How about other Web archives?

Wayback Syntax - challenge

1. Closest version: 2009 of the URL [youtube.com](https://www.youtube.com) in Internet Archive
2. List all archived versions of the website [gov.uk](https://www.gov.uk) in UK Web Archive

Wayback Syntax – challenge 1

<http://web.archive.org/web/20090000000000/youtube.com>

← → ↻ ⓘ Not secure | web.archive.org/web/20090101094601/http://www.youtube.com/

INTERNET ARCHIVE
Wayback Machine
559,233 captures
28 Apr 2008 - 27 Aug 2018

Go NOV JAN FEB
2007 01 2009 2010

YouTube Broadcast Yourself
Worldwide | English

Sign Up | QuickList (0) | Help | Sign In

Home Videos Channels Community

Search Upload

Promoted Videos

 [I KISSED A DOG \(A K...](#)
omovies

 [NBA highlights from...](#)
NBA

 [How to apply blush ...](#)
SelfMagazine

 [A New Plan | Carpet...](#)
carpetbros

Advertisement

Want to personalize this homepage?
[Sign up for a YouTube Account](#)
[Sign in with your Google Account](#) ?

What's New

 [YouTube Greeting Cards](#)
Send video greeting cards to your friends and family

 [Captions and Subtitles](#)
Add multi-track captions and subtitles to your videos

 [Video Annotations](#)
Add interactive commentary and links to your videos

[2008: The Year in Video](#)
It has been a busy and exciting year here at YouTube. With 13 hours of video uploaded every minute, hundreds of millions of views a day and 23 country-specific versions of the site, apart from anyt...
[Read more in our Blog](#)

Featured Videos

[See More Featured Videos](#)

Featured | Most Viewed | Most Discussed | Top Favored

 **Pork and Beans**
Pork & Beans official dance contest at www.dancejam.com Watch the official vide...
★★★★★ 15,883,838 views weezer

 **Christian the Lion - the full story (in HQ)**
© Beckmann Group. TO MAKE A DONATION AND HELP THE WORLD'S WILDLIFE
★★★★★ 18,583,712 views TadManly

 **Frozen Grand Central**
From <http://www.ImprovEverywhere.com>, over 200 people freeze in place on cue in
★★★★★ 14,660,521 views ImprovEverywhere

Wayback Syntax – challenge 2

`http://webarchive.org.uk/wayback/archive/*/gov.uk`

← → ↻ Secure | https://www.webarchive.org.uk/wayback/archive/*/gov.uk



You are here: [Home](#) > Open UK Web Archive

Provided by: 

Enter Web Address:

Searched for [http://gov.uk](#) Set Anchor Window: 0 Results

Search Results for 1 Jan, 1996 - 31 Dec, 2018

Jan 1996 - Dec 1996	Jan 1997 - Dec 1997	Jan 1998 - Dec 1998	Jan 1999 - Dec 1999	Jan 2000 - Dec 2000	Jan 2001 - Dec 2001	Jan 2002 - Dec 2002	Jan 2003 - Dec 2003	Jan 2004 - Dec 2004	Jan 2005 - Dec 2005	Jan 2006 - Dec 2006	Jan 2007 - Dec 2007	Jan 2008 - Dec 2008	Jan 2009 - Dec 2009	Jan 2010 - Dec 2010	Jan 2011 - Dec 2011	Jan 2012 - Dec 2012	Jan 2013 - Dec 2013	Jan 2014 - Dec 2014	Jan 2015 - Dec 2015	Jan 2016 - Dec 2016	Jan 2017 - Dec 2017	Jan 2018 - Dec 2018	Jan 2019 - Dec 2019
0 pages	0 pages	0 pages	0 pages	0 pages	0 pages	0 pages	0 pages	0 pages	0 pages	0 pages	0 pages	0 pages	0 pages	0 pages	0 pages	8 pages	62 pages	35 pages	28 pages				
																	15 Apr. 2013	23 Jan. 2014	4 Jan. 2016	1 Apr. 2018			
																	15 Apr. 2013	18 Feb. 2014	4 Jan. 2016	1 Apr. 2018			
																	15 Apr. 2013	18 Feb. 2014	5 Jan. 2016	6 Jun. 2018			
																	24 Apr. 2013	7 Apr. 2014	5 Feb. 2016	6 Jun. 2018			
																	24 Apr. 2013	11 Apr. 2014	22 Feb. 2016	26 Jun. 2018			
																	1 Aug. 2013	24 Apr. 2014	24 Feb. 2016	26 Jun. 2018			
																	5 Aug. 2013	24 Apr. 2014	25 Feb. 2016	27 Jun. 2018			
																	5 Aug. 2013	11 Jun. 2014	1 Apr. 2016	28 Jun. 2018			
																		11 Jun. 2014	1 Apr. 2016	28 Jun. 2018			
																		14 Jul. 2014	20 May. 2016	28 Jun. 2018			
																		14 Jul. 2014	4 Jun. 2016	20 Jun. 2018			

Application Programming Interfaces (APIs)

API – Application Programming Interface



Automatic access

Easy integration

Fast development of new applications

No need to understand core code

Web Archives APIs

Arquivo.pt API

- + Search by text or by URL
- Only works on Arquivo.pt



Memento TimeTravel API

- + Search in several Web archives
- Only search by URL



Arquivo.pt API



Arquivo.pt API



arquivo.pt/textsearch

Automatic access

URL search

Text search

Metadata search

Arquivo.pt API



arquivo.pt/textsearch

API response in **JSON** format.

Arquivo.pt API use case



http://contamehistorias.inesctec.pt/arquivopt/?lang_code=en



Explore examples

Economy and Society

Crise Oriente Médio Política Internacional	Crise Coreia do Norte Política Internacional	Crise Crimeia Política Internacional	II Guerra Mundial História
--	--	--	--------------------------------------

• • • • •

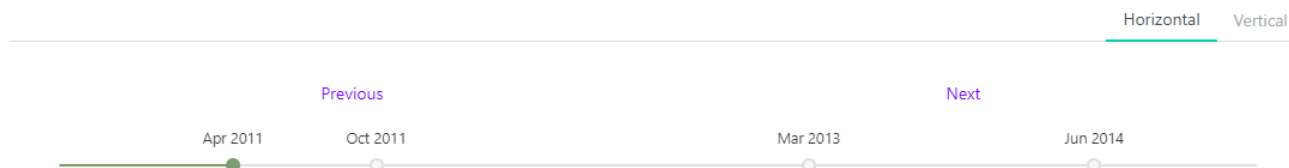
Politics, Justice and Journalism

BES Justiça e Economia	Portugal Telecom Justiça e Economia	Zeinal Bava Justiça e Economia	Oliveira e Costa Justiça e Economia
----------------------------------	---	--	---

Arquivo.pt API use case



Narrative sobre Barack Obama during the last 10 years



- Notícias relevantes entre 25, June 2009 e 26, April 2011

Barack Obama e Lula da Silva querem acordo climático.

Sarah Jessica Parker vai trabalhar para Barack Obama. positive

Michelle e Barack Obama lideram caça aos ovos da Páscoa na Casa Branca.

Lady Gaga ultrapassa Barack Obama no Facebook. negative

Barack Obama homenageia Bruce Springsteen na Casa Branca. positive

Barack Obama visita loja gourmet em Miami.

Michelle e Barack Obama aproveitam as férias para namorar. negative

Eva Longoria deslumbra em jantar de gala com Barack Obama. positive

Cavaco convida Barack Obama para visita oficial a Portugal.

Barack Obama alvo de ataque religioso nos Estados Unidos.

Barack Obama entre os mais poderosos do mundo.

Barack Obama aterra em Lisboa.

Barack Obama faz visita surpresa ao Afeganistão.

Barack Obama chega ao Havai para as férias de Natal.

Barack Obama decreta estado de emergência sanitária nos EUA.

Filhas de Barack Obama impedidas de usar Facebook. negative

Raul Castro disposto a dialogar com Barack Obama.

Barack Obama de visita ao Brasil.

Barack Obama recandidata-se a presidente dos Estados Unidos.

Barack Obama anuncia recandidatura na Internet.

URL search

(Arquivo.pt API)

The New York Times

[nytimes.com](https://www.nytimes.com)

URL search request



[arquivo.pt/textsearch?versionHistory=nytimes.com
&prettyPrint=true](http://arquivo.pt/textsearch?versionHistory=nytimes.com&prettyPrint=true)

List archived versions of the URL:

nytimes.com

offset=0 (first result)

maxItems=50 (number of results)

URL search response



```
{
  "title": "The New York Times - Breaking News, World News & Multimedia",
  "originalURL": "http://www.nytimes.com/",
  "linkToArchive": "http://arquivo.pt/wayback/20141223183313/http://www.nytimes.com/",
  "tstamp": "20141223183313",
  "contentLength": "38701",
  "digest": "HCUQ77BX4OWLNAU6VLMF7CV6RHIT3ZJ5",
  "mimeType": "text/html",
  "linkToScreenshot": "http://arquivo.pt/screenshot/?url=http%3A%2F%2Farquivo.pt%2FnoFrame%2Freplay%2F20141223183313/http://www.nytimes.com/",
  "date": "1419359593",
  "encoding": "UTF-8",
  "linkToNoFrame": "http://arquivo.pt/noFrame/replay/20141223183313/http://www.nytimes.com/",
  "status": "200",
  "collection": "FAWP1920141223",
  "linkToExtractedText": "http://arquivo.pt/textextracted?m=http%3A%2F%2Fwww.nytimes.com%2F%2F20141223183313/http://www.nytimes.com/",
  "linkToMetadata": "http://arquivo.pt/textsearch?metadata=http%3A%2F%2Fwww.nytimes.com%2F%2F20141223183313/http://www.nytimes.com/"
}
```

URL search request parameter: **offset**



[arquivo.pt/textsearch?versionHistory=nytimes.com
&offset=50](http://arquivo.pt/textsearch?versionHistory=nytimes.com&offset=50)

offset=50 (firstresult)

maxItems=50 (number of results)

URL search request: maxItems



[arquivo.pt/textsearch?versionHistory=nytimes.com](http://arquivo.pt/textsearch?versionHistory=nytimes.com&maxItems=1000&from=2012&to=2014)
[&maxItems=1000](#)
[&from=2012](#)
[&to=2014](#)

maxItems=1000 (number of results)

URL search request: from



[arquivo.pt/textsearch?versionHistory=nytimes.com
&from=20100224174130](http://arquivo.pt/textsearch?versionHistory=nytimes.com&from=20100224174130)

List archived versions of URL

nytimes.com

With date equal or after:

2010 February 24, at 17h41m30s

URL search request: from to



[arquivo.pt/textsearch?versionHistory=nytimes.com
&from=20100224174130&to=20150223183001](http://arquivo.pt/textsearch?versionHistory=nytimes.com&from=20100224174130&to=20150223183001)

List archived versions of URL

nytimes.com

With date between:

2010 February 24, at 17h41m30s

And

2015 February 23, at 18h30m01s

URL search request: fields



[arquivo.pt/textsearch?versionHistory=nytimes.com
&fields=originalURL,tstamp](https://arquivo.pt/textsearch?versionHistory=nytimes.com&fields=originalURL,tstamp)

Filter response to show only

URL of the page

Timestamp (**tstamp**) when it was preserved

URL search response: fields



```
{  
  "originalURL": "http://nytimes.com/",  
  "tstamp": "20161231190553"  
},  
{  
  "originalURL": "https://www.nytimes.com/",  
  "tstamp": "20161231180537"  
},  
{  
  "originalURL": "http://www.nytimes.com/",  
  "tstamp": "20161231180218"  
}
```

Text search

(Arquivo.pt API)



Text search request



[arquivo.pt/textsearch?q=euro 2004
&prettyPrint=true](http://arquivo.pt/textsearch?q=euro%202004&prettyPrint=true)

Search words **euro** and **2004**.

offset=0 (first result)

maxItems=50 (number of results)

Text search request: **expression**



arquivo.pt/textsearch?q=“euro 2004”

Search results with expression

“euro2004”





Text search request: **exclude word**



arquivo.pt/textsearch?q=euro 2004 -currency

Search results with words

euro 2004

without word

currency

Text search request: **type**



[arquivo.pt/textsearch?q=euro 2004&type=pdf](http://arquivo.pt/textsearch?q=euro%202004&type=pdf)

Search results with words

euro 2004

In files of type:

PDF

Text search response: summary of results

```
{  
  "serviceName": "Arquivo.pt - the Portuguese web-archive",  
  "linkToService": "http://arquivo.pt",  
  "next_page": "http://arquivo.pt/textsearch?q=euro%202004&prettyPrint=true&offset=50",  
  "previous_page": "http://arquivo.pt/textsearch?q=euro%202004&prettyPrint=true&offset=0",  
  "total_items": "21664141",  
  "request_parameters": {  
    "q": "euro 2004",  
    "prettyPrint": "true"  
  },  
}
```

Text search response: result item

```
{
  "title": "Euro 2004 - Estádios do Euro 2004",
  "originalURL": "http://helder2004.no.sapo.pt/",
  "linkToArchive": "http://arquivo.pt/wayback/20091217183109/http://helder2004.no.sapo.pt/",
  "tstamp": "20091217183109",
  "contentLength": "1718",
  "digest": "7bb196359aef86a45767a34b21a78a97",
  "mimeType": "text/html",
  "linkToScreenshot": "http://arquivo.pt/screenshot/?url=http%3A%2F%2Farquivo.pt%2FnoFrame%2Freplay%2F20091217183109",
  "date": "1261074669",
  "encoding": "windows-1252",
  "linkToNoFrame": "http://arquivo.pt/noFrame/replay/20091217183109/http://helder2004.no.sapo.pt/",
  "snippet": "<em>Euro</em> <em>2004</em> - Est&aacute;dios do <em>Euro</em>
  <em>2004</em> &nbsp; Aqui Pode Encontrar Todos Os Palcos Do <em>Euro</em> <em>2004</em>
  &nbsp; &nbsp; ENTRAR<span class=\"ellipsis\"> ... </span>",
  "collection": "AWP6",
  "linkToExtractedText": "http://arquivo.pt/textextracted?m=http%3A%2F%2Fhelder2004.no.sapo.pt%2F%2F20091217183109",
  "linkToMetadata": "http://arquivo.pt/textsearch?metadata=http%3A%2F%2Fhelder2004.no.sapo.pt%2F%2F20091217183109"
}
```

Text search response field: linkToMetadata

```
{
  "title": "Euro 2004 - Estádios do Euro 2004",
  "originalURL": "http://helder2004.no.sapo.pt/",
  "linkToArchive": "http://arquivo.pt/wayback/20091217183109/http://helder2004.no.sapo.pt/",
  "tstamp": "20091217183109",
  "contentLength": "1718",
  "digest": "7bb196359aef86a45767a34b21a78a97",
  "mimeType": "text/html",
  "linkToScreenshot": "http://arquivo.pt/screenshot/?url=http%3A%2F%2Farquivo.pt%2FnoFrame%2Freplay%2F20091217183109",
  "date": "1261074669",
  "encoding": "windows-1252",
  "linkToNoFrame": "http://arquivo.pt/noFrame/replay/20091217183109/http://helder2004.no.sapo.pt/",
  "snippet": "<em>Euro</em> <em>2004</em> - Est&aaacute;dios do <em>Euro</em>
  <em>2004</em> &nbsp; Aqui Pode Encontrar Todos Os Palcos Do <em>Euro</em> <em>2004</em>
  &nbsp; &nbsp; ENTRAR<span class=\"ellipsis\"> ... </span>",
  "collection": "AWP6",
  "linkToExtractedText": "http://arquivo.pt/textextracted?m=http%3A%2F%2Fhelder2004.no.sapo.pt%2F%2F20091217183109",
  "linkToMetadata": "http://arquivo.pt/textsearch?metadata=http%3A%2F%2Fhelder2004.no.sapo.pt%2F%2F20091217183109"
}
```

Metadata search response

```
{  
  "title": "Euro 2004 - Estádios do Euro 2004",  
  "originalURL": "http://helder2004.no.sapo.pt/",  
  "linkToArchive": "http://arquivo.pt/wayback/20091217183109/http://helder2004.no.sapo.pt/",  
  "tstamp": "20091217183109",  
  "contentLength": "1203",  
  "digest": "EYUZNPAREQBWSZLOMRM7FKJSHJXNQWTZ",  
  "mimeType": "text/html",  
  "linkToScreenshot": "http://arquivo.pt/screenshot/?url=http%3A%2F%2Farquivo.pt%2FnoFrame%2Freplay%2F20091217183109",  
  "date": "1261074669",  
  "encoding": "windows-1252",  
  "linkToNoFrame": "http://arquivo.pt/noFrame/replay/20091217183109/http://helder2004.no.sapo.pt/",  
  "status": "200",  
  "collection": "AWP6",  
  "linkToExtractedText": "http://arquivo.pt/textextracted?m=http%3A%2F%2Fhelder2004.no.sapo.pt%2F%2F20091217183109",  
  "linkToMetadata": "http://arquivo.pt/textsearch?metadata=http%3A%2F%2Fhelder2004.no.sapo.pt%2F%2F20091217183109",  
  "filename": "IAH-20091217182045-00726-awp01.fccn.pt.arc.gz",  
  "offset": "99075262"  
}
```

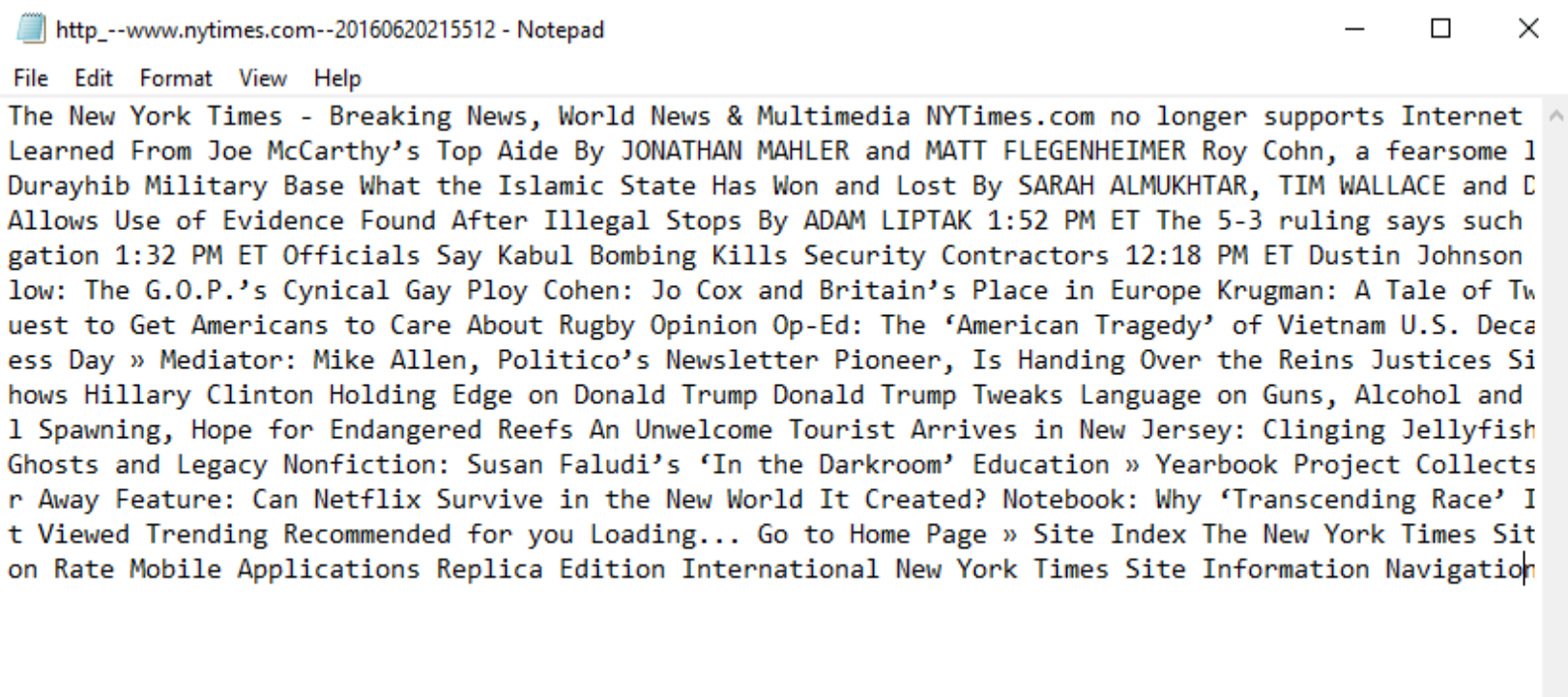

Text search response: linkToExtractedText

```
{
  "title": "Euro 2004 - Estádios do Euro 2004",
  "originalURL": "http://helder2004.no.sapo.pt/",
  "linkToArchive": "http://arquivo.pt/wayback/20091217183109/http://helder2004.no.sapo.pt/",
  "tstamp": "20091217183109",
  "contentLength": "1718",
  "digest": "7bb196359aef86a45767a34b21a78a97",
  "mimeType": "text/html",
  "linkToScreenshot": "http://arquivo.pt/screenshot/?url=http%3A%2F%2Farquivo.pt%2FnoFrame%2Freplay%2F20091217183109",
  "date": "1261074669",
  "encoding": "windows-1252",
  "linkToNoFrame": "http://arquivo.pt/noFrame/replay/20091217183109/http://helder2004.no.sapo.pt/",
  "snippet": "<em>Euro</em> <em>2004</em> - Est&acute;dios do <em>Euro</em>
<em>2004</em> &nbsp; Aqui Pode Encontrar Todos Os Palcos Do <em>Euro</em> <em>2004</em>
&nbsp; &nbsp; ENTRAR<span class=\"ellipsis\"> ... </span>",
  "collection": "AWP6",
  "linkToExtractedText": "http://arquivo.pt/textextracted?m=http%3A%2F%2Fhelder2004.no.sapo.pt%2F%2F20091217183109",
  "linkToMetadata": "http://arquivo.pt/textsearch?metadata=http%3A%2F%2Fhelder2004.no.sapo.pt%2F%2F20091217183109"
}
```


linkToExtractedText: downloads as txt file



linkToExtractedText: txt file



Arquivo.pt: application programming example

github.com/arquivo/example-api

Click on example.html

```
<html>
<head>
  <link rel="stylesheet" href="https://maxcdn.bootstrapcdn.com/bootstrap/4.0.0/css/bootstrap.min.css">
  <meta name="viewport" content="width=device-width, initial-scale=1.0">
</head>
<body>
<div id="content"></div>
<script>
  function handler(response) {
    for (var i = 0; i < response.response_items.length; i++) {
      var item = response.response_items[i];
      document.getElementById("content").innerHTML += "<br/>"
      + "<div class='col offset-sm-2 col-sm-8 alert alert-primary' style='word-wrap:break-word'>"
      + "<a href='"+item.linkToArchive+"'><h4>"+ item.title+"</h4></a>"
      + "<strong><h5 class='text-dark'>"+ item.tstamp.substr(0,4) +"</strong></h5>"
      + "<h5 class='text-muted'><a>"+ item.originalURL+"</a></h5>";
      + "</div><br/>" ;
    }
  }
</script>
<script src="http://arquivo.pt/textsearch?q=Barack%20Obama&maxItems=5&itemsPerSite=1&callback=handler"></script>
</body>
</html>
```

Head of example.html

```
<head>  
  <link rel="stylesheet" href="https://maxcdn.bootstrapcdn.com/bootstrap/4.0.0/css/bootstrap.min.css">  
  <meta name="viewport" content="width=device-width, initial-scale=1.0">  
</head>
```

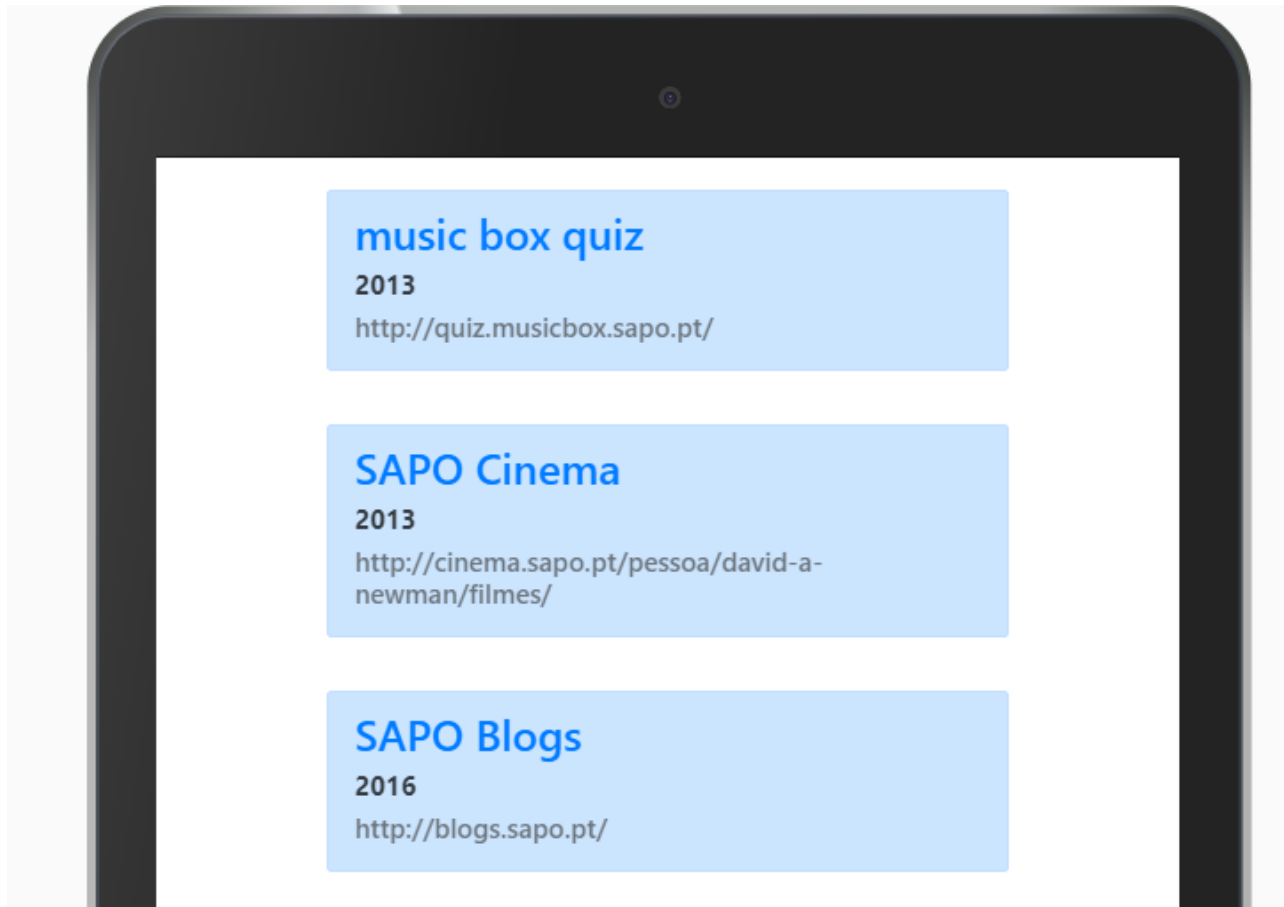
Handler function

```
<div id="content"></div>
<script>
function handler(response) {
  for (var i = 0; i < response.response_items.length; i++) {
    var item = response.response_items[i];
    document.getElementById("content").innerHTML += "<br/>"
    + "<div class=\"col offset-sm-2 col-sm-8 alert alert-primary\" style=\"word-wrap:break-word\">"
    + "<a href=\"" + item.linkToArchive + "><h4>" + item.title + "</h4></a>"
    + "<strong><h5 class=\"text-dark\">" + item.tstamp.substr(0,4) + "</strong></h5>"
    + "<h5 class=\"text-muted\"><a>" + item.originalURL + "</a></h5>";
    + "</div><br/>" ;
  }
}
</script>
```

Call to Arquivo.pt API

```
<script src="http://arquivo.pt/textsearch?q=Barack%20Obama&maxItems=5&itemsPerSite=1  
&callback=handler">  
</script>
```

Arquivo.pt application example



Arquivo.pt API –challenge

Update the previous example

1. Show 15 results instead of only 5
2. List **500 versions** of the URL [nytimes.com](https://www.nytimes.com) between the years of **2010** and **2011**

1. Show 15 results instead of only 5

```
<script src = "arquivo.pt/textsearch?q=Barack Obama &maxItems=15&callback=handler">  
</script>
```

2. List 500 versions of the URL nytimes.com between the years of 2010 and 2011

```
<script src="http://arquivo.pt/textsearch?versionHistory=nytimes.com  
&from=2010&to=2011&maxItems=500&callback=handler"></script>
```

Memento TimeTravel API



Memento TimeTravel 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.

These prior versions of web pages are named **Mementos**. **Mementos** can be found in **web archives** or in systems that support versioning such as **wikis** and **revision control systems**.”

Memento TimeTravel API



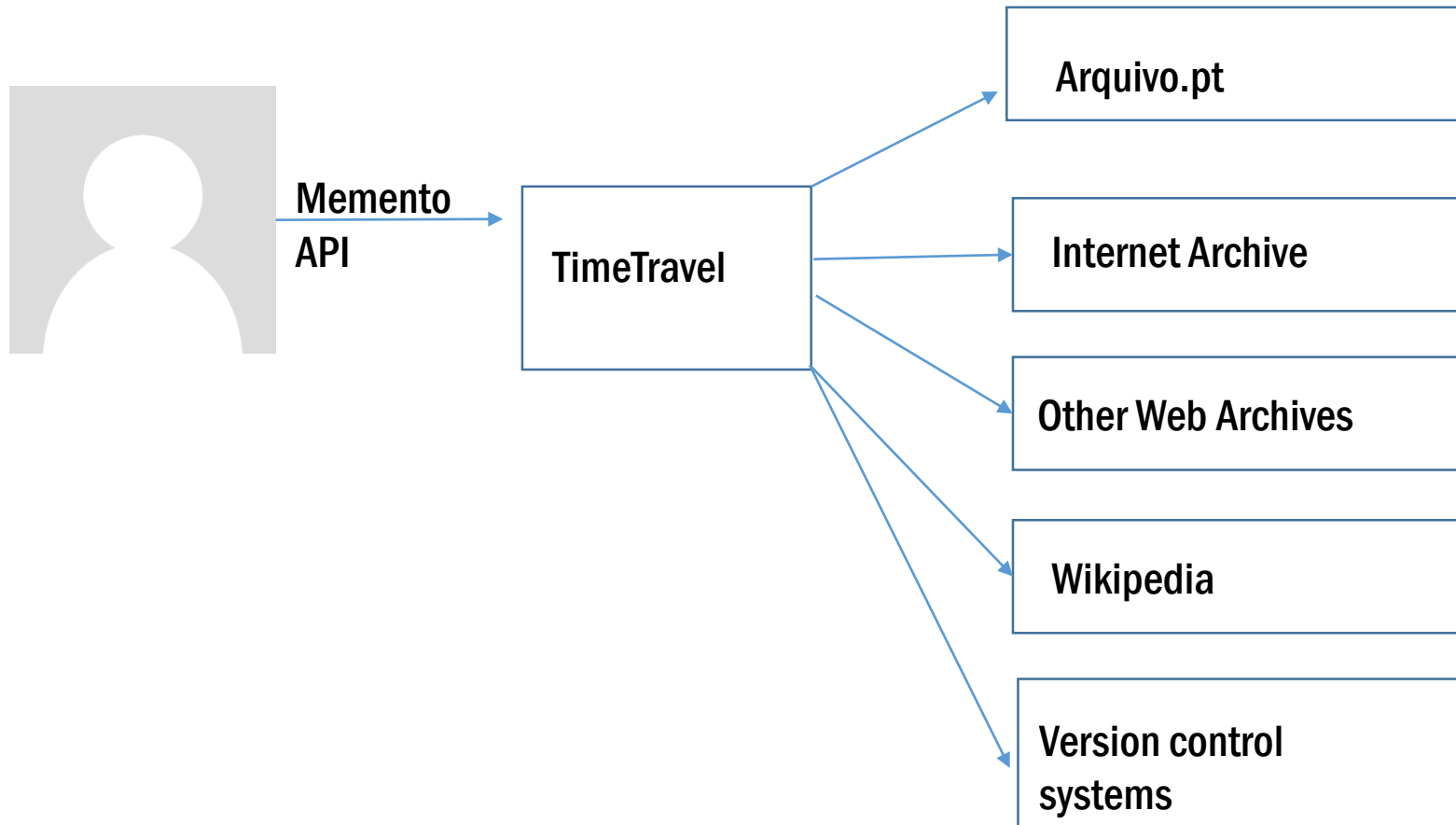
<http://timetravel.mementoweb.org/guide/api/>

Memento interoperability protocol.

<https://mementoweb.org/guide/rfc/>

<https://tools.ietf.org/html/rfc7089>

Memento TimeTravel API



Memento TimeTravel JSON API



[http://timetravel.mementoweb.org/api/json/
20140429175654/http://nytimes.com](http://timetravel.mementoweb.org/api/json/20140429175654/http://nytimes.com)

Search for **mementos** of the URL

<http://nytimes.com>

Near the date:

2014 April 29 at 17:56:54

```
"mementos":{
  "next":{
    "datetime":"2014-04-29T22:33:16Z",
    "uri":[
      "https://swap.stanford.edu/20140429223316/http://www.nytimes.com/"
    ]
  },
  "last":{
    "datetime":"2018-05-30T14:00:33Z",
    "uri":[
      "http://webarchive.proni.gov.uk/20180530140033/https://nytimes.com/"
    ]
  },
  "prev":{
    "datetime":"2014-04-25T09:59:17Z",
    "uri":[
      "https://swap.stanford.edu/20140425095917/https://www.nytimes.com/"
    ]
  },
  "first":{
    "datetime":"1996-11-12T18:15:13Z",
    "uri":[
      "https://web.archive.org/web/19961112181513/http://www.nytimes.com:80/",
      "http://wayback.archive-it.org/all/19961112181513/http://www.nytimes.com/",
      "http://web.archive.bibalex.org:80/web/19961112181513/http://www.nytimes.com/"
    ]
  },
  "closest":{
    "datetime":"2014-04-29T17:56:54Z",
    "uri":[
      "http://arquivo.pt/wayback/20140429175654/http://www.nytimes.com/"
    ]
  }
}
```


Memento TimeTravel: **closest** memento

```
"closest":{  
  "datetime":"2014-04-29T17:56:54Z",  
  "uri":[  
    "http://arquivo.pt/wayback/20140429175654/http://www.nytimes.com/"  
  ]  
}
```

Memento TimeTravel JSON API



[http://timetravel.mementoweb.org/timemap/json/
http://nytimes.com](http://timetravel.mementoweb.org/timemap/json/http://nytimes.com)

Search for **timemaps** for the URL

<http://nytimes.com>

Memento TimeTravel JSON API

timemap response list



```
{
  "original_uri": "http://europa.eu",
  "timegate_uri": "http://timetravel.mementoweb.org/timegate/http://europa.eu",
  "timemap_index": [
    {
      "uri": "https://web.archive.org/web/timemap/link/http://europa.eu",
      "memento_compliant": "yes",
      "archive_id": "ia"
    },
    {
      "uri": "http://webarchive.proni.gov.uk/timemap/http://europa.eu",
      "memento_compliant": "yes",
      "archive_id": "pron"
    },
    {
      "uri": "http://web.archive.bibalex.org/web/timemap/link/http://europa.eu",
      "memento_compliant": "yes",
      "archive_id": "ba"
    },
    {
      "uri": "http://timetravel.mementoweb.org/webcite/timemap/link/http://europa.eu",
      "memento_compliant": "no",
      "archive_id": "webcite"
    },
  ],
}
```

Memento TimeTravel JSON API

timemap response



<https://web.archive.org/web/19971008182708/http://www.sapo.pt:80/>;

rel="first memento";

datetime="Wed, 08 Oct 1997 18:27:08 GMT",

<<https://web.archive.org/web/19971210144509/http://www.sapo.pt:80/>>;

rel="memento";

datetime="Wed, 10 Dec 1997 14:45:09 GMT",

Image Search API

 [Advanced image search](#)

between: and:

[Web](#) [Images](#) Remove explicit 



caras.sapo.pt/fam...
17 October, 2012



www.sabado.pt/Mul...
9 November, 2012



caras.sapo.pt/fam...
3 October, 2012



www.gluvox.com/ta...
23 June, 2009



www.meiosepublici...
26 October, 2012



caras.sapo.pt/fam...
17 October, 2012



caras.sapo.pt/fam...
17 October, 2012



jornalpositivo.pt/
16 October, 2012



sol.sapo.pt:80/bl...
23 October, 2008



caras.sapo.pt/fam...
16 October, 2012



caras.sapo.pt/new...
12 November, 2012



www.asbeiras.pt/a...
12 November, 2012



caras.sapo.pt/fam...
24 November, 2012



www.sapo.pt:80/
5 November, 2008



caras.sapo.pt/fam...
17 November, 2012

Fernando Melo <fernando.melo@fccn.pt>