



ARCHIVE.PT

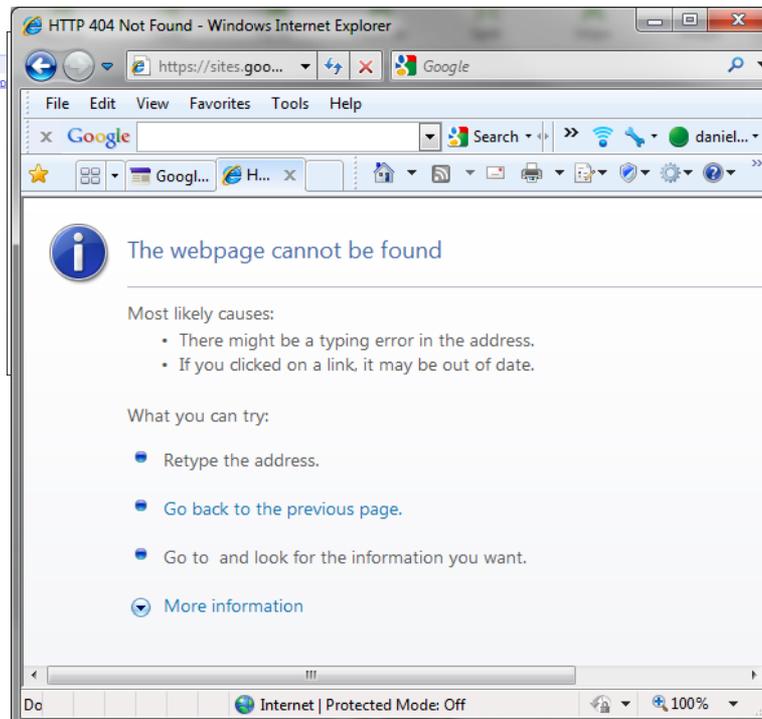
# A Memory Infrastructure

By Daniel Gomes

Daniel Gomes  
[daniel.gomes@fccn.pt](mailto:daniel.gomes@fccn.pt)



# However, all these valuable information quickly disappears





ARCHIVE.PT

80%

Disappear or change within  
1 year

The right to be  
**forgotten** is  
naturally granted  
on the Web



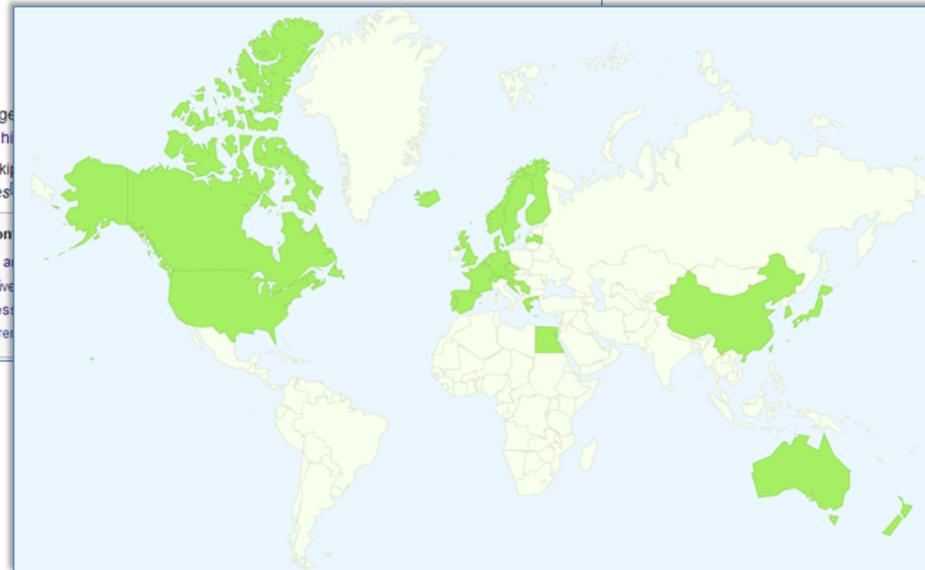
The right to be  
**remembered** on  
the Web is a  
challenge



# 77 web archiving initiatives across the world to preserve Humanity's online memory



The screenshot shows a Wikipedia article page. On the left is the Wikipedia navigation sidebar with the logo and various links. The main content area has a title 'List of Web archiving initiatives' and a sub-header 'From Wikipedia, the free encyclopedia'. Below the title, there is a table with a 'Contents' column and a list of items: '1 Web a', '2 Archi', '3 Acces', and '4 Refer'.





Most web archives  
are part  
of the IIPC





PORTUGUESE  
WEB ARCHIVE

# The Portuguese Web Archive project started in 2008

## Portuguese Web Archive

### Welcome to the Tomba project: the Portuguese web archive

Publishing tools, such as Blogger, enabled people with limited technical skills to become web publishers. Never before in the history of mankind so much information was published. However, it was never so ephemeral. Web documents such as news, blogs or discussion forums are valuable descriptions of our times, but most of them will not last longer than one year.

If we do not archive the current web contents, the future generations could witness an information gap in our days.

The [Internet Archive](#) collects and stores contents from the world-wide web. However, it is difficult for a single organization to archive the web exhaustively while satisfying all needs, because the web is permanently changing and many contents disappear before they can be archived.

As a result, several countries are creating their own national archives to ensure the preservation of contents of historical relevance to their cultures.

Portugal is now beginning its national web archiving initiative with the Tomba project at [FCCN](#) (National Foundation for Scientific Computing).

#### Contents

1. [Welcome to the Tomba project: the Portuguese web archive](#)



It was announced  
in 2012

A screenshot of the Portuguese Web Archive website. The page features a search bar with the text "Search the Archive" and a link to "Advanced search". Below the search bar, there is a section titled "Search and access pages of the past" with a recycling icon, stating "See or revisit pages that have already disappeared." and "There are more than 1000 million contents archived between 1996 and 2011 at your disposal." with a link to "Know the project". The main content area is titled "Some important events recorded in our archive" and lists four items: José Saramago (1998 Nobel Prize in Literature), Expo '98 (1998 Lisbon World Exposition), UEFA Euro 2004 (2004 UEFA European Football Championship), and Luís Figo (Soccer player Figo nominated FIFA World Player of the Year). The page also includes a language selector (Português, English) and a Help link in the top right corner, and an "experimental" banner in the top left corner.

experimental

Português English Help

 PORTUGUESE  
WEB ARCHIVE

[Advanced search](#)

 **Search and access pages of the past**  
See or revisit pages that have already disappeared.  
There are more than 1000 million contents archived between [1996](#) and [2011](#) at your disposal.  
[Know the project](#)

Some important events recorded in our archive  
Browse through events that made history

 **José Saramago**  
1998 Nobel Prize in Literature  
Personal webpage in 2008

 **Expo '98**  
1998 Lisbon World Exposition  
Post-event webpage

 **UEFA Euro 2004**  
2004 UEFA European Football Championship  
Web blog

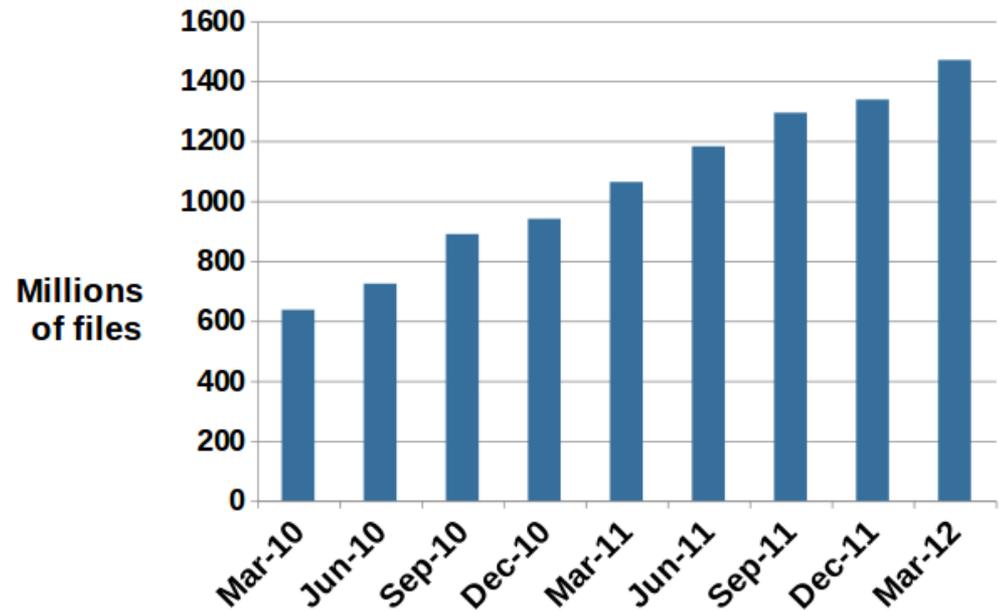
 **Luís Figo**  
Soccer player Figo nominated FIFA World Player of the Year  
News article in the sport newspaper "O Jogo" in 2001

[See all the highlights](#)

# Archive.pt in a nutshell

- A free preservation service provided to web authors
  - ◆ Only preserves publicly available information
  - ◆ Mainly related to the Portuguese community
  - ◆ Respects imposed exclusions ([robotstxt.org](http://robotstxt.org))
  - ◆ Opt-out option for authors
  
- A past-web search engine available to web users
  - ◆ A complement for live-web search engines
  - ◆ A “Google” for the past!

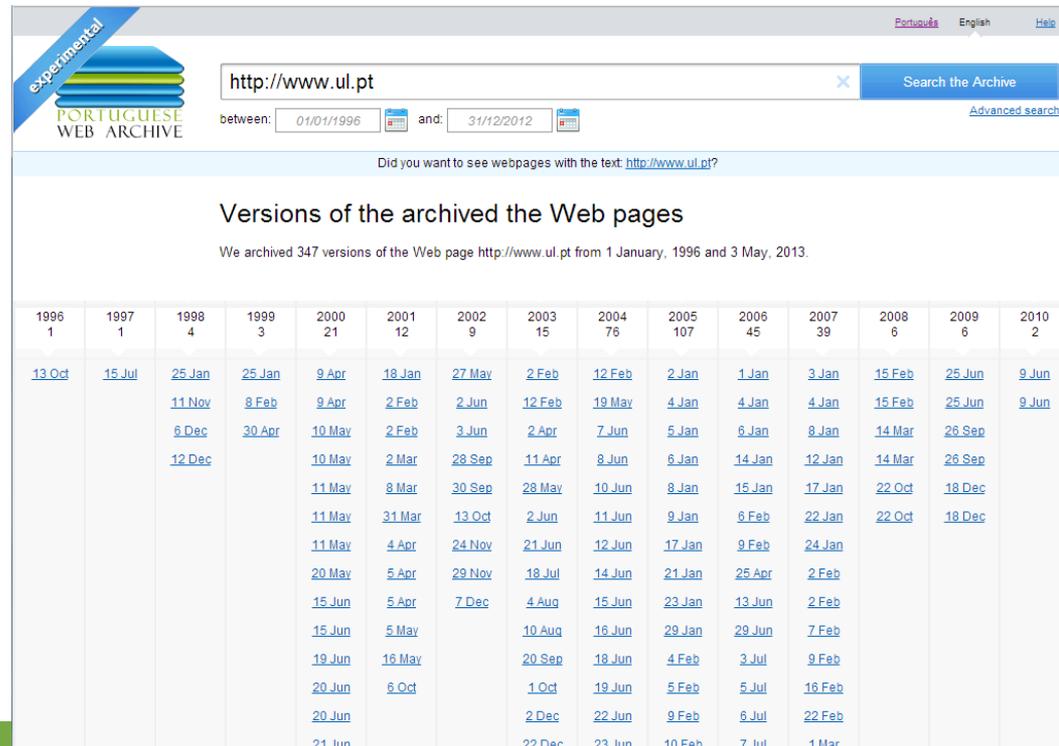
1 700 million archived files  
since 1996 (57 TB)  
...and growing



How can **anyone**  
find web archived  
information?



# Using URL search like the Internet Archive Wayback Machine



experimental

PORTUGUESE WEB ARCHIVE

Português English Help

http://www.ul.pt

Search the Archive

Advanced search

between: 01/01/1996 and: 31/12/2012

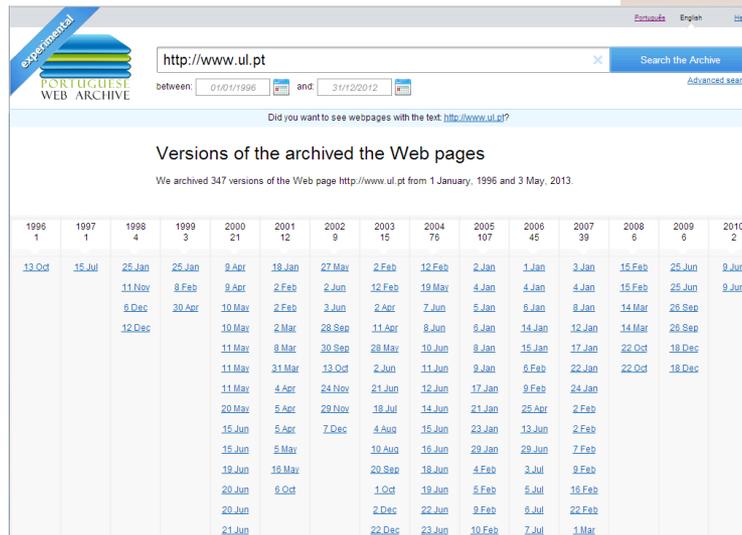
Did you want to see webpages with the text: <http://www.ul.pt?>

Versions of the archived the Web pages

We archived 347 versions of the Web page <http://www.ul.pt> from 1 January, 1996 and 3 May, 2013.

| 1996                   | 1997                   | 1998                   | 1999                   | 2000                   | 2001                   | 2002                   | 2003                   | 2004                   | 2005                   | 2006                   | 2007                   | 2008                   | 2009                   | 2010                  |
|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|-----------------------|
| 1                      | 1                      | 4                      | 3                      | 21                     | 12                     | 9                      | 15                     | 76                     | 107                    | 45                     | 39                     | 6                      | 6                      | 2                     |
| <a href="#">13 Oct</a> | <a href="#">15 Jul</a> | <a href="#">25 Jan</a> | <a href="#">25 Jan</a> | <a href="#">9 Apr</a>  | <a href="#">18 Jan</a> | <a href="#">27 May</a> | <a href="#">2 Feb</a>  | <a href="#">12 Feb</a> | <a href="#">2 Jan</a>  | <a href="#">1 Jan</a>  | <a href="#">3 Jan</a>  | <a href="#">15 Feb</a> | <a href="#">25 Jun</a> | <a href="#">9 Jun</a> |
|                        |                        | <a href="#">11 Nov</a> | <a href="#">8 Feb</a>  | <a href="#">9 Apr</a>  | <a href="#">2 Feb</a>  | <a href="#">2 Jun</a>  | <a href="#">12 Feb</a> | <a href="#">19 May</a> | <a href="#">4 Jan</a>  | <a href="#">4 Jan</a>  | <a href="#">4 Jan</a>  | <a href="#">15 Feb</a> | <a href="#">25 Jun</a> | <a href="#">9 Jun</a> |
|                        |                        | <a href="#">6 Dec</a>  | <a href="#">30 Apr</a> | <a href="#">10 May</a> | <a href="#">2 Feb</a>  | <a href="#">3 Jun</a>  | <a href="#">2 Apr</a>  | <a href="#">7 Jun</a>  | <a href="#">5 Jan</a>  | <a href="#">6 Jan</a>  | <a href="#">8 Jan</a>  | <a href="#">14 Mar</a> | <a href="#">26 Sep</a> |                       |
|                        |                        | <a href="#">12 Dec</a> |                        | <a href="#">10 May</a> | <a href="#">2 Mar</a>  | <a href="#">28 Sep</a> | <a href="#">11 Apr</a> | <a href="#">8 Jun</a>  | <a href="#">6 Jan</a>  | <a href="#">14 Jan</a> | <a href="#">12 Jan</a> | <a href="#">14 Mar</a> | <a href="#">26 Sep</a> |                       |
|                        |                        |                        |                        | <a href="#">11 May</a> | <a href="#">8 Mar</a>  | <a href="#">30 Sep</a> | <a href="#">28 May</a> | <a href="#">10 Jun</a> | <a href="#">8 Jan</a>  | <a href="#">15 Jan</a> | <a href="#">17 Jan</a> | <a href="#">22 Oct</a> | <a href="#">18 Dec</a> |                       |
|                        |                        |                        |                        | <a href="#">11 May</a> | <a href="#">31 Mar</a> | <a href="#">13 Oct</a> | <a href="#">2 Jun</a>  | <a href="#">11 Jun</a> | <a href="#">9 Jan</a>  | <a href="#">6 Feb</a>  | <a href="#">22 Jan</a> | <a href="#">22 Oct</a> | <a href="#">18 Dec</a> |                       |
|                        |                        |                        |                        | <a href="#">11 May</a> | <a href="#">4 Apr</a>  | <a href="#">24 Nov</a> | <a href="#">21 Jun</a> | <a href="#">12 Jun</a> | <a href="#">17 Jan</a> | <a href="#">9 Feb</a>  | <a href="#">24 Jan</a> |                        |                        |                       |
|                        |                        |                        |                        | <a href="#">20 May</a> | <a href="#">5 Apr</a>  | <a href="#">29 Nov</a> | <a href="#">18 Jul</a> | <a href="#">14 Jun</a> | <a href="#">21 Jan</a> | <a href="#">25 Apr</a> | <a href="#">2 Feb</a>  |                        |                        |                       |
|                        |                        |                        |                        | <a href="#">15 Jun</a> | <a href="#">5 Apr</a>  | <a href="#">7 Dec</a>  | <a href="#">4 Aug</a>  | <a href="#">15 Jun</a> | <a href="#">23 Jan</a> | <a href="#">13 Jun</a> | <a href="#">2 Feb</a>  |                        |                        |                       |
|                        |                        |                        |                        | <a href="#">15 Jun</a> | <a href="#">5 May</a>  |                        | <a href="#">10 Aug</a> | <a href="#">16 Jun</a> | <a href="#">29 Jan</a> | <a href="#">29 Jun</a> | <a href="#">7 Feb</a>  |                        |                        |                       |
|                        |                        |                        |                        | <a href="#">19 Jun</a> | <a href="#">16 May</a> |                        | <a href="#">20 Sep</a> | <a href="#">18 Jun</a> | <a href="#">4 Feb</a>  | <a href="#">3 Jul</a>  | <a href="#">9 Feb</a>  |                        |                        |                       |
|                        |                        |                        |                        | <a href="#">20 Jun</a> | <a href="#">6 Oct</a>  |                        | <a href="#">1 Oct</a>  | <a href="#">19 Jun</a> | <a href="#">5 Feb</a>  | <a href="#">5 Jul</a>  | <a href="#">16 Feb</a> |                        |                        |                       |
|                        |                        |                        |                        | <a href="#">20 Jun</a> |                        |                        | <a href="#">2 Dec</a>  | <a href="#">22 Jun</a> | <a href="#">9 Feb</a>  | <a href="#">6 Jul</a>  | <a href="#">22 Feb</a> |                        |                        |                       |
|                        |                        |                        |                        | <a href="#">21 Jun</a> |                        |                        | <a href="#">22 Dec</a> | <a href="#">23 Jun</a> | <a href="#">10 Feb</a> | <a href="#">7 Jul</a>  | <a href="#">1 Mar</a>  |                        |                        |                       |

# Problem with URL search



experimental

PORTUGUESE WEB ARCHIVE

between 01/01/1996 and 31/12/2012

Search the Archive

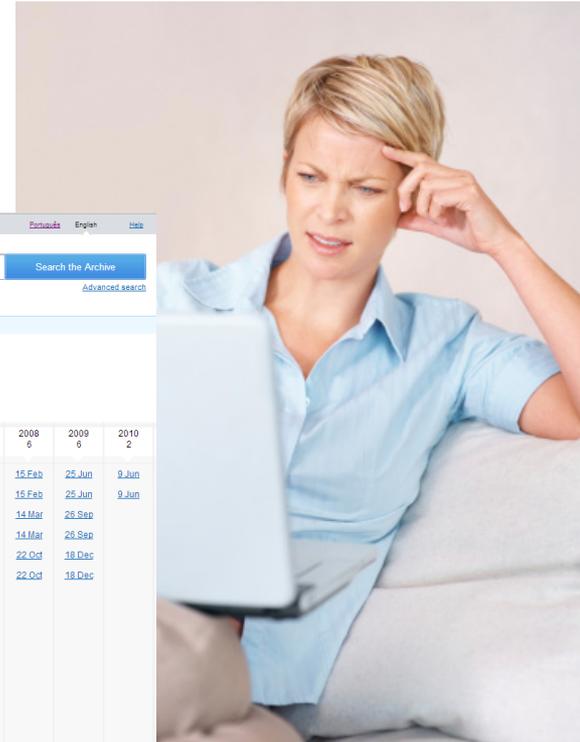
Advanced search

Did you want to see webpages with the text: <http://www.ul.pt>?

Versions of the archived the Web pages

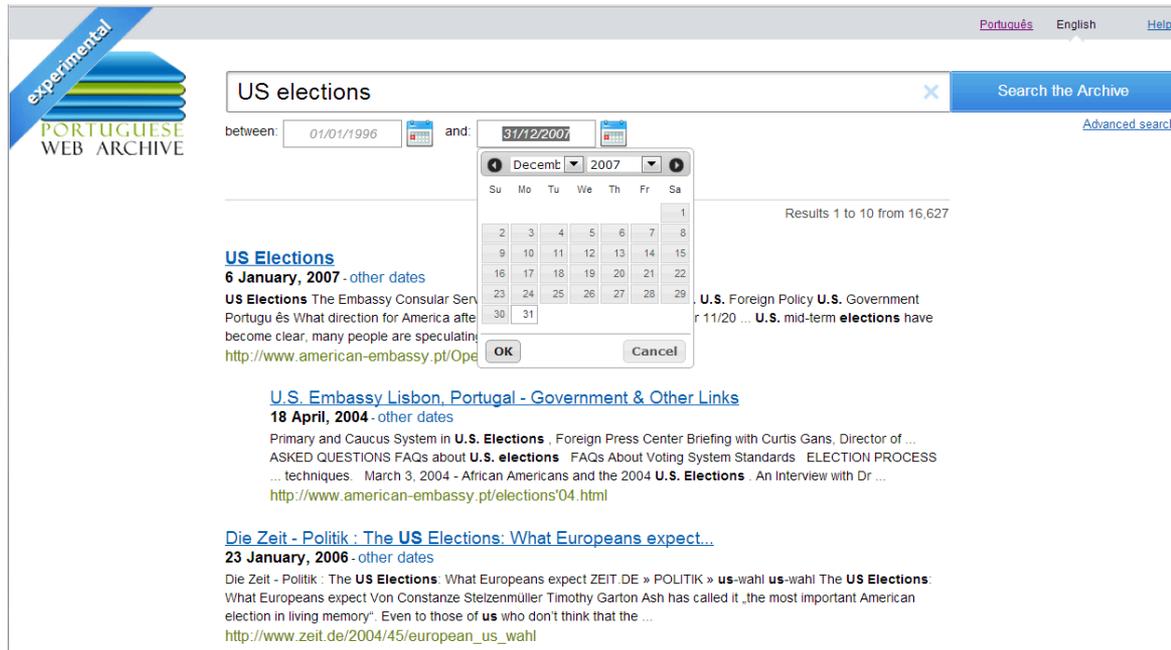
We archived 347 versions of the Web page <http://www.ul.pt> from 1 January, 1996 and 3 May, 2013.

| 1996   | 1997   | 1998   | 1999   | 2000   | 2001   | 2002   | 2003   | 2004   | 2005   | 2006   | 2007   | 2008   | 2009   | 2010  |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|-------|
| 13 Oct | 15 Jul | 25 Jan | 25 Jan | 9 Apr  | 18 Jan | 27 Mar | 2 Feb  | 12 Feb | 2 Jan  | 1 Jan  | 3 Jan  | 15 Feb | 25 Jun | 9 Jun |
|        |        | 11 Nov | 8 Feb  | 9 Apr  | 2 Feb  | 2 Jun  | 12 Feb | 19 Mar | 4 Jan  | 4 Jan  | 4 Jan  | 15 Feb | 25 Jun | 9 Jun |
|        |        | 8 Dec  | 30 Apr | 10 Mar | 2 Feb  | 3 Jun  | 2 Apr  | 7 Jun  | 5 Jan  | 6 Jan  | 8 Jan  | 14 Mar | 26 Sep |       |
|        |        | 12 Dec |        | 10 Mar | 2 Mar  | 28 Sep | 11 Apr | 8 Jun  | 6 Jan  | 14 Jan | 12 Jan | 14 Mar | 26 Sep |       |
|        |        |        |        | 11 Mar | 8 Mar  | 30 Sep | 28 Mar | 19 Jun | 8 Jan  | 15 Jan | 17 Jan | 22 Oct | 18 Dec |       |
|        |        |        |        | 11 Mar | 31 Mar | 13 Oct | 2 Jun  | 11 Jun | 9 Jan  | 5 Feb  | 22 Jan | 22 Oct | 18 Dec |       |
|        |        |        |        | 11 Mar | 4 Apr  | 24 Nov | 21 Jun | 12 Jun | 17 Jan | 9 Feb  | 24 Jan |        |        |       |
|        |        |        |        | 20 Mar | 5 Apr  | 29 Nov | 18 Jul | 14 Jun | 21 Jan | 25 Apr | 2 Feb  |        |        |       |
|        |        |        |        | 15 Jun | 5 Apr  | 7 Dec  | 4 Aug  | 15 Jun | 23 Jan | 13 Jun | 2 Feb  |        |        |       |
|        |        |        |        | 15 Jun | 5 May  |        | 10 Aug | 15 Jun | 29 Jan | 29 Jun | 7 Feb  |        |        |       |
|        |        |        |        | 19 Jun | 16 Mar |        | 20 Sep | 18 Jun | 4 Feb  | 3 Jul  | 9 Feb  |        |        |       |
|        |        |        |        | 20 Jun | 6 Oct  |        | 1 Oct  | 19 Jun | 5 Feb  | 5 Jul  | 16 Feb |        |        |       |
|        |        |        |        | 20 Jun |        |        | 2 Dec  | 22 Jun | 9 Feb  | 6 Jul  | 22 Feb |        |        |       |
|        |        |        |        | 21 Jun |        |        | 22 Dec | 23 Jun | 10 Feb | 7 Jul  | 1 Mar  |        |        |       |



→ Users **do not know** the URL from the past that **contained** the information that they need.

Archive.pt also provides **full-text search**  
over 1.2 billion web files  
archived  
since 1996



experimental  
PORTUGUESE  
WEB ARCHIVE

Português English Help

US elections  Search the Archive

between: 01/01/1996 and: 31/12/2007

Results 1 to 10 from 16,627

**US Elections**  
**6 January, 2007** - other dates  
US Elections The Embassy Consular Ser...  
Português What direction for America afte...  
become clear, many people are speculatin...  
<http://www.american-embassy.pt/Op...>

**U.S. Embassy Lisbon, Portugal - Government & Other Links**  
**18 April, 2004** - other dates  
Primary and Caucus System in **U.S. Elections** , Foreign Press Center Briefing with Curtis Gans, Director of ...  
ASKED QUESTIONS FAQs about **U.S. elections** FAQs About Voting System Standards ELECTION PROCESS  
... techniques. March 3, 2004 - African Americans and the 2004 **U.S. Elections** . An Interview with Dr ...  
<http://www.american-embassy.pt/elections'04.html>

**Die Zeit - Politik : The US Elections: What Europeans expect...**  
**23 January, 2006** - other dates  
Die Zeit - Politik : The **US Elections**: What Europeans expect ZEIT.DE » POLITIK » us-wahl us-wahl! The **US Elections**:  
What Europeans expect Von Constanze Steizenmüller Timothy Garton Ash has called it „the most important American  
election in living memory“. Even to those of us who don't think that the ...  
[http://www.zeit.de/2004/45/european\\_us\\_wahl](http://www.zeit.de/2004/45/european_us_wahl)



# Archive.pt is used by international users

|     | Country / Territory | Sessions | % Sessions |
|-----|---------------------|----------|------------|
| 1.  | Portugal            | 88,298   | 77.57%     |
| 2.  | Brazil              | 6,868    | 6.03%      |
| 3.  | United States       | 4,522    | 3.97%      |
| 4.  | Japan               | 3,615    | 3.18%      |
| 5.  | (not set)           | 1,609    | 1.41%      |
| 6.  | United Kingdom      | 1,316    | 1.16%      |
| 7.  | Germany             | 1,154    | 1.01%      |
| 8.  | France              | 809      | 0.71%      |
| 9.  | Spain               | 614      | 0.54%      |
| 10. | Angola              | 488      | 0.43%      |

- Archived web pages written in several languages
- Includes all Portuguese speaking domains (.AO, .MZ, .CV) except Brazil
- 22% of the Sessions were not from Portugal



# Web archives.

## For Who? For What?

- Journalist revisits past news
- Historian analyses digital documents
- Professor recovers presentation
- Manager documents project
- Web user recovers page from broken link
- ...
- All of us.

Access  
mechanisms for  
automatically  
processing web-  
archived data



# OpenSearch programming interface access to extend functionality

[My favorites](#) ▾ | [Sign in](#)

 **pwa-technologies**  
PWA preserves today's knowledge for future generations.

[Project Home](#) | [Downloads](#) | [Wiki](#) | [Issues](#) | [Source](#)

Search  Current pages ▾ for

**OpenSearch**  
*OpenSearch API.* Updated Mar 14, 2012 by [migoo...](#)

### 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 defined at [http://www.opensearch.org/Specifications/OpenSearch/1.1/Draft\\_5](http://www.opensearch.org/Specifications/OpenSearch/1.1/Draft_5).

It also follows the OpenSearch Time extension (Draft 1) namespace at [http://www.opensearch.org/Specifications/OpenSearch/Extensions/Time/1.0/Draft\\_1](http://www.opensearch.org/Specifications/OpenSearch/Extensions/Time/1.0/Draft_1) that describes how to set temporal search parameters.

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

### Query

The OpenSearch URL must contain the query and search parameters.

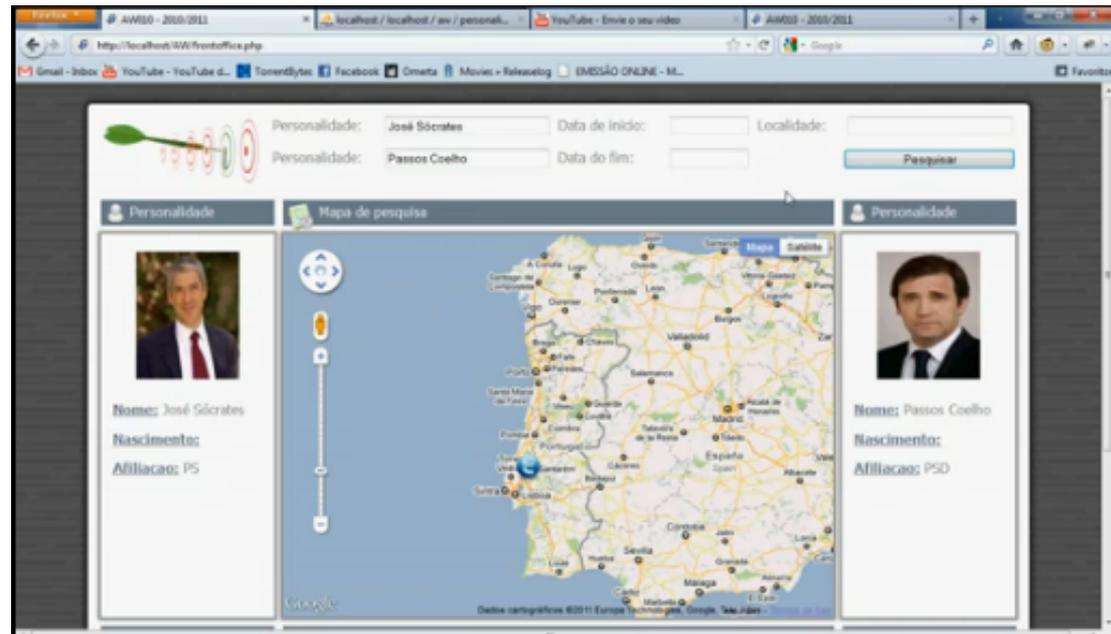
Parameters:

# OpenSearch to develop innovative services over web-archived information

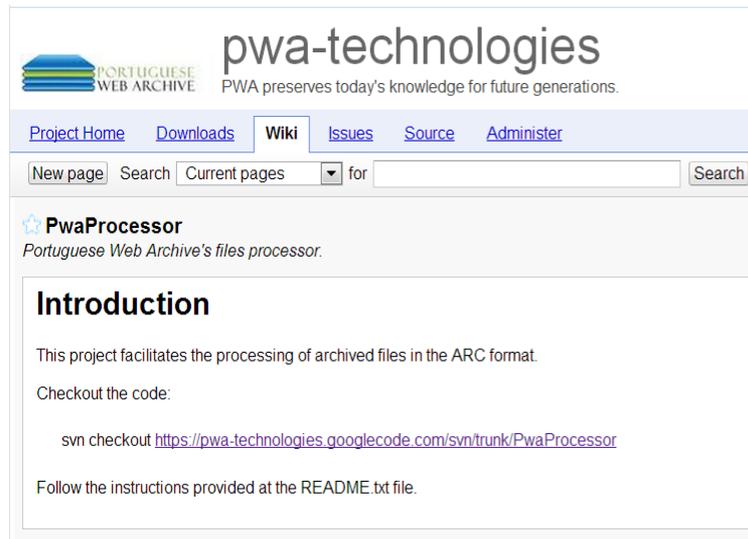
Web Applications course to develop information

aggregators about:

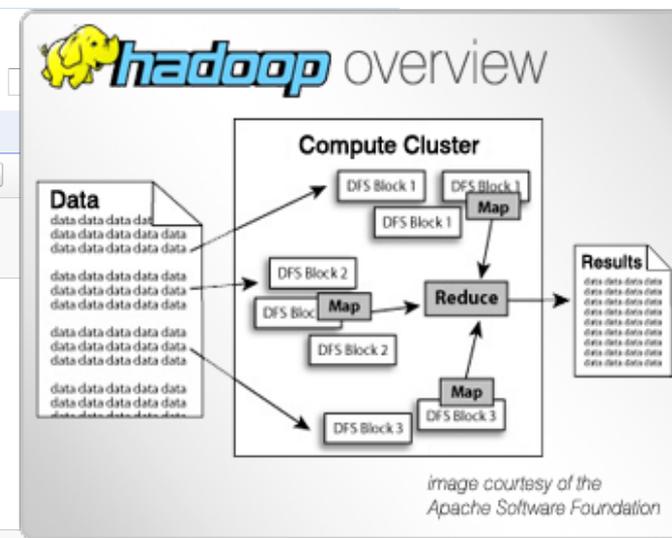
- Politicians (2011)
- Football (2012)



# Big Data analytics over web-archived data through our large-scale processing platform (Hadoop cluster)



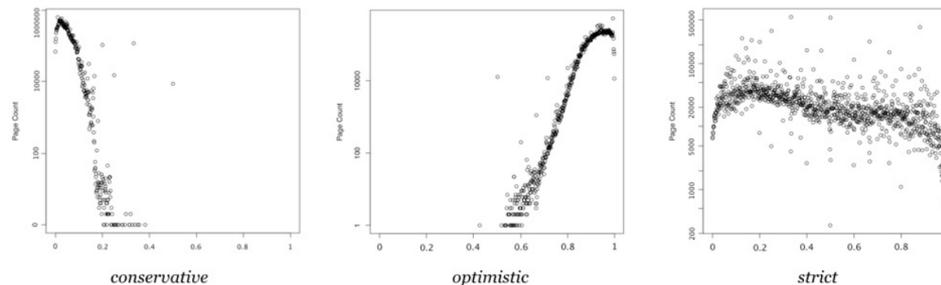
The screenshot shows the project page for PwaProcessor on the pwa-technologies website. The page includes a navigation menu with links for Project Home, Downloads, Wiki, Issues, Source, and Administer. Below the navigation is a search bar and a section for PwaProcessor, which is described as "Portuguese Web Archive's files processor". The main content area features an "Introduction" section with the text: "This project facilitates the processing of archived files in the ARC format. Checkout the code: svn checkout <https://pwa-technologies.googlecode.com/svn/trunk/PwaProcessor> Follow the instructions provided at the README.txt file."



Hadoop cluster used to automatically measure the accessibility of the Web for people with disabilities

## Results

*rates versus page count distribution*

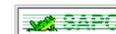
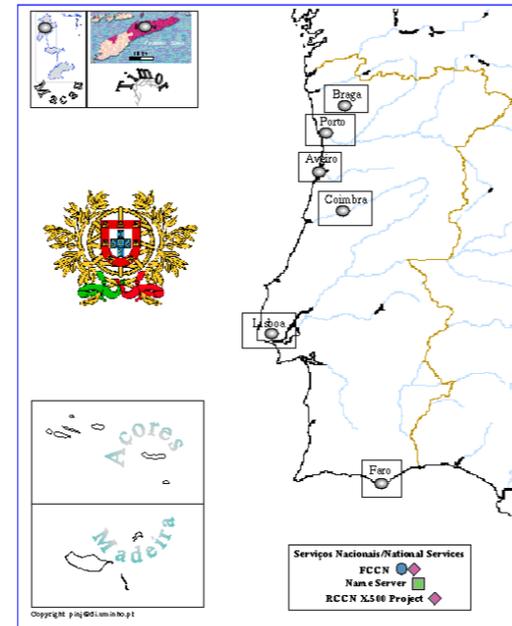


In Rui Lopes, Daniel Gomes, Luís Carrico, Web Not For All: A Large Scale Study of Web Accessibility, 2010.



# The first Portuguese web page (90's)

«Home Page» de Portugal / Portugal Home Page



[Servidor de Apontadores Portugueses](#)

|  |                                   |   |                       |
|--|-----------------------------------|---|-----------------------|
| <a href="#">Acerca de Portugal</a>           | <a href="#">Portugal Cultural</a> | <a href="#">Conferências e outros eventos</a> | <a href="#">Ajuda</a> |
| <a href="#">About Portugal</a>               | <a href="#">Cultural events</a>   |   |                       |
| <a href="#">Conferences and other events</a> | <a href="#">Help</a>              |   |                       |

 Free Speech Online  
Blue Ribbon Campaign

Se não possuir suporte gráfico, prima a [aqui](#) / If you don't have graphics support, click [here](#)



[Europe Home Page](#) [EC Home Page](#) [World Home Page](#)

# International events

2011 Egyptian  
revolution: a web-  
based revolution, a  
web-archived  
revolution



**O CASTENDO**

TERRAS DE PENALVA ONDE «A LIBERDADE É A COMPREENSÃO DA NECESSIDADE»

[» por Lenirix](#) | [home](#) | [posts recentes](#)

Domingo, 13 de Fevereiro de 2011

**O Povo do Egito derrubou Mubarak**



**mais sobre mim**



lenirix  
www.lenirix.pt

**pesquisar neste blog**

  
OK

**Março 2011**

| DOM | SEG | TER | QUA | QUI | SEX | SAB |
|-----|-----|-----|-----|-----|-----|-----|
|     | 1   | 2   | 3   | 4   | 5   | 6   |
| 8   | 7   | 8   | 9   | 10  | 11  | 12  |
| 13  | 14  | 15  | 16  | 17  | 18  | 19  |
| 20  | 21  | 22  | 23  | 24  | 25  | 26  |
| 27  | 28  | 29  | 30  | 31  |     |     |

**posts recentes**

- [Cobardia: Declaração de Gue...](#)
- [Não à intervenção militar...](#)
- [Dia de indignação e luta...](#)
- [Aves de Boto e Manuel de F...](#)
- [PSG 1-1 Benfica](#)
- [F. Garcia: Insubmisão e...](#)
- [Deuses, heróis e bússolas de...](#)
- [O amor pela verdade \(II\)](#)
- [Entradas](#)
- [O amor pela verdade \(II\)](#)
- [Pela honestidade](#)
- [O amor pela verdade \(I\)](#)
- [Benfica 1-1 Porto Moniz](#)
- [Contra a preparação de tra...](#)
- [Os primeiros voluntários...](#)

# Regional events

## Results for the 2001 Portuguese elections



# Open access Science: publications, test collections, data sets, tools. (1998-2010)

Linguateca project:  
resources for the  
computational  
processing of the  
Portuguese language

## Linguateca

[Structure](#)  
[Team](#)  
[Overview](#)  
[Access to resources](#)  
[Evaluation contests](#)  
**Catalogue of resources**

- [Corpora](#)
- [Lexica](#)
- [Encyclopedias](#)
- [Ontologies](#)
- [Services](#)
- [IR collections](#)
- [Course materials](#)
- [Texts in Portuguese](#)
- [Media](#)
- [Other](#)

[Catalogue of tools](#)  
[Catalogue of actors](#)  
[Catalogue of publications](#)  
[Interesting information](#)  
[FAQ](#)  
[Gramateca](#)

[Português](#)  
[Norsk](#)



Comments and suggestions

Please use the navigation options: on the left handside menu (clicking on the links on the right handside does not update the left handside).

## Portuguese language resources

[Linguateca](#)

Last updated 24 July 2003

This page gives you an overview of Portuguese language resources mentioned, or directly accessible, on the Internet.

The aim is to make the resources available to everyone involved in the computational processing of Portuguese, and also to facilitate communication between the various developers and users of the resources.

We take the opportunity to thank everyone who takes the effort to contribute to this site, by providing links to other sites and/or offering resources that we in turn may make accessible.

### Conventions

All the references are presented in alphabetic order, using the names and language chosen by the respective authors.

#### Symbols used in these pages

|   |   |
|---|---|
|   | Resources available in CD format                                  |
|  | Resources available in floppy disk                                |
|  | Searchable <i>online</i> resources                                |
|  | Downloadable <i>online</i> resources                              |
|  | Resources available <i>online</i> , fee required for the service. |
|  | Text integrally available   |
|  | Audio resources   |

Go to the [Portuguese page](#) to see all the links.

[Send questions, comments and suggestions](#)

# Previous governmental portals (2011)

New  
government  
erased previous  
content



The screenshot shows the official website of the Portuguese Government in 2011. The header features the Portuguese coat of arms and the text "GOVERNO DE PORTUGAL". Navigation links include "GOVERNO", "CIDADÃOS", "EMPRESAS", "Portal Móvel", "Mapa do Site", and "Contactos". A search bar is located in the top right corner.

The main content area is titled "XVIII Governo Constitucional" and includes a sub-navigation menu: "PRIMEIRO-MINISTRO", "GOVERNO", "COMUNICAÇÃO", "CONSULTAS PÚBLICAS", "NOTÍCIAS", and "PORTUGAL".

The left sidebar contains a "GOVERNO INTERACTIVO" section with links for "Subscreva a Newsletter", "RSS@gov.pt", and "Contacte o Governo". Below this is a "MULTIMÉDIA" section with a video player and a "Ver galerias" link. At the bottom of the sidebar is a "SÍTIOS EM DESTAQUE" section featuring the "Portal da Língua Portuguesa".

The main content area displays several news items:

- Primeira Loja Móvel do Cidadão** (2011-05-19): A news item about the inauguration of the first mobile citizen shop in Palmela, allowing rural populations to access services like the Citizen Card and municipal services.
- Eliminados obstáculos ao acesso a mais de 60 profissões** (2011-05-19): A news item about the Council of Ministers approving a Decree-Law to simplify access to over 60 professions, removing barriers and promoting public interest.
- Contração do PIB em linha com previsões do Governo** (2011-05-13): A news item where the Prime Minister affirmed that the decrease in Gross Domestic Product in the first quarter of the year, according to INE data, is in line with government forecasts and reflects the consequences of budgetary measures.

The right sidebar contains a "PRIMEIRO-MINISTRO" section with a link to "PM: resposta à moção de censura ao Governo, AR 10 Março 2011". Below this is a "CONSELHO DE MINISTROS" section with the coat of arms and a link to "Comunicado de 2 de Junho de 2011". At the bottom of the sidebar is a "TWITTER" section with the text "Sem conteúdos".

The official gazette of the Republic of Portugal is published exclusively online since 2006 (e.g. DRE.pt contains all the Laws)

- 4 copies **exclusively** prepared and **printed for preservation**
- Printed documents are usually digitized and put online for preservation?!

Define o Novo Regime de Publicação Exclusivamente Electrónica do *Diário da Assembleia da República* e Novas Regras para o Uso de Novas Tecnologias de Informação e Comunicação no Trabalho Parlamentar

Resolução da Assembleia da República n.º 68/2003, de 7 de Agosto

A Assembleia da República resolve, nos termos do n.º 5 do artigo 166.º da Constituição, o seguinte:

Artigo 1.º

*Diário da Assembleia da República*

1 - A partir de 15 de Setembro de 2003, a 1.ª série do *Diário da Assembleia da República* passa a ser exclusivamente disseminada em formato electrónico através do portal da Assembleia da República na Internet.

2 - A 2.ª série do *Diário da Assembleia da República* deixará igualmente de ser publicada em suporte tradicional, devendo ser adoptadas todas as medidas necessárias para que a respectiva publicação electrónica integral ocorra no mais curto prazo.

3 - A edição electrónica do *Diário da Assembleia da República* faz fé plena e a publicação dos actos através dela realizada vale para todos os efeitos legais e regimentais, devendo ser utilizado mecanismo que assinala, quando apropriado, a respectiva data e hora de colocação em leitura pública.

4 - Os serviços preparam, editam e depositam na Biblioteca da Assembleia da República e na Biblioteca Nacional quatro exemplares de uma versão impressa das duas séries do *Diário*, preparada unicamente para tal efeito.

Web archived  
DRE.pt  
since 2004



The screenshot shows the homepage of the Diário da República Eletrónico (DRE.pt). At the top, there is a green and red banner with the Portuguese coat of arms and the text "DIÁRIO DA REPÚBLICA ELETRÓNICO". Below this, the main content area features a large watermark of the coat of arms. On the left side, there is a navigation menu with sections: "Acesso Gratuito", "I Série" (containing links for "Imagens", "Sumários", "Datas de Distribuição", "Sumários por e-mail", and "Diplomas Numerados"), "Portal Jurídico", and "União Europeia". On the right side, there is a section for "Acesso por Assinatura" and a list of "III Série" items including "Legislação", "Actos da Administração Pública", "Concursos Públicos", and "Actos Societários". At the bottom, there are links for "Jornais Oficiais", "Buscas no D.R.E.", "Lingua Oficial Portuguesa", and "Instalação e Utilização".



A consulta das Bases de Dados do DRE não dispensa a consulta do D.R. original  
Linha Azul: 808 200 110 | e-mail: [dre@incm.pt](mailto:dre@incm.pt) | [Precos para 2004](#)

# Printed publications are also web archived

→ Newspaper front pages on 24 May 2009

## Em Destaque

### Capas dos Jornais de Hoje



[Diário de Notícias](#)



[Global](#)



[24 Horas](#)



[Metro](#)



[Destak](#)



[O Primeiro de Janeiro](#)



[Diário Económico](#)



[Jornal de Negócios](#)



[Jornal de notícias](#)



[Público](#)



[Correio da manhã](#)



[Jornal A Bola](#)



[Record](#)



[O Jogo](#)

Old photos from the  
pre-Digital Era that  
were published  
online (1966)



Praça do Comércio  
10 de Junho de 1966

# Historical documents with added value (meta-data)

Blog post with digitization of the republican newspaper (1915)



SEGUNDA-FEIRA, JULHO 26, 2010

A DEFESA DA REPÚBLICA - in A CORJA

ALMANAQUE REPUBLICANO

Itinerário, tábua ou fragmentos da Alma Republicana

Artur B. Mendonça José M. Martins

Vale do Mondego, Coimbra (Portugal)

A In anque Republicano

facebook

Cria o teu cartão de visita

ETIQUETAS

- 18 de Janeiro de 1934

ARQUIVO

ARQUIVO

E-Mail

e-mail

CENTENÁRIO DA REPÚBLICA 1910-2010

BLOGS

- 100 Anos República (Alcácer)
- 100 Anos República F. Foz
- 31 de Janeiro
- Abril de Novo

SAÍDA QUANTOS

A DEFESA DA REPÚBLICA

A CORJA. Semanário republicano anti-clerical independente (Liberdade, Justiça, Verdade, Progresso) - [Ano 1, nº1 (6 de Fevereiro 1915) ao nº 25 (16 de Agosto 1915)], Coimbra; **Administrador**, Anibal Reis [nº22, M. Simões]; **Secretário**, Mário de Brito [nº10, J. L. Frazão]; **Director**: José Peixoto de Alarcão [nº21, Fernandes Martins]; **Colaboradores**: A. Batista Rama, Afonso Duarte (poema), Alfredo Pimenta (poema), António Correia de Oliveira (poema), António Nobre (poema), Baldaque da Silva, Coelho Neto, Ernesto Almeida, Fernandes Martins, Fernando de Araújo, Guerra Junqueiro (poema), J. Peixoto de Alarcão, J. Pestana Júnior, João de Deus (poema), José Figueiredo Júnior [idem in, "A Revolta"], Ribeiro de Carvalho (poema); **Administração**, Rua Dr. João Jacinto, 38, Coimbra; **Redacção**, Couraça de Lisboa, 10, Coimbra;

Even  
insignificant  
individual  
information.  
Teenager blogged:  
“math test went bad”

---

## Blog de uma adolescente

[« março 2006 | Main | maio 2006 »](#)

---

abril 29, 2006

### **Fim de semana**

Isto está uma seca, tive que estudar ciências! Pois é, a minha única nega. Tou tão deprimida... o teste de matemática correu-me tão mal!! Sabia aquilo tudo e na hora de-me uma branca. Agora é estudar, estudar, estudar, estudar..... Espero que o vosso fim de semana seja bem melhor que o meu inclusive o feriado, que também ficarei em casa a estudar!!  
\*\*\*\*\*Beijocas\*\*\*\*\*

Posted by IFilipinha2000 at 06:57 PM | [Comentários: \(1\)](#)

---

abril 27, 2006

### **Testes e amigas**

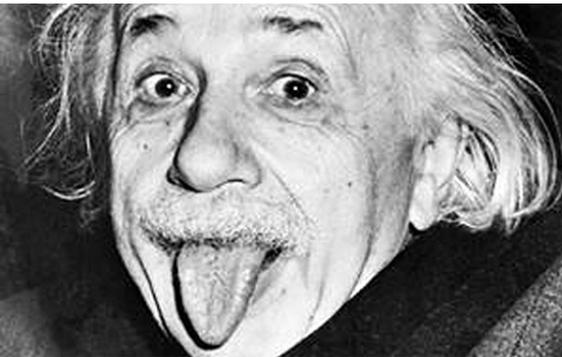
Estou furiosa!! Estudei bué para matemática e correu-me super mal o teste. Deu-me uma branca. Agora não posso tirar nenhuma nega às outras disciplinas. A minha colega quer esquecer o rapaz de que gosta, mas ela gosta tanto dele.... Sabem, por um lado acho que ela devia esquece-lo, mas ela já o tentou esquecer tanta vez que mais vale lutar por ele. Quem sabe se com ela ele não muda a sua atitude? Com os amiguinhos gozões dele é que é um bocadinho difícil, mas nada é impossível.  
\*\*\*\*\*Beijocas\*\*\*\*\*

# What if this teenager will become a...

→ Famous singer?  
(who used to be a  
"nerd")

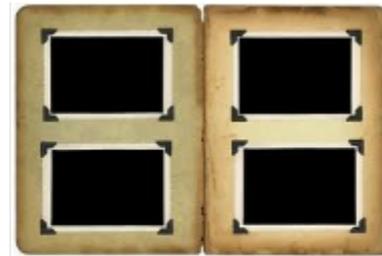
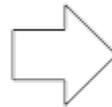
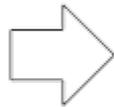


→ Physics Nobel?  
(that was not that  
good at math)



# Alert!

People are not preserving their digital data



Web archives may be the **only** source of memories to many people



Collaborate  
with *Archive.pt*



We need **your** collaboration!



For non-techies:  
suggest the archive  
of interesting sites

[arquivo.pt/sugerir](http://arquivo.pt/sugerir)

A screenshot of the Portuguese Web Archive website's "Suggest a Portuguese site to be archived" page. The page features a navigation menu with "Home", "Collaboration", "FAQ", "About", "Tools", "Dissemination", and "Search the past". A search bar is located in the top right corner. The main content area includes a heading "Suggest a Portuguese site to be archived" and instructions for users to provide the address of the site and a description. A text input field contains "http://". On the right side, there is a "News" section with three entries: "Nov 10, 2014 Portuguese Web Archive: a Memory Infrastructure @DLM2014", "Jun 27, 2014 Our new article on Web Archiving", and "Jun 27, 2014 System administrator for the Portuguese Web Archive". A "More news..." link is also present. At the bottom right, there is a "Like" button for the page.



For techies: All our software is freely available and you can improve it

pwa-technologies at [code.google.com](https://code.google.com)

A screenshot of the project page for "pwa-technologies" on code.google.com. The page header includes the "PORTUGUESE WEB ARCHIVE" logo and the text "PWA preserves today's knowledge for future generations." Below the header is a navigation menu with links for "Project Home", "Downloads", "Wiki", "Issues", "Source", and "Administer". The main content area is divided into two columns. The left column contains "Project Information" (with a "Recommend this on Google" button), "Starred by 3 users" (with a "Project feeds" link), "Code license" (with a "GNU Lesser GPL" link), "Labels" (with links for "Web", "Archive", "Service", and "WebArchive"), and "Members" (with a list of names and email addresses). The right column contains a description of the project's main goal, a detailed paragraph about the development of the PWA IR system, and a paragraph about the PWA search engine's public service and research platform capabilities. A "Main features" link is located at the bottom of the right column.

# For web authors: follow recommendations to publish preservable sites

## Recommendations for authors to enable web archiving

This page presents recommendations to create web pages that can be efficiently archived and accessed across time. It is impossible to preserve the web without the authors collaboration.

*Last update on September 16, 2010.*

A web archive works similarly to a web search engine like Google. Therefore, the presented recommendations contribute to create contents that can be correctly processed of both these types of systems.

The following recommendations were classified in *fundamental* and *advisable*. They were written with the purpose of helping any author. [Please contact us if you have any doubt.](#)

### Web site organization

- [One link for each content](#) (fundamental)
- [Crawler-friendly homepage](#) (fundamental)
- [Maintain address to the same content across time](#) (advisable)
- [Robots Exclusion Protocol to indicate access restrictions](#) (advisable)

[arquivo.pt/recommendations](http://arquivo.pt/recommendations)



For everybody:  
disseminate  
Archive.pt

[arquivo.pt/news](http://arquivo.pt/news)



# Final message

Web archives are crucial  
infrastructures for  
Humanity's Memory





ARCHIVE.PT



Thank you!

[daniel.gomes@fccn.pt](mailto:daniel.gomes@fccn.pt)