

Arquivo da Web Portuguesa

Uma infraestrutura de suporte
à investigação científica

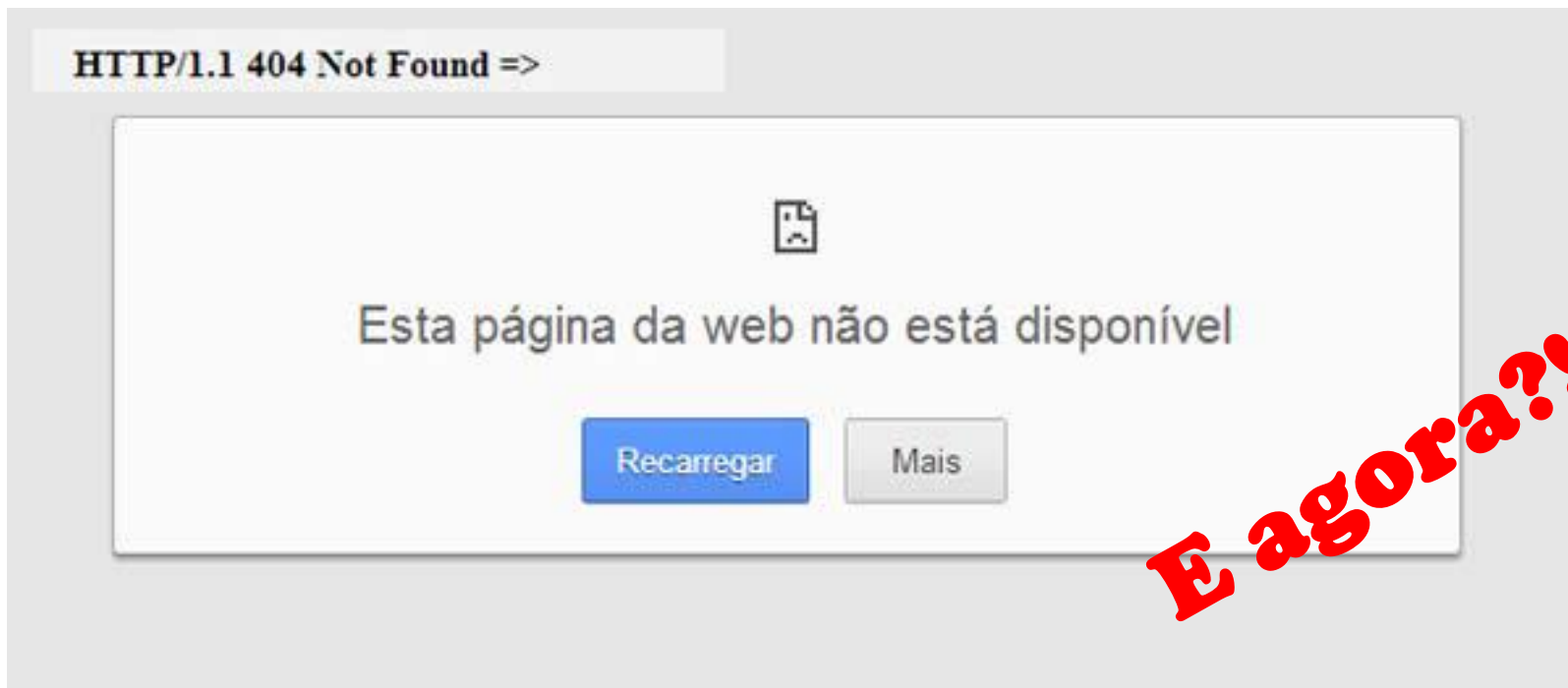
Hugo Viana
hugo.viana@fccn.pt

DataStorm Workshop'2015

O esforço de preservar o que nasceu **Efêmero**

- 50 dias - 50% dos documentos mudam
(Cho and Garcia-Molina. 2000);
- 1 ano - 80% tornam-se inacessíveis
(Ntoulas, Cho and Olson. 2004);
- 27 meses - 13% das referências web desapareceram
(<http://webcitation.org/>. 2007);

Upsss... 404, outra vez!




Um solução válida


← → ↻ Press to search Arquivo.pt(pt)

arquivo.pt - Arquivo da Web Portuguesa: pesquisa sobre o passado

arquivo.pt/searchTests.jsp - Internet Archive: Pesquisa do Arquivo da Web Portuguesa




ARQUIVO DA WEB
PORTUGUESA


 Pesquise e aceda a páginas do passado
Veja ou reveja páginas que já desapareceram.
São mais de 1200 milhões de conteúdos arquivados entre [1996](#) e [2012](#)
que estão ao seu dispor.
[Conheça o projeto](#)

[Pesquisas avançadas](#)


Alguns momentos relevantes registados no arquivo
Navegue por momentos que ficaram para a história.




[José Saramago](#)
Prémio da Literatura 1998
Página pessoal em 2008



[Expo '98](#)
Exposição Mundial de 1998 em Lisboa
Página pós-evento



[UEFA Euro 2004](#)
Campeonato Europeu de Futebol 2004 em Portugal
Web blog



[Luís Figo](#)
Figo eleito melhor jogador do mundo
Notícia do jornal "O Jogo" em 2001

[Ver todos os destaques](#)

Ajudas

- [Ajuda](#)
- [Perguntas frequentes](#)
- [Contacto](#)

Sobre


- [Objetivos](#)
- [Publicações](#)
- [Vídeos](#)
- [Termos e Condições](#)

Colaboração


- [Sugira um sítio web](#)
- [Respeite recomendações](#)
- [Formeça conteúdos](#)
- [Divulgue o Arquivo.pt](#)
- [Projetos de I&D](#)

Notícias

- [Twitter](#)
- [Facebook](#)
- [LinkedIn](#)
- [RSS](#)




FCCN
Fundação para a Ciência e Tecnologia
Instituto de Telecomunicações



GOVERNO DE
PORTUGAL


MINISTÉRIO DA EDUCAÇÃO
E CIÊNCIA

POS - CONECTIVIDADE
Programa Nacional de Investimento em Infraestruturas



FCT
Fundação para a Ciência e Tecnologia

FCCN

 ARQUIVO.PT

Exemplo: <http://a9.com/-/opensearch/>

[search >](#)[help](#) | [preferenc](#)

- [Overview](#)
- [Open Search Home](#)
- [Frequently Asked Questions](#)
- [See Submitted Search Feeds](#)
- [Submit Your Search to A9.com](#)
- [Specification](#)
- [Specifications and developer resources for OpenSearch are now hosted at \[opensearch.org\]\(http://opensearch.org\)](#)

OpenSearch

OpenSearch is a set of **simple formats** for the **sharing of search results**. Any website that has a search feature can make their results available in OpenSearch™ format. Other tools can then read those search results. For example, here is an aggregator that brings together search results from many websites:



OpenSearch is an open format with a [Creative Commons license](#). Rather than reinventing the wheel, it uses the simple and very popular syndication formats RSS and Atom, along with a document describing the search engine.

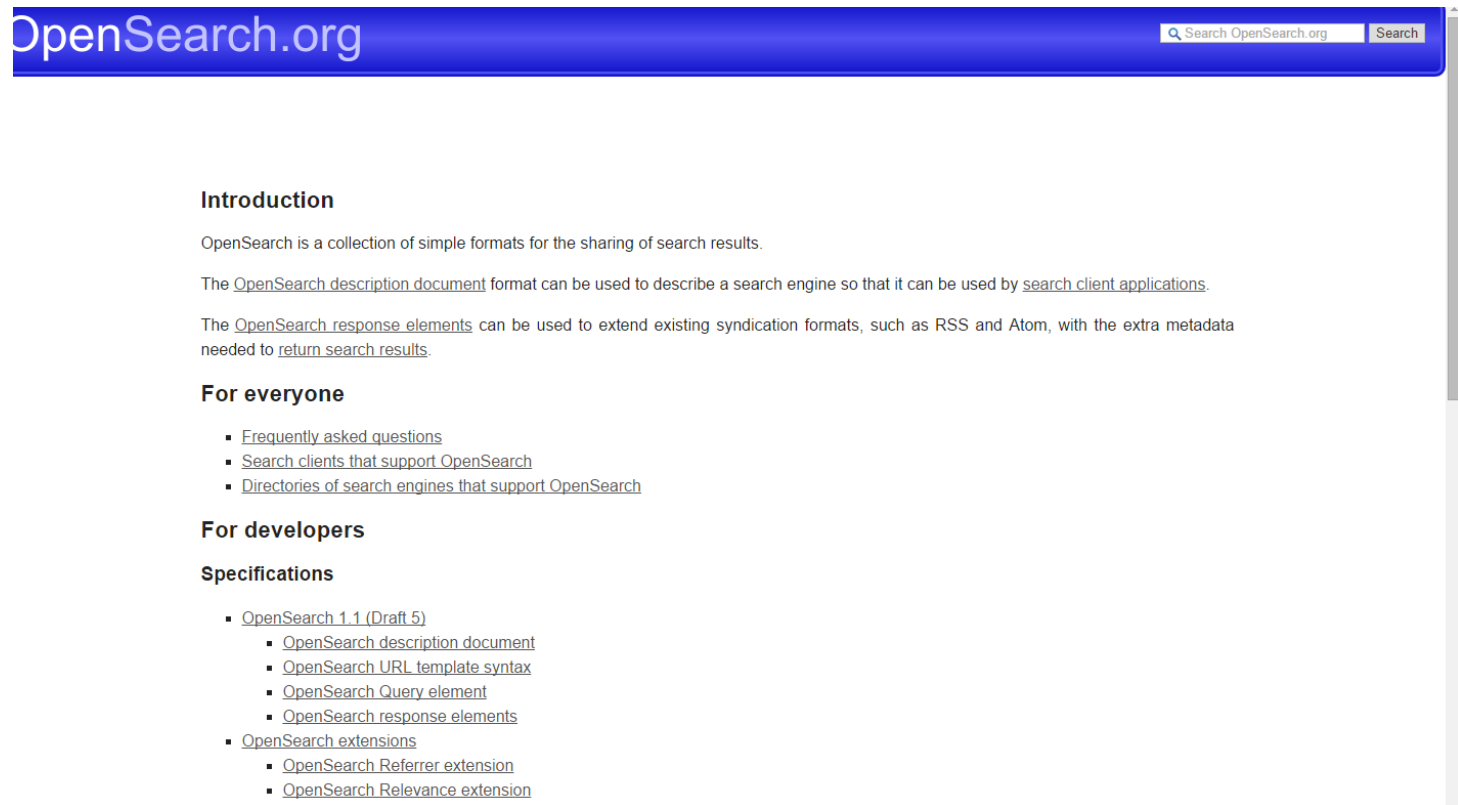
Searchers: [browse OpenSearch feeds](#)

Developers: start at the [FAQ](#), [How-to](#), [Specifications](#), or [Submit your search](#).

[about A9.com](#) | [clickriver.com](#) | [jobs](#) | [developers](#) | [privacy](#) | [conditions of use](#) | [feedback](#)

Exemplo

http://www.opensearch.org/Home



Exemplos de páginas **arquivadas**

Eventos Internacionais

Revolução Egípcia de 2011: revolução baseada na web, revolução arquivada da web



A primeira página portuguesa (90's)



PSD

(psd2011.com, 2011)

PASSOS COELHO
ALERTA PORTUGUESES PARA ESTAREM ATENTOS

PASSOS COELHO AGENDA NOTÍCIAS CANDIDATOS PROGRAMA GALERIA CONTACTE-NOS PESQUISA

Passos Coelho preparado para ser Primeiro-ministro
Durante uma hora o candidato do PSD a Primeiro-Ministro, Passos Coelho, mostrou que tem capacidade e conhecimento para enfrentar a crise agravada nos seis anos de governo de Sócrates e fazer Portugal crescer e criar empregos.
O debate revelou que Passos Coelho e o PSD têm o melhor...
[ler mais >](#)

PROGRAMA ELEITORAL
EU APOIO PASSOS COELHO
SITE INSTITUCIONAL PSD.PT
SUBSCREVA A NEWSLETTER
AGENDA
22.05 - Passos Coelho visita Leiria
22.05 - Passos Coelho visita Alcobça – Almoço

ÚLTIMAS NOTÍCIAS
PSD defende o Serviço Nacional de Saúde
Em declarações à imprensa Passos Coelho diz que está a defender o Serviço Nacional de Saúde [ler mais >](#)

ENVOLVA-SE

FOTOS VÍDEOS RÁDIO

Site Japonês


(*psd2011.com, 2011*)



Evolução das páginas ao longo dos anos

SAPO.PT 2015

ATENÇÃO Este site utiliza cookies. Ao navegar no site estará a consentir a sua utilização. [Saiba mais sobre o uso de cookies.](#)



Insira o texto a pesquisar

Q

PT

MAIL

CARROS

CASAS

FOLHETOS

VOUCHER

EMPREGO

VÍDEOS

BLOGS

RESTAURANTES

JORNAIS

LIFESTYLE

MAG

SAPO BIZ

INÍCIO


ATUALIDADE

DESPORTO


ECONOMIA

VIDA

TECNOLOGIA


 **SAPO**MAG

O seu novo site de cinema, televisão e entretenimento




CINEMA

Estes atores resistem à imagem de "meninos bonitos"




TV

Como foram escolhidos os atores de "A Guerra dos Tronos"




FILMES

Quando o cinema desceu ao tamanho de uma formiga




TV

Tem pouco tempo? Tome nota destas séries




EUROGRUPO CONCORDA COM EMPRÉSTIMO INTERCALAR À GRÉCIA

SIC



MENSALIDADES, PRAZOS E TAXAS FIXAS. FINANCIAMENTO A 100%.



TAN desde **8,90%** e TAEG desde **10,7%**.

SAPO.PT 2011

1 de Janeiro — Faça do SAPO a sua homepage

Acesso: Internet SAPO | Meo Fibra | Meo | Telepac | TMN | TDT

Área de Cliente SAPO

Web | Imagens | Vídeos **Novo!** | Notícias | Blogs | Produtos | PAi | PBi

Insira o texto a pesquisar Pesquisar

Mail | Blogs | Carros | Casas | Fotos | Mapas | Vídeos | Notícias | Messenger Todo o SAPO ▾

Qual o clube que José Mourinho passou a comandar em 2009? Entre no Jogo do Ano 2010!

Destaques | Desporto | Economia | Vida | Tecnologia | Local Vídeos

Fontes: DD | DE | Lusa | RTP | SIC | Sol



Mulheres no poder
192 países, apenas nove mulheres presidentes (SAPO)

São 192 os Estados com assento nas Nações Unidas e a maioria são repúblicas, mas o ano 2011 começa

Mulheres no poder
192 países, apenas nove mulheres presidentes (SAPO)

Transportes
Utentes criticam aumento dos preços dos transportes (Sol)

Argentina
Sismo de 6,9 na escala de Richter abalou Argentina (SIC)

Ano Novo
Balanço de atentado em Alexandria ascende a 21 mortos (SAPO)

 **Mail**  **Tempo**  **Horóscopo**  **Emprego**

personalizar página

HTTP Status 404 - /wayback /20110101160218 /about:blank

type Status report

message /wayback/20110101160218/about:blank

description The requested resource (/wayback/20110101160218/about:blank) is not available.

SAPO.PT 2003



**Multimedia**

PESQUISAR

[E-Mail GRATUITO](#) | [Bolsa](#) | [Carro](#) | [Casa](#) | [Farmácias](#) | [Hoje](#) | [Horóscopo](#) | [Salas de Cinema](#) | [Top SAPO](#) | [Postais](#) | [Fóruns](#) | [Opinião](#) | [WebChat](#) | [IRCChat](#) | [Jogos](#) | [Pesquisa Mundial](#) | [Leilões](#)

29 de Fevereiro

CANAIS

[CNL](#)
[Banda Larga](#)
[Carnaval](#)
[Desporto](#)
[Dia do Pai](#)
[Educação](#)
[Economia](#)
[Emprego](#)
[Especial Óscares](#)
[Família](#)
[Hoje](#)
[Horóscopo](#)
[Jogos](#)
[Jornais do Mundo](#)
[Lazer](#)
[Leilões](#)
[Saúde e Medicina](#)
[Milénio](#)
[Namorados](#)
[Netcupões](#)
[Notícias](#)

>> DESTAQUES DO DIA

NOTÍCIAS	ÚLTIMA HORA!	DESTAQUES
<ul style="list-style-type: none">▶ O famoso jogo Metal Slug vai ter mais uma sequência▶ Moçambique: Comunidade portuguesa no país a salvo	<ul style="list-style-type: none">▶ PSD: Helena S. Osório comenta o Congresso.▶ Envie já um postal ao pai.▶ Virgindade: sim ou não?▶ Ouça aqui a Rádio Voxx	<ul style="list-style-type: none">▶ Licite já a bola do Pelé.▶ Vota: Troféus Jetnet 2000.▶ Dia do Pai: Clique já!▶ Visite o Canal TeK.▶ Bolsa: Comentário-29/2 ESDdealer (8h40).

>> [Clique e ganhe uma viagem ao Carnaval do Rio!](#)

>> DIRECTÓRIO

[Ajuda](#)

Novidades Congressos , Eventos Sociais	Comércio, Indústria e Serviços Serviços , Web Design , Saúde , Lojas
Ensino e Investigação Univs. , Institutos , Escolas	Páginas Pessoais Páginas Pessoais , Lista de E.Mails
Comunicação Social Jornais , Rádios , Televisão	Sociedade e Cultura Museus , Hospitais , Governo

SAPO.PT 1997

[Informações](#)[Correio Electrónico](#)[TOP SAPO](#)[Novidades](#)[Imagens Satélite](#)[Pesquisar](#)[Opções](#)

Procura pelo E-Mail de alguém ? Já conhece a [base de E-Mails](#) do SAPO ?

- **Novidades**

[Novos Links](#), [Congressos](#), ...

- **Ensino e Investigação**

[Universidades](#), [Institutos](#), [Escolas](#), ...

- **Comunicação Social**

[Jornais](#), [Rádios](#), [Televisão](#), ...

- **Entretenimento**

[Desportos](#), [Fora de Casa](#), [Música](#), ...

- **Serviços de Informação**

[Software](#), [Mailing Lists](#), [IRC](#), ...

- **Comércio, Indústria e Serviços**

[Serviços](#), [Informática](#), [Saúde](#), [Lojas](#), ...

- **Páginas Pessoais**

[Páginas pessoais](#), [Lista de E-Mails](#)

- **Sociedade e Cultura**

[Museus](#), [Hospitais](#), [Religião](#), [Governo](#), ...

- **Regional**

[Câmaras Municipais](#), [Turismo](#), [Timor](#), ...

- **Computadores e Internet**

[Docs](#), [Web Designers](#), [Software](#), [ISPs](#), ...



SAPO.PT

Histórico de versões arquivadas



entre: e:

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

Versões da página web guardadas no arquivo

Foram gravadas no arquivo 1.930 versões da página <http://sapo.pt> entre 1 Janeiro, 1996 e 1 Julho, 2015.

2000 48	2001 51	2002 20	2003 29	2004 197	2005 415	2006 75	2007 139	2008 6	2009 7	2010 255	2011 370	2012 163	2013 125
29 Fev	5 Jan	24 Jan	5 Fev	16 Fev	1 Jan	1 Jan	2 Jan	1 Jan	20 Mai	26 Mar	1 Jan	1 Jan	10 Jan
29 Fev	6 Jan	6 Fev	10 Fev	19 Mar	2 Jan	1 Jan	5 Jan	14 Mar	24 Jun	1 Abr	2 Jan	2 Jan	11 Jan
3 Mar	7 Jan	30 Mar	19 Fev	5 Abr	3 Jan	2 Jan	7 Jan	20 Mar	26 Set	5 Abr	3 Jan	3 Jan	12 Jan
4 Mar	8 Jan	1 Abr	20 Fev	20 Mai	4 Jan	2 Jan	7 Jan	21 Out	26 Set	8 Abr	4 Jan	4 Jan	15 Jan
4 Mar	19 Jan	29 Mai	24 Mar	3 Jun	4 Jan	5 Jan	9 Jan	22 Out	26 Set	9 Abr	5 Jan	5 Jan	17 Jan
5 Mar	24 Jan	30 Mai	12 Abr	9 Jun	5 Jan	6 Jan	11 Jan	22 Out	18 Dez	12 Abr	6 Jan	6 Jan	18 Jan
7 Mar	10 Fev	4 Jun	19 Abr	9 Jun	5 Jan	10 Jan	12 Jan		18 Dez	13 Abr	7 Jan	7 Jan	19 Jan
11 Mar	24 Fev	5 Jun	22 Abr	11 Jun	6 Jan	10 Jan	14 Jan			16 Abr	8 Jan	8 Jan	22 Jan
8 Abr	26 Fev	6 Jun	24 Abr	12 Jun	7 Jan	11 Jan	16 Jan			19 Abr	9 Jan	9 Jan	23 Jan
10 Mai	1 Mar	25 Jul	26 Abr	12 Jun	7 Jan	12 Jan	16 Jan			22 Abr	10 Jan	10 Jan	26 Jan
10 Mai	2 Mar	27 Jul	2 Jun	13 Jun	9 Jan	13 Jan	17 Jan			30 Abr	11 Jan	11 Jan	29 Jan
11 Mai	5 Abr	6 Ago	3 Jun	14 Jun	10 Jan	15 Jan	19 Jan			2 Mai	12 Jan	12 Jan	31 Jan



Como pesquisar em Arquivo.pt?

Acesso Humano	Acesso Automático
Pesquisa através de texto integral	Sintaxe para referenciar páginas do <i>Arquivo.pt</i>
Pesquisa por URL	Opensearch (API)

Acesso Humano

Pesquisa por Texto Integral



Search interface for Arquivo da Web Portuguesa. The search bar contains the text "sapo". To the right of the search bar is a blue button labeled "Pesquisar no Arquivo". Below the search bar, there are two date input fields: "entre: 01/01/1996" and "e: 31/12/2014". To the right of these fields is a link labeled "Pesquisa avançada".

Resultados 1 a 10 de 270.798.576

[music box quiz](#)

8 Novembro, 2013 - outras datas

meses. ConstruÃdo pelo SAPO MÃsicas via music box Powered by last.fm OK ...

<http://quiz.musicbox.sapo.pt/>

[SAPO Cinema](#)

15 Novembro, 2013 - outras datas

SAPO Cinema Cinema Mobile Menu de navegação Pesquisa Geral Pesquisa de Cartaz Início Estreias & ... Realização Ator / Atriz Escolher Limpar filtros Partilha siga o sapo cinema nas redes Facebook SAPO Twitter SAPO Google + SAPO Video SAPO Termos de pesquisa Pesquisar Home Estreias e Cartaz Atualidade ...

<http://cinema.sapo.pt/pessoa/david-a-newman/filmes/>

[SAPO Cinema](#)

8 Novembro, 2013 - outras datas

SAPO Cinema Cinema Mobile Menu de navegação Pesquisa Geral Pesquisa de Cartaz Início Estreias & ... a SIC. Adaptado de Sapo Saber a 23-01-2011 ? comentários Listagem Mais antigo Mais recente Ordem ... Participação Realização Escolher Limpar filtros Partilha siga o sapo cinema nas redes Facebook SAPO ...

<http://cinema.sapo.pt/pessoa/diana-chaves/filmes/17365?ordem=antigos>

[SAPO - Portugal Online!](#)

8 Junho, 2010 - outras datas

SAPO - Portugal Online! Saltar para: Pesquisa [1] , Lista de Serviços [2] , Notícias [3] ou Destaques SAPO [4] SAPO.pt Pesquisa SAPO Web Imagens Notícias Blogs Produtos Directório PAi PBi Pesquisar: Onde: Pesquisar Serviços Mail Blogs Carros Casas Fotos Mapas Vídeos Notícias Messenger Todo o SAPO ...

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

[SAPO: Portugal Online](#)

Acesso Humano

Pesquisa por Texto Integral

Pesquisa Avançada

- Frase
- Exclusão
- Intervalo de datas
- Formato
- Restrição por site
- Resultados por página

Pesquisa Avançada
Defina os detalhes da sua pesquisa utilizando as opções abaixo.

Pesquisar páginas por: [Pesquisar no Arquivo](#)

Palavras

Tenham estas palavras:
ex.: expo 98

Tenham esta expressão ou frase:
ex.: pavilhão da utopia

Não tenham estas palavras:
ex.: concerto

Data

Entre: e
Ordenadas por:

Formato

Ver as páginas com o formato:

Sítio Web

Com este endereço:
ex.: www.arquivo.pt

Acesso Humano

Pesquisa por URL



sapo.pt

entre: 01/01/1996 31/12/2014

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

Versões da página web guardadas no arquivo

Foram gravadas no arquivo 1.930 versões da página <http://sapo.pt> entre 1 Janeiro, 1996 e 1 Julho, 2015.

2000 48	2001 51	2002 20	2003 29	2004 197	2005 415	2006 75	2007 139	2008 6	2009 7	2010 255	2011 370	2012 163	2013 125
29 Fev	5 Jan	24 Jan	5 Fev	18 Fev	1 Jan	1 Jan	2 Jan	1 Jan	20 Mai	26 Mar	1 Jan	1 Jan	10 Jan
29 Fev	6 Jan	6 Fev	10 Fev	19 Mar	2 Jan	1 Jan	5 Jan	14 Mar	24 Jun	1 Abr	2 Jan	2 Jan	11 Jan
3 Mar	7 Jan	30 Mar	19 Fev	5 Abr	3 Jan	2 Jan	7 Jan	20 Mar	26 Set	5 Abr	3 Jan	3 Jan	12 Jan
4 Mar	8 Jan	1 Abr	20 Fev	20 Mai	4 Jan	2 Jan	7 Jan	21 Out	26 Set	8 Abr	4 Jan	4 Jan	15 Jan
4 Mar	19 Jan	29 Mai	24 Mar	3 Jun	4 Jan	5 Jan	9 Jan	22 Out	26 Set	9 Abr	5 Jan	5 Jan	17 Jan
5 Mar	24 Jan	30 Mai	12 Abr	9 Jun	5 Jan	6 Jan	11 Jan	22 Out	18 Dez	12 Abr	6 Jan	6 Jan	18 Jan
7 Mar	10 Fev	4 Jun	19 Abr	9 Jun	5 Jan	10 Jan	12 Jan		18 Dez	13 Abr	7 Jan	7 Jan	19 Jan
11 Mar	24 Fev	5 Jun	22 Abr	11 Jun	6 Jan	10 Jan	14 Jan			16 Abr	8 Jan	8 Jan	22 Jan
8 Abr	26 Fev	6 Jun	24 Abr	12 Jun	7 Jan	11 Jan	16 Jan			19 Abr	9 Jan	9 Jan	23 Jan
10 Mai	1 Mar	25 Jul	28 Abr	12 Jun	7 Jan	12 Jan	16 Jan			22 Abr	10 Jan	10 Jan	28 Jan
10 Mai	2 Mar	27 Jul	2 Jun	13 Jun	9 Jan	13 Jan	17 Jan			30 Abr	11 Jan	11 Jan	29 Jan
11 Mai	5 Abr	9 Ago	3 Jun	14 Jun	10 Jan	15 Jan	19 Jan			2 Mai	12 Jan	12 Jan	31 Jan

Acesso Automático

Sintaxe URL para referenciar páginas

<http://arquivo.pt/wayback/wayback/20100405144904/http://www.sapo.pt/noticias/desporto>



Acesso Automático: Opensearch

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

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<rss xmlns:opensearch="http://a9.com/-/spec/opensearch/1.1/" xmlns:pwa="http://
<channel>
<title>PWA Search Engine</title>
<description>PWA search results for query: http://sapo.pt</description>
<link>http://archive.pt</link>
<opensearch:totalResults>67809</opensearch:totalResults>
<opensearch:startIndex>0</opensearch:startIndex>
<opensearch:itemsPerPage>10</opensearch:itemsPerPage>
<opensearch:Query role="request" searchTerms="http://sapo.pt" startPage="1"/>
<item>
<title>http://sapo.pt/</title>
<link>http://arquivo.pt/wayback/wayback/id2702index18</link>
<source url="http://sapo.pt/">Original URL of http://sapo.pt/</source>
<pwa:id>2702</pwa:id>
<pwa:index>18</pwa:index>
<pwa:digest>d41d8cd98f00b204e9800998ecf8427e</pwa:digest>
<pwa:tstamp>20090926171317000</pwa:tstamp>
<pwa:contentLength>162</pwa:contentLength>
<pwa:primaryType>text</pwa:primaryType>
<pwa:subType>html</pwa:subType>
</item>
```

Arquitetura do *Arquivo.pt*

1º Passo

Recolher
páginas web

Indexar
coleção
recolhida

2º Passo

Tornar páginas
Pesquisáveis

Colocar
em
produção

3º Passo

Tornar páginas
acessíveis

1º Passo: Recolher páginas web

Arquivo.pt tem 3 tipos de recolhas:

→ AWP – Recolha feitas com intervalos de 3 em 3 meses.

Ex: *<http://www.caleida.pt/saramago/>*

→ FAWP – Recolha feita todos os dias.

Ex: *<http://www.dn.pt/inicio/default.aspx>*

→ EAWP – Recolha feita em ocasiões extraordinárias.

Ex: *.eu*

1º Passo: Recolher páginas web

Quantidade de informação recolhida e indexada

Tipo	Espaço (TB)
AWP	21,8
FAWP	53,1
EAWP	5,9
<i>Internet Archive</i>	2,0
Total	83,3

Tabela 1 – Espaço ocupado (ARCS)

Tipo	Espaço (TB)
AWP	11,7
FAWP	5,9
EAWP	1,8
<i>Internet Archive</i>	1,1
Total	20,8

Tabela 2 – Espaço ocupado (Índices)

Quer ter acesso às coleções?

**É fácil. Fale connosco e o Arquivo.pt
agradece!**

2º Passo: Tornar páginas pesquisáveis

Indexação: Cluster Hadoop

Características:

- Hadoop-Cluster v0.14.5
- CentOS 6
- Cluster com 11 computadores
- Com 77 TB de espaço em disco rígido
- 352 GB de memória

3º Passo: Tornar páginas acessíveis

Arquitetura do Sistema de Pesquisa

3º Passo: Tornar páginas acessíveis

Arquitetura do Sistema de Pesquisa

Broker

Processamento de
pedidos ao **Arquivo.pt**

QueryServer

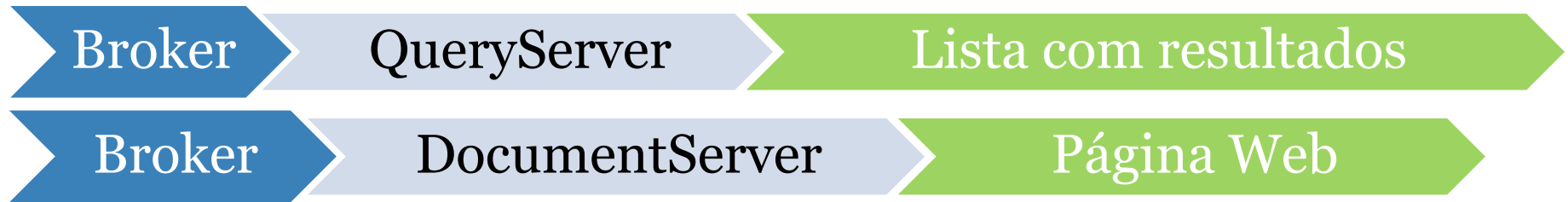
Lista com resultados

DocumentServer

Páginas arquivadas

3º Passo: Tornar páginas acessíveis

Arquitetura do Sistema de Pesquisa



3º Passo: Tornar páginas acessíveis

Arquitetura do Sistema de Pesquisa: **Broker**



3º Passo: Tornar páginas acessíveis

Arquitetura do Sistema de Pesquisa: QueryServer



Logo of Arquivo da Web Portuguesa

Search interface for the Arquivo da Web Portuguesa system. The search bar contains "sapo.pt". The date range is set from "01/01/1996" to "31/12/2014". The button "Pesquisar no Arquivo" is visible. A link for "Pesquisa avançada" is also present.

Pretenda ver os resultados das páginas que contém o texto: <http://sapo.pt>?

Versões da página web guardadas no arquivo

Foram gravadas no arquivo 1.930 versões da página <http://sapo.pt> entre 1 Janeiro, 1996 e 1 Julho, 2015.

2000 48	2001 51	2002 20	2003 29	2004 197	2005 415	2006 75	2007 139	2008 6	2009 7	2010 255	2011 370	2012 163	2013 125
29 Fev	5 Jan	24 Jan	5 Fev	16 Fev	1 Jan	1 Jan	2 Jan	1 Jan	20 Mai	26 Mar	1 Jan	1 Jan	10 Jan
29 Fev	6 Jan	6 Fev	10 Fev	19 Mar	2 Jan	1 Jan	5 Jan	14 Mar	24 Jun	1 Abr	2 Jan	2 Jan	11 Jan
3 Mar	7 Jan	30 Mar	19 Fev	5 Abr	3 Jan	2 Jan	7 Jan	20 Mar	26 Set	5 Abr	3 Jan	3 Jan	12 Jan
4 Mar	8 Jan	1 Abr	20 Fev	20 Mai	4 Jan	2 Jan	7 Jan	21 Out	26 Set	8 Abr	4 Jan	4 Jan	15 Jan
4 Mar	19 Jan	29 Mai	24 Mar	3 Jun	4 Jan	5 Jan	9 Jan	22 Out	26 Set	9 Abr	5 Jan	5 Jan	17 Jan
5 Mar	24 Jan	30 Mai	12 Abr	9 Jun	5 Jan	6 Jan	11 Jan	22 Out	18 Dez	12 Abr	6 Jan	6 Jan	18 Jan
7 Mar	10 Fev	4 Jun	19 Abr	9 Jun	5 Jan	10 Jan	12 Jan		18 Dez	13 Abr	7 Jan	7 Jan	19 Jan
11 Mar	24 Fev	5 Jun	22 Abr	11 Jun	6 Jan	10 Jan	14 Jan			16 Abr	8 Jan	8 Jan	22 Jan
8 Abr	26 Fev	6 Jun	24 Abr	12 Jun	7 Jan	11 Jan	16 Jan			19 Abr	9 Jan	9 Jan	23 Jan
10 Mai	1 Mar	25 Jul	26 Abr	12 Jun	7 Jan	12 Jan	16 Jan			22 Abr	10 Jan	10 Jan	26 Jan
10 Mai	2 Mar	27 Jul	2 Jun	13 Jun	9 Jan	13 Jan	17 Jan			30 Abr	11 Jan	11 Jan	29 Jan
11 Mai	5 Abr	9 Ago	3 Jun	14 Jun	10 Jan	15 Jan	19 Jan			2 Mai	12 Jan	12 Jan	31 Jan

3º Passo: Tornar páginas acessíveis

Arquitetura do Sistema de Pesquisa: QueryServer



Search interface showing the query "sapo" entered in the search bar. The interface includes a "Pesquisar no Arquivo" button and a "Pesquisa avançada" link. Date filters are set for "entre: 01/01/1996" and "e: 31/12/2014".

Resultados 1 a 10 de 270.798.576

[music box quiz](#)

8 Novembro, 2013 - outras datas

meses. Construído pelo SAPO MÃsicas via music box Powered by last.fm OK ...

<http://quiz.musicbox.sapo.pt/>

[SAPO Cinema](#)

15 Novembro, 2013 - outras datas

SAPO Cinema Cinema Mobile Menu de navegação Pesquisa Geral Pesquisa de Cartaz Início Estreias & ... Realização Ator / Atriz Escolher Limpar filtros Partilha siga o **sapo** cinema nas redes Facebook SAPO Twitter SAPO Google + SAPO Video SAPO Termos de pesquisa Pesquisar Home Estreias e Cartaz Atualidade ...

<http://cinema.sapo.pt/pessoa/david-a-newman/filmes/>

[SAPO Cinema](#)

8 Novembro, 2013 - outras datas

SAPO Cinema Cinema Mobile Menu de navegação Pesquisa Geral Pesquisa de Cartaz Início Estreias & ... a SIC. Adaptado de Sapo Saber a 23-01-2011 ? comentários Listagem Mais antigo Mais recente Ordem ... Participação Realização Escolher Limpar filtros Partilha siga o **sapo** cinema nas redes Facebook SAPO ...

<http://cinema.sapo.pt/pessoa/diana-chaves/filmes/17365?ordem=antigos>

[SAPO - Portugal Online!](#)

8 Junho, 2010 - outras datas

SAPO - Portugal Online! Saltar para: Pesquisa [1] , Lista de Serviços [2] , Notícias [3] ou Destaques SAPO [4] SAPO.pt Pesquisa SAPO Web Imagens Notícias Blogs Produtos Directório PAI PBI Pesquisar: Onde: Pesquisar Serviços Mail Blogs Carros Casas Fotos Mapas Vídeos Notícias Messenger Todo o SAPO ...

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

[SAPO: Portugal Online](#)

3º Passo: Tornar páginas acessíveis

Arquitetura do Sistema de Pesquisa: DocumentServer

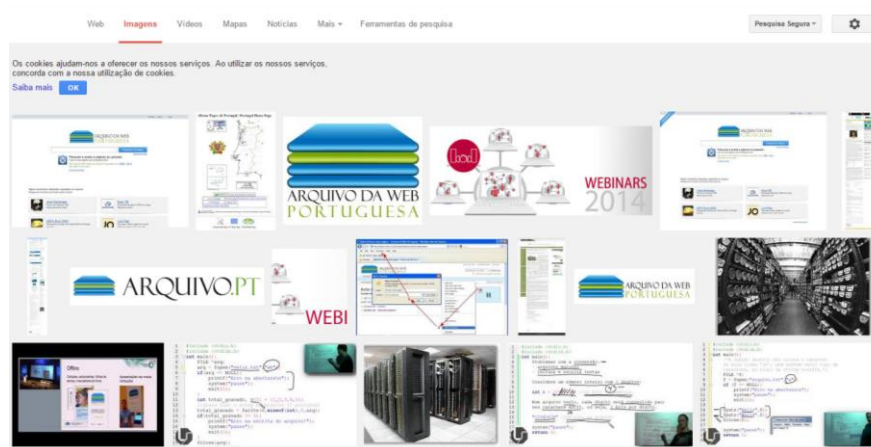


Trabalho Futuro

→ Suportar protocolo *Memento*?



→ Pesquisa por imagens?



Investigue utilizando o Arquivo.pt

Dê ideias para melhorar o serviço.

Divulgue o Arquivo.pt!

Muito Obrigado!

Hugo Viana
hugo.viana@fccn.pt