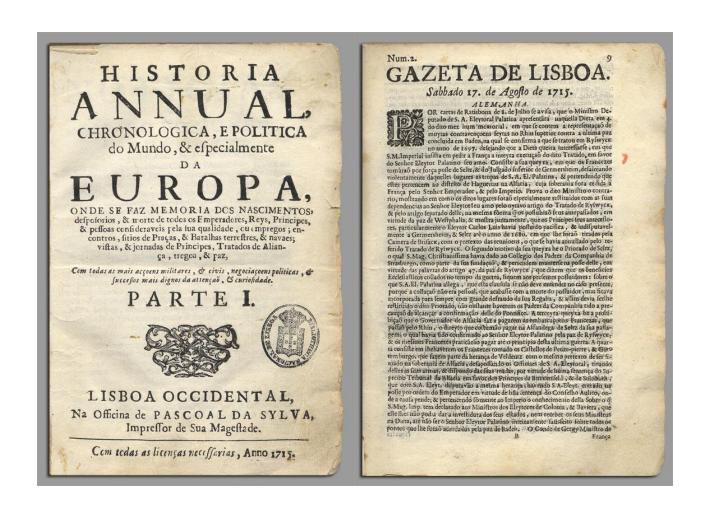


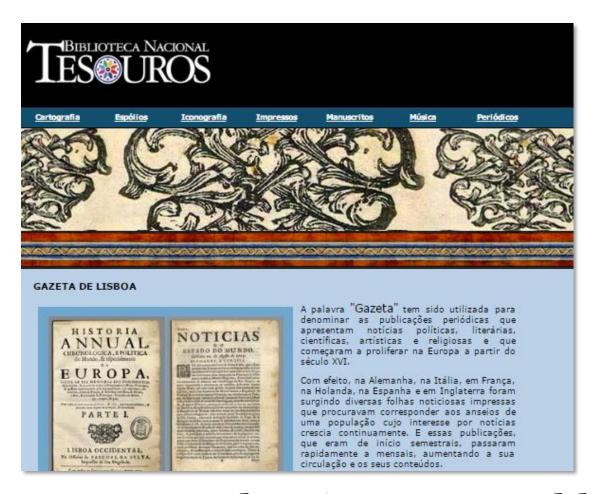
Arquivo.pt

Blogs que ficam para a História Daniel Gomes

"Gazeta de Lisboa" foi 1º jornal **impresso** português, iniciado em **1715**

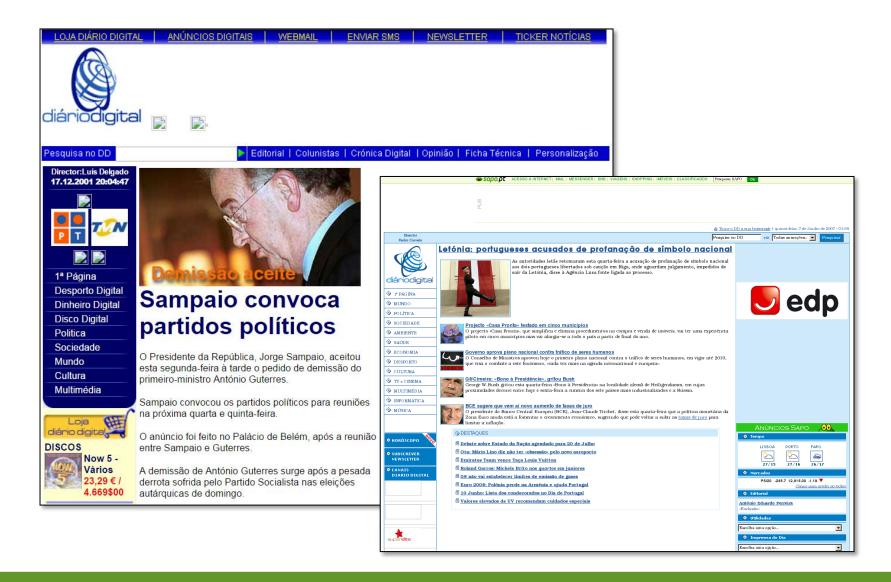


"Gazeta de Lisboa" foi suspenso em 1762.

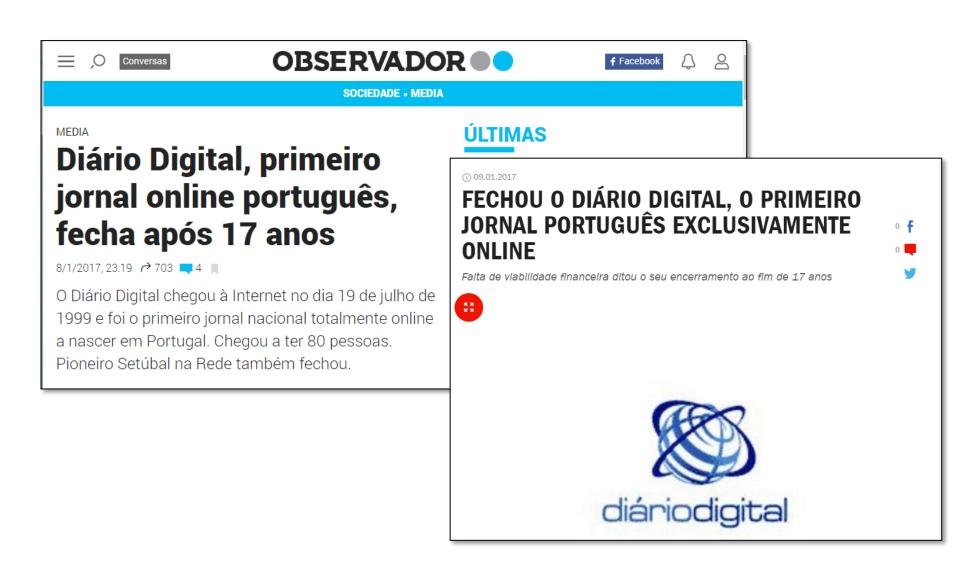


300 anos depois as suas publicações continuam acessíveis

"Diário Digital" foi o 1º jornal português **exclusivamente online,** iniciado em **1999**



"Diário Digital" desapareceu em 2017



Após apenas 17 anos, como acedemos às publicações do "Diário Digital"?



Página não encontrada.



Facto:

A Web tem substituído a Imprensa

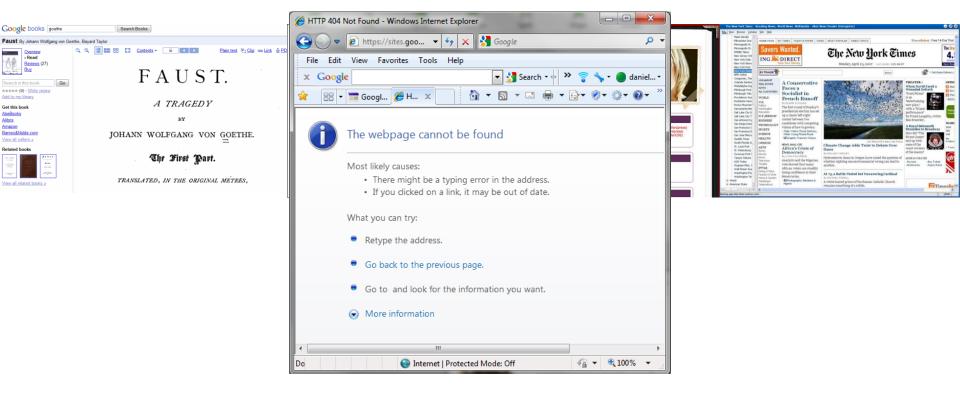


Blogs pessoais Galerias de fotografias (publicadas em Blogs)

E-books (derivados de textos de Blogs) Notícias (Blogs temáticos)



Contudo, esta valiosa informação desaparece rapidamente





80%

das páginas da Web desaparecem ou mudam passado apenas 1 ano.

São também necessários Arquivos para a Web!

Preservar a Imprensa











Preservar a Web















Serviço **gratuito** de preservação fornecido a autores da Web

- Só preserva informação de acesso público
- Respeita restrições de recolha impostas (robotstxt.org)
 - Autores podem abdicar do serviço de preservação gratuito após a recolha da página.

3 900 000 000

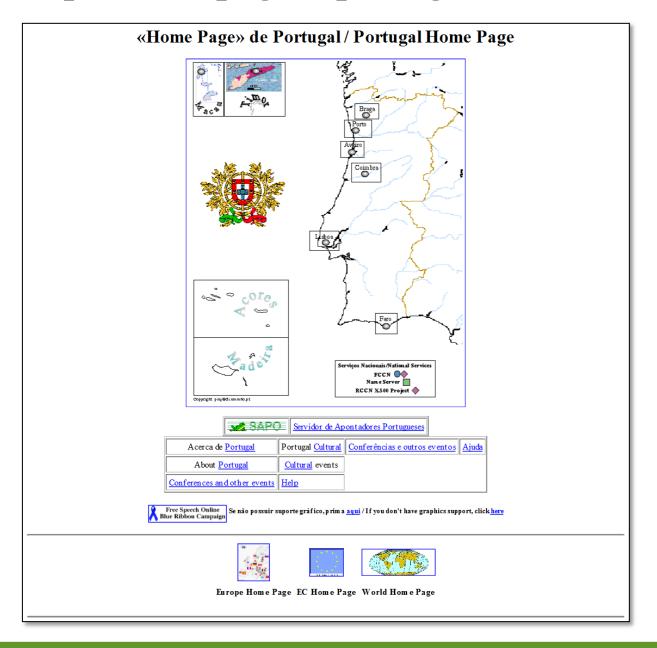
3,9 mil milhões de ficheiros da web preservados

...desde 1996



Exemplos de páginas preservadas

A primeira página portuguesa (90's)



Eventos nacionais



Resultados das eleições portuguesas de 2001

Eventos internacionais



Revolução Egípcia de 2011: revolução baseada na web, revolução arquivada da web Os artefactos originais são páginas da Web!

Alguns conteúdos são únicos no mundo

1^a página da Biblioteca do Congresso (1996)

Obtida a partir de um CD-ROM anexo a um livro



Choose a topic below, see what's new, or search our Web pages and Gopher menus.

General Information and Publications

Find out about the Library and its mission, information for visitors, publications (including Library Associates and Civilization Magazine), employment opportunities, and other general information.

Government, Congress, and Law

Search THOMAS (legislative information), access services of the Law Library of Congress (including the Global Legal Information Network), or locate government information.

Research and Collections Services

Browse historical collections for the National Digital Library (American Memory), visit Library Reading Rooms, access special services for persons with disabilities, and read about Library of Congress cataloging, acquisitions, and preservation operations and policy.

Copyrigh

Learn about the U.S. Copyright Office and the registration process, access copyright information circulars and form letters, and read about many other copyright-related topics.

Library of Congress Online Services

Search Library of Congress databases and online catalog (including LOCIS) or connect to the Library's Gopher (LC MARVEL).

Events and Exhibits

Read about Library events, conferences and seminars or view electronic versions of major exhibits.

Explore the Interne

Search the Internet, browse topical collections of Internet resources organized by Library of Congress subject specialists, and learn more about the Internet and the World Wide Web.



Library of Congress

Comments: lcweb@loc.gov (02/09/96)



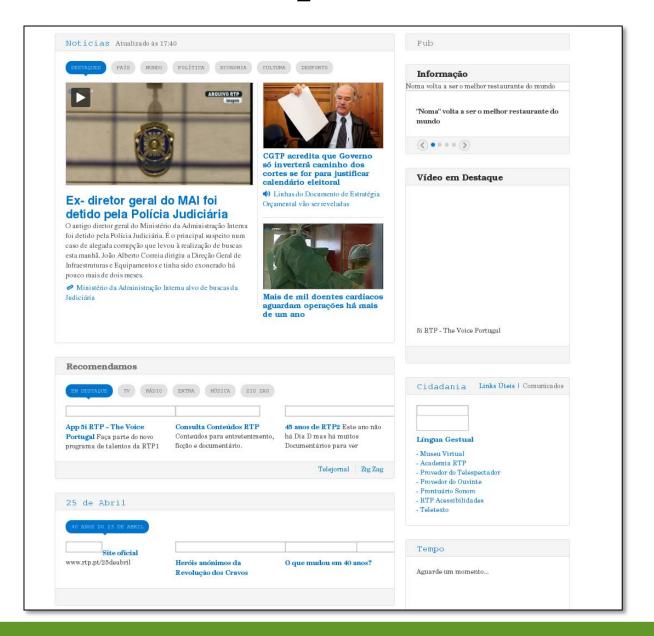
Como **pesquisar** informação arquivada?



Pesquisa por URL: Histórico de um endereço



www.RTP.pt (2014)



www.RTP.pt (2006)



www.RTP.pt (2003)



Internacional Nacional Desporto Economia Cultura Dossier Iraque

GuiaTV

RTPVideo

Superliga

Teletexto



guerra no iraque

- Começou a grande ofensiva militar sobre Bagdade
- Pentágono oficializa início da ofensiva
- Confirmada a segunda vitima mortal do lado norte-americano
- Forças iraquianas incendiaram sete poços de petróleo

🔔 Devido à quantidade de acessos o formato do site foi temporariamente alterado. 🕰

Bagdade permanece debaixo de fogo



Pelo sétimo dia consecutivo da operação "Liberdade Iraquiana", a capital iraquiana acordou ao som de novos bombardeamentos aéreos. Há informações sobre 650 mortos iraquianos em Najaf.

ver noticia -->



Representantes da aliança anglo-americana fazem balanço da situação no Iraque

ver noticia -->



Saddam Hussein apela às milícias para que combatam o inimigo

O presidente iraquiano, Saddam Hussein, apelou, terça-feira, ao corpo das milícias 'Fedayin' para que se juntem e lutem contra as forças da coligação anglo-americana no Iraque.

ver noticia

Noticias

- 2008-08-26-08:15:01
 Prisioneiros norteamericanos terão sido
 executados New York
 Times
- 2008-08-26 08-10:00 Iraquianos dispararam sobre população em Bassorá
- Bagdade permanece debaixo de fogo
- Revista de imprensa diária - ACTUALIZADA
- 2008-08-26 08:24:59
 Coligação quer controlar
 sul do Iraque antes de
 atacar Bagdad
- 2008-08-26 08:00:00
 Coreia do Norte adverte
 que Japão se arrisca à
 "auto-destruição"
- 2008-08-26 07:55:01
 Combates em Najaf
 causaram cerca de 650
 mortos iraquianos
- Pneumonia atípica faz encerrar escolas em Hong Kong
- 2008-08-26 07:40:00
 Portugal disponível para
 participar na
 reconstrução do Iraque

www.RTP.pt (1999)

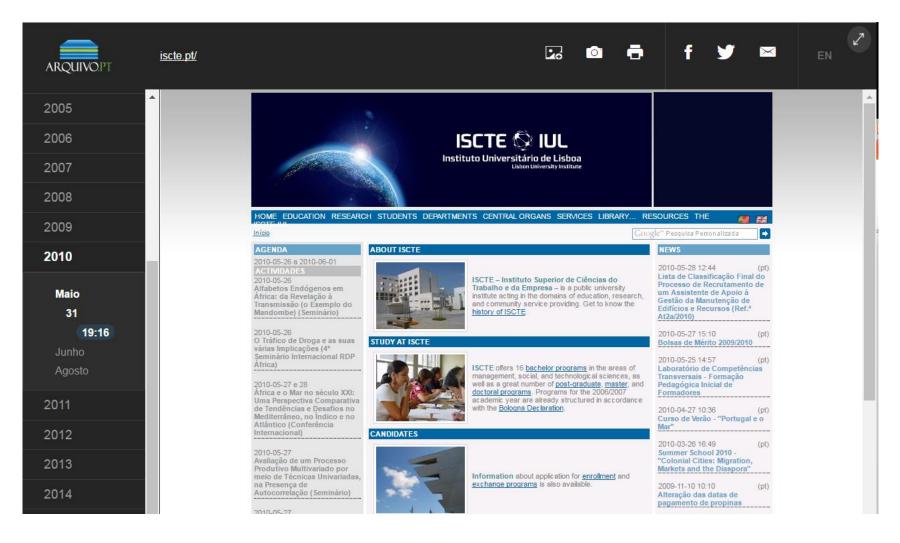


www.RTP.pt (1998)



Reprodução de páginas preservadas:

Navegar entre versões e seguir ligações



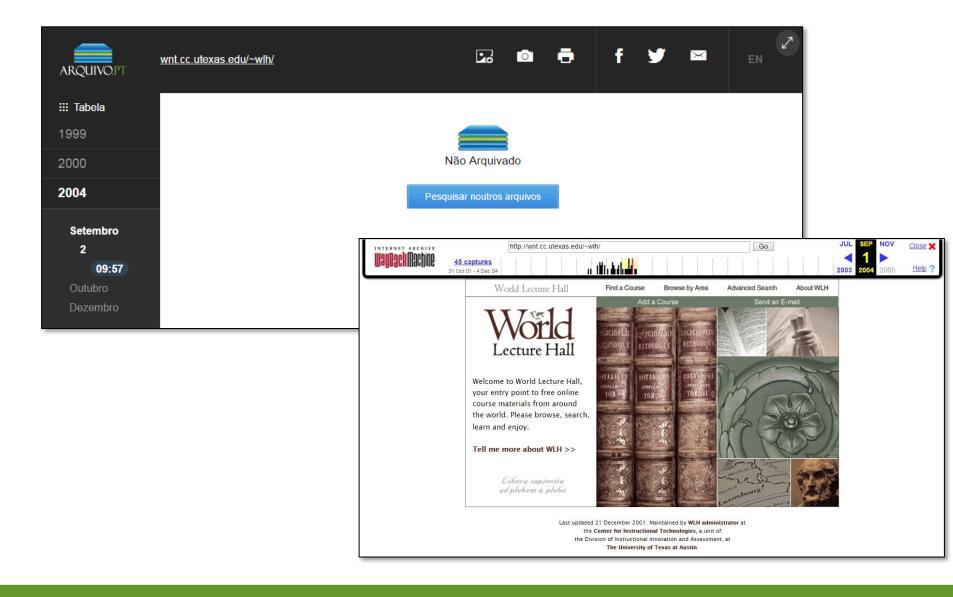
Guardar página completa em vez de

"Print Screen"



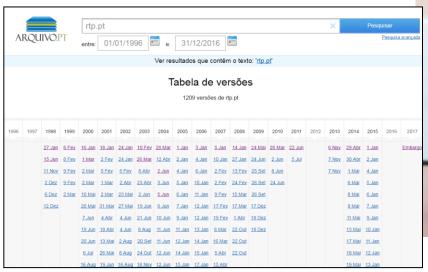


Pesquisar noutros arquivos





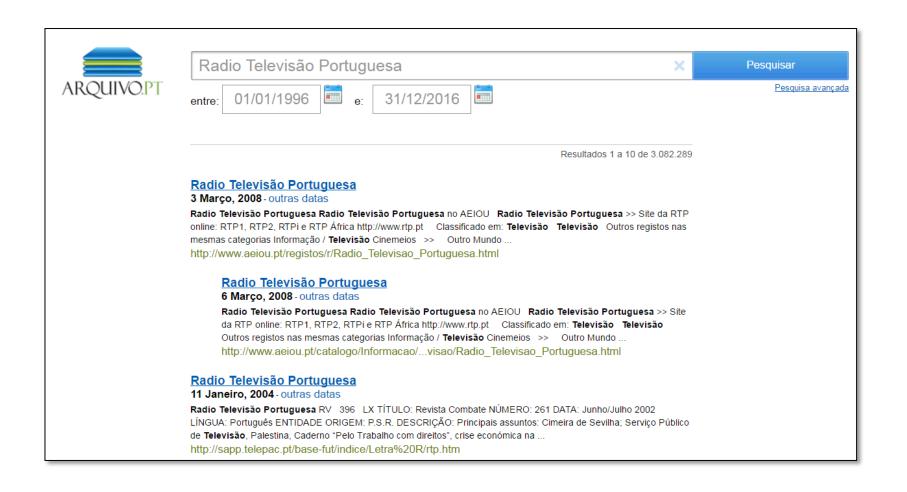
Problema com pesquisa por URL



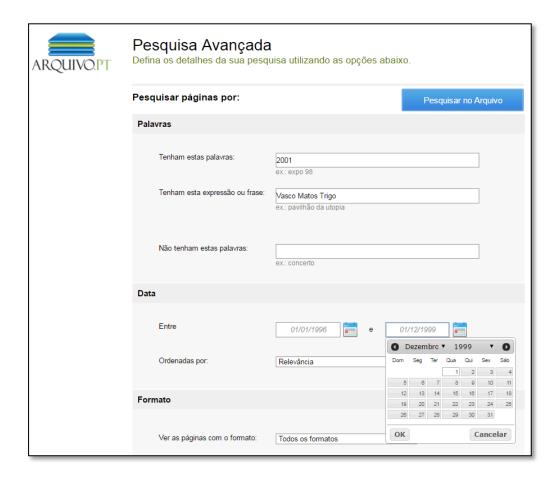


Os utilizadores não sabem o **endereço exacto** da página que continha a informação que necessitam.

Inovação do Arquivo.pt: **pesquise textos** de páginas do passado



Pesquisa avançada



- 2001
- "Vasco Matos Trigo"
- Antes de 1999

2001: programa do passado acerca do futuro

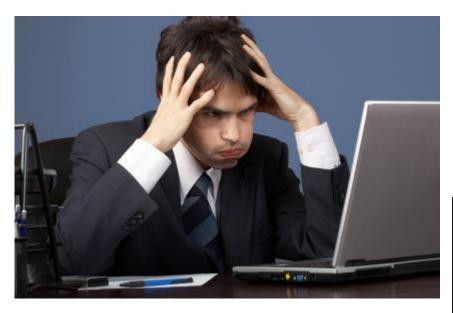




Arquivo.pt

Casos de uso

Gestão: documentação/auditoria de projeto concluído



"Estava tudo no site"

Linguateca

Estrutura Equipa

Apresentação Acesso a recursos

- AC/DC
- CETEMPúblico
- CETENFolha
 CHAVE
- COMPARA
- Corpógrafo
- CorTrad
 Esfinge
- Floresta Sintá(c)tica
- GeoNetPT
 PAPEL
- PANTERA
- PoNTE
- ReEscreve
- REPENTINO
- Repositório
- WebJspell
- <u>WPT</u>

 Avaliação conjunta

Catálogo de recursos
Catálogo de ferramentas
Catálogo de atores
Catálogo de publicações
Informação interessante

Informação interessante
Perguntas já respondidas
Gramateca

Englis

Repositório de recursos, publicações e programas

Linguateca

Neste espaço providenciamos uma "estante electrónica" para todos quantos quiserem ter os seus trabalhos relacionados com o processamento computacional da lingua portuguesa acessiveis directamente da rede.

Basta enviarem-nos a versão electrônica, indicando a referência que queiram que apareça no nosso catálogo de publicações, ou, no caso de programas ou outros recursos, enviarem-nos o respectivo programa e documentação básica.

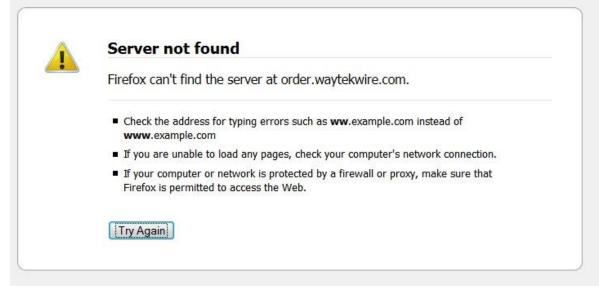
De momento, temos os seguinte recursos importantes (por ordem de disponibilização), em que indicamos também, com gratidão, o responsável pela sua cedência, e a data a partir da qual ficaram disponíveis no Repositório

- Regras de expressões bilingues inglês-português, Logos Corporation (Anabela Barreiro, Janeiro de 2001)
- Base de Dados para Sintetizador de Fala em Português Europeu, FEUP/IPB-BD (João Paulo Teixeira, Abril de 2001)
- <u>Dicionário Portuguez-Latino de Pedro José da Fonseca</u> (Digitalizado por Ricardo Ueda Karpischek, Julho de 2003)
- Colecção FolhaRIcol (Vera Lúcia Strube de Lima, Agosto de 2003)
- Corpus TeMário (Lucia H. Machado Rino, Outubro de 2003)
- Colecção WBR-99 (Berthier Ribeiro-Neto, Nivio Ziviani, Dezembro de 2003)
- <u>Port4NooJ: recursos linguísticos para o português no formato NooJ</u> (Anabela Barreiro, Novembro de 2007)
- Léxico-gramática dos adjectivos intransitivos humanos do português europeu (Paula Carvalho, Dezembro de 2008)
- Erros ortográficos em português de Portugal e sua correcção (José Carlos Medeiros, Agosto de 2009)
- Memórias de tradução de português de Portugal para inglês britânico (Março de 2010)
- Correcção ortográfica de português (Abril de 2010)
- ReLi, corno de resenhas de livros (Cláudia Freitas, Maio de 2013)

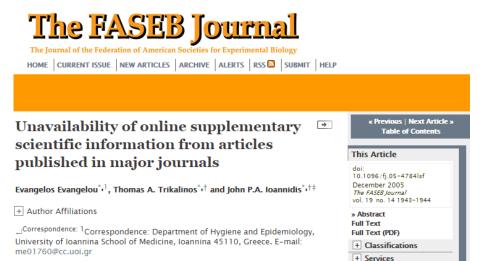
Portfólio/CV

"O meu melhor trabalho foi um site que já não existe."

• 60% dos sites desaparecem após apenas 2 anos.

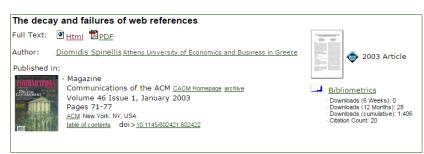


Investigação em todas as áreas: recursos indisponíveis



Malaysian Journal of Library & Information Science, Vol. 16, no. 3, December 2011: 17-29

Death of web citations: a serious alarm for authors



Persistence and Decay of Web Citations Used in Theses and Dissertations Available at the Sokoine National Agricultural Library, Tanzania

ACADEMIC JOURNAL ARTICLE

By Sife, Alfred S.; Bernard, Ronald

International Journal of Education and Development using Information and Communication Technology , Vol. 9, No. 2 , July 1, 2013

JOURNAL OF THE ASSOCIATION FOR

INFORMATION SCIENCE AND TECHNOLOGY

Research Article

Link decay in leading information science journals

Dion Hoe-Lian Goh and Peng Kin Ng

Article first published online: 3 NOV 2006 DOI: 10.1002/asi.20513

Copyright © 2007 Wiley Periodicals, Inc., A Wiley Company

Issue



Journal of the American Society for Information Science and Technology

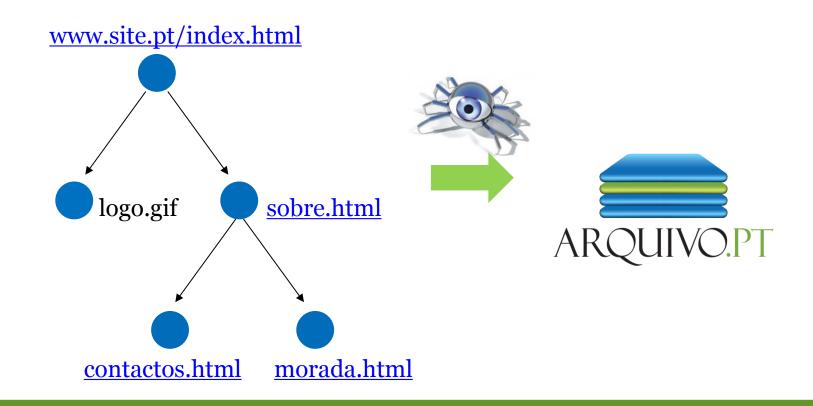
Volume 58, Issue 1, pages 15–24, 1 January 2007



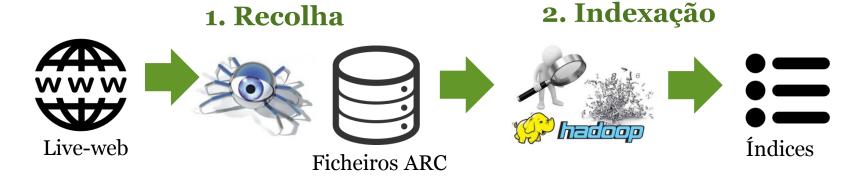
Como funciona o Arquivo.pt?

Informação é recolhida **automaticamente** da Web para ser preservada

- Conjunto de endereços iniciais (e.g. páginas de entrada)
- Recolha automática e iterativa de ficheiros hiperligados



Pesquisa análoga aos motores de busca mas com **reprodução** das páginas **preservadas**





4. Reprodução e Acesso





3. Pesquisa



Insira palavra ou URL × Pesquisar



Arquivo.pt

O que fazer para que um site **seja preservado**

Informação é Património!

- Os sites são Património
- A informação publicada nos sites é valiosa
- **Preservar informação** não é "coisa de informático"

Sugera site para ser preservado quando é **criado** ou **descoberto**

- arquivo.pt/sugerir
- Formulário simples e público
- Basta a página de entrada



Tais como sites relacionados com as eleições



Em fundo: Gabinete do Presidente da República

[Noticias] [Candidatos] [Presidência] [Campanha] [Resultados] [Votação Virtual] [Newsgroup] [Fotos da vitória]



Este serviço foi consultado 11416 vezes.

Comentários e Sugestões para <u>Lusa</u> e <u>Telepac</u>. Realização <u>Telepac</u> e <u>Tinta Invisível</u>

Telepac.pt, 1996

PSD

(psd2011.com, 2011)



Site japonês (psd2011.com, 2015)



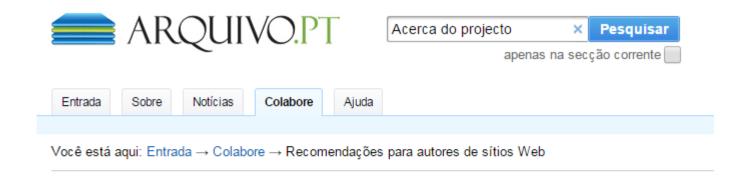
PS (ps.pt/legislativas2011/, 2011)



PS (ps.pt/legislativas2011/, 2015)



Siga e divulgue recomendações para criar sites preserváveis



Recomendações para a criação de sítios web preserváveis

Apresenta recomendações para que as páginas de um sítio Web possam ser corretamente arquivadas e acedidas ao longo dos anos. É impossível preservar a web portuguesa sem a colaboração dos autores.

Última atualização em 2015-08-03.

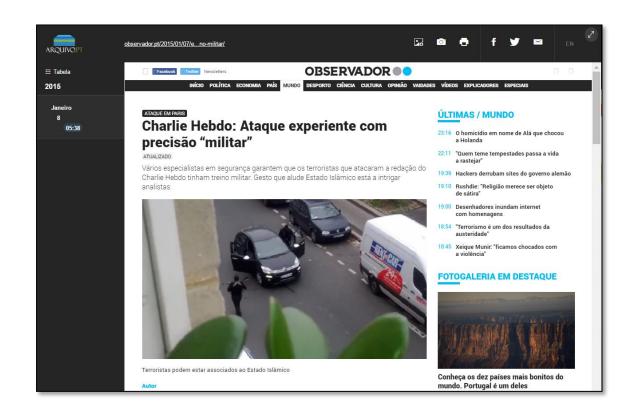
arquivo.pt/recomendacoes

Incluindo boas velhas práticas de publicação

Data de arquivo: 8 Janeiro 5:38, 2015

Data de publicação (observador.pt):

?



Omissão de data de publicação é um problema grave e comum. **Contexto temporal é essencial** para interpretar informação.

Publicar sempre a data de publicação



Especialmente em páginas de notícias

Verificar o ficheiro *robots.txt* (Robots Exclusion Protocol)



Autor pode proibir acesso automático a partes do seu site (<u>www.robotstxt.org</u>)

http://www.dn.pt: site original

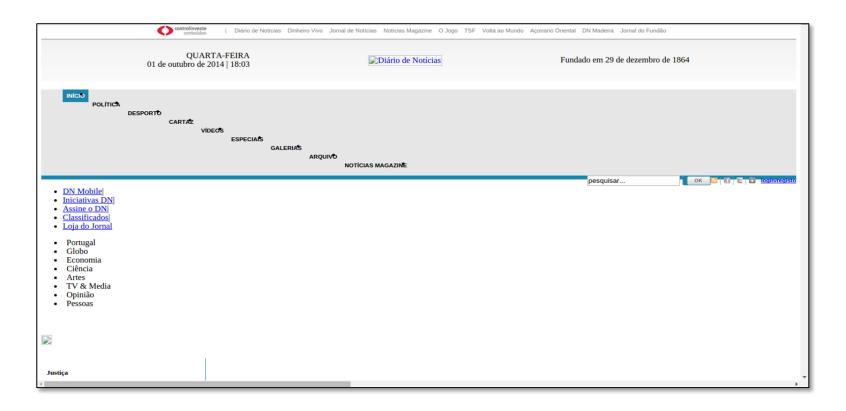


http://www.dn.pt/robots.txt

```
User-agent: *
Disallow: /common/scripts/
Disallow: /common/css/
Disallow: /admin/
Disallow: /search/
Sitemap: http://www.dn.pt/google_news.ashx
```

http://www.dn.pt

Consequência da regra de exclusão: página preservada de forma **incompleta**



Não deixe que aconteça o mesmo problema com o seu site!

Permitir preservação **gratuita** pelo Arquivo.pt no robots.txt

User-agent: Arquivo-web-crawler Disallow:



Blogs que **ficam** para História Blogs que **fazem** a História

Estórias escritas por alguns



História descrita por todos



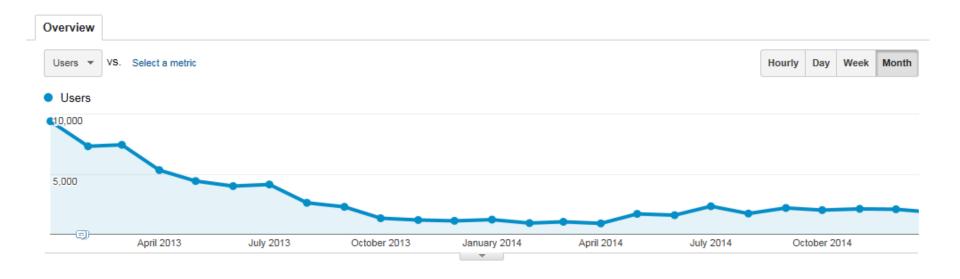
Preservar **blogs** concede um lugar na História ao "cidadão comum"



Divulgar o Arquivo.pt

Precisa de ser conhecido para ser utilizado

Temos poucos utilizadores



E poucos recursos para divulgação.

Precisamos da **Vossa** ajuda para **divulgar** o Arquivo.pt!



Online



Disseminar as nossas notícias arquivo.pt/news



Inscreva-se no grupo de divulgação agora para ajudar a divulgar!

- Só para informações importantes
 (ex. formações, novas coleções e funções, colaborações, bolsas).
- Pode sair quando quiser

Ajude a divulgar o Arquivo.pt para que seja útil a mais pessoas!

Inscreva-se para poder receber e divulgar novidades acerca do Arquivo.pt (ex. novas coleções, projetos, colaborações, bolsas).

Enviaremos apenas as informações mais importantes e pode desinscrever-se quando quiser.

| Nome | Email | Instituição |
|------|-------|-------------|
| | | |
| | | |
| | | |
| | | |
| + | | |
| | | |



Arquivo.pt: em suma

- Serviço gratuito de preservação fornecido aos autores da Web
- Um "Google" para o passado!
 - Complementa a pesquisa sobre a Web do presente
- Infraestrutura para investigação