



Melhorando a Wikipédia com o Arquivo.pt

Jornadas FCCN

6 Abril 2024

pedro.gomes@fccn.pt



ARQUIVO.PT



Arquivo.pt **preserva** conteúdos históricos **publicados online**

GALILEO Versão 2.0

Janeiro 1993

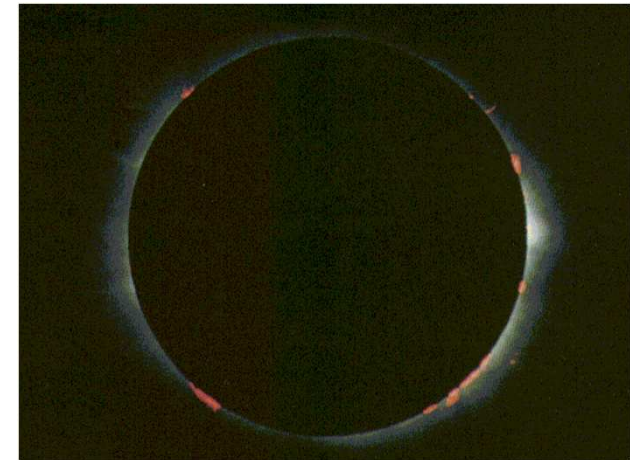
Manual de utilização e sugestões de exploração	AUTORES Programação: João Veloso Manual: Elisa Prata Pina e M. Augusta Patrício Orientação: Carlos Fiolhais
--	--

Departamento de Física da Universidade de Coimbra

ÍNDICE

- [1. Apresentação sumária do programa](#)
- [2. Características do equipamento](#)
- [3. Configuração mínima do computador](#)
- [4. Ficheiros do programa](#)

nautilus.fis.uc.pt- 1993
(página mais antiga)



spacelink.nasa.gov – 1992
(imagem mais antiga)

visitmadeira.com

ARQUIVO.PT
visitmadeira.com 5 Março às 21h35, 2023

Madeira. Tão tua

ONDE IR O QUE FAZER A ACONTECER ONDE FICAR TRAVEL INFO WISHLIST ♥ PT ▾

Madeira. Tão Tua
Paisagens tão nossas.
VER VÍDEO

Menu

Tabela

- 2022 ▾
- 2023 ▾
- Janeiro ▾
- Fevereiro ▾
- Março ▾
- 5 Março 21h35, 2023
- 7 Março 21h14, 2023
- 12 Março 20h49, 2023

Pesquise páginas do passado

Menu ARQUIVO.PT Opções

Q josé saramago Pesquisar

1996 1 Jan 2021 27 Abr

Páginas Imagens Pesquisa avançada

Cerca de 47.893.029 resultados desde 1996 a 2021

[→ asintermitenciasdamorte.blogs.sapo.pt](https://asintermitenciasdamorte.blogs.sapo.pt)

AS INTERMITÊNCIAS DA MORTE – JOSÉ SARAMAGO

7 Dezembro 2017

AS INTERMITÊNCIAS DA MORTE – **JOSÉ SARAMAGO** As Intermitências da Morte, de **José Saramago**. Leya BIS – livros de bolso 15 Jan 09 AS INTERMITÊNCIAS DA MORTE As Intermitências da Morte **José Saramago** «No dia seguinte ninguém morreu.» Assim começa este romance de **José Saramago**. Colocada a hipótese, o ...

[→ josesaramago.org](https://josesaramago.org)

Fundação José Saramago | José Saramago

4 Fevereiro 2016

Fundação **José Saramago** | **José Saramago** Início A Fundação Quem somos Onde estamos Conteúdos expositivos Declaração de Princípios Informações oficiais e de gestão **José Saramago** Autobiografia Biografia de **José Saramago** Cronobiografia Distinções Nobel Obra Bibliografia ativa Artigos Conferências Outras ...

Pesquise imagens do passado

























Menu ARQUIVO.PT Opções

Q José saramago Pesquisar

1991 6 Ago 2024 2 Abr

Páginas Imagens Narrativa Pesquisa avançada

Cerca de 4.616.008 resultados desde 1991 até 2024

 → http://sr.globalvoicesonline... 13 Janeiro 2011	 → http://geneall.net/pt/nom... 2 Agosto 2017	 → http://islakokotero.blogsp... 26 Dezembro 2009	 → http://teknomatika.blogsp... 11 Maio 2020	 → http://aeiou.caras.pt/pole... 18 Dezembro 2009	 → http://jf-loures.pt/editais-... 30 Agosto 2019	 → http://pt.globalvoicesonli... 13 Janeiro 2011	 → http://web.bn.pt/index.ph... 13 Outubro 2018
 → https://sv.wikiloc.com/spa... 22 Agosto 2021	 → http://www.rtve.es// 11 Dezembro 2018	 → http://teknomatika.blogsp... 11 Maio 2020	 → https://academiagalega.o... 25 Março 2019	 → http://aeiou.visao.pt/JL 20 Julho 2010	 → http://activa.aeiou.pt/artig... 13 Agosto 2010	 → http://www.hojelusofonia... 13 Agosto 2015	 → https://novaziadaonda.w... 8 Fevereiro 2016
 → http://evene.lefigaro.fr/ce...	 → http://noticias.sapo.pt/not...	 → http://ohequenao.wordpr...	 → http://p3.publico.pt/cultur...	 → https://capeiaarraiana.wo...	 → http://www.iea.usp.br/pes...	 → https://eusouempreende...	 → https://dcbm.uaig.pt:80/

Pesquise o **histórico** de um endereço

Menu ARQUIVO.PT

Q ipv.pt Pesquisar

1991 6 Ago 2022 25 Mai

Páginas Imagens Narrativa Pesquisa avançada

Cerca de 338 resultados desde 1991 até 2022

Tabela Lista

5	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
			18 Jan	25 Jan	18 Out	19 Abr	22 Jan	2 Fev	5 Abr	1 Jan	8 Fev	2 Fev	4 Jan	23 Jun	6 Jun	22 Jan	22 Jan	7 Nov	30 Set	23 Abr	9 Jan	22 Fev	7 Abr	13 Mar	23 Jan	8 Jan	1 Abr
			19 Fev	2 Fev		15 Mai	28 Mar	16 Fev	10 Mai	22 Jan	21 Fev	10 Fev	8 Fev	26 Set	4 Ago	19 Mai		12 Nov	28 Nov	15 Mai	10 Fev	6 Jun	13 Abr	18 Mar	4 Fev	11 Jan	25 Mai
			2 Dez	8 Fev		7 Jul	29 Mai	29 Mar	5 Jun	23 Jan	10 Abr	17 Fev	14 Fev	18 Dez		2 Jul				13 Set	6 Jun	6 Ago	20 Abr	24 Mar	11 Fev	12 Fev	
			12 Dez	29 Abr		20 Jul	4 Jun	21 Abr	11 Jun	5 Fev	13 Jun	25 Fev	12 Mar							24 Nov	8 Jun	13 Dez	2 Jun	24 Mar	27 Fev	12 Fev	
				4 Mai		26 Set	27 Set	5 Jun	11 Jun	8 Fev	15 Jun	6 Mar	21 Out								9 Jun		5 Jun	22 Mai	6 Mar	14 Fev	
						27 Nov	29 Set	20 Jun	12 Jun	10 Fev	1 Jul	14 Mar									6 Nov		19 Jul	1 Set	11 Mar	17 Fev	
							20 Nov	19 Out	14 Jun	17 Fev	4 Jul	23 Mar									10 Nov		30 Jul	1 Set	15 Mar	3 Mar	
								26 Out	16 Jun	24 Fev	5 Jul	10 Abr											3 Ago	6 Set	25 Mar	24 Jun	
								23 Dez	18 Jun	26 Fev	18 Jul	10 Mai											13 Out	18 Nov	26 Mar	24 Jun	
								25 Dez	26 Jun	27 Fev	19 Jul	9 Jun											27 Out	8 Dez	28 Mar	25 Jun	
								26 Dez	7 Jul	28 Fev	13 Ago	10 Jun											31 Out	8 Dez	8 Abr	25 Jun	
									10 Jul	3 Mar	20 Ago	24 Jun											7 Nov	8 Dez	5 Mai	25 Jun	
									13 Jul	4 Mar	25 Ago	10 Ago											27 Nov	20 Dez	2 Jun	25 Jun	

Navegue por
fccn.pt em 1996

Menu ARQUIVO.PT Opções

fccn.pt 13 Outubro às 14h56, 1996

Fundação para a Computação Científica Nacional

A Fundação tem a responsabilidade de gerir e desenvolver a RCCN, Rede da Comunidade Científica Nacional, a rede académica portuguesa.

- [RCCN](#)

Outras actividades: [SCCN](#) | [CCCN](#) | [ICCN](#) | [ACCN](#)

- [Eventos](#)
- [Publicações](#)
- [Serviços disponíveis](#)

 [Servidores WWW das instituições ligadas à RCCN](#)  [Home Page de Portugal](#)

 [Estatísticas de utilização deste servidor](#)
5685 visitantes desde 95-11-22

Fundação para a Computação Científica Nacional (FCCN)
Av. Brasil, 101
1799 Lisboa Codex
Tel. +351 1 8481906
Fax. +351 1 8472167
email. info@fccn.pt

Este servidor é mantido por: *Florentino dos Santos Gameiro - flo@rccn.net*
Última alteração: *Wednesday, 10-Jan-96 18:13:25 CUT*

 "A promoção de infraestruturas no domínio da computação científica gerindo-as, nuns casos, e doando-as noutros, é o principal objectivo da [Fundação](#), que actua predominantemente no âmbito das Universidades, Centros e Laboratórios de I&D."

SavePageNow

Grave páginas no momento

Menu

ARQUIVO.PT

■ SavePageNow Grava páginas no Arquivo.pt

Gravar

ex. www.fccn.pt

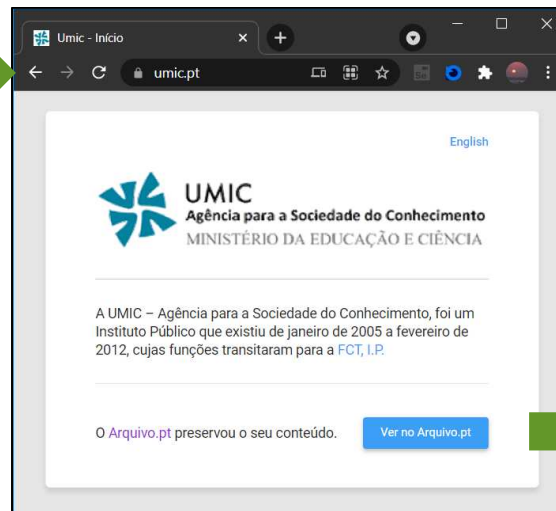
O serviço SavePageNow permite arquivar uma página no exato momento em que o utilizador faz o pedido.

Os conteúdos arquivados serão posteriormente integrados no acervo do Arquivo.pt.

[Saber mais](#)

arquivo.pt/savepagenow

Não mate o seu site histórico! Preserve-o no **Memorial do Arquivo.pt**: arquivo.pt/memorial



Antes que o seu site “abandonado” seja atacado, envie-o para o Memorial e evite problemas de segurança, custos e perdas de informação!

CitationSaver Preserva citações a conteúdos online

Os documentos citam conteúdos online que desaparecem rapidamente.

O CitationSaver preserva o conteúdo das ligações citadas (ex. páginas web citadas num livro) para que possam ser recuperadas mais tarde a partir do Arquivo.pt. [Saber mais](#)

Submeta um documento e o CitationSaver preservará as ligações nele citadas:

URL	Ficheiro	Texto
Insira o URL do documento:		
<input type="text"/>		
Formatos aceites: PDF, TXT, HTML. Tamanho máximo: 100 MB		
O seu e-mail (opcional):		
<input type="text"/>		
Guardar ligações citadas		

Sugira o seu website para ser preservado

Sugerir websites para serem preservados

Última atualização em 16 de Outubro de 2020 às 11:22

Partilhe esta página: arquivo.pt/sugerir

***Required**

Email address *

Your email address _____

Endereços dos sites (1 por linha) *

Your answer _____

Descrição dos sites

Your answer _____

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

Arquivo.pt é um serviço **internacional** e **interdisciplinar**

Metade dos **utilizadores são internacionais**

Country	Users	% Users
1. Portugal	46,891	46.56%
2. United States	26,373	26.19%
3. Brazil	2,266	2.25%
4. Russia	2,234	2.22%
5. United Kingdom	2,231	2.22%
6. Japan	2,172	2.16%
7. Canada	1,237	1.23%
8. Mozambique	1,213	1.20%

Fonte de dados para **Inteligência Artificial**

Generating a European Portuguese BERT Based Model Using Content from Arquivo.pt Archive

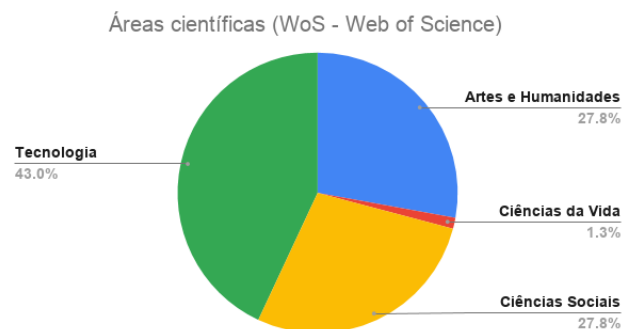
[Nuno Miquelina](#), [Paulo Quaresma](#) & [Vitor Beires Nogueira](#)

Conference paper | [First Online: 21 November 2022](#)

608 Accesses

Part of the [Lecture Notes in Computer Science](#) book series (LNCS, volume 13756)

Suporta trabalhos em **várias áreas**



Mais de 1 000 **referências académicas**

Google Scholar search results for "Arquivo.pt" | "Portuguese web archive". The page shows 1,020 results. The first result is "Major Minors-Ontological representation of minorities by newspapers" by P.J.P. Martins, L.J.A.D. Costa, and J.C. Ramalho, published in 2021. The second result is "Automatic Classification of Stigmatizing Articles of Mental Illness: The Case of Portuguese Online Newspapers" by J.L. Oliveira, published in 2022.

Catálogo de serviços do Arquivo.pt

Última atualização em 19 de Setembro de 2023 às 10:20

Públicos

- [Pesquisa e acesso](#) sobre a informação preservada da web
- [Interfaces para Programação de Aplicações \(APIs\)](#)
- [Sugira websites](#) para serem preservados
- [SavePageNow](#): archive imediatamente páginas da web
- [Integração de coleções históricas de dados da web](#)
- [Formação](#) acerca de preservação da web
- [Dados abertos](#) que listam informação arquivada da web relativa a vários temas
- [CitationSaver](#): extrai os links contidos em documentos e arquiva as correspondentes páginas
- [Arquivo404](#): mostra páginas preservadas em vez de páginas não encontradas

Para entidades parceiras

- [Memorial](#): preserve a informação do seu antigo website antes de o desativar
- [Arquivo a-pedido](#) de websites em alta-qualidade
- [Criação online de coleções e exposições temáticas](#)
- [Exposições itinerantes de cartazes](#) que exibem páginas históricas da web

Todos os **serviços** em
arquivo.pt/catalogo



WIKIPÉDIA

A enciclopèdia lliure



List of Web archiving initiatives

From Wikipedia, the free encyclopedia



This article's **use of external links may not follow Wikipedia's policies or guidelines**. Please improve this article by removing excessive or inappropriate external links, and converting useful links where appropriate into footnote references. *(June 2019)* (*Learn how and when to remove this template message*)

This article contains a list of **Web archiving initiatives** worldwide. For easier reading, the information is divided in three tables: web archiving initiatives, archived data, and access methods.

This Wikipedia page was originally generated from the results obtained for the research paper *A survey on web archiving initiatives*^[2] published by the *Arquivo.pt*^[2] (the Portuguese web-archive) team.



Map of Web archiving initiatives worldwide in April 2021^[1]

Web archiving initiatives [edit]

Name ↕	Country ↕	Creation Year ↕	Technologies ↕	Number of Employees		Comments
				Full-time ↕	Part-time ↕	
						The End of Term <i>Web Archive</i> ^[2] captures and saves U.S. Government federal government websites / <i>web.archive.org</i> is the 1 <i>initiative</i>

Contents hide

(Top)

- Web archiving initiatives
- Archived data
- Access methods
- See also
- References
- External links



WIKIPÉDIA
A enciclopédia livre



ChatGPT

Wikipédia vs ChatGPT

- Grande parte da fonte de dados do ChatGPT foi os textos da Wikipédia.
- Existe uma restrição temporal no ChatGPT.
- ChatGPT não contém as fontes da informação de onde as respostas foram retiradas.



en.wikipedia.org/wiki/List_of_Web_archiving_initiatives

List of web archiving initiatives

From Wikipedia, the free encyclopedia

This article's use of external links may not follow Wikipedia's policies or guidelines. Please improve this article by removing excessive or inappropriate external links, and converting useful links where appropriate into footnote references. (June 2019) *(Learn how and when to remove this template message)*

This article contains a list of Web archiving initiatives worldwide. For easier reading, the information is divided in three tables: web archiving initiatives, archived data, and access methods.

This Wikipedia page was originally generated from the results obtained for the research paper *A survey on web archiving initiatives*^[1] published by the Arquivo.pt^[2] (the Portuguese web-archive) team.

Contents [hide]

- Web archiving initiatives
- Archived data
- Access methods
- See also
- References
- External links

Map of Web archiving initiatives worldwide in April, 2021. Data#

Link Results

Check My Links is supported by Logfile, Easily ingest, stream, search and dashboard structured logs with Logfile.

Links: 1328 Queue: 0

Caching: FALSE Method: GET

Valid links: 1176

Valid redirecting links: 142

Warnings: 0

Invalid links: 10

39. [Web Archiving Project \(WARP\)](#), [The National Diet Library, Japan](#)" on 2013-11-17.
40. [National Library of Korea - OASIS \(Online Archiving & Searching\)](#) on 2013-10-31. [Oasis.go.kr](#), 2013-08-01. Archived from [the original](#) **404** on 2013-10-31.
41. [WebART \(Web Archive Retrieval Tools\)"](#).
42. ["Latvijas Nacionālā bibliotēka - Rasmošana"](#).
43. ["New Zealand Web Archive"](#). Natlib.govt.nz. Retrieved 2021-02-26.
44. ["Netarkivet"](#). [Nasjonalbiblioteket](#) (in Norwegian Bokmål). Retrieved 2019-12-25.
45. ["The National Library of Norway"](#). [IIPC](#). Retrieved 2019-12-25.
46. ["Arquivo.pt - the Portuguese web-archive: search pages from the past"](#). Scientific Computing (FCGN). 13 August 2013. Retrieved 13 August 2013.
47. ["Web archive of Cacak 404 ^{\[german dead link\]}"](#). digital.cacak.dis.rs
48. ["Web Archive Singapore"](#). Was.ni.sg. Retrieved 2013-11-17.
49. Digital Resources (Digital Resources Archive of the University Library in Bratislava)^[1]
50. ["Slovenian Web Archive"](#). National and University Library of Slovenia. Retrieved 2018-02-02.
51. [Biblioteca Nacional de España. "Archivo de la web española"](#). Archived from [the original](#) **404** on 2014-02-23. Retrieved 2014-02-20.
52. [National Library of Catalonia \(16 November 2012\). "PADICAT: The Web Archive of Catalonia"](#). National Library of Catalonia. Retrieved 16 November 2012.
53. [Kai Oswald Seidler. "Basque Digital Heritage Archive \(ONDARENET\)"](#). euskadi.net. Archived from [the original](#) on 2012-12-20. Retrieved 2013-11-17.
54. ["Kulturarw3 - Kungliga biblioteket"](#) (in Swedish). Kb.se. 2020-01-01. Retrieved 2021-05-04.
55. [AAW Designs. "Aleph Archives"](#). aleph-archives.com. Retrieved 2013-11-17.
56. ["Expatriate Archive Centre Blog Archive"](#). xpatarchive.com. Retrieved 2020-02-03.
57. ["Web Archiving Bucket"](#). webarchivingbucket.com. Retrieved 2013-11-17.
58. ["Web Archive Switzerland"](#). E-helvetica.nb.admin.ch. Retrieved 2013-11-17.
59. ["NTU Web Archiving System, NTUAS"](#) **404** [ntu.edu.tw](#). Retrieved 2013-11-17.

25% dos links externos referenciados na Wikipedia portuguesa estavam quebrados.



Problema de Content Drift

A ligação pode referenciar um conteúdo ainda disponível, mas este pode já não ser o que se pretendia referenciar.



Mantenha os antigos domínios

2005@Arquivo.pt

The screenshot shows the website 'ARQUIVO.PT' for 'vivapraia.com' dated '9 Abril às 09:13, 2005'. The page features a navigation menu on the left with a 'Tabela' button and a list of years from 2005 to 2013. The main content area includes a header with 'VivaPraia - WAP' and 'VivaPraia - SMS', a 'NOTÍCIAS' section with a cartoon character, and several article thumbnails with titles like 'A qualidade das zonas balneares', 'O que é o VivaPraia', 'Como se deve utilizar a praia?', 'Top Praia', 'InfoViva', 'Fórum', 'Bandeira Azul', 'Links >', 'Aprender com o VivaPraia', 'Colabore conosco', 'Queres saber mais sobre a praia?', 'A praia mais próxima de si', 'Escolha o concelho', and 'Qualidade da sua Praia'. A footer contains logos for 'INATOR', 'INSTITUTO DA A.G.A.', 'Instituto do Ambiente', and 'Produzido por: VIBRAM'.

vivapraia.com

2021@Arquivo.pt

The screenshot shows the website 'ARQUIVO.PT' for 'vivapraia.com' dated '12 Janeiro às 14:02, 2021'. The page features a navigation menu on the left with a 'Tabela' button and a list of years from 2013 to 2021. The main content area includes a header with 'ROVAN 青岛润方机械有限公司' and a phone number '400-0639-5'. Below the header is a navigation bar with links: '网站首页', '公司简介', '产品展示', '新闻资讯', '施工案例', '荣誉资质', '常见问题', and '联系我们'. A search bar is present with the text '请输入你要查询的关键词' and a '搜索' button. The main content area shows a large image of workers in a factory setting. The footer contains the date '12 Janeiro às 14:02'.

revisionista.pt



Revisionista.pt

Des-cobre as notícias



Dezenas de artigos noticiosos são discretamente **editados**, **atualizados** ou **corrigidos** diariamente

Por vezes **sem qualquer referência** à versão original ou ao **que foi alterado**.

The goal of **Revisionista.pt** is revealing these edits.

Futebol – Liga nega demissão de Proença (dn.pt)

26 de Setembro de 2017, 20:55 Correção Não Visível Comparar

Futebol - Liga nega demissão de Proença ~~mas admite eleições antecipadas~~ que continua até 2019

mas admite eleições antecipadas	<u>que continua até 2019</u>
Mas divergências sobre alteração de estatutos podem levar a eleições antecipadas	<u>havia</u>
há	<u>podia</u>
pede	<u>uma proposta estatutária do Paços de Ferreira mas o clube da capital do móvel acabou por retirar</u>

8

~~Futebol~~ – Liga nega demissão de Proença ~~mas admite eleições antecipadas~~ que continua até 2019

Arquivada em Mon, 25 Sep 2017

17:21:32 GMT

O texto adicionado surge destacado a verde.

O texto ~~removido~~ surge destacado a vermelho.

"Proença nem se demitiu nem ameaçou demitir-se", diz assessor da Liga de clubes ao DN

António Barroso, assessor da Liga, esclarece ao DN que Pedro Proença nunca pensou sair do organismo. ~~Mas divergências sobre alteração de estatutos podem levar a eleições antecipadas~~

A demissão de Pedro Proença da Liga de Clubes, avançada esta tarde pelo Jornal de Notícias, foi negada pela assessoria de imprensa do organismo ao DN. "O presidente da Liga não apresentou a

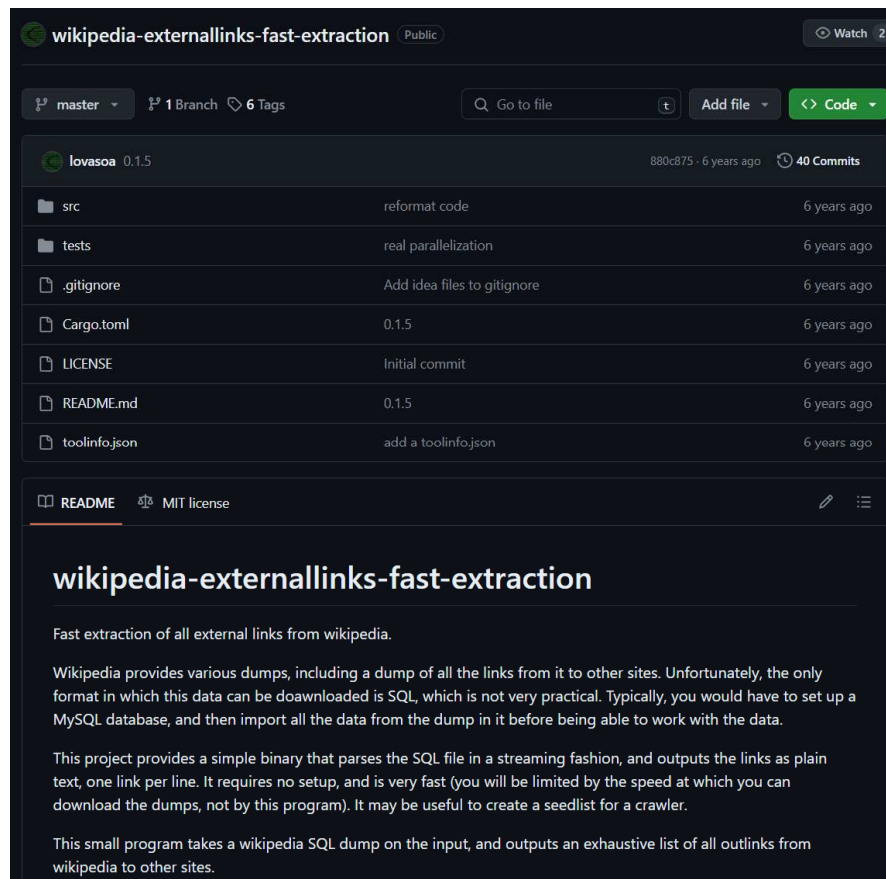
Arquivo.pt lançou um projeto para preservar as referências online

1. Arquivo.pt preservou as páginas referenciadas

- A Wikipédia portuguesa contém cerca de **1 milhão de artigos** e **em média são editadas 140 páginas por dia**.
- É necessário um método automático para obter toda a informação de forma periódica.



github.com/lovasoa/wikipedia-externallinks-fast-extraction



The screenshot shows the GitHub repository page for 'wikipedia-externallinks-fast-extraction'. The repository is public and has 2 watchers. It is currently on the 'master' branch, with 1 branch and 6 tags. The repository was last updated 6 years ago (commit 880c875). There are 40 commits in total.

File	Commit Message	Time
src	reformat code	6 years ago
tests	real parallelization	6 years ago
.gitignore	Add idea files to gitignore	6 years ago
Cargo.toml	0.1.5	6 years ago
LICENSE	Initial commit	6 years ago
README.md	0.1.5	6 years ago
toolinfo.json	add a toolinfo.json	6 years ago

[README](#) [MIT license](#)

wikipedia-externallinks-fast-extraction

Fast extraction of all external links from wikipedia.

Wikipedia provides various dumps, including a dump of all the links from it to other sites. Unfortunately, the only format in which this data can be doawnloaded is SQL, which is not very practical. Typically, you would have to set up a MySQL database, and then import all the data from the dump in it before being able to work with the data.

This project provides a simple binary that parses the SQL file in a streaming fashion, and outputs the links as plain text, one link per line. It requires no setup, and is very fast (you will be limited by the speed at which you can download the dumps, not by this program). It may be useful to create a seedlist for a crawler.

This small program takes a wikipedia SQL dump on the input, and outputs an exhaustive list of all outlinks from wikipedia to other sites.

dumps.wikimedia.org

Wikimedia Downloads

If you are reading this on Wikimedia servers, please note that we have rate limited downloaders and we are capping the number of per-ip connections to 2. This will help to ensure that everyone can access the files with reasonable download times. Clients that try to evade these limits may be blocked. Our mirror sites do not have this cap.

Data downloads

The Wikimedia Foundation is requesting help to ensure that as many copies as possible are available of all Wikimedia database dumps. Please **volunteer to host a mirror** if you have access to sufficient storage and bandwidth.

Database backup dumps

A complete copy of all Wikimedia wikis, in the form of wikitext source and metadata embedded in XML. A number of raw database tables in SQL form are also available.


These snapshots are provided at the very least monthly and usually twice a month. If you are a regular user of these dumps, please consider subscribing to [xmldatadumps-l](#) for regular updates.

dumps.wikimedia.org/ptwiki/latest

File Name	Date	Time	Size
ptwiki-latest-abstract2.xml.gz	04-Apr-2024	03:38	22988644
ptwiki-latest-abstract2.xml.gz-rss.xml	04-Apr-2024	03:39	763
ptwiki-latest-abstract3.xml.gz	04-Apr-2024	03:27	19717616
ptwiki-latest-abstract3.xml.gz-rss.xml	04-Apr-2024	03:39	763
ptwiki-latest-abstract4.xml.gz	04-Apr-2024	03:23	17392620
ptwiki-latest-abstract4.xml.gz-rss.xml	04-Apr-2024	03:39	763
ptwiki-latest-abstract5.xml.gz	04-Apr-2024	03:39	21336707
ptwiki-latest-abstract5.xml.gz-rss.xml	04-Apr-2024	03:39	763
ptwiki-latest-abstract6.xml.gz	04-Apr-2024	03:37	23938175
ptwiki-latest-abstract6.xml.gz-rss.xml	04-Apr-2024	03:39	763
ptwiki-latest-all-titles-in-ns0.gz	04-Apr-2024	02:41	11459775
ptwiki-latest-all-titles-in-ns0.gz-rss.xml	04-Apr-2024	02:41	775
ptwiki-latest-all-titles.gz	04-Apr-2024	02:41	32081413
ptwiki-latest-all-titles.gz-rss.xml	04-Apr-2024	02:41	754
ptwiki-latest-babel.sql.gz	01-Apr-2024	16:15	65221
ptwiki-latest-babel.sql.gz-rss.xml	04-Apr-2024	02:41	751
ptwiki-latest-category.sql.gz	01-Apr-2024	16:10	4851843
ptwiki-latest-category.sql.gz-rss.xml	04-Apr-2024	02:40	760
ptwiki-latest-categorylinks.sql.gz	01-Apr-2024	16:05	233582130
ptwiki-latest-categorylinks.sql.gz-rss.xml	04-Apr-2024	02:40	775
ptwiki-latest-change_tag.sql.gz	01-Apr-2024	16:14	172249288
ptwiki-latest-change_tag.sql.gz-rss.xml	04-Apr-2024	02:40	766
ptwiki-latest-change_tag_def.sql.gz	01-Apr-2024	16:14	3488
ptwiki-latest-change_tag_def.sql.gz-rss.xml	04-Apr-2024	02:40	778
ptwiki-latest-externallinks.sql.gz	01-Apr-2024	16:08	361912282
ptwiki-latest-externallinks.sql.gz-rss.xml	04-Apr-2024	02:40	775

Obtivemos **14 milhões** de ligações

Recolha dos Link Externos da Wikipédia

- **Nome:** EAWP42
 - **Dia de recolha:** 15 de Fevereiro de 2023
 - **Número de ligações:** 14 milhões de URLs
 - **Total de ficheiros:** 12 milhões de ficheiros
 - **Total espaço em disco:** 856 GB
- 

Será que podemos atualizar automaticamente as citações da Wikipédia?



2. Consertar automaticamente os links quebrados

- Existem **robôs de software** que adicionam automaticamente ligações para versões arquivadas se o link estiver quebrado:
 - [Pywikibot](#)
 - [Wayback Medic](#)
 - [InternetArchiveBot](#)



Problemas

- É necessário diversas permissões para fazer alterações de páginas da wikipédia.
- O único projeto com atualizações frequentes é o InternetArchiveBot. Contudo, é um projeto otimizado para infraestrutura do Internet Archive e para a Wikipédia EN.

Problemas

Operation status [\[edit \]](#)

For the most up to date information see [the run pages](#) or [Wiki Operations Summary on Airtable](#)

- ● InternetArchiveBot is currently running on 300+ Wikimedia wikis.
- ● We have moved the management interface to a new server. Please start using iabot.wmcloud.org instead of iabot.toolforge.org. Please [let us know](#) if anything broke during this process.
- ● Testing is stalled on Alemannisch Wikipedia (als), Asturian Wikipedia (ast), and Japanese Wikipedia (ja).
- ● Bot is approved but disabled indefinitely pending software improvements on French Wikipedia (fr), MediaWiki.org, Norwegian Nynorsk Wikipedia (nn), Polish Wikipedia (pl), and Portuguese Wikipedia (pt).

Last updated: 20:54, 5 November 2023 (UTC)

— · — · —

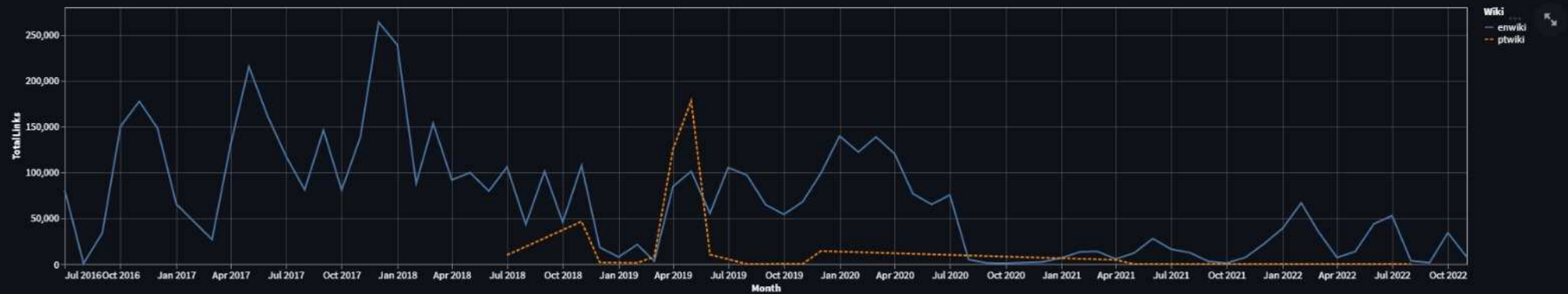
InternetArchiveBot

Monthly Links Edits on Selected Wikis

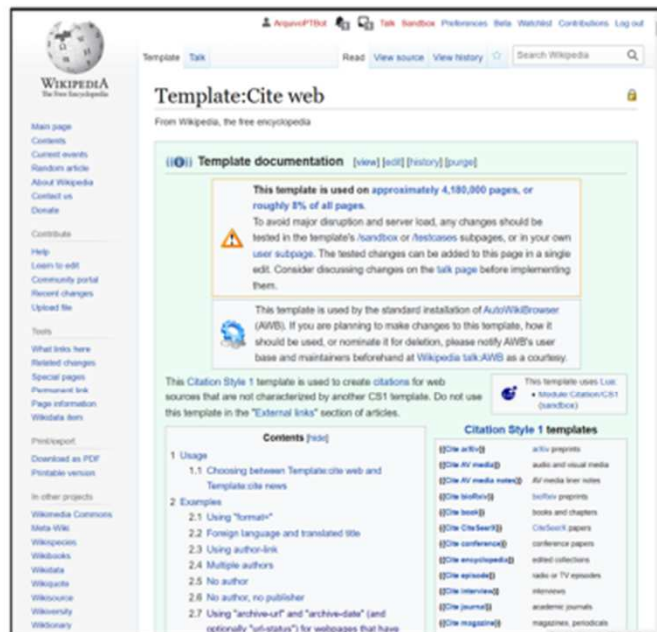
Select Wikis to compare:

enwiki × ptwiki ×

TotalLinks TotalEdits DeadLinks DeadEdits Data



Como citar Arquivo da Web nos links da Wikipédia



Using "archive-url" and "archive-date" (and optionally "url-status") for webpages that have been archived

By default, if "archive-url" is used, `|url-status=dead` and the main link is to the archive:

- `{{cite web |url=http://www.incb.org/pdf/e/list/green.pdf |title=List of psychotropic substances under international control |date=2005-04-30 |publisher=International Narcotics Control Board |access-date=2022-02-20 |archive-url=https://web.archive.org/web/20050907150136/http://www.incb.org/pdf/e/list/gr |archive-date=2005-09-07}}`

Electronic

Artigo Discussão

Ler Editar Ver histórico Ferramentas ▼

Origem: Wikipédia, a enciclopédia livre.

Electronic é um álbum de remixes da banda Madredeus, lançado em 2002. Atingiu a certificação de Disco de Ouro, por vendas superiores a 20 mil cópias em Portugal.^[1]

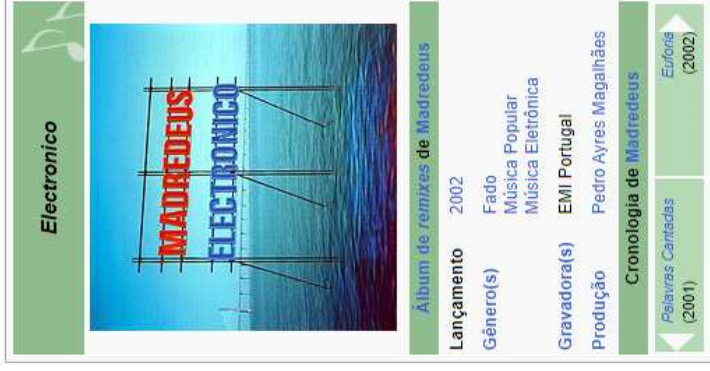
Este é um disco remisturado por diversos produtores, músicos e DJs de diversos pontos do mundo. Entre eles estão **Craig Armstrong**, **Telepopmusik**, Banzai Republic, Sunday Best e ainda Raiph Myerz.^{[2][3]}

Faixas

- "Haja O Que Houver" [Lux Mix] - 5:08
- "Vem (Além de Toda a Solidão)" [Alpha Remix] - 3:13
- "Ecos Na Catedral" [Sunday Best Remix] - 4:13
- "O Paraíso" [Afro Buscemi Mix] - 5:32
- "O Mar" [Moana Maru Remix] - 7:33
- "O Sonho" - 5:32
- "A Lira/Solidão No Oceano" [Afterlife Mix] - 5:32
- "A Andorinha da Primavera" - 5:37
- "Oxalá" [Telepopmix] - 5:18
- "Ao Longe O Mar" [Astro Reflect e Chill Mix] - 8:13
- "Ainda Insect" [Harmonic Mix] - 6:49
- "Anseio (Fuga Apressada) - Anseio 2" [Craig Armstrong Remix] - 4:38
- "Guitarra" [Manitoba Remix] - 8:04

Referências

- ↑ PÚBLICO. «"Madrdeus Electronic" é disco de ouro». PÚBLICO. Consultado em 14 de fevereiro de 2023
- ↑ Antunes, Marisa Caetano (14 de junho de 2002). «"Madrdeus Electronic"». SIC Online. Consultado em 14 de fevereiro de 2023. Arquivado do original em 25 de junho de 2002
- ↑ Guedes, Nuno Miguel (13 de junho de 2002). «VERSÕES - Electricidade de Portugal». Visão. Consultado em 14 de fevereiro de 2023. Arquivado do original em 27 de setembro de 2002



Parâmetro *url* e *arquivourl*

A editar Electronico

Adicionar línguas

Artigo Discussão

Ler Editar Editar código-fonte Ver histórico

- Não insira textos protegidos por [direitos de autor](#), pois serão removidos. Garanta a [verificabilidade](#) das informações: [cite](#) sempre [fontes confiáveis](#).
- Escreva em português formal sem alterar as [versões do idioma](#).
- Saiba [o que não fazer](#) e evite [erros comuns ao editar](#). Veja também as nossas [políticas e recomendações](#) e a [página de ajuda](#).

```
}}

''''''Electronico'''''' é um [[álbum]] de remixes da banda [[Madredeus]], lançado em [[2002]]. Atingiu a certificação de Disco de Ouro, por vendas superiores a 20 mil cópias em Portugal.<ref>{{Citar web|ultimo=PÚBLICO|url=https://www.publico.pt/2002/09/25/culturaipsilon/noticia/madredeus-electronico-e-disco-de-ouro-180044|titulo="Madredeus Electronico" é disco de ouro|acessodata=2023-02-14|website=PÚBLICO|língua=pt}}</ref>

Este é um disco remisturado por diversos produtores, músicos e [[DJ]]s de diversos pontos do mundo. Entre eles estão [[Craig Armstrong]], [[Telepopmusik]], Banzai Republic, Sunday Best e ainda Ralph Myerz.<ref>{{Citar web|ultimo=Antunes|primeiro=Marisa Caetano|url=http://www.sic.pt/article1665visual4.html|titulo="Madredeus Electronico"|data=14-06-2002|acessodata=2023-02-14|website=SIC Online|arquivourl=https://arquivo.pt/wayback/20020625051351/http://www.sic.pt/article1665visual4.html|arquivodata=2002-06-25|urlmorta=sim}}</ref><ref>{{Citar web|ultimo=Guedes|primeiro=Nuno
```

Conclusões

- Wikipédia é um dos sites mais confiáveis e com mais informação do mundo.
 - Nem tudo são histórias felizes, mas conseguimos extrair sempre algo bom.
- 