

Search the Past Web with Arquivo.pt

Catalog of Arquivo.pt services <u>daniel.gomes@fccn.pt</u>





ARQUIVO.PT

Who are we?



Free online service to research the Past Web Preserves **publicly accessible** information related with:

- Portugal
- **Research** and **Education** (international)

Governmental service provided by Foundation for Science and Technology (Portugal) A digital research infrastructure

Arquivo.pt begun in 2007

Menu	ARQUIVO.PT	Opções
	arquivo-web.fccn.pt/portuguese-web-archive?set_language=en 16 Março às 04:52, 2008	
I Tabela		Site Map Accessibility Contact
		Search
		only in current section
0000	Home Crawler Team	
2008	You are here: Home	English Português
Março 💽		
16 Março às 04:52	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 information was published. However, it was never so ephemeral. Web documents such as news, blogs or discussion forums a 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 GINTER Archive collects and stores contents from the world-wide web. However, it is difficult for a single organization	re valuable descriptions of our
	while satisfying all needs, because the web is permanently changing and many contents disappear before they can be archiv	
	As a result, several countries are creating their own national archives to ensure the preservation of contents of historical rele	evance to their cultures.
	Portugal is now beginning its national web archiving initiative with the Tomba project at 🔒 FCCN (National Foundation for Scie	entific Computing). 🔒 Send this Print this
	Fundação para a Computação Científica Nacional	
	The Plone [®] CMS — Open Source Content Management System is <u>@</u> 2000-2008 by the Plone Foundation Plone [®] and the Plone logo are registered trademarks of the Plone Foundation. Distributed under the GNU G	
	Powered by Plone Valid XHTML Valid CSS Section 508 WCAG	

Arquivo.pt is used word-wide



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. 풍 United Kingdom	2,231	2.22%
6. 🦲 Japan	2,172	2.16%
7. 💽 Canada	1,237	1.23%
8. 🔚 Mozambique	1,213	1.20%
9. 💶 India	902	0.90%
10. 🥅 Germany	894	0.89%

- 53% of users are international
- User Interfaces and documentation also in English
- Combined with Google Translate enables cross-lingual access to preserved content

Arquivo.pt preserves national and international historical web content

GALII	LEO Versão 2.0	
Janeiro 1993		
	Manual de utilização e sugestões de exploração	<i>AUTORES</i> Programação: João Veloso Manual: Elisa Prata Pina e M. Augusta Patrício Orientação: Carlos Fiolhais
ÍNDICE	Departamen	to de Física da Universidade de Coimbra
2. Cara 3. Cont	sentação sumária do programa cterísticas do equipamento ũguração mínima do computador eiros do programa	

(oldest page)



spacelink.nasa.gov – 1992 (oldest image)



How to **search** and **access** the preserved information?

Search texts from the past in any language

M	enu ARQUIVO.PT)	Options
	Q Ambasada RP	Search	n
	1996 1 Jan		2021 12Nov
	Pages Images Narrative	Advanced s	search
	→ berlin.msz.gov.pl/pl		
	Ambasada RP w Berlinie		
	30 March 2019		
	Ambasada RP w Berlinie Masz wyłączoną obsługę języka JavaScript Polski Deutsch RSS Wersja tekstowa Być wiernym Ojczyźnie Rzeczypospolitej Polskiej Aktualności Ambasada Współpraca dwustronna się zwrócić? Zasady realizacji projektów polonijnych Ambasada Ambasada Ambasada RP w		
	→ dublin.msz.gov.pl/pl		
	Ambasada RP w Dublinie		
	10 June 2017		
	Ambasada RP w Dublinie Masz wyłączoną obsługę języka JavaScript English Polski RSS Wersja tekstowa Polsce – służyć, Europę Świat rozumiać Aktuelności Ambasada Wanólarzeg dwystrenne – WYDZIAŁ EM KONSULARNYM – Ambasada – Ambasada PD w		

Świat – rozumieć Aktualności **Ambasada** Współpraca dwustronna ... WYDZIAŁEM KONSULARNYM **Ambasada Ambasada RP** w Dublinie 5 Ailesbury Road Ballsbridge D04 W221 Dublin ...

Search the **history** of a web address

Menu						ARQU	IVO.PT							
	Q bbc.co.uk										Search	Search		
	1996 1 Jan											2021 12Nov		
	Pag		Narrative								Advanced se			
	Cerca	a de 2.538 resu	Itados desde 19	96 a 2021										
•	00100					^Ⅲ Table	≣ List						Þ	
2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	
15 Feb	4 May	29 May	4 Jan	21 Jan	10 Jan	5 Sep	1 Jan	1 Jan	20 Jan	1 Jan	1 Jan	1 Jan	3 Jan	
14 Mar	20 May	30 May	6 Jan	27 Jan	11 Jan	7 Sep	2 Jan	2 Jan	9 Feb	1 Jan	2 Jan	1 Jan	3 Jan	
22 Oct	20 May	4 Aug	12 Jan	31 Jan	12 Jan	23 Sep	3 Jan	4 Jan	10 Feb	2 Jan	3 Jan	2 Jan	3 Jan	
22 Oct	11 Jun		12 Jan	11 Mar	15 Jan	24 Sep	4 Jan	8 Jan	3 Apr	2 Jan	4 Jan	3 Jan	4 Jan	
23 Oct	22 Jun		14 Jan	12 Mar	17 Jan	26 Sep	5 Jan	8 Jan	4 Apr	2 Jan	5 Jan	4 Jan	4 Jan	
	6 Aug		21 Jan	13 Mar	18 Jan	27 Sep	6 Jan	17 Jan	5 Apr	3 Jan	6 Jan	5 Jan	4 Jan	
	6 Aug		21 May	14 Mar	19 Jan	28 Sep	8 Jan	17 Jan	6 Apr	3 Jan	7 Jan	6 Jan	6 Jan	
	3 Sep		1 Jul	15 Mar	22 Jan	30 Sep	10 Jan	21 Jan	7 Apr	3 Jan	8 Jan	7 Jan	6 Jan	
	3 Sep		1 Dez	16 Mar	23 Jan	4 Oct	10 Jan	24 Jan	8 Apr	3 Jan	9 Jan	8 Jan	6 Jan	
	4 Sep		2 Dez	17 Mar	26 Jan	5 Oct	12 Jan	28 Jan	9 Apr	3 Jan	10 Jan	9 Jan	6 Jan	
	8 Sep			18 Mar	29 Jan	6 Oct	17 Jan	30 Jan	10 Apr	3 Jan	11 Jan	10 Jan	8 Jan	

Browse <u>bbc.co.uk</u> 10 years ago





Search **images** from the past





S Visit

Title: Warsaw city centre

Alt text: Warsaw city centre

Caption: Poland's education system has seen some extraordinary changes in recent years. Following a huge drive in the nation's Katowice =6 1001+ Wrocław University of Science and Technology Wrocław Read mor... URL: timeshighereducation.com/sites/default/files/styles/the_breaking_news_image_style/public/warsaw_city_c entre.jpg?tick=NYFWGvNj Resolution: jpg 620: x413 Capture date: 7 June 2019 at 02:28

Page

Title: Best universities in Poland | Times Higher Education (THE)

URL: timeshighereducation.com/student/best-universities/best-universities-poland Capture date: 7 June 2019 at 02:28

Application **P**rogramming Interfaces to develop new applications



Arquivo.pt APIs

APIs

Daniel Gomes edited this page 11 days ago · 33 revisions

APIs specific to Arquivo.pt that enable the full exploration of its functions

- Arquivo.pt API (Full-text & URL search)
- Image Search API v1.1 (beta version)

APIs based on international standards to enable interoperability among web archives and code reuse

- CDX-server API (URL search)
- Memento API (URL search)



Arquivo.pt API in action: TellMeStories.pt

Home Search	About	Português
tell me stories Create a news narrative about any subject		
Q Barack Obama		
advanced options		
	Narrative	Related terms
Narrative about Barack Obama during the last 10 years		
in the second se		P
2012 2013 2014 2015 2016 2017 10/2011 03/2013 01/2015 11/2016 12/2019 12/2019	2018	2019
 Relevant news between 03/01/2011 and 03/10/2011 Fotos: Barack Obama chega ao Havai para as férias de Natal Barack Obama decreta estado de emergência sanitária nos EUA 		
 Barack Obama faz visita surpresa ao Afeganistão Dilma Rousseff vai receber Barack Obama a 19 de Março 		
• Barack Obama recandidata-se a presidente dos Estados Unidos • Barack Obama anuncia recandidatura à Casa Branca • Barack Obama confirma morte de Bin Laden		
 Barack Obarna comma morte de Bin Laden Barack Obarna não vai revelar imagens do corpo de Bin Laden Barack Obarna defende investigação sobre ligações de Bin Laden ao Paquistão 		
 Barack Obama decende investigição sobre ingrições de bin fader do raquistão Barack Obama hoje na coroação de Souto de Moura em Washington Barack e Michelle Obama em visita oficial a Inglaterra 		
* Barack Obama recebe príncipe Carlos na Casa Branca		

Automatically generates narratives about any subject based on online news from the Past.

Winners of the Arquivo.pt Award 2018.

Learn the API through the arquivo.pt/tutorial





Collections and automatic selection

Public list of Arquivo.pt collections

List of the collections preserved by Arquivo.pt (publicly available) 🛛 🖈 🖻 📀

⊞

File Edit View Insert Format Data Tools Add-ons Help Last edit was made on 10 September by SAW - FCCN Serviços e Arquivos Web

A2 • fx =IMPORTRANGE("https://docs.google.com/spreadsheets/d/1hfLyq9N2ZJEd1hra2V00PI24jTy42ByR6M667rG-oHo/edit", "Summary!A2:G")

	A	В	С	D	E	F	G	н
1	Collection ID	Description	Collection Start date	Collection End date	Total files (Number of URLs collected)	Total seeds (Preserved sites)	Total volume of WARCS files (TB)	Collection in Production? (URL, Page and Image indexed)
2	AWP1	1st Complete crawl of the Portuguese web, mainly from the .PT domain, in 2008.	2008-02-12	2008-03-06	56,046,288	154,787	1.60	
3	AWP2	2nd Complete crawl of the Portuguese web, mainly from the .PT domain, in 2008.	2008-03-11	2008-05-30	48,718,404	-	1.60	
4	AWP3	3rd Complete crawl of the Portuguese web, mainly from the .PT domain, in 2008.	2008-10-21	2008-12-10	51,863,006	193,294	2.00	
5	AWP4	4th Complete crawl of the Portuguese web, mainly from the .PT domain, in 2009.	2009-05-01	2009-05-31	68,776,707	366,880	2.50	
6	AWP5	5th Complete crawl of the Portuguese web, mainly from the .PT domain, in 2009.	2009-10-01	2009-10-31	119,135,566	373,323	3.80	
7	AWP6	6th Complete crawl of the Portuguese web, mainly from the .PT domain, in 2009.	2009-12-01	2009-12-31	118,810,364	340,018	3.50	
8	AWP7	7th Complete crawl of the Portuguese web, mainly from the .PT domain, in 2010.	2010-05-01	2010-05-31	87,988,812	389,957	2.90	
9	AWP8	Incremental crawl of the Portuguese web web, mainly from the .PT domain, in 2010. The AWP8 crawl is incremental because it was performed using DeDuplicator (http://landsbokasafn.github.io/DeDup licator/) taking the content of AWP7 as baseline. Thus, the files that remained unchanged from the AWP7 complete crawl were not archived (duplicated) on the AWP8 incremental crawl.	2010-08-01	2010-08-31	75,771,317	411,562	1.90	

arquivo.pt/collections

Experiments to preserve European Union web content





Horizon 2020 European Union funding for Research & Innovation

Developed a low-cost methodology to automatically **select** and preserve online information about given topics

Preserving Websites Of Research & Development Projects

Daniel Bicho Foundation for Science and Technology: Arquivo.pt Av. do Brasil, 101 1700-066 Lisboa, Portugal daniel.bicho@fccn.pt Daniel Gomes Foundation for Science and Technology: Arquivo.pt Av. do Brasil, 101 1700-066 Lisboa, Portugal daniel.gomes@fccn.pt

International Conference on Digital Preservation 2016, available at arquivo.pt/publications

Existing open datasets are complemented with links to URLs and web archives

title	endDate	projectUrl	totalCost	URLsBingSearch	ArchivedProjectURLs
MARine MAnagement and Ecosystem Dynamics under climate change	2019-09-30	https://www.marmaed.uio.no/	4073903,28		https://arquivo.pt/wayback/https://www.marmaed.ui o.no/
Services to enhance innovation management capacities of SMEs in England, Northern Ireland and Wales 2015 -2016	2016-12-31	http://www.enterprise-europe.co.uk/	2336311,17		https://arquivo.pt/wayback/http://www.enterprise- europe.co.uk/
Developing novel tools and technologies to assess the				https://www.findaphd.com/phds/project/assess-the-safety-and-efficacy-of-renal-regenerative- medicine-therapies-using-advanced-imaging-technologies-renaltoolbox-project/?p101521, https://www.liverpool.ac.uk/mathematical-sciences/staff/rachel-bearon/research/, https://www.liverpool.ac.uk/veterinary-science/staff/lorenzo-ressel/research/, https://www.cyanagen.com/grants/, https://umexpert.um.edu.my/selvee.html, https://iroffice.msu.ac.th/th/index.php/scholarships/128-european-funding-and-fellowship- opportunities, https://www.ukrmp.org.uk/reports_and_papers/imaging-technologies-for-monitoring- the-safety-efficacy-and-mechanisms-of-action-of-cell-based-regenerative-medicine-therapies-in- models-of-kidney-disease/, https://onlinelibrary.wiley.com/doi/full/10.1002/jev2.12033, https://www.autocrat.eu/author/admin/, https://translational- medicine.biomedcentral.com/articles/10.1186/s12967-019-1994-0, https://onlinelibrary.wiley.com/doi/full-xml/10.1002/jev2.12033, https://renaltoolbox.org/imaging- technologies-to-evaluate-safety-and-efficacy/, https://www.findaphd.com/phds/project/developing- mathematical-tools-for-image-analysis-to-evaluate-the-effects-of-regenerative-medicine-therapies- on-kidney-function-renaltoolbox-project/?p101518, https://en.unito.it/research/research-projects-and- initiatives/international-research-projects/marie-sklodowska-curie, https://region8.healthresearch.ph/index.php/11-announcement/165-europe-asean-research-funding- and-fellowship-opportunities, https://www.mdpi.com/journal/jcm/special_issues/Osteoarthritis_Management_and_Regenerative_St rategies, https://www.omicsonline.org/, https://www.omicsonline.org/scholarly/methods-and- techniques-in-molecular-biology-journals-articles-ppts-list.php, https://royalsocietypublishing.org/doi/10.1098/rstb.2017.0229,	https://arquivo.pt/wayback/https://www.findaphd.co m/phds/project/assess-the-safety-and-efficacy-of- renal-regenerative-medicine-therapies-using- advanced-imaging-technologies-renaltoolbox- project/?p101521, https://arquivo.pt/wayback/https://www.liverpool.ac .uk/mathematical-sciences/staff/rachel- bearon/research/, https://arquivo.pt/wayback/https://www.liverpool.ac .uk/veterinary-science/staff/lorenzo- ressel/research/, https://arquivo.pt/wayback/https://www.cyanagen.c om/grants/, https://arquivo.pt/wayback/https://umexpert.um.ed u.my/selvee.html, https://arquivo.pt/wayback/https://iroffice.msu.ac.th /th/index.php/scholarships/128-european-funding- and-fellowship-opportunities, https://arquivo.pt/wayback/https://www.ukrmp.org.u k/reports_and_papers/imaging-technologies-for- monitoring-the-safety-efficacy-and-mechanisms- of-action-of-cell-based-regenerative-medicine-
safety and efficacy of cell-based regenerative medicine				https://uctranslationalmedicine.org/mtm-final-symposium-2021/,	therapies-in-models-of-kidney-disease/,
therapies, focusing on kidney disease	2022-10-31			https://www.nature.com/articles/s41536-021-00154-y, https://europepmc.org/article/MED/32292894	https://arquivo.pt/wayback/https://onlinelibrary.wile



Cordis-h2020organizationsComplementedByArquivoPT.xlsx

Dataset contendo a listagem das organizações dos projetos financiados pelo programa H2020 (Horizon 2020 - European Commission



Open datasets about thematic collections







arquivo.pt/dadosabertos



Tools to support web preservation

Suggest websites to be preserved

- Public form
- Any domain
- Any language
- Submit home page
- arquivo.pt/suggest

Suggest websites to be preserved

International websites are welcome!

sawfccn@gmail.com Switch accounts	ć
*Required	
Email *	
Your email address	
Website (1 per line) *	
Your answer	
Description of the Websites	

SavePageNow: Archive a webpage immediately

■ SavePageNow Record	d webpages on Arquivo.pt	
Page address		Record
e.g. www.fccn.pt		
The SavePageNow service	e allows a page to be archived at the exact m	oment the user makes the request.
The archived contents are	later integrated in the Arquivo.pt collection.	
Learn more		



Missing content? Complete the page!





Missing files obtained from the live-web/archives are then integrated to improve the quality of the archived page

Collaborative curation: Give it a try!

Don't kill your historical website! Preserve it in the **Arquivo.pt Memorial**: <u>arquivo.pt/memorial</u>



arquivo.pt/arquivo404 fix the broken links in your website!



Page not found at live website



Explore the Past Web

Create Web Design museums



WebDesignMuseum.org

Online exhibitions of web-archived documents

Memória dos Museus na Web

Websites dos museus portugueses (1996-2021)

Lista de museus em Portugal Sobre Home



Ciência



Museu Nacional de História Natural e da



Museu Nacional de Arte Contemporânea - Museu do Chiado



Museu Nacional dos Coches



Museu Nacional Soares

arquivo.pt/expos



Museu Nacional Machado de Castro

Memória dos Museus na Web

Websites dos museus portugueses (1996-2021)

Lista de museus em Portugal Sobre Home

Museu Nacional de Arte Antiga



Página de entrada do website do Museu Nacional de Arte Antiga (museudearteantiga.pt, 2013)

Memória do sítio Web

As primeiras páginas Web do Museu Nacional de Arte Antiga (MNAA) remontam a 1997, integradas no portal do Ministério da Cultura, na secção do Instituto Português de Museus.

dos Reis

Celebrate anniversaries



arquivo.pt/indielisboa

Custom re-search systems



ARQUIVO.PT collection:EAWP19 מוות Q Search 6 Au Narrative Pages Images Advanced search → Did you mean: מוות collection:ep19 About 14.111 results from 1991 to 2023 israblog.nana10.co.il/blogread.asp?blog=849548&year=2015&month=1 מוות וקוקייה 18 December 2018 מוות וקוקייה התחברות לבלוג צור בלוג נוסף לקבועים בבלוג זה 🛛 בכל הבלוגים 🛛 שם בלוג/בלוגר חפש ... 🕇 לפתיחת בלוג חדש מוות וקוקייה תובנות מעיניני דיומא כינוי: מוות וקוקייה תמונה מלאו כאן את כתובת ... מפורסמים נכתב על ידי מוות וקוקייה , 31/1/2015 00:21 בקטגוריות שפמים , פסימי , שירי ילדים , תוכים הצג israblog.nana10.co.il/blogread.asp?blog=849548&catcode=208582&catdesc=%C3%A9%C3%B9%C3%A5 מוות וקוקייה 31 December 2018 מוות וקוקייה התחברות לבלוג צור בלוג נוסף לקבועים בבלוג זה 🛛 בכל הבלוגים 🛛 שם בלוג/בלוגר חפש ... לפתיחת בלוג חדש מוות וקוקייה תובנות מעיניני דיומא כינוי: 10/9/2015 , א לקטעים בבלוגים אחרים בקטגוריה זו לחצו . כן לא שחור לבן ביי נכתב על ידי מוות וקוקייה (10/9/2015 - מוות וקוקייה (10/9/2015 - מוות וקוקייה) או העובריה זו לחצו . כן א שחור לבן ביי נכתב על ידי מוות וקוקייה (10/9/2015 - מוות וקוקייה) או העובריה מוות וקוקייה (10/9/2015 - מוות וקוקייה) או העובריה מוות וקוקייה (10/9/2015 - מוות וקוקייה) או העובריה מוות ו www.youtube.com/watch?v=eLoXtUGV2dY&feature=applinks YouTube - מוות קליני - הסיפור האמיתי 13 November 2018 - YouTube Ignorar navegação RU Iniciar sessão Pesquisar A carregar ... parar A carregar... Fila de espera de visualizações alin קליני - הסיפור האמיתי ... 20/06/2012 בשנים האחרונות רבים הדיווחים על חוויות של אנשים שחוו מוות קליני וטוענים שנשמתם יצאה

.. מהגוף

arquivo.pt/israblog

Meet the Winners of the Arquivo.pt **Annual** Awards





Support of the President of the Republic of Portugal



Attend Training courses about web preservation and research

New ways of searching the past (module A)

– Any Internet user

Publish preservable information on the web (B)

– Web authors

Automatic processing of information preserved from the Web (C)

Developers, Researchers

Do-it-yourself web archiving (D)

– Researchers, Librarians, Archivists

Videos in English also available!



arquivo.pt/training

Read this book 🙂

Daniel Gomes · Elena Demidova Jane Winters · Thomas Risse *Editors*

The Past Web

Exploring Web Archives

Description Springer

Preprint version available in open access at: <u>arquivo.pt/book</u>

List of web archiving initiatives

Glam			Sub	omit your program pro	posal for the (GLAM W	Viki Conference	[Help with translations!]	
Main menu [hide]	List of We	eb archivin	g initiat	ives			🕱 1 language 🗸		
Main page	Article Talk		0				Read Edit View history	Tools [hide]	
Contents	From Wikipedia, the free er	ncvclopedia						Actions	
Current events								Move	
Random article							elines. Please improve this	General	
About Wikipedia			-	r inappropriate external lir 9) (Learn how and when to rei		-	I links where appropriate into	What links here	
Contact us								Related changes	
Donate			-	dwide. For easier reading		is	State 1 to 1	Special pages	
Switch to old look		, i i i i i i i i i i i i i i i i i i i		data, and access method				Permanent link	
Contribute				sults obtained for the rese		vey		Page information	
Help	on web archiving ini	<i>tiatives</i> ⁽²⁾ published b	y the Arquivo.pt i	Ithe Portuguese web-ar	chive) team.			Cite this page	
Learn to edit	Web archivin	g initiatives	odit 1					Wikidata item	
Community portal	web aremvin	g initiatives [euitj			-		Edit	
Recent changes						Ν	Map of Web archiving initiatives worldwide in April 🛱	interlanguage	
Upload file							2021 ^[1]	links	
Languages								Print/export	
Language links are at the top of the					Numbe			Download as PDF	
page across from the title.	Name	Country	Creation	 Technologies 	Employees		Comments	Printable version	
			Year		Full-time +	Part-	+	r fillable version	
						time			
Contents [hide]							The End of Term Web Archive & captures and		
(Тор)							saves U.S. Government federal government websites (.gov, .mil, etc) in the Legislative,		
Web archiving initiatives							Executive, or Judicial branches of the		
Archived data		government at the end of presidential							
Access methods							administrations. Beginning in 2008, the EOT		
See also	End of Term Web	United States	2008	Heritrix, Wayback		6-10	has thus far preserved websites from		
References	Archive						administration changes in 2008, 2012 and		
External links							2016, and is currently gearing up for the 2020 transition. Project partners ∠ include CA		
							Digital Library, Internet Archive, Library of		
							Congress, George Washington University,		
							Stanford University University of North Tayaa		

Keep in touch!

Subscribe our channels:

- arquivo.pt/subscribe: mailing list in English
- <u>arquivo.pt/news</u>: social networks and videos