

**Dia mundial da preservação digital**

7 novembro 2024

CitationSaver preserva citações  
para conteúdos online

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



# Arquivo.pt preserva conteúdos históricos publicados online

***GALILEO*** Versão 2.0

Janeiro 1993

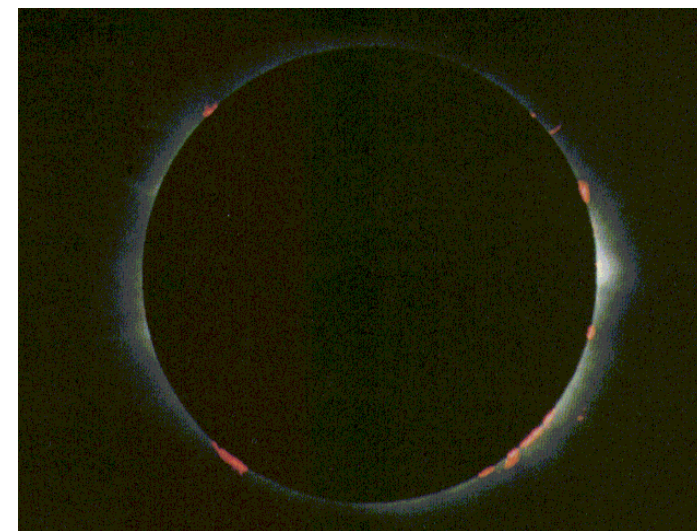
Manual de utilização e sugestões de exploração	<b><i>AUTORES</i></b> <b>Programação:</b> João Veloso <b>Manual:</b> Elisa Prata Pina e M. Augusta Patrício <b>Orientação:</b> Carlos Fiolhais
--	---

Departamento de Física da Universidade de Coimbra

***ÍNDICE***

<a href="#">1. Apresentação sumária do programa</a>
<a href="#">2. Características do equipamento</a>
<a href="#">3. Configuração mínima do computador</a>
<a href="#">4. Ficheiros do programa</a>

[nautilus.fis.uc.pt](http://nautilus.fis.uc.pt)- 1993  
(página mais antiga)



[spacelink.nasa.gov](http://spacelink.nasa.gov) – 1992  
(imagem mais antiga)

# Pesquisar texto para o passado

The screenshot shows the ARQUIVO.PT search interface. At the top, there is a dark blue header with 'Menu' on the left, 'ARQUIVO.PT' in the center, and 'Options' on the right. Below the header is a search bar containing the text 'Bibliothèque nationale de France' and a 'Search' button. To the left of the search bar is a date range selector showing '1991 6 Aug' and '2024 3 Apr' with a horizontal line and a dot in between. Below the date range are three buttons: 'Pages', 'Images', and 'Narrative'. To the right of the date range is an 'Advanced search' button. Below the search bar, it says 'About 1.385.409 results from 1991 to 2024'. There are two search results displayed, both from 'www.bnf.fr'. The first result is titled 'Bibliothèque nationale de France' and dated '3 October 2009'. The second result is also titled 'Bibliothèque nationale de France' and dated '28 June 2009'. Both results have a snippet of text below the title. At the bottom of the second result, there is a button that says 'More results from www.bnf.fr'.

Menu Options

ARQUIVO.PT

Q Bibliothèque nationale de France Search

1991 6 Aug 2024 3 Apr

Pages Images Narrative Advanced search

About 1.385.409 results from 1991 to 2024

[www.bnf.fr](http://www.bnf.fr)  
**Bibliothèque nationale de France**  
3 October 2009  
Bibliothèque nationale de France Aller au contenu Vie de la Bibliothèque Actualités culturelles ... Découvrir le site web Écrire à la BnF Bibliothèque nationale de France La Bibliothèque Actualités ... Offres d'emploi Presse Lettres d'information Crédits © Bibliothèque nationale de France, 30 ...

[www.bnf.fr](http://www.bnf.fr)  
**Bibliothèque nationale de France**  
28 June 2009  
Bibliothèque nationale de France Aller au contenu Vie de la Bibliothèque Actualités culturelles ... Découvrir le site web Écrire à la BnF Bibliothèque nationale de France La Bibliothèque Actualités ... 'information Crédits © Bibliothèque nationale de France, 25 juin 2009 ...

[More results from www.bnf.fr](#)

# Pesquisar imagens para o passado

Menu Options

ARQUIVO.PT















---

Q  Search

1991 ● 2024  
6 Aug ● 3 Apr

Pages Images Narrative Advanced search

About 12.663.508 results from 1991 to 2024

 → <a href="http://www.crl-bourgogne...">http://www.crl-bourgogne...</a> 15 September 2009	 → <a href="http://www.google.com/p...">http://www.google.com/p...</a> 23 December 2010	 → <a href="http://www.europeana-ne...">http://www.europeana-ne...</a> 22 November 2014	 → <a href="http://www.photo-passion...">http://www.photo-passion...</a> 9 July 2009	 → <a href="https://www.sgdl.org/">https://www.sgdl.org/</a> 14 September 2019	 → <a href="http://www.europeanareg...">http://www.europeanareg...</a> 26 November 2014	 → <a href="https://br.depositphotos.c...">https://br.depositphotos.c...</a> 14 January 2020
 → <a href="https://www.ekium.eu/fr/a...">https://www.ekium.eu/fr/a...</a> 25 September 2019	 → <a href="http://www.google.com/p...">http://www.google.com/p...</a> 23 December 2010	 → <a href="http://www.lexilogos.com...">http://www.lexilogos.com...</a> 9 July 2009	 → <a href="http://www.camertola.pl/...">http://www.camertola.pl/...</a> 22 January 2012	 → <a href="https://www.genesispark...">https://www.genesispark....</a> 11 September 2019	 → <a href="http://www.photo-passion...">http://www.photo-passion...</a> 9 July 2009	 → <a href="http://www.europeana-ne...">http://www.europeana-ne...</a> 22 November 2014

# Pesquisar histórico de um endereço

Menu ARQUIVO.PT

Q pnf.fr Search

1991 6 Aug 2024 3 Apr

Pages Images Narrative Advanced search

About 202 results from 1991 to 2024

← Table List →

2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
							14 Mar	28 Jun					6 Sep		9 Jan	6 Jun	6 Apr	14 Mar	12 Jan	11 Jan	19 Jan	24 Jan	
							22 Oct	2 Oct					24 Nov		15 May	20 Jun	12 Jul	14 Mar	12 Jan	11 Jan	25 Jan	24 Jan	
								3 Oct					26 Nov		20 May	3 Aug	22 Jul	18 Mar	23 Jan	14 Jan	25 Jan	12 Feb	
																	19 Sep	20 Mar	23 Jan	18 Jan	25 Jan	15 Feb	
																	27 Sep	20 Mar	1 Feb	4 Mar	25 Jan	18 Feb	
																	28 Sep	27 Mar	5 Feb	4 Mar	11 Apr	18 Feb	
																	1 Oct	27 Mar	5 Feb	30 Apr	16 Apr	21 Mar	
																	4 Oct	2 Apr	7 Feb	30 Apr	16 Apr	21 Mar	
																	5 Nov	17 Apr	14 Feb	30 Apr	16 Apr	21 Mar	
																		17 Apr	7 Mar	2 Jun	24 Apr		

# O 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%

Somos uma **fonte de dados** para a IA

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)

Temos mais de 1,000 **referências académicas**

The screenshot shows a Google Scholar search interface. The search bar contains the query "Arquivo.pt"|"Portuguese web archive". The results page is titled "Articles" and shows "Page 6 of about 1,020 results (0.05 sec)". There are two main article entries visible. The first entry is titled "Major Minors-Ontological representation of minorities by newspapers" and includes a PDF link to "uminho.pt". The second entry is titled "Automatic Classification of Stigmatizing Articles of Mental Illness: The Case of Portuguese Online Newspapers" and includes a link to "books.google.com". The interface also shows filters for "Any time", "Sort by relevance", and "Any type".

Todos os serviços:  
[arquivo.pt/catalogo](http://arquivo.pt/catalogo)

## Serviço públicos

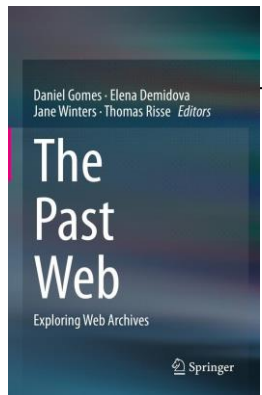
- [Pesquisa e acesso](#) sobre a informação preservada da web
- [Interfaces para Programação de Aplicações \(APIs\)](#)
- [Sugira websites](#) para serem preservados
- [SavePageNow](#): archive imediatamente páginas da web
- [Integração de coleções históricas de dados da web](#)
- [Formação](#) acerca 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](#)

Novo Serviço  
**CitationSaver**

# Problema: as **citações científicas** para recursos online podem ficar **partidas**



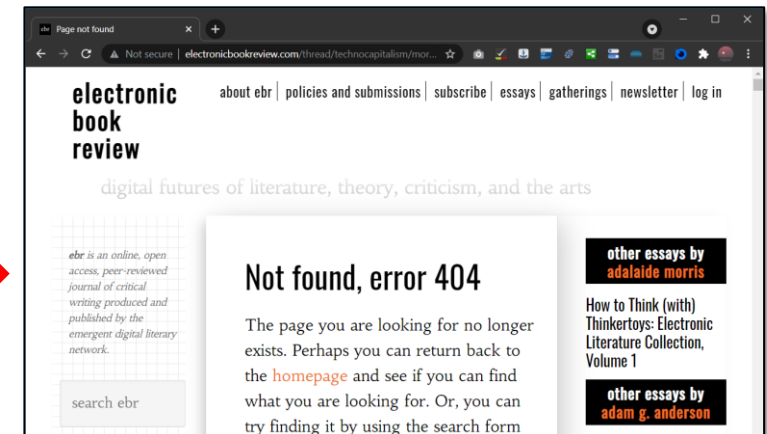
214

A. Helmond and F. van der Vlist

Helmond A, Nieborg DB, van der Vlist FN (2019) Facebook's evolution: development of a platform-as-infrastructure. *Internet Hist* 3(2):123–146. <https://doi.org/10.1080/24701475.2019.1593667>

Helmond A, van der Vlist FN (2019) Social media and platform historiography: challenges and opportunities. *TMG – J Media Hist* 22(1):6–34. <https://www.tmgonline.nl/articles/434/>

Kirschenbaum MG (2003) Virtuality and VRML: software studies after Manovich. *Electronic book review*. Retrieved from: <http://www.electronicbookreview.com/thread/technocapitalism/morememory>



As publicações perdem valor científico porque passa a ser impossível reproduzir as experiências.

# Um problema grave que afeta todas as áreas da ciência

**The FASEB Journal**  
The Journal of the Federation of American Societies for Experimental Biology  
HOME | CURRENT ISSUE | NEW ARTICLES | ARCHIVE | ALERTS | RSS | SUBMIT | HELP

**Unavailability of online supplementary scientific information from articles published in major journals**

Evangelos Evangelou<sup>\*,1</sup>, Thomas A. Trikalinos<sup>\*,†</sup> and John P.A. Ioannidis<sup>\*,†‡</sup>

+ Author Affiliations  
\_\*Correspondence: <sup>1</sup>Correspondence: Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina 45110, Greece. E-mail: me01760@cc.uoi.gr

« Previous | Next Article »  
Table of Contents

**This Article**  
doi: 10.1096/fj.05-47841sf  
December 2005  
The FASEB Journal  
vol. 19 no. 14 1943-1944

» Abstract  
Full Text  
Full Text (PDF)  
+ Classifications  
+ Services

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

**Death of web citations: a serious alarm for authors**

**The decay and failures of web references**

Full Text: [Html](#) [PDF](#)

Author: [Diomidis Spinellis](#) Athens University of Economics and Business in Greece

Published in:

- Magazine  
Communications of the ACM [CACM Homepage](#) [archive](#)  
Volume 46 Issue 1, January 2003  
Pages 71-77  
ACM New York, NY, USA  
[table of contents](#) doi > [10.1145/802421.602422](#)



 2003 Article

 **Bibliometrics**

- Downloads (6 Weeks): 0
- Downloads (12 Months): 28
- Downloads (cumulative): 1,406
- Citation Count: 20

**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

# CitationSaver preserva as citações nos documentos



[arquivo.pt/services/citationsaver](http://arquivo.pt/services/citationsaver)

## 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
<p data-bbox="262 768 690 801"><b>Insira o URL do documento:</b></p> <input data-bbox="262 818 2201 919" type="text" value="https://sobre.arquivo.pt/wp-content/uploads/SearchingImagesWebArchiveDSAA-202final.pdf"/> <p data-bbox="282 933 1116 966">Formatos aceites: PDF, TXT, HTML. Tamanho máximo: 100 MB</p> <p data-bbox="262 1011 631 1043"><b>O seu e-mail (opcional):</b></p> <input data-bbox="262 1061 2201 1168" type="text" value="pedro.gomes@fccn.pt"/> <p data-bbox="262 1200 812 1282"><b>Guardar ligações citadas</b></p>		

## 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
<p data-bbox="249 763 1080 801">Carregue um documento a partir do seu computador:</p> <div data-bbox="242 811 2216 915"><span data-bbox="267 843 588 881">Selecione um ficheiro</span> <input data-bbox="624 843 772 881" type="text" value="Plan.pdf"/></div> <p data-bbox="267 929 1110 961">Formatos aceites: PDF, TXT, HTML. Tamanho máximo: 100 MB</p> <p data-bbox="249 1011 621 1048">O seu e-mail (opcional):</p> <div data-bbox="249 1058 2216 1168"><input data-bbox="267 1096 606 1133" type="text" value="pedro.gomes@fccn.pt"/></div> <div data-bbox="249 1200 805 1282"><span data-bbox="293 1219 761 1256">Guardar ligações citadas</span></div>		

## CitationSaver Preserva citações a conteúdos online

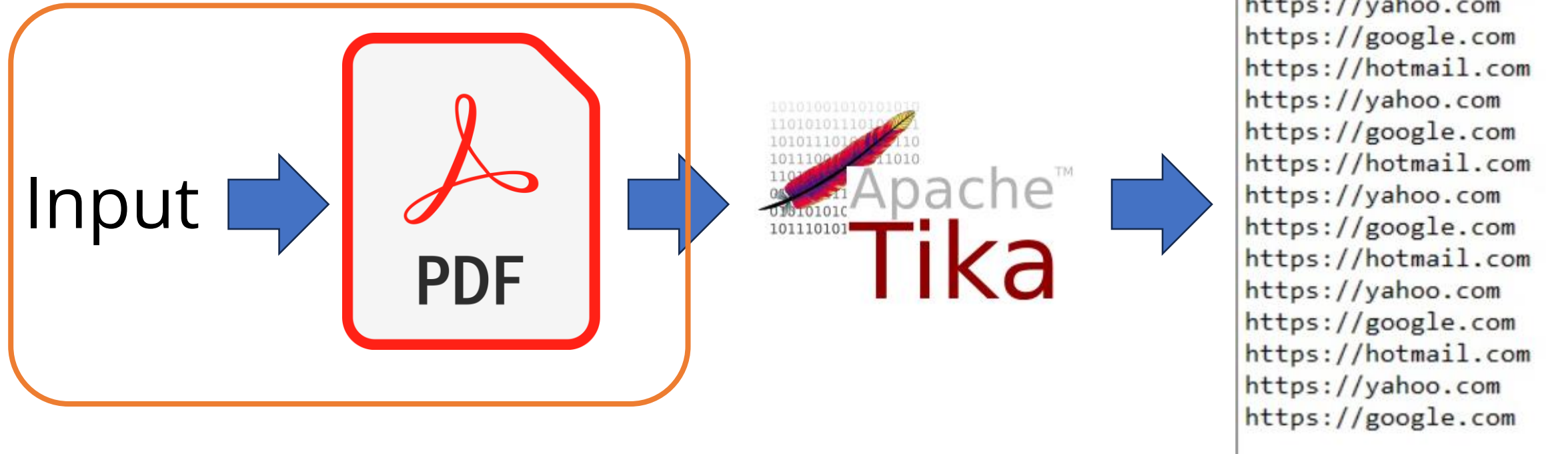
Os documentos citam conteúdos online que desaparecem rapidamente.

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](#)

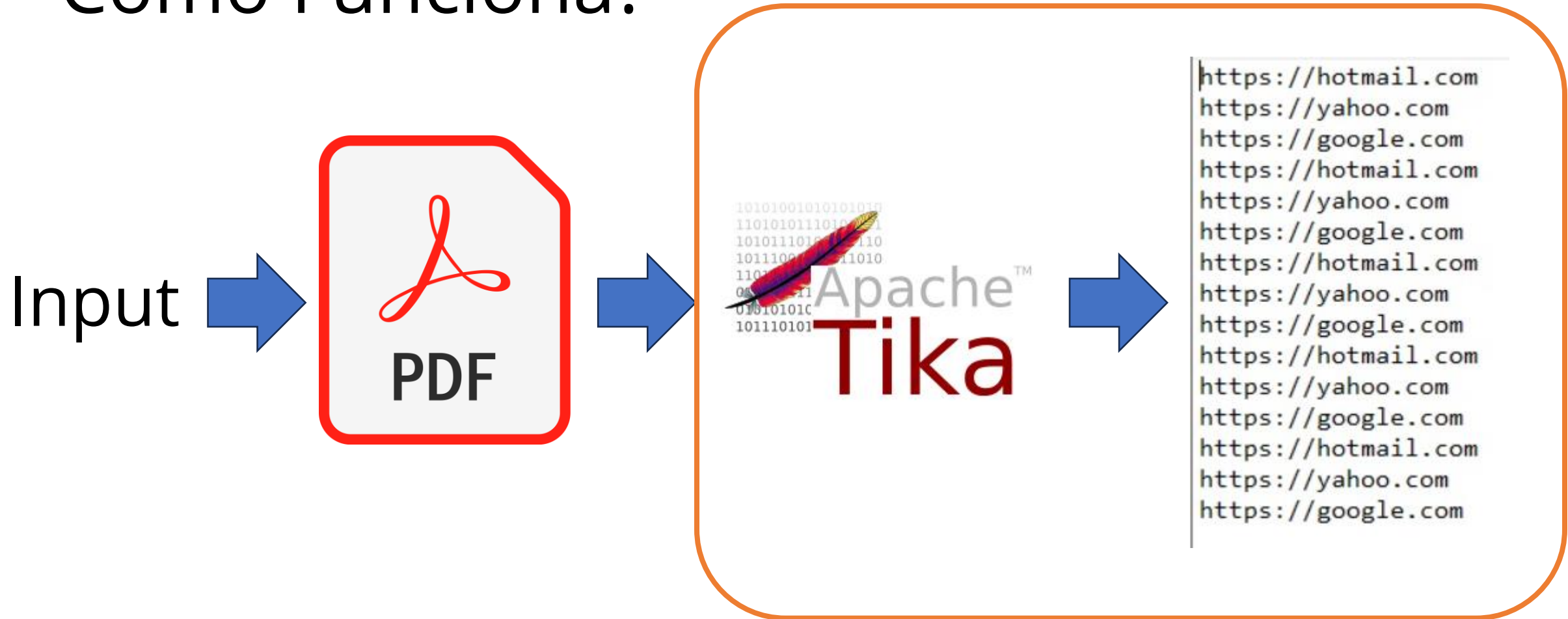
Submeta um documento e o CitationSaver preservará as ligações nele citadas:

URL	Ficheiro	Texto
<p data-bbox="466 604 856 629"><b>Insira um texto contendo URLs:</b></p> <div data-bbox="466 644 2020 1068" style="border: 1px solid #ccc; padding: 10px;"><p data-bbox="486 668 631 686">REFERENCES</p><p data-bbox="486 696 1319 811">[1] D. Gomes, "Web archives as research infrastructure for digital societies: the case study of Arquivo.pt," <i>Archeion</i>, vol. 123, pp. 46-85, Nov. 2022. [Online]. Available: <a href="https://www.ejournals.eu/Archeion/2022/123/art/22601/">https://www.ejournals.eu/Archeion/2022/123/art/22601/</a></p><p data-bbox="486 815 1251 901">[2] S. Brin and L. Page, "The Anatomy of a Large-Scale Hypertextual Web Search Engine," <i>Computer Networks</i>, vol. 30, pp. 107-117, 1998. [Online]. Available: <a href="http://www-db.stanford.edu/backrub/google.html">http://www-db.stanford.edu/backrub/google.html</a></p><p data-bbox="486 929 1319 1015">[3] L. A. Barroso, J. Dean, and U. Holzle, "Web search for a planet: The Google cluster architecture," <i>IEEE Micro</i>, vol. 23, no. 2, pp. 22-28, Mar. 2003.</p><p data-bbox="486 1019 1251 1048">[4] F. Chang, J. Dean, S. Ghemawat, W. C. Hsieh, D. A.</p></div> <p data-bbox="466 1096 766 1122"><b>O seu e-mail (opcional):</b></p> <div data-bbox="466 1136 2020 1222" style="border: 1px solid #ccc; padding: 5px;"><input data-bbox="466 1136 2020 1222" type="text" value="pedro.gomes@fccn.pt"/></div> <div data-bbox="466 1250 907 1308" style="background-color: #007bff; color: white; padding: 10px; text-align: center; width: fit-content; margin: 10px auto;"><b>Guardar ligações citadas</b></div>		

# Como Funciona?

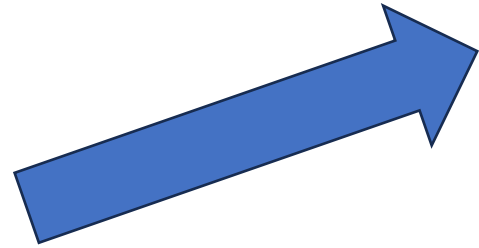


# Como Funciona?



# Como Funciona?

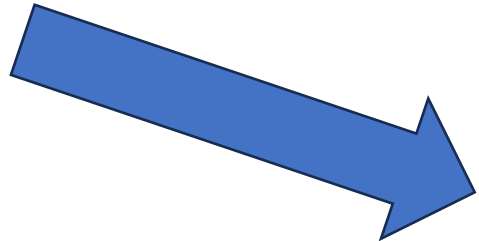
```
https://hotmail.com  
https://yahoo.com  
https://google.com  
https://hotmail.com  
https://yahoo.com  
https://google.com  
https://hotmail.com  
https://yahoo.com  
https://google.com  
https://hotmail.com  
https://yahoo.com  
https://google.com  
https://hotmail.com  
https://yahoo.com  
https://google.com
```



Heritrix



Brozzler



Browsertrix

# CitationSaver **limitações**

- Dependemos dos utilizadores para encontrar documentos científicos.
- Só pode ser submetido um documento de cada vez.



Para resolver as limitações  
o Arquivo.pt integrou APIs externas

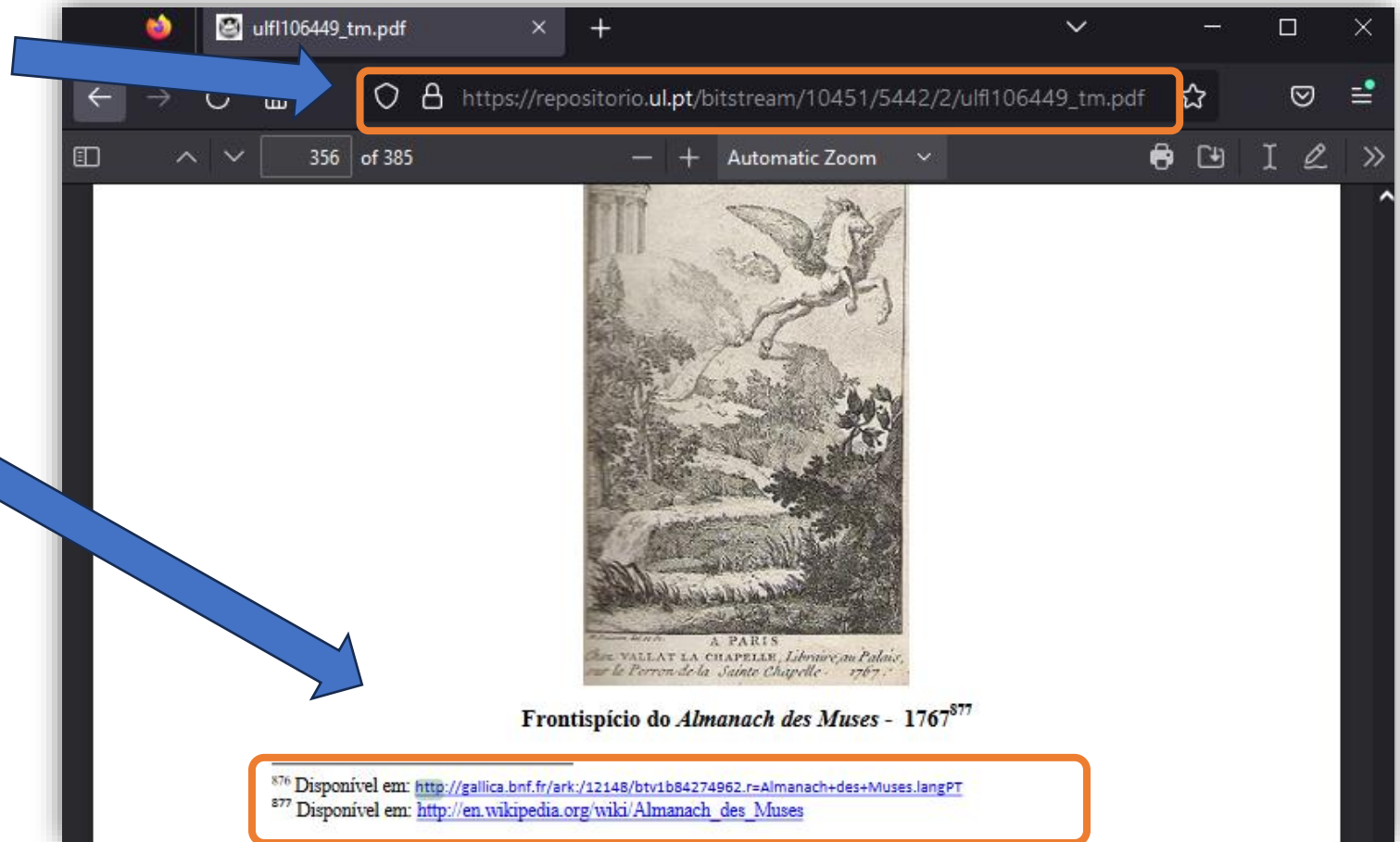
**Arquivo.pt**

**RCAAP API  
PTCRIS SciProj API  
CienciaVitae API**



# RCAAP (Repositórios Científicos de Acesso Aberto de Portugal) disponibiliza uma API com acesso a PDFs

1. Através da API, extraímos a lista de todos os URLs do PDFs.
2. Download de cada PDF.
3. Extraímos os URLs das citações usando o serviço CitationSaver.
4. Recolhemos as citações.



# Citação extraída e preservada

bibliotheca Augustana - presen. X

https://arquivo.pt/wayback/20160320234516/https://www.hs-augsburg.de/~harsch/Chronologia/Lsante01/Horatiu

Menu

ARQUIVO.PT

Options

hs-augsburg.de/~harsch/Chronologia/Lsante01/Horatiu/hor\_c000.html 20 March at 23h45, 2016

<<< [introductio](#)

BIBLIOTHECA AUGUSTANA

Q. Horatius Flaccus  
65 - 8 a. Chr. n.

2016

March

20 March 23h45, 2016

2022

Table

Carminum libri IV

23 a. Chr. n.

Textus:  
Horaz, Oden und Epoden  
ed. B. Kytzler, Stuttgart 1978

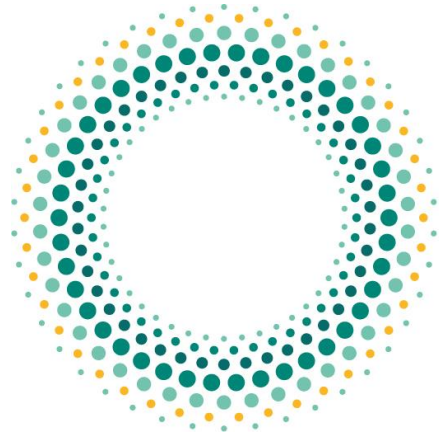
Liber primus

[Carmen I](#) Maecenas atavis edite regibus  
[Carmen II](#) Iam satis terris nivi atque dirae  
[Carmen III](#) Sic te diva potens Cypri

# Dados obtidos usando a RCAAP API

- 151 466 **PDFs**
- 2 197 292 **URLs**
- 9.1 TB **Dados recolhidos**





# Dia mundial da preservação digital

7 novembro 2024

[arquivo.pt/services/citationsaver](https://arquivo.pt/services/citationsaver)

[contacto@arquivo.pt](mailto:contacto@arquivo.pt)



ARQUIVO.PT