



Research the Past Web using Web archives

Daniel.Gomes@fccn.pt

Daniel.Bicho@fccn.pt

Fernando.Melo@fccn.pt

Tutorial outline

Research the Past Web using Web archives

1. Search and access

- The Past Web: examples and use cases
- Public online services

2. Publish and preserve

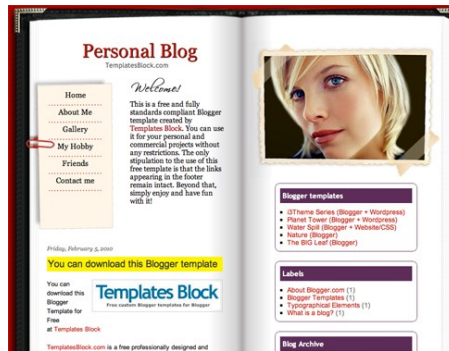
- Recommendations to publish preservable information
- How to preserve information collected from the Web

3. Automatic processing

- Interoperability protocols
- Application Programming Interfaces (API)

4. Your precious [feedback](#): please take your notes

Fact: Web replaced Press



Personal blogs

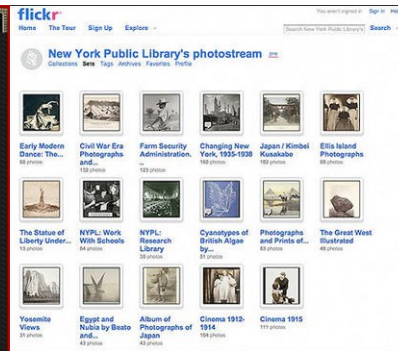


Photo galleries

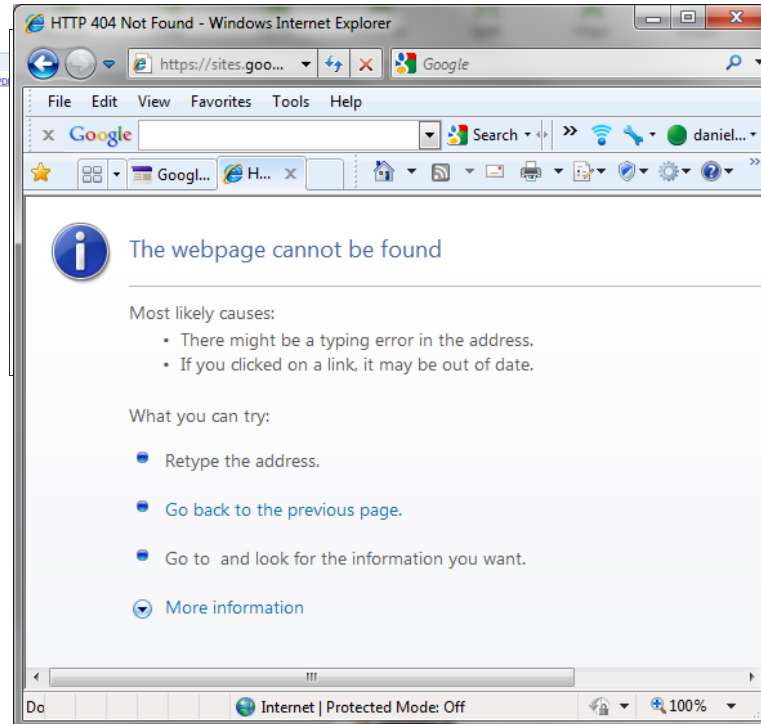


E-books



News

However, this valuable information quickly vanishes



80%

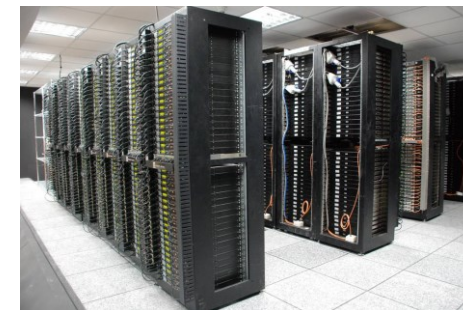
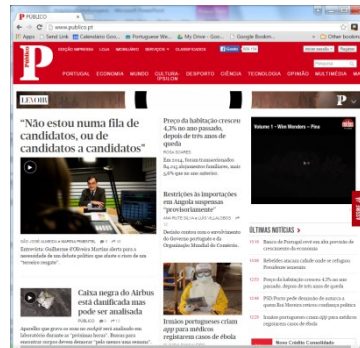
**of web pages
change their content or disappear
within 1 year**

Archives are also require to preserve information published online

Preserve the Press



Preserve the Web



[Archive.org](https://archive.org): the first web archive



Brewster Kahle founded the Internet Archive in 1996, a non-profit organization with the mission of providing “Universal access to all knowledge”.

Web archiving initiatives around the World (August 2018):

86



WIKIPEDIA
The Free Encyclopedia

- Main page
- Contents
- Featured content
- Current events
- Random article
- Donate to Wikipedia
- Wikipedia store

Interaction

- Help
- About Wikipedia
- Community portal
- Recent changes
- Contact page

Tools

- What links here
- Related changes
- Upload file
- Special pages
- Permanent link
- Page information
- Wikidata item
- Cite this page

Print/export

- Create a book
- Download as PDF
- Printable version

Dcgomes77   [Talk](#) [Sandbox](#) [Preferences](#) [Beta](#) [Watchlist](#) [Contributions](#) [Log out](#)

Article [Talk](#) [Read](#) [Edit](#) [View history](#)  [More](#)

List of Web archiving initiatives

From Wikipedia, the free encyclopedia

This page 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](#) (the Portuguese web-archive) team.

Contents [\[hide\]](#)

- [Web archiving initiatives](#)
- [Archived data](#)
- [Access methods](#)
- [References](#)
- [External links](#)



Map of Web archiving initiatives worldwide in June, 2014

Web archiving initiatives [\[edit \]](#)

Name	Country	Creation Year	Technologies	Number of Employees		Comments
				Full-time	Part-time	
Australia's Web Archive ^[2]	Australia	1996	PANDORA Digital Archiving System (PANDAS) , NLA Trove , HTTrack .	10	>5	The PANDORA Archive which takes a selective approach is a collaborative program of 11 agencies that provide an estimate average monthly staffing equivalent to around 10 FTE. IT support provided by the National Library of Australia : 0.25 person-month. Whole .au domain harvests have been conducted annually since 2005 in collaboration with the Internet

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)**

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. 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 our users are international

User Interfaces and documentation also in English

Combined with Google Translate enables **cross-lingual access** to preserved content

The Past Web

Examples of web-archived pages

Historical web pages: Tim Berners Lee personal page and W3C in the 90s

Tim Berners-Lee



Bio

Tim Berners-Lee is the inventor of the World Wide Web, an internet-based hypermedia initiative for global information sharing. He wrote the first Web clients and server and defined the URL, HTTP and HTML specifications on which the web depends while working at [CERN](#), the European Particle Physics Laboratory, in late 1990.

Tim is now the Director of the [W3 Consortium](#), an open forum of compa mission to realize the full potential of the Web. He is a Principal Research Computer Science ([LCS](#))at the Massachusetts Institute of Technology ([MIT](#))

Before going to CERN, Tim was a founding director of Image Computer S consultant in hardware and software system design, real-time communicat processing; and a principal engineer with Plessey Telecommunications, in of Oxford University.

Tim is married to Nancy Carlson. They have two children, born 1991 and

www.w3.org/pub/WWW/People/Berners-Lee/, 1996
@Arquivo.pt

www.w3.org, 1996
@Arquivo.pt



The World Wide Web Consortium

The World Wide Web is the universe of network-accessible information. The [World Wide Web Consortium](#) exists to realize the full potential of the Web.

W3C works with the global community to produce [specifications](#) and [reference software](#). W3C is funded by industrial [members](#) but its products are freely available to all. The Consortium is run by [MIT LCS](#) and by [INRIA](#), in collaboration with [CERN](#) where the web originated. Please see the [list of members](#) to learn about individual members and visit their World Wide Web sites.

- [W3C Activity areas and directions](#)
- [How to contact W3C](#)
- [Frequently Asked Questions about W3C](#)
- [W3C meetings, newsletter, mailing lists \[W3C Members only\]](#)
- [Help](#)

In this document:


- [News and Updates](#)
- [Web Specifications and Development Areas](#)
- [W3C Software](#)
- [The World Wide Web and the Web Community](#)
- [Getting involved with the W3C](#)

No Google, no Facebook, you had to navigate through the links

European Home Page

Free Speech Online
Blue Ribbon Campaign

This is a very simple European Map of WWW/NIR sites. Please click on any flag to go to the (sensitive map) Home Page of that country.
If you don't have graphics support, here is [CERN's list](#) of sites organized by geography. If you want to see another european Home Page, try this [one](#).
Try the new [Beautiful Cultural European Home Pages](#) (with travel tips).

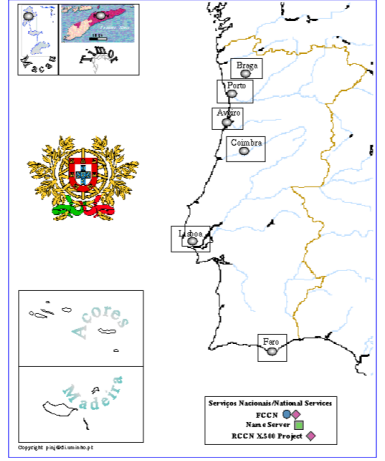


Copyright 1996/1997 @U.uminho.pt

s700.uminho.pt/europa.html, 1996

@Arquivo.pt

«Home Page» de Portugal/ Portugal Home Page



SAPO - Servidor de Apontadores Portugueses

[Acerca de Portugal](#) [Portugal Cultural](#) [Conferências e outros eventos](#) [Ajuda](#)
[About Portugal](#) [Cultural events](#)
[Conferences and other events](#) [Help](#)

Free Speech Online
Blue Ribbon Campaign

Se não possuir suporte gráfico, prima aqui / If you don't have graphics support, click [here](#)

[Europe Home Page](#) [EC Home Page](#) [World Home Page](#)

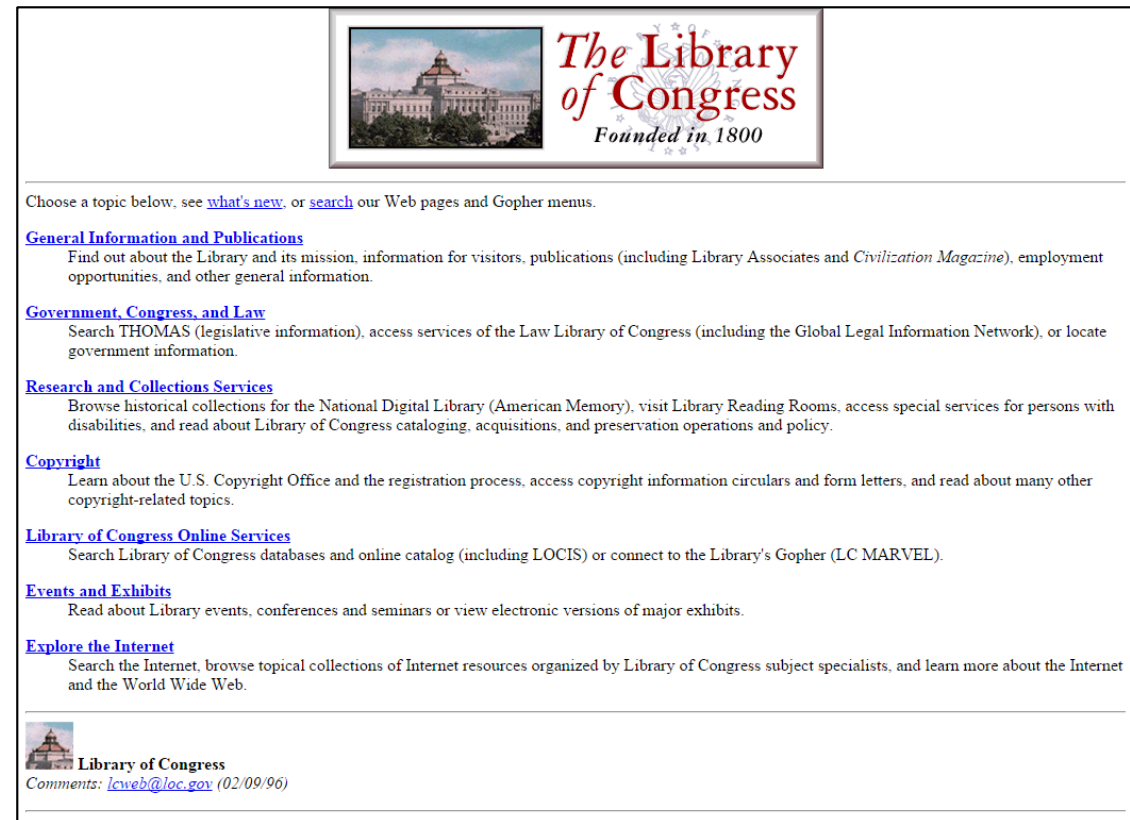
s700.uminho.pt/homepage-pt.html, 1996

@Arquivo.pt

Web-born information **NOT** collected online: web pages from CD-ROMs (digital archeology)

First web page of the Library of Congress (1996).

Obtained from a CD-ROM
attached to a printed book.



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.

[Copyright](#)
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 Internet](#)
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)

Printed information, re-published online

Today's newspapers
Today's press covers

Kiosko.NET

Monday, 11/Aug/2014

Home Africa Asia-Pacific Europe Latin America USA Canada Archive

Europe ▾ Belgium Denmark France Germany Greece Ireland Italy Netherlands Norway Poland Portugal Russia Spain Sweden Switzerland Turkey United Kingdom

Canada ▾ USA ▾ Alabama Arizona California Colorado Florida Georgia Illinois Indiana Maryland Massachusetts Michigan Minnesota Missouri New Jersey New York Ohio Oregon Pennsylvania Texas Washington Washington D.C.



Send in the warplanes, Tories urge Cameron

Refugees find path out of siege in Iraq

Ebola outbreak traced to child in Guinea

La UE reclama una investigación del contrabando en Gibraltar

Erdogan obtiene el respaldo para

U.S. Stocks Only Dented By Tensions

Kurdish Fight in Northern Iraq

USA TODAY 08.11.14

Death on a dirt track clouds NASCAR image

ISRAEL, HAMAS TO TRY AGAIN

U.S. airstrikes bring calm to Iraq

Defiant Maliki clings to power

Yardinler mountain siege, head into Kurdistan

Drugs, deception, and a disappearance

Donors boost public colleges

ALL TOO OFTEN, THEY FIND A GUN

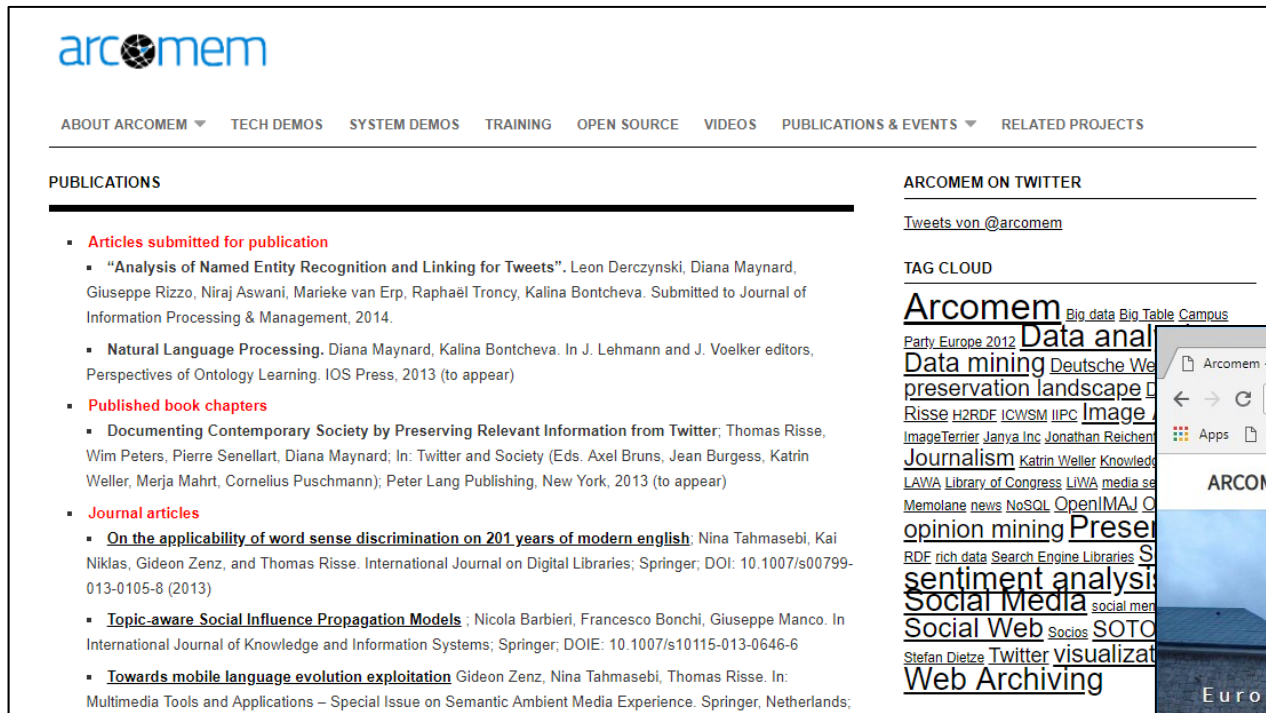
Old printed information, published online, enriched with user-generated meta-data

The screenshot shows the website 'Almanaque Republicano'. At the top is a banner with the title 'Almanaque Republicano' and a coat of arms. Below the banner, the date 'SEGUNDAFEIRA, JULHO 26, 2010' and the article title 'A DEFESA DA REPÚBLICA - in A CORJA' are displayed. A central image shows a historical document titled 'A DEFESA DA REPÚBLICA' with the heading 'SAIBAM QUANTOS'. To the left of the document is a sidebar with 'ARQUIVO' and 'E-Mail at - m' buttons. To the right is a sidebar with 'ALMANAQUE REPUBLICANO' and 'Itinerário, tábuas ou documentos da A. R. de M. Menção José M. Martins'. Below the document is a red-bordered box containing the following text: 'A CORJA. Semanário republicano anti-clerical independente (Liberdade, Justiça, Verdade, Progresso) - [Ano I, nº1 (6 de Fevereiro 1915) ao nº 25 (16 de Agosto 1915)], Coimbra; **Administrador**, Aníbal Reis [nº22, M. Simões]; **Secretário**, Mário de Brito [nº10, J. L. Frazão]; **Director**: José Peixoto de Alarcão [nº21, Fernandes Martins]; **Colaboradores**: A. Batista Rama, Afonso Duarte (poema), Alfredo Pimenta (poema), António Correia de Oliveira (poema), António Nobre (poema), Baldaque da Silva, Coelho Neto, Ernesto Almeida, Fernandes Martins, Fernando de Araújo, Guerra Junqueiro (poema), J. Peixoto de Alarcão, J. Pestana Júnior, João de Deus (poema), José Figueiredo Júnior [idem in, "A Revolta"], Ribeiro de Carvalho (poema); **Administração**, Rua Dr. João Jacinto, 38, Coimbra; **Redacção**, Couraça de Lisboa, 10, Coimbra'. Below this text is another red-bordered box with the text 'Almanaque Republicano' and 'Cria o teu cartão de visita'. At the bottom right, there is a 'ETIQUETAS' section with a list: '18 de Janeiro de 1934'. On the far left, there is a 'CENTENÁRIO DA REPÚBLICA 1910-2010' logo and a 'BLOGS' section with a list of links.

Published
on print in 1915

Enriched
online in 2010

Open-access research data and publications



The screenshot shows the Arcocomem website's 'PUBLICATIONS' page. The header includes the Arcocomem logo and navigation links: ABOUT ARCOMEM, TECH DEMOS, SYSTEM DEMOS, TRAINING, OPEN SOURCE, VIDEOS, PUBLICATIONS & EVENTS, and RELATED PROJECTS. The 'PUBLICATIONS' section is titled and contains a list of articles. To the right, there are sections for 'ARCOMEM ON TWITTER' (with a link to 'Tweets von @arcocomem') and 'TAG CLOUD'.

arcocomem

ABOUT ARCOMEM ▾ TECH DEMOS SYSTEM DEMOS TRAINING OPEN SOURCE VIDEOS PUBLICATIONS & EVENTS ▾ RELATED PROJECTS

PUBLICATIONS

- **Articles submitted for publication**
 - "Analysis of Named Entity Recognition and Linking for Tweets". Leon Derczynski, Diana Maynard, Giuseppe Rizzo, Niraj Aswani, Marieke van Erp, Raphaël Troncy, Kalina Bontcheva. Submitted to Journal of Information Processing & Management, 2014.
 - **Natural Language Processing**. Diana Maynard, Kalina Bontcheva. In J. Lehmann and J. Voelker editors, Perspectives of Ontology Learning. IOS Press, 2013 (to appear)
- **Published book chapters**
 - **Documenting Contemporary Society by Preserving Relevant Information from Twitter**; Thomas Risse, Wim Peters, Pierre Senellart, Diana Maynard; In: Twitter and Society (Eds. Axel Bruns, Jean Burgess, Katrin Weller, Merja Mahrt, Cornelius Puschmann); Peter Lang Publishing, New York, 2013 (to appear)
- **Journal articles**
 - **On the applicability of word sense discrimination on 201 years of modern english**; Nina Tahmasebi, Kai Niklas, Gideon Zenz, and Thomas Risse. International Journal on Digital Libraries; Springer; DOI: 10.1007/s00799-013-0105-8 (2013)
 - **Topic-aware Social Influence Propagation Models**; Nicola Barbieri, Francesco Bonchi, Giuseppe Manco. In International Journal of Knowledge and Information Systems; Springer; DOIE: 10.1007/s10115-013-0646-6
 - **Towards mobile language evolution exploitation** Gideon Zenz, Nina Tahmasebi, Thomas Risse. In: Multimedia Tools and Applications – Special Issue on Semantic Ambient Media Experience. Springer, Netherlands;

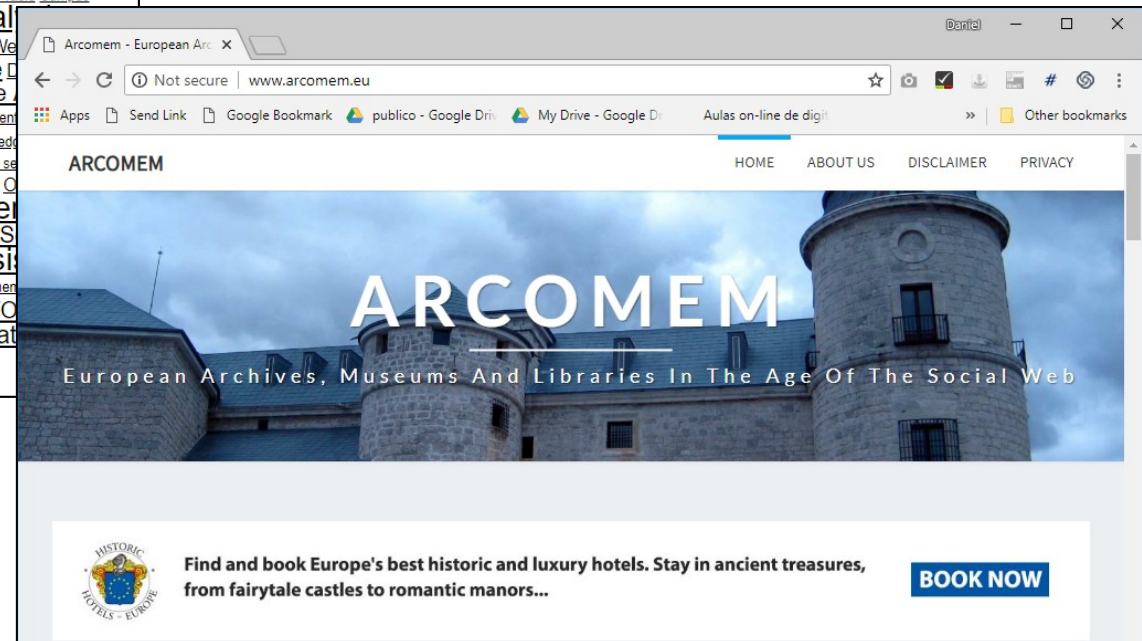
ARCOMEM ON TWITTER

[Tweets von @arcocomem](#)

TAG CLOUD

Arcocomem [Big data](#) [Big Table](#) [Campus](#)
[Party Europe 2012](#) [Data anal](#)
[Data mining](#) [Deutsche We](#)
[preservation landscape](#) [D](#)
[Risse H2RDF](#) [ICWSM](#) [IIPC](#) [Image](#)
[ImageTerrier](#) [Janya Inc](#) [Jonathan Reichen](#)
[Journalism](#) [Katrin Weller](#) [Knowled](#)
[LAWA](#) [Library of Congress](#) [LIWA](#) [media se](#)
[Memolane news](#) [NoSQL](#) [OpenIMAJ](#) [O](#)
[opinion mining](#) [Preser](#)
[RDF rich data](#) [Search Engine Libraries](#) [S](#)
[sentiment analysis](#)
[Social Media](#) [social med](#)
[Social Web](#) [Socios](#) [SOTO](#)
[Stefan Dietze](#) [Twitter visualizat](#)
[Web Archiving](#)

arcocomem.eu, 2014
@Internet Archive



The screenshot shows a browser window displaying the Arcocomem website. The page features a large banner with the text 'ARCOMEM European Archives, Museums And Libraries In The Age Of The Social Web'. Below the banner, there is a section for 'HISTORIC HOTELS - EUROPE' with the text 'Find and book Europe's best historic and luxury hotels. Stay in ancient treasures, from fairytale castles to romantic manors...' and a 'BOOK NOW' button.

ARCOCOMEM

HOME ABOUT US DISCLAIMER PRIVACY

European Archives, Museums And Libraries In The Age Of The Social Web

HISTORIC HOTELS - EUROPE

Find and book Europe's best historic and luxury hotels. Stay in ancient treasures, from fairytale castles to romantic manors...

BOOK NOW

arcocomem.eu, 2018

The Past Web

Use cases

Fix citations to broken links that jeopardize Science in all areas

Data sets and tools applied in experiments are online.

When cited resources disappear, experiments cannot be reproduced

- Not Reproducible, Not Science.

Web archives preserve online scientific information.

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

Death of web citations: a serious alarm for authors

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

« Previous | Next Article »
Table of Contents

This Article

doi:
10.1096/fj.05-4784lsf
December 2005
The FASEB Journal
vol. 19 no. 14 1943-1944

» Abstract
Full Text
Full Text (PDF)

+ Classifications

+ Services

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

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

Management: documentation/audit of finished project



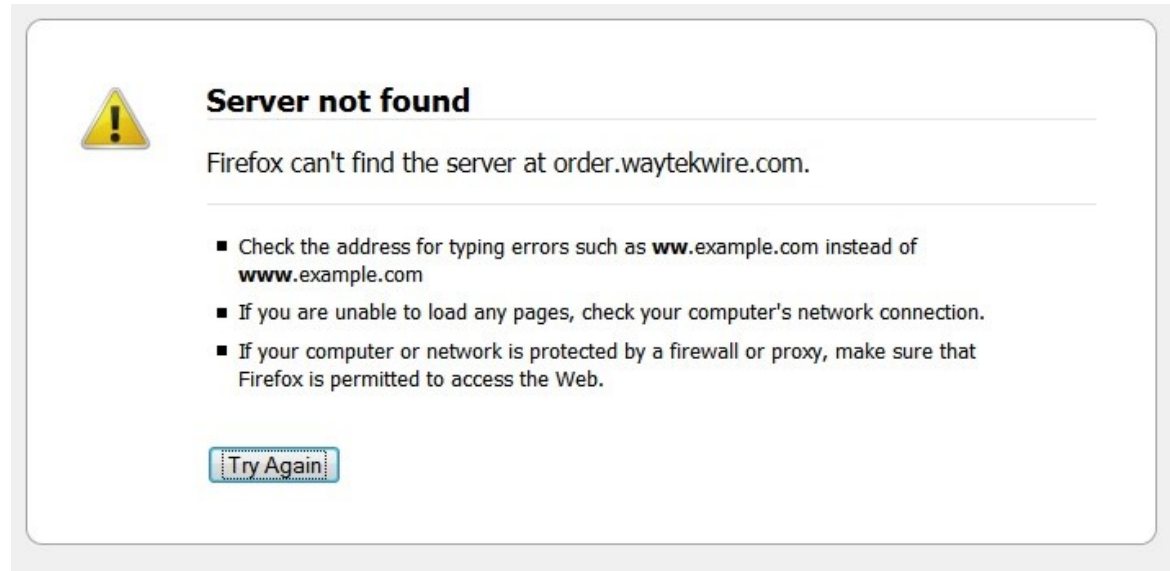
“Everything was on the website”

<p>Linguateca</p> <p>Structure Team Overview Access to resources Evaluation contests Catalogue of resources Catalogue of tools Catalogue of actors Catalogue of publications</p> <ul style="list-style-type: none">▪ Publication search▪ Publication addition▪ SUPeRB <p>Interesting information FAQ Gramateca</p> <hr/> <p>Portugués Norsk</p> <hr/> <p> Comments and suggestions</p> <p><small>Please use the navigation options on the left handside frame (clicking on the links on the right handside does not update the left handside).</small></p>	<p>Publications on the computational processing of portuguese</p> <p>Linguateca</p> <hr/> <p>Our goal is to gather a collection of bibliographic references on the computational Processing of portuguese in this pages. We appreciate any contribution that you as an author or reader can give us, by providing us with additional information or to direct us to pages with relevant content.</p> <p>To look for publications we suggest the use of our search interface. To contribute with new references you can use this form. To use SUPeRB as your publications manager see here.</p> <p>We currently list 2375 references under the following categories:</p> <ul style="list-style-type: none">• Publications in journals (184 ordered alphabetically, by date; In BibTeX: orderer alfabetically, by date)• Books (1 publications, orderer alphabetically, by date; In BibTeX: orderer alphabetically, by date)• Book chapters (189 publications, orderer alphabetically, by date; In BibTeX: orderer alphabetically, by date)• Book (Author) (57 publications, orderer alphabetically, by date; In BibTeX: orderer alphabetically, by date)• Book (Compilation) (53 publications, orderer alphabetically, by date; In BibTeX: orderer alphabetically, by date)• teshis (152 publications, orderer alphabetically, by date; In BibTeX: orderer alphabetically, by date)• Articles in international conferences (1134 publications, orderer alphabetically, by date; In BibTeX: orderer alphabetically, by date)• Reports and monographies (192 publications, orderer alphabetically, by date; In BibTeX: orderer alphabetically, by date)• Other academical work (3 publications, orderer alphabetically, by date; In BibTeX: orderer alphabetically, by date)• Publications only available on the web (95 publications, orderer alphabetically, by date; In BibTeX: orderer alphabetically, by date)• Presentations (297 publications, orderer alphabetically, by date; In BibTeX: orderer alphabetically, by date) <hr/> <p>Publications inserted last month.</p> <hr/> <p>Last update 30 November 2015. Comments e suggestions</p>
--	---

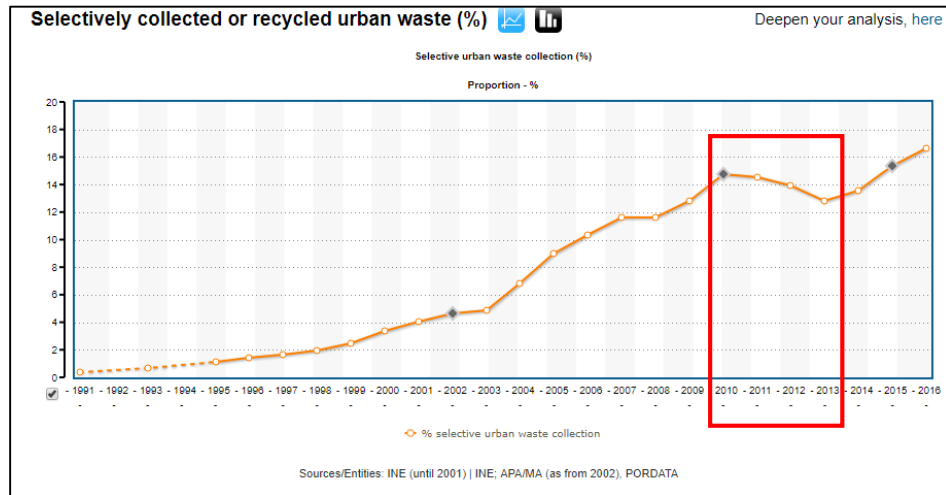
Maintain Portfolio/CV

“My best work was a website that no longer exists...”

60% of the websites disappear just after 2 years.



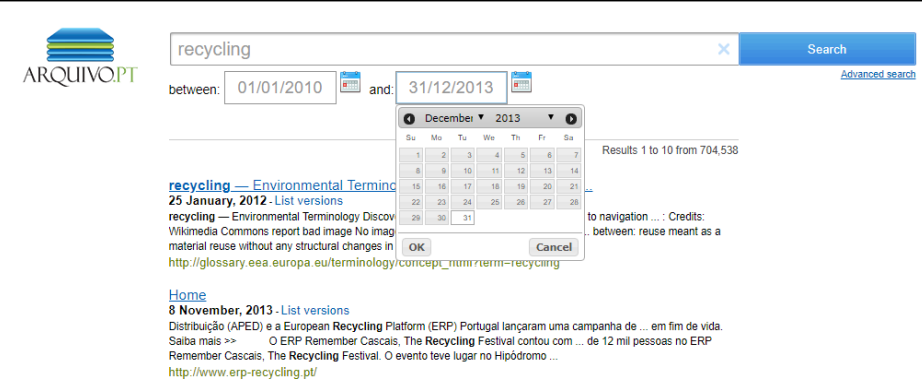
Support fast qualitative analysis for historical research



PORDATA: quantitative analysis

**Statistical event:
Recycling decrease 2010-2013**

Why?



ARQUIVO.PT

recycling [Advanced search](#)

between: and:

December 2013

Su Mo Tu We Th Fr Sa

1 2 3 4 5 6 7

8 9 10 11 12 13 14

15 16 17 18 19 20 21

22 23 24 25 26 27 28

29 30 31

Results 1 to 10 from 704,538

to navigation ... : Credits: between: reuse meant as a

[recycling — Environmental Terminology](#)
25 January, 2012 - List versions

[recycling — Environmental Terminology](#) Discover Wikimedia Commons report bad image No image material reuse without any structural changes in http://glossary.eea.europa.eu/terminology/concept_term=recycling

[Home](#)

8 November, 2013 - List versions

Distribuição (APED) e a European Recycling Platform (ERP) Portugal lançaram uma campanha de ... em fim de vida. Saiba mais >> O ERP Remember Cascals. The Recycling Festival contou com ... de 12 mil pessoas no ERP Remember Cascals. The Recycling Festival. O evento teve lugar no Hipódromo ... <http://www.erp-recycling.pt/>

Arquivo.pt: qualitative analysis

**Search for news related to recycling between
2010-2013**

Search the Past Web

Anyone can do it.

URL Search: Internet Archive Wayback Machine

[Archive.org](https://archive.org)

World wide content

URL search

The screenshot shows the Internet Archive Wayback Machine interface. At the top left is the logo "INTERNET ARCHIVE WayBackMachine". To the right, it says "Explore more than 335 billion web pages saved over time". Below this is a search bar containing "http://nytimes.com/" and a "DONATE" button. The main content area displays "Saved 189,548 times between November 12, 1996 and August 13, 2018." with links to "Summary of nytimes.com" and "Site Map of nytimes.com". A bar chart shows the number of saves per year from 2000 to 2018, with 2001 highlighted in yellow. Below the chart is a calendar for March 2013, with the date "Fri, 01 Mar 2013 01:20:45 GMT" and a note "(why: focused_crawls, elpais.com.uy, top_news)". The calendar shows the days of the month, with several dates circled in blue: 4, 5, 6, 7, 8, 18, 26, 24, 30, and 26.

New York Times news on February 14th, 2009

<h1>The New York Times</h1> <p>Saturday, February 14, 2009 Last Update: 1:11 AM ET</p>			
<input type="text"/> <input type="button" value="Search"/>		Try Our EXTRA Home Page	
		 Get Home Delivery Personalize Your Weather	
<p>JOBS REAL ESTATE AUTOS ALL CLASSIFIEDS</p> <p>WORLD U.S. POLITICS N.Y./REGION BUSINESS TECHNOLOGY SPORTS SCIENCE HEALTH OPINION ARTS Books Movies Music</p>	<h2>Stimulus Plan Sets New Limits on Executive Pay</h2> <p>By EDMUND L. ANDREWS and ERIC DASH</p> <p>A provision buried in the stimulus bill passed by Congress would impose restrictions on executive bonuses at companies reaping bailout money.</p> <h3>Stimulus Plan Receives Final Approval in Congress</h3>	 <p style="text-align: center;"><small>Stephen Farrell/The New York Times</small></p> <h2>Iraqi Militants Become Patrons of the Arts</h2> <p>By STEVEN LEE MYERS</p>	
		<h2>OPINION »</h2> <p>TALK SHOW When Updike and Cheever Came to Visit Dick Cavett recalls the rare thrill of having two great writers as guests on the same show.</p>  <ul style="list-style-type: none">· Collins: Stock a Cabinet·  Comments· Editorial: California Prisons· Op-Ed: No Price on Love· Op-Ed: Baby Bailout· Op-Ed: Stagflation· Opinionator: Renditions	
		<h2>MOVIES »</h2> <p>Lesser-Known Mobsters "Gomorra" takes an unsentimental look at Italian organized crime.</p>	<h2>BUSINESS »</h2> <p>A Stress Test for the Latest Bailout Plan The stress test finds out how crippled big banks would be</p>

New York Times (2005)

meet face-to-face with top employers

The New York Times

ON THE WEB

UPDATED MONDAY, MAY 2, 2005 9:34 PM ET | [Personalize Your Weather](#)

Sign up for a Business Gold Card from American Express and get a free Times News Tracker or Premium Crosswords subscription!

SEARCH SPONSORED BY [Open a](#)

Site Search: [Go to Advanced Search/Archive](#)

NYT Since 1996 Search

LOG IN REGISTER NOW. It's Free!

JOB MARKET

REAL ESTATE

AUTOS *NEW*

NEWS

[International](#)

[National](#)

[Washington](#)

[New York/Region](#)

[Business](#)

[Technology](#)

[Science](#)

[Health](#)

[Sports](#)

[Education](#)

[Weather](#)

[Obituaries](#)

[NYT Front Page](#)

[Corrections](#)

OPINION

[Plan to Transform Brooklyn Waterfront Advances at City Hall](#)

By DIANE CARDWELL 6:28 PM ET

Officials agreed on a plan that would transform the north Brooklyn waterfront into a neighborhood of residential towers.

[African Strain of Polio Virus Hits Indonesia](#)

By DONALD G. McNEIL Jr. 9:17 PM ET

The case of polio indicates that an outbreak spreading from Nigeria has reached the world's fourth-most-populous country.

[Enlarge This Image](#)



Rahav Segev for The New York Times

Rock Fans Embrace the Random

Real Estate

AUDIO SLIDE SHOW



[A Concerted Effort](#)

A 300-square-foot studio with a baby grand Steinway.

Go to: [Article](#) | [Real Estate](#)

New York Times on 9/11, 2001

The New York Times on the Web

UPDATED TUESDAY, SEPTEMBER 11, 2001 4:43 PM ET | [Personalize Your Weather](#)

Search Past 30 Days
[Go to Advanced Search](#)

[Sign Up](#)
[Log In](#)

[World Trade Center Topped in Attack](#)

[Thousands Feared Dead in New York and Washington](#)

By JAMES BARRON

In what appeared to be coordinated attacks, hijacked jetliners crashed into both towers of the World Trade Center and the Pentagon Tuesday morning. [Go to Article](#)

[Slide Show: Destruction in New York](#)

[In Washington, an Attack on a Symbol of American Power](#)

By DAVID STOUT

Shortly after the attacks in New York, another aircraft crashed into the Pentagon, the massive building that is the very symbol of the American defense establishment.

[Throughout U.S., Shock After Attacks](#)
 [Bush Vows to Hunt Down Terrorists](#)



TECHNOLOGY

[Dell on Top](#)

Everyone is feeling the slump, but Michael Dell says his company can tolerate it better than most.



Q&A

[Ottawa, Canada](#)

susz99 asks: Does anyone have any suggestions for things to do in Ottawa for a group of about 20?

[answer the question](#)

[More News From Reuters](#)

[More News From AP](#)

New York Times on the early days of the Web

Welcome! First-time users [register here](#) or [take a QuickTour](#) of the site.

NEW
AUTO
CLASSIFIEDS

The New York Times
ON THE WEB

knicks^{NEW!}
celtics
one on one
newyork.boston.com

"All the News That's Fit to Print"

Tuesday, November 12, 1996

SECTIONS

- Front Page ● CyberTimes
- Politics ● Business
- Editorials/Letters ● Op-Ed
- Arts & Leisure ● Travel
- Real Estate ● Job Market
- Diversions ● Web Specials

NEWS BY CATEGORY

CLASSIFIEDS | FORUMS

SERVICES | SEARCH

TABLE OF CONTENTS



Crowds Looking for Food in Zaire

LATE NEWS UPDATE

Two Aircraft Collide
In Air Near Delhi,
Killing About 350

IN CYBERTIMES

Europe Betting on
Self-Regulation to
Control the Internet

[Help Center](#) | [Low Graphics](#)

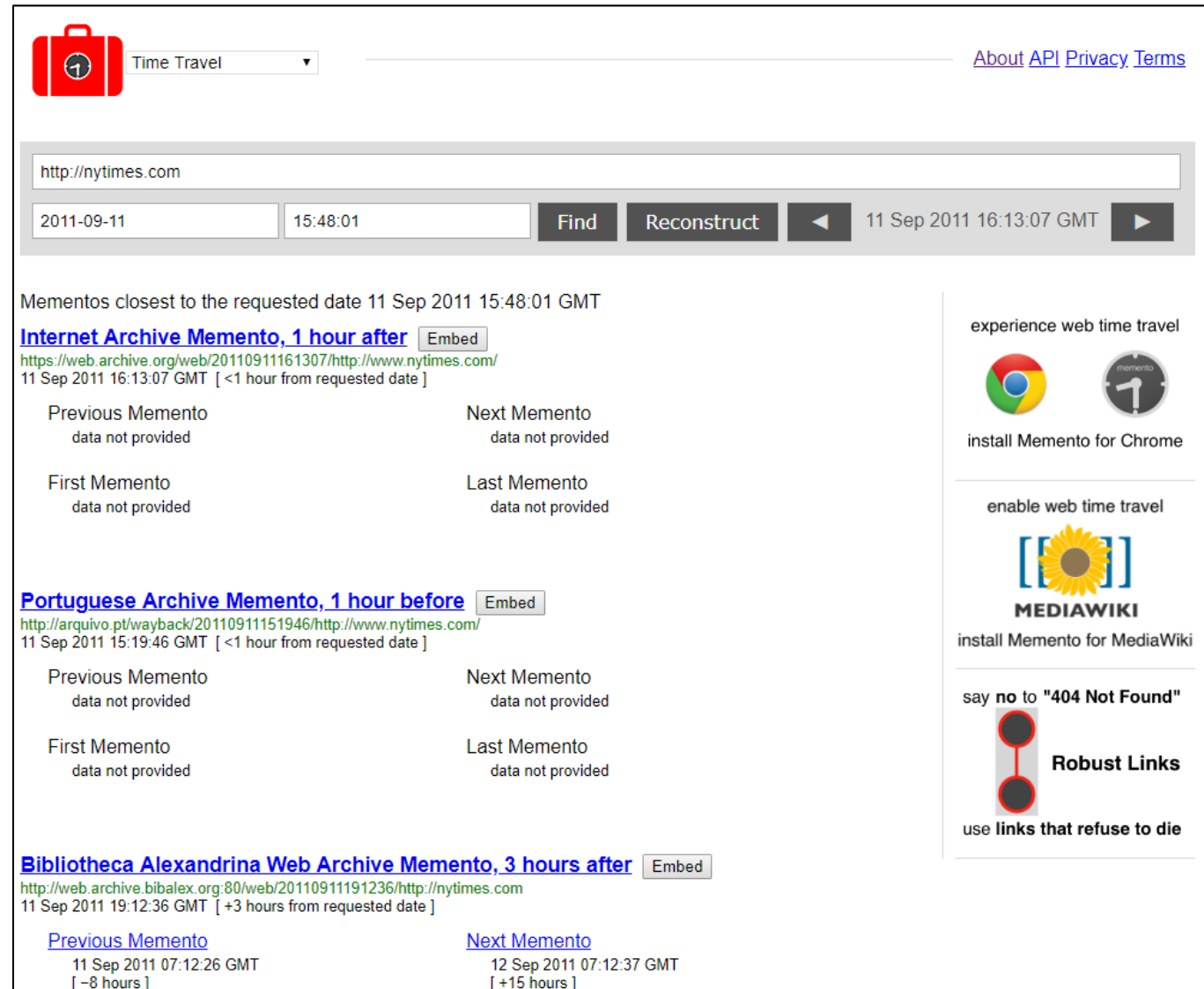
PLEASE OPEN YOUR WINDOW TO THE WIDTH OF THIS LINE OF TEXT.

[Copyright 1996 The New York Times](#)

URL search across several web archives

Memento Time Travel Portal

TimeTravel.mementoweb.org



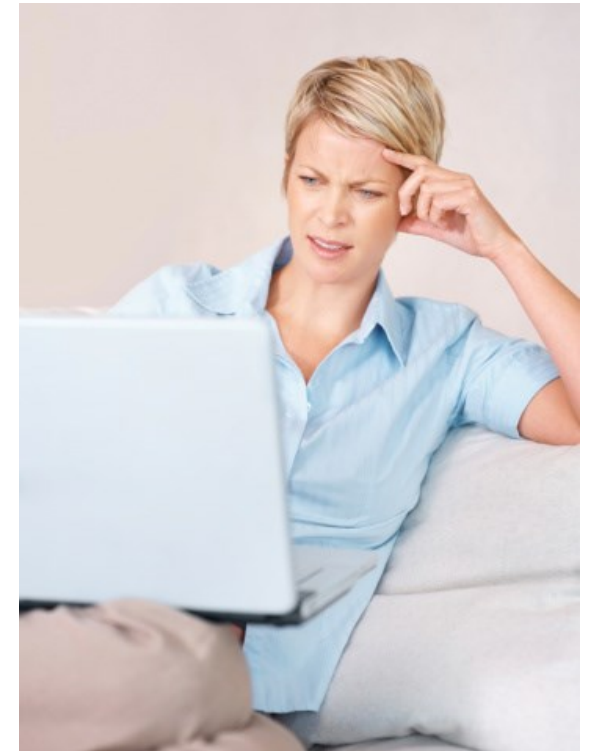
The screenshot displays the Memento Time Travel Portal interface. At the top left is a red suitcase icon with a clock face. To its right is a dropdown menu labeled "Time Travel". On the far right are links for "About API Privacy Terms". Below this is a search bar containing the URL "http://nytimes.com". Underneath the search bar are two input fields: the first contains "2011-09-11" and the second contains "15:48:01". To the right of these fields are buttons for "Find" and "Reconstruct", followed by a left arrow, the text "11 Sep 2011 16:13:07 GMT", and a right arrow. The main content area shows "Mementos closest to the requested date 11 Sep 2011 15:48:01 GMT". The first result is "Internet Archive Memento, 1 hour after" with an "Embed" button. Below it is the URL "https://web.archive.org/web/20110911161307/http://www.nytimes.com/" and the timestamp "11 Sep 2011 16:13:07 GMT [<1 hour from requested date]". To the right of this result are labels for "Previous Memento" (data not provided) and "Next Memento" (data not provided). The second result is "Portuguese Archive Memento, 1 hour before" with an "Embed" button. Below it is the URL "http://arquivo.pt/wayback/20110911151946/http://www.nytimes.com/" and the timestamp "11 Sep 2011 15:19:46 GMT [<1 hour from requested date]". To the right of this result are labels for "Previous Memento" (data not provided) and "Next Memento" (data not provided). The third result is "Bibliotheca Alexandrina Web Archive Memento, 3 hours after" with an "Embed" button. Below it is the URL "http://web.archive.bibalex.org:80/web/20110911191236/http://nytimes.com" and the timestamp "11 Sep 2011 19:12:36 GMT [+3 hours from requested date]". To the right of this result are labels for "Previous Memento" (data not provided) and "Next Memento" (data not provided). At the bottom of the results are two more entries: "Previous Memento" with timestamp "11 Sep 2011 07:12:26 GMT [-8 hours]" and "Next Memento" with timestamp "12 Sep 2011 07:12:37 GMT [+15 hours]". On the right side of the page, there are three promotional banners. The first is for "experience web time travel" with Chrome and Memento logos and the text "install Memento for Chrome". The second is for "enable web time travel" with the MediaWiki logo and the text "install Memento for MediaWiki". The third is for "say no to '404 Not Found'" with the Robust Links logo and the text "use links that refuse to die".

Limitation of URL search

Users do not know the exact URL of the web page of the Past that contained the information that they search for.

Search for: Michael Jackson concert before 1998.

Which was the address of the web page?



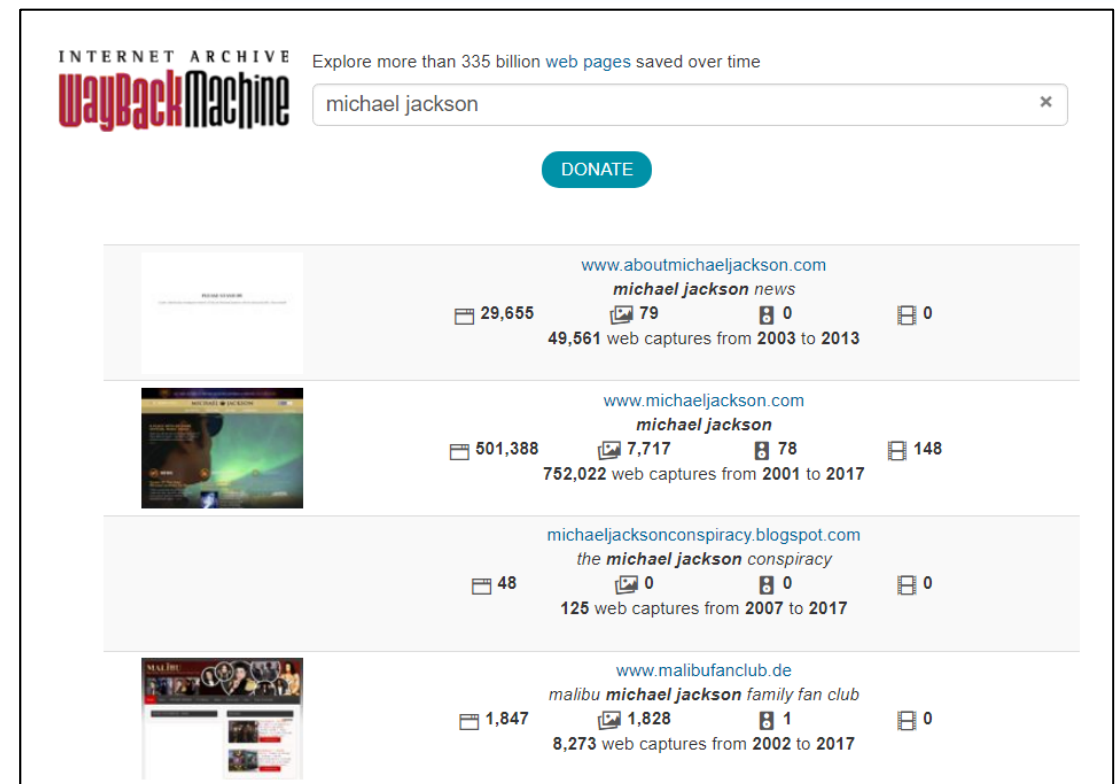
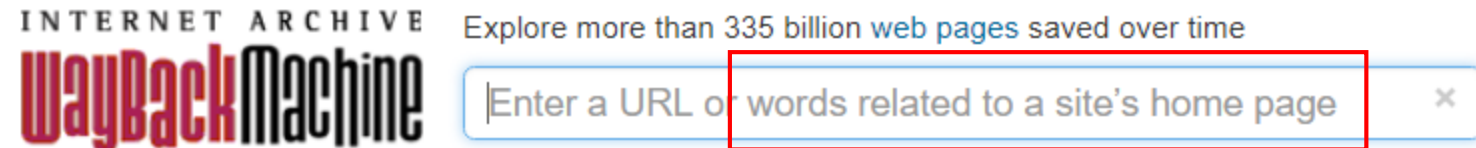
Keyword search related to home pages of websites

[Archive.org](https://archive.org)

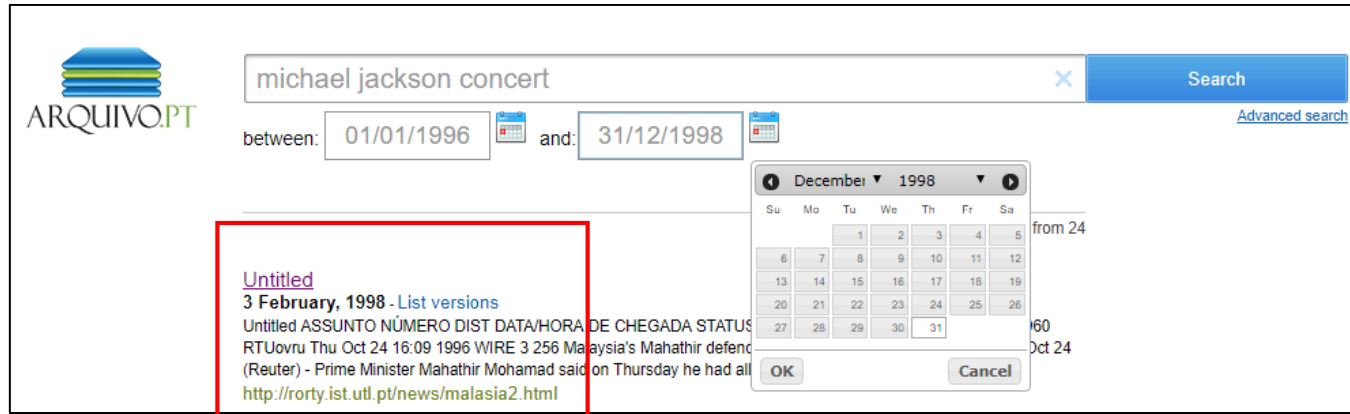
Keywords from the links to the homepages of websites.

Search results do not include:

- Home pages that do not receive links
- Internal webpages (not home pages)
- Cannot define time span for search



Full-text search over all pages with Arquivo.pt



ARQUIVO.PT

michael jackson concert

Search

Advanced search

between: 01/01/1996 and: 31/12/1998

December 1998

Su Mo Tu We Th Fr Sa

1 2 3 4 5 from 24

6 7 8 9 10 11 12

13 14 15 16 17 18 19

20 21 22 23 24 25 26

27 28 29 30 31 60 Oct 24

OK Cancel

[Untitled](#)

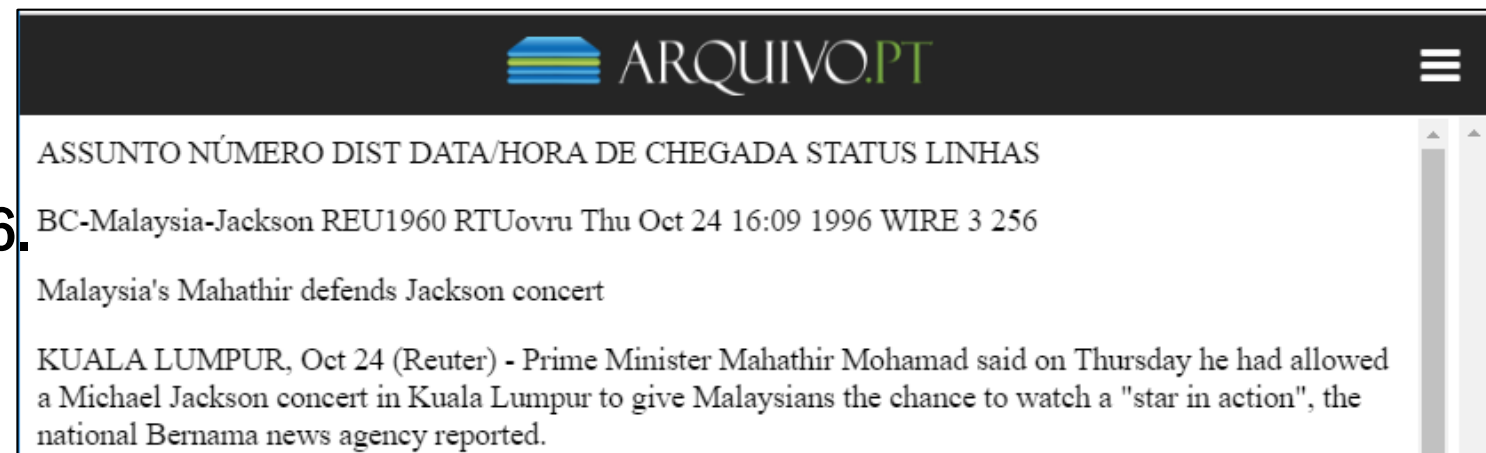
[3 February, 1998 - List versions](#)

Untitled ASSUNTO NÚMERO DIST DATA/HORA DE CHEGADA STATUS LINHAS
RTUovru Thu Oct 24 16:09 1996 WIRE 3 256 Malaysia's Mahathir defends
(Reuter) - Prime Minister Mahathir Mohamad said on Thursday he had all
<http://rorty.ist.utl.pt/news/malasia2.html>

Not site home page

“Untitled”

There was a controversial Michael Jackson concert in Malaysia in 1996.



ARQUIVO.PT

ASSUNTO NÚMERO DIST DATA/HORA DE CHEGADA STATUS LINHAS

BC-Malaysia-Jackson REU1960 RTUovru Thu Oct 24 16:09 1996 WIRE 3 256

Malaysia's Mahathir defends Jackson concert

KUALA LUMPUR, Oct 24 (Reuter) - Prime Minister Mahathir Mohamad said on Thursday he had allowed a Michael Jackson concert in Kuala Lumpur to give Malaysians the chance to watch a "star in action", the national Bernama news agency reported.



Advanced Search

Refine the details of your search using the options bellow.

Search pages by:

Search

Words

With these words:

Portugal

ex.: group draw

With this phrase:

michael jackson

ex.: euro 2004

Without any of these words:

albums

ex.: rugby

Date

Between

01/01/1996



and

01/12/2017



Sort by:

All formats

Adobe PDF (.pdf)

Adobe PostScript (.ps)

HTML (.htm, .html)

Microsoft Excel (.xls)

Microsoft PowerPoint (.ppt)

Microsoft Word (.doc)

Rich Text Format (.rtf)

Format

Show the pages in the format:

All formats

**Advanced search:
did Michael Jackson ever came
to Portugal?**

Results in Portuguese can be automatically translated to answer our question.

ARQUIVO.PT

Portugal "michael jackson" -albums Search

between: 01/01/1996 and: 01/12/2017

Advanced search

[Revista TIME Michael Jackson - Abitureiras \(Santarém\) - Livr...](#)
10 April, 2015 - List versions
Rochas Ornamentais Portuguesas - 3 Volumes » Fotos: Fotos de Revista TIME Michael Jackson - Abitureiras > Revista TIME Michael Jackson Classificados grátis - Portugal Copyright © 2015 TIME Michael Jackson - Abitureiras (Santarém) - Livros - Revistas Revista TIME Michael Jackson - Abitureiras (Santarém) - Livros - Revistas Revista TIME Michael Jackson - Abitureiras (Santarém) - Livros - Revistas
<http://abitureiras.olx.pt/revista-time-michael-jackson-iid-460216091>

[Michael Jackson – Wikipédia, a enciclopédia livre](#)
11 April, 2015 - List versions
Michael Jackson – Wikipédia, a enciclopédia livre Michael Jackson Origem: Wikipédia, a enciclopédia livre. Michael Jackson. Nota: Para outros significados, veja Michael Jackson. Michael Jackson durante performance ao vivo de Wanna Be ...
http://pt.wikipedia.org/wiki/Michael_Jackson

Michael Jackson – Wikipé x

Not secure | arquivo.pt/wayback/20150411005521/http://pt.wikipedia.org/wiki/Michael_Jackson

Apps Send Link Google Bookmark publico - Google Drive portugal 2/3

ARQUIVO.PT

originais de Michael foram remixadas por Justin Timberlake e Kevin Antunes (produtor de Rihanna) ao longo de um trabalho de anos no estúdio, sendo liberados pela Epic Records em conjunto com o espólio de Michael Jackson. Apesar de mais de 60 canções serem usadas para a The Immortal World Tour, a versão simples contendo 15 faixas e versão Deluxe com do

terá duas versões a

Em Portugal

Michael Jackson actuou apenas uma vez em Portugal.

O espectáculo foi a 26 de Setembro de 1992 no Estádio Jo

duas vezes, dizendo "I love you" ("amo vocês") a começar e

No Brasil

Michael esteve três vezes no Brasil²⁶.

A primeira vez em que Jackson veio ao Brasil foi em setem

grupo Jackson 5, durante uma turnê pela América Latina. N

irmãos no então

tando-se em São

Paulo, Brasília, Porto Alegre, Salvador, Rio de Janeiro e Belo Horizonte", além de se apresentar para a Rede Tupi.

Back Alt+Left Arrow

Forward Alt+Right Arrow

Reload Ctrl+R

Save as... Ctrl+S

Print... Ctrl+P

Cast...

Translate to English

Show Anchors

View page source Ctrl+U

View frame source

Reload frame

Inspect Ctrl+Shift+I

Did Michael Jackson ever come to Portugal? Yes, in 1992.

The screenshot shows a web browser window with the URL `arquivo.pt/wayback/20150411005521/http://pt.wikipedia.org/wiki/Michael_Jack...`. The page content includes a paragraph about a CD release and a section titled "In Portugal" which states: "Michael Jackson performed only once in Portugal . The show was the 26 of September of 1992 at the Estádio José Alvalade in the capital Lisbon . The singer only spoke to the audience twice, saying "I love you" ("I love you") the start and "peace" ("Peace") in farewell. ³³

Access the Past Web

Access web pages from the Past

Navigate along past versions

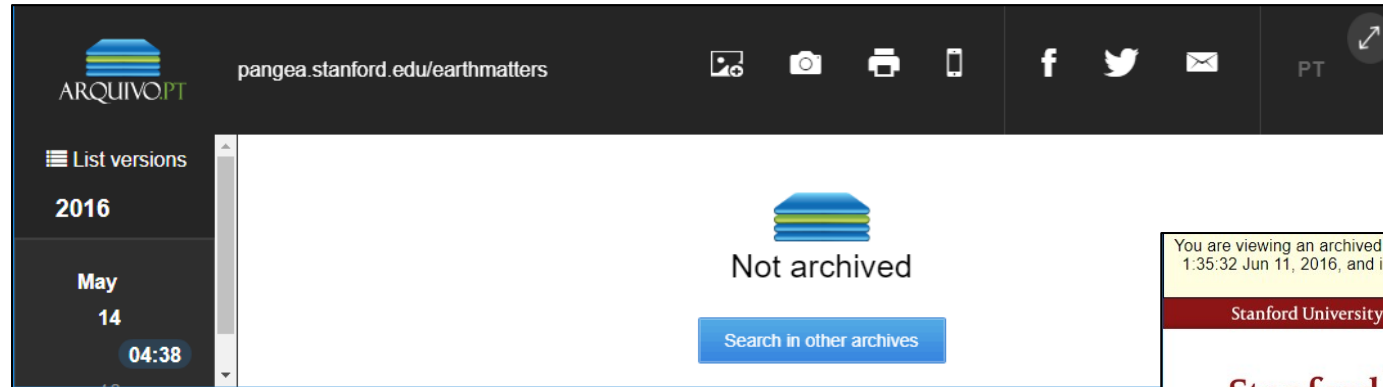
The screenshot displays the ARQUIVOPT interface for navigating through past versions of the Greenpeace International website. The interface is divided into two main sections:

- Left Sidebar (Navigation):**
 - Top: ARQUIVOPT logo and "List versions" link.
 - Year selection: 2008, 2009, 2010, 2011, 2012, 2013.
 - Month selection: February, March, October.
 - Date selection: February 15, 04:35; June 23, 22:16, 22:40; September; December.
- Main Content Area (Website Preview):**
 - Top: ARQUIVOPT logo, URL (www.greenpeace.org/international/), and social media icons.
 - Header: GREENPEACE International logo and search bar.
 - Content: "Welcome to Greenpeace International" with a navigation menu (HOME, About Greenpeace, What we do, Greenpeace victories, Blogs) and a "GO TO YOUR NATIONAL SITE" dropdown.
 - Bottom: "What we do: the issues we work on worldwide" section with sub-sections like "Stop climate change" and "Protect ancient forests", each with an "Enlarge Image" and "Read More" link.

Browse links across pages

- Destiny version with the closest date to the origin page

Navigate between web archives



ARQUIVOPT pangea.stanford.edu/earthmatters

List versions

2016

May 14 04:38

Not archived

Search in other archives



You are viewing an archived web page, collected at the request of [Stanford University Archives](#) using [Archive-It](#). This page was captured on 1:35:32 Jun 11, 2016, and is part of the [Stanford University Website Collection](#) collection. The information on this web page may be out of date. See [All versions](#) of this archived page.

Stanford University Login with your SUNetID

Stanford | SCHOOL OF EARTH, ENERGY & ENVIRONMENTAL SCIENCES

ABOUT ▾ PEOPLE ▾ ACADEMICS ▾ RESEARCH ▾ NEWS & EVENTS ▾ RESOURCES ▾ Search this site...

earthmatters

science and insights for people who care about Earth, its resources and its environment

JUNE 9, 2016

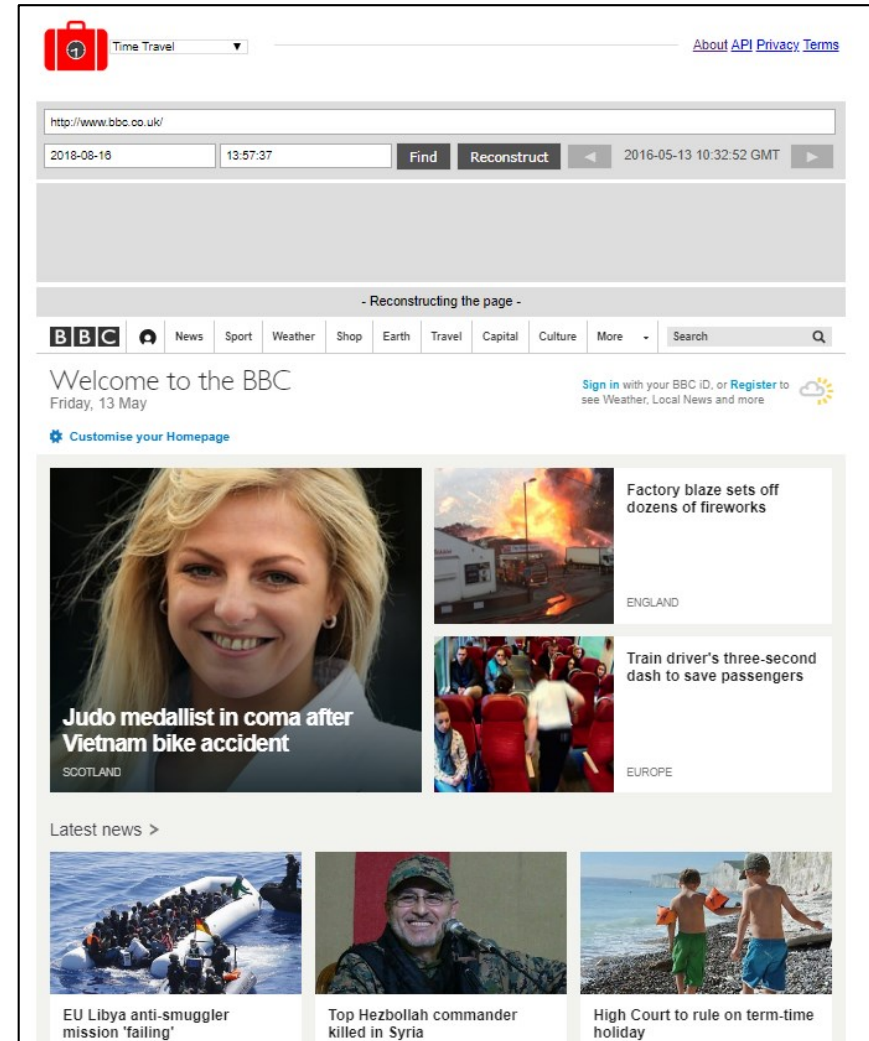
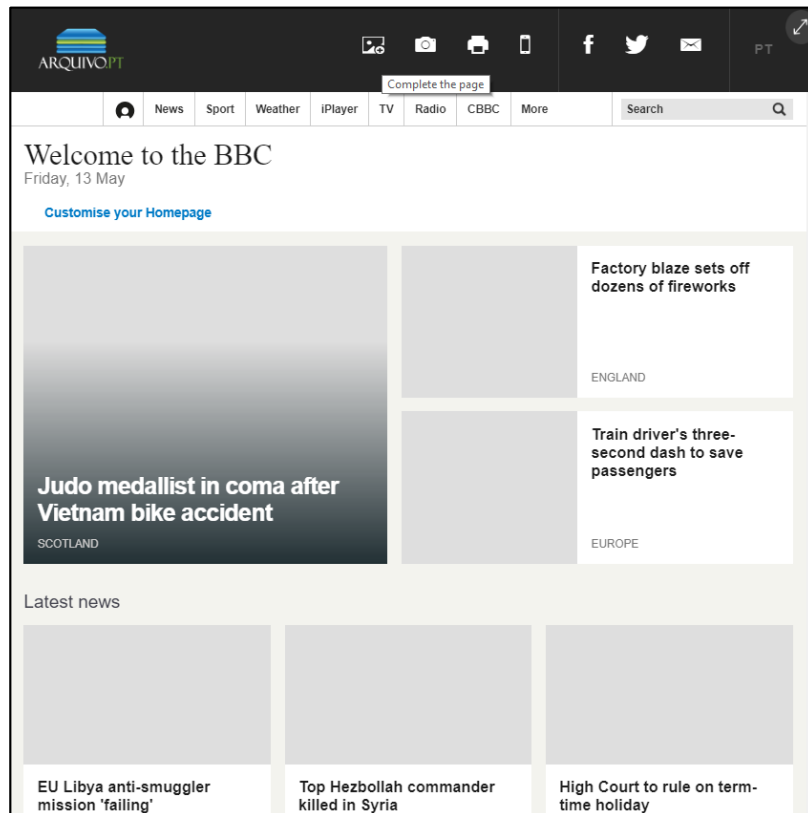
Stanford researchers calculate groundwater levels from satellite data

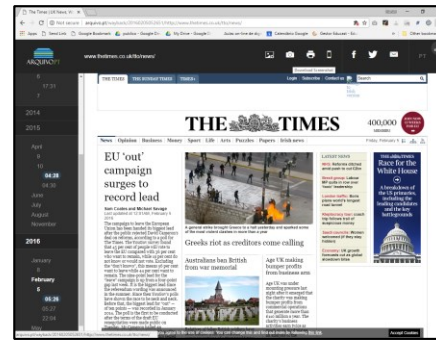
STANFORD SCHOOL OF EARTH, ENERGY & ENVIRONMENTAL SCIENCES

A new computer algorithm that can "fill in" underground water levels in areas where quality data is not available could lead to improved models of groundwater flow in regions

Powered by:
TimeTravel.memento.org

Complete page with resources from external web archives





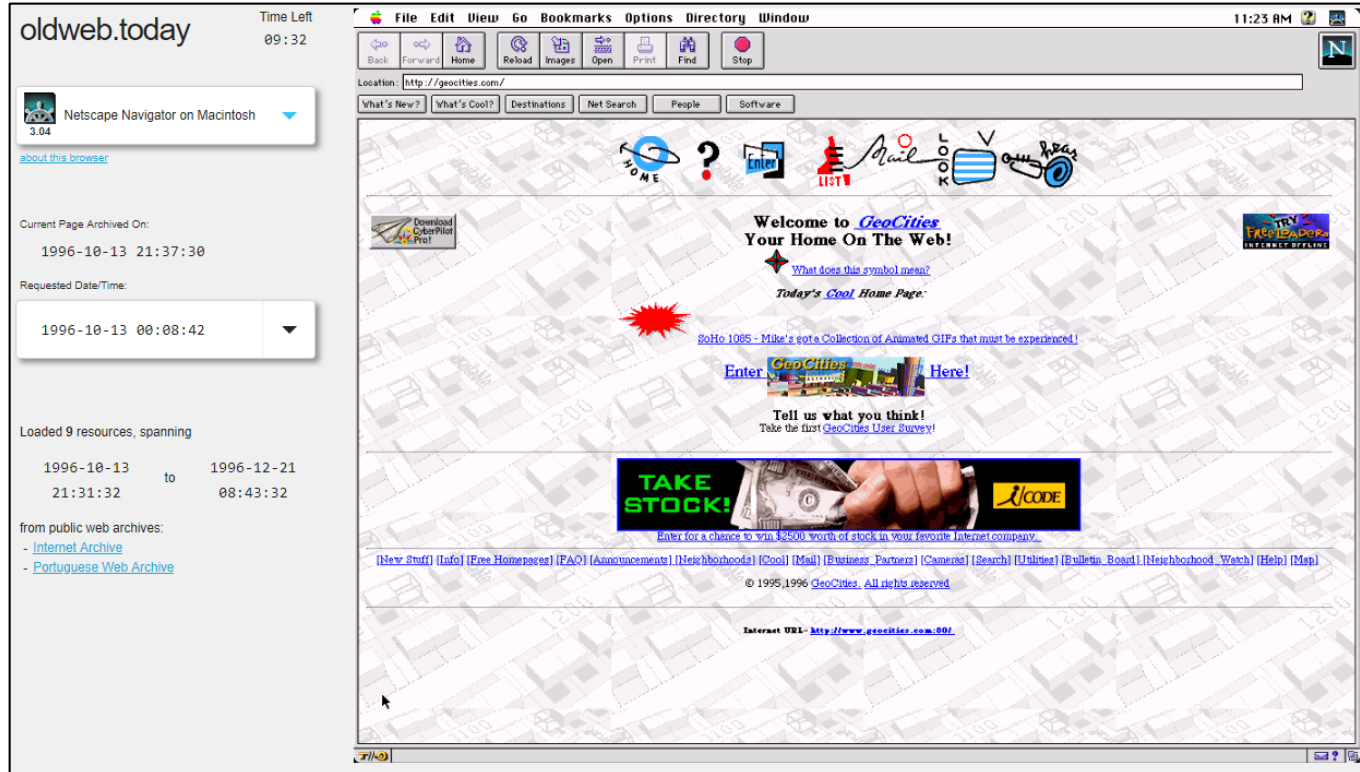
Save full page instead of partial
“Print Screen”

Embed in documentation (e.g.
thesis, papers)

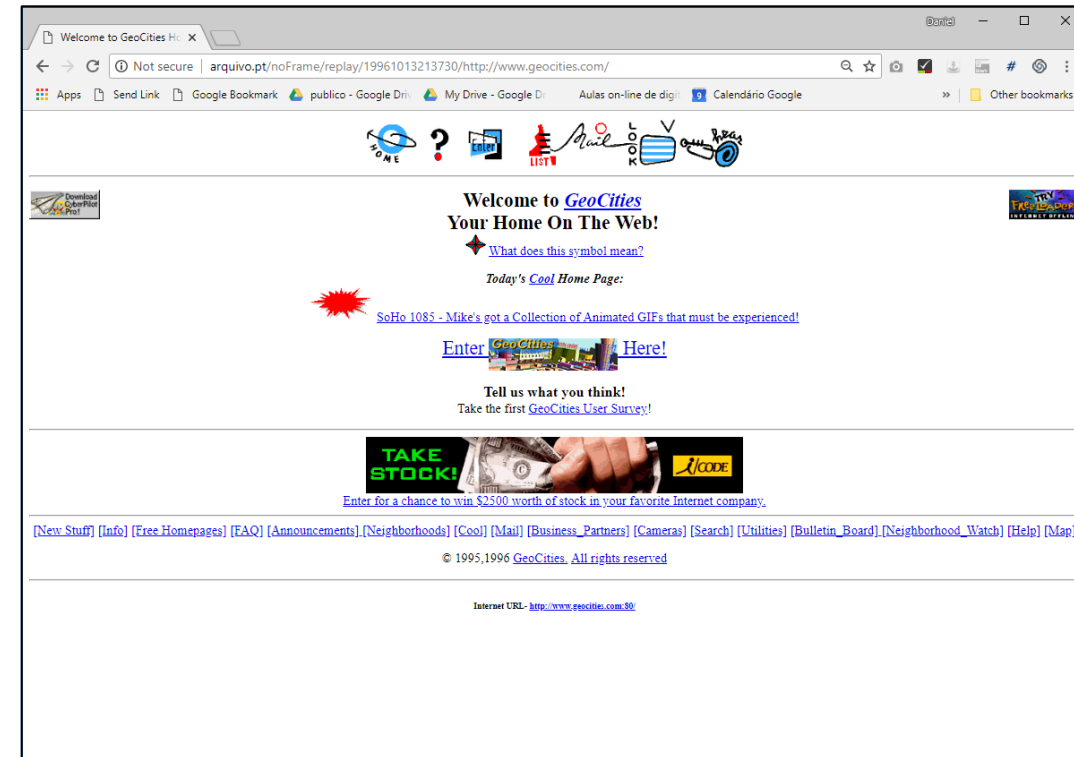
Keep local copies

Oldweb.today: geocities.com, 1996

1997 browser:
Netscape Navigator v.3.04, Macintosh



2018 browser:
Google Chrome v.68, Windows 10



Search and Access the Past Web

Search the Present -> Web Search Engines

Search the Past -> Web Archives!

Free online services

- [Archive.org](https://archive.org): URL search
- [TimeTravel.mementoweb.org](https://timetravel.mementoweb.org): search across web archives
- [Arquivo.pt](https://arquivo.pt): full text search
- [Oldweb.today](https://oldweb.today): access using old browsers

Tutorial outline

Research the Past Web using Web archives

1. Search and access

- The Past Web: examples and use cases
- Public online services

2. Publish and preserve

- Recommendations to publish preservable information
- How to preserve information collected from the Web

3. Automatic processing

- Protocols
- Application Programming Interfaces (API)