Using Wayback Machine for Research

Nicholas Taylor
Repository Development Group
What Is the WAYBACK MACHINE?
WABAC Machine?
Internet Archive’s Wayback Machine

Surf the web as it was
not one, but many Wayback Machines

- open source software to “replay” web archives
  - rewrites links to point to archived resources
  - allows for temporal navigation within archive
- used by many web archiving institutions
  - 33 out of 62 initiatives listed on Wikipedia
Government of Canada Web Archive

Library and Archives Canada
www.collectionscanada.gc.ca

Français | Home | Contact Us | Help | Search | canada.gc.ca
---|---|---|---|---|---
Home > Government of Canada Web Archive > Versions

Government of Canada Web Archive

Versions

URL: http://www.accord-treaty.gc.ca/main.asp?language=0

2006-12-09
2007-11-14

Date Created: 2007-08-09
Date Modified: 2007-10-17

Top of Page
Fernando Couto - O grande capitão. Entrou na prova como titular, mas não resistiria à revolução de Zidane, na sequência da derrota com a Grécia. A partir daí, Ricardo Carvalho mostrou ser dono incontestável da defesa portuguesa. A Couto só voltaria a ter (curto) espaço, a reforçar a defesa nos últimos minutos dos jogos com a Espanha e Holanda.

Figo - Sobre ele pesava a responsabilidade de conduzir a equipa portuguesa. Acusou também o peso da época esgotante que teve no Real Madrid, não conseguindo manter o ritmo nos 90 minutos. Reagiu mal às substituições, comportou-se como um “menino mimado” na conferência de imprensa antes do jogo com a Espanha; jogou - não de raiva, mas com prazer - com a Holanda, merecendo o troféu que o distinguiu como melhor jogador. De qualquer forma, decisivo na solidão da equipa portuguesa, formando um meio-campo de luxo com Costinha, Deco, Maniche e Cristiano Ronaldo.

França - Uma equipa nos limites. A sua solidão competitiva, a qualidade técnica de seus guerreiros, deveu-se, como a ideia de que a qualquer momento...
BioSingapore - An Industry Association 
Representing Singapore's Biosciences Enterprises

BioSingapore was established in May 2004 as an industry association for life sciences businesses in Singapore with the support of the Economic Development Board of Singapore.

BioSingapore was formally launched on 12 October 2004 at Biomedical Asia 2004. The launch is seen as an important milestone in the development of Singapore's biomedical sector.
## Web Archive Singapore

### Search Results for Jan 1, 2006 - Jan 1, 2013

<table>
<thead>
<tr>
<th>Year</th>
<th>Pages</th>
<th>Dates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 page</td>
<td>Aug 13, 2007</td>
</tr>
<tr>
<td>2010</td>
<td>0 pages</td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>0 pages</td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td>0 pages</td>
<td></td>
</tr>
</tbody>
</table>
Catalonian Web Archive
California Water Districts Web Archive

Title: BAWSCA
Archival URL: http://webarchives.cdlib.org/srv1/6685n33/ftp://bawasca.org/
Date captured: 04/23/2012 11:03 PM
About this archive

This document is an archived copy for study and research. The original may be available at http://bawasca.org/

About BAWSCA  Board of Directors  Agendas & Documents  Annual Survey  Water Supply  Links to Water Agencies  News & Press  Contact Us

Search

Highlights

Long Term Reliable Water Supply Strategy
A reliable supply of water is required to support the health, safety, employment, and economic opportunities of the existing and expected future residents in the BAWSCA service area. BAWSCA is working with its member agencies to develop a strategy that ensures reliable supplies of water will be available when and where needed through 2035. → more

Water Supply Agreement Finalized
The Water Supply Agreement between the City and County of San Francisco and Wholesale Customers has been finalized. → more

2012 Silicon Valley Water Conservation Awards
The Silicon Valley Water Conservation Awards recognize organizations, agencies, businesses and individuals whose programs and leadership have advanced water conservation in Silicon Valley. → more

The Bay Area Water Supply and Conservation Agency represents the interests of 24 cities and water districts, and two private utilities, that purchase water wholesale from the San Francisco regional water system. These entities provide water to 1.7 million people, businesses and community organizations in Alameda, Santa Clara and San Mateo counties.

Our goals are to ensure:

- A reliable water supply
- High-quality water
- A fair price
FRIDAY, JULY 11, 2008

9 Month Check Up - Interesting

Nate had his 9 month check up yesterday. It was interesting.
First off, we had to wait in the drs office for over an hour before we were seen. Seriously, kids are NOT that patient. Nate was very very good, considering.
We finally saw the dr, and Nate wasn't impressed. He immediately looked like he was going to start crying when he saw her.
He didn't cry, but looked like he was on high alert the whole time.
We talked to her about the problems we have been having at meal times. IE: WILL NOT EAT. Nates newest thing is just to spit pretty much every ounce of food you try to feed him, out at you. He has pretty much become a human food sprinkler.ICK!
She said that kids go through stages like this, albeit, normally at later ages.

This is where it gets crazy. During this conversation, the big A word was brought up. AUTISM. Wait, WHAT?!
I am still naure what the fuck she was talking about, for continue...
LIMITATIONS AND WORKAROUNDS

Common
How Much Information? 2003

Summary

Information Flow

Stored Information

Wrap-up

Executive Summary

I. Summary of Findings

II. Method

III. Stored Information

A. Paper
B. Film
C. Magnetic
D. Optical

IV. Information Flows

A. Broadcasting
B. Telephony
C. Internet

V. Qualifications

A. Limitations

How much new information is created each year? Nearly all created information is stored in four physical media: print, film, magnetic, and optical, and seen or heard in four information flows: through electronic channels, telephone, radio, and TV, and the Internet. This study of information storage and flows analyzes the year 2002 in order to estimate the annual size of the stock of new information recorded in storage media, and heard or seen each year in information flows. Where reliable data was available we have compared the 2002 findings to those of our 2000 study (which used 1999 data) in order to describe a few trends in the growth rate of information.

1. Print, film, magnetic, and optical storage media produced about 5 exabytes of new information in 2002. Ninety-two percent of the new information was stored on magnetic media, mostly in hard disks.

   How big is five exabytes? If digitized with full formatting, the seventeen million books in the Library of Congress contain about 136 terabytes of information, five exabytes of information is equivalent in size to the information contained in 57,000 new libraries the size of the Library of Congress book collections.

   Hard disks store most new information. Ninety-two percent of new information is stored on magnetic media, primarily hard disks. Film represents 7% of the total, paper 0.01%, and optical media 0.002%.

   The United States produces about 40% of the world's new stored information, including 33% of the world's new printed information, 30% of the world's new film titles, 40% of the world's information stored on optical media, and about 50% of the information stored on magnetic media.

   How much new information per person? According to the Population Reference Bureau, the world population is 6.3 billion, thus almost 800 MB of recorded information is produced per person each year. It would take about 30 feet of books to store the equivalent of 800 MB of information on paper.
workaround: hide the banner
limitation: AJAX-enabled sites
limitation: AJAX-enabled sites
workaround: disable JavaScript
limitation: nav menu link errors
workaround: insert live site URL in archive
workaround: insert live site URL in archive
workaround: insert live site URL in archive
limitation: no full-text search
workaround: none yet, but R&D ongoing
Basic MECHANICS
## structure of a Wayback Machine URL

<table>
<thead>
<tr>
<th>Wayback Machine URL</th>
<th>collection</th>
<th>date/timestamp (YYYYMMDDHHMMSS)</th>
<th>URL of archived resource</th>
</tr>
</thead>
</table>
URL-based access
URL-based access

United States Senator Orrin Hatch

Committee Assignments

Senator Hatch is the ranking member of the Senate Committee on Finance, where he has been a leading advocate for policies to encourage savings and investment, such as the Capital Formation Act, and many other pro-growth tax bills. He has proposed a tax credit for families to defray adoption expenses, extension of the research tax credit, repeal of the estate tax,
### Search Results for Jan 1, 2003 - Dec 31, 2012

<table>
<thead>
<tr>
<th>Year</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pages</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>4</td>
</tr>
</tbody>
</table>

- Apr 6, 2011
- May 5, 2011
- Jun 2, 2011
- Jul 7, 2011
- Aug 4, 2011
- Aug 15, 2011
- Sep 6, 2011
- Oct 7, 2011
- Nov 3, 2011
- Dec 7, 2011
- 2012
- Jan 5, 2012
- Feb 1, 2012
- Mar 7, 2012
- Apr 4, 2012
date wildcarding
You are viewing a Web site, archived on 5:03:55 Jun 03, 2011, that is part of a Library of Congress Web Archive Collection.
External links, forms, and search boxes may not function within this collection. [hide]

http://civilwarexpo.tripod.com/

Nominator review

American Civil War Sesquicentennial Traveling Educational Exposition

The American Civil War Sesquicentennial Traveling Educational Exposition is designed to celebrate & educate the 150th Anniversary of the events that lead to the War of the Rebellion (Northern), the War of Northern Aggression (Southern) into Reconstruction.

The ACW was fought for many reasons and under many names. We look to these next six years to breakdown those reasons and help everyone learn from the mistakes & lessons of the past.

George Santanyana (b. 1863 d. 1952), Spanish born American Philosopher Once stated.

“Those who do not learn from history are doomed to repeat it”
document wildcarding
Strategies for

FINDING MISSING RESOURCES
removed or moved?

- don’t start with the archive
- missing resources have often just moved (Klein & Nelson, 2010)
- Synchronicity for Firefox helps find new location
  - scrapes archived version for “fingerprint” keywords; uses them to query search engines
MementoFox

The Library of Congress

Get It Online... Words, Pictures & Sound

Find It...

- Search Our Catalog
- Advanced Search
- Research Centers
- Index A-Z
- Site Map
- Ask a Librarian
- Today in History

Especially For...

- Researchers
- Law Researchers
- Librarians & Archivists
- Teachers
- Publishers
- Persons with Disabilities
- Blind Persons

Information & Services

- US Copyright Office
- NDIPP
- About the Library
- Visit Tour
- Philanthropy
- Jobs & Fellowships
- Shop

More Online Collections...

News & Events

Veterans History Project
Collecting stories of war & service

- Civil War Symposium
  Conference at Library Nov. 12-14
- Featured Cybercast
  War Letters

Complete News & Events

The Library of Congress

Contact Us

Please Read Our Legal Notices
find archives for a site whose URL has changed

- website URL changed recently
- historical URL is unknown
- **solution**: use search engine to find historical URL then apply it in the archive
Federal IT Dashboard
check Internet Archive’s Wayback Machine
IA Wayback coverage goes back to July 2010
LCWA only goes back to June 2011
use search engine to find historical URL
use search engine to find historical URL

Federal IT Dashboard
www.itdashboard.gov

IT Dashboard | The White House
www.whitehouse.gov/open/innovations/it-dashboard

IT Dashboard | drupal.org
drupal.org/node/1103308

Performance Dashboards | USA.gov
www.usa.gov/Contact/Engagement/Dashboards.shtml

IT Dashboard from the Office of the National Coordinator...
Open Government Initiative
TRANSPARENCY • PARTICIPATION • COLLABORATION


IT Dashboard

Federal US Chief Information Officer (CIO), Office of Management and Budget

Analyze and Evaluate a Portfolio of Over 7,000 Federal Information Technology Investments

Welcome to IT Dashboard

[Video player]
note the redirect from http://it.usaspending.gov/
append URL to IA Wayback URL
append URL to LC Wayback URL

<table>
<thead>
<tr>
<th>Year</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pages</td>
<td>0 pages</td>
<td>0 pages</td>
<td>0 pages</td>
<td>0 pages</td>
<td>0 pages</td>
<td>0 pages</td>
<td>12 pages</td>
<td>6 pages</td>
<td>43 pages</td>
<td>24 pages</td>
</tr>
</tbody>
</table>

 Wayback 1.6

Searched for [http://it.usaspending.gov/](http://it.usaspending.gov/)
find archives for a site whose URL has changed

- congressional committee hearings archive
- live site URL doesn’t work in archive
- **solution**: find a site in the archive that would link to the desired site, then navigate to contemporaneous snapshot
hearings archive only spans 2001-2006
hearings archive URL changed in 2011

<table>
<thead>
<tr>
<th>Year</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pages</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>15</td>
<td>2</td>
</tr>
</tbody>
</table>

Search Results for Jan 1, 2003 - Dec 31, 2012

- Jan 6, 2011
- Jan 16, 2011
- Feb 3, 2011
- Feb 11, 2011
- Mar 3, 2011
- Apr 6, 2011
- May 4, 2011
- Jun 1, 2011
- Jul 5, 2011
- Aug 4, 2011
- Aug 15, 2011
- Sep 8, 2011
- Oct 7, 2011
- Nov 2, 2011
- Dec 7, 2011
- Jan 5, 2012
- Feb 1, 2012
- Mar 7, 2012
- Apr 4, 2012
- May 9, 2012

Set Anchor Window: none
### Search Results for Jan 1, 2003 - Dec 31, 2012

|-------|----------|----------|----------|----------|----------|----------|----------|----------|----------|

**Search Results for Jan 1, 2003 - Dec 31, 2012**

- Set Anchor Window: none
- 541 Results
You are viewing a Web site, archived on 05:24 Nov 05, 2010, that is part of a Library of Congress Web Archive Collection. External links, forms, and search boxes may not function within this collection. 

http://www.energycommerce.house.gov/

If you are viewing a Library of Congress Web archive, this site has been archived as part of a Library of Congress Web Archive Collection. External links, forms, and search boxes may not function within this collection.

Snapshot from prior to site change
navigate to appropriate section
navigate to appropriate section
find archives for a previously accessible webpage

- records currently stored in password-protected part of site may have previously been publicly-accessible
- conceptual site organization lasts longer than exact link construction
- **solution**: figure out where desired resource would be on the live site, then navigate to analogous section on archived site
location of resources on live site
location of resources on live site
authentication required
check the site in the archive
navigate to an individual capture
Events Calendar

Below is a list of upcoming events related to healthcare transformation. These events may range from webcasts to conferences to committee hearings on Capitol Hill.

View only... View All  go

July 14, 2004
Washington, DC
Newt Gingrich Testified at House Subcommittee Hearing
Type of event: Committee Hearings | Time: 2:00 PM

July 17, 2004
Seattle, WA
Newt Gingrich to participate on panel at the National Governors Association Annual Meeting
Type of event: Public Policy Initiatives | Time: 2:30 PM - 4:00 PM

July 21, 2004
Washington, DC
Newt Gingrich to Provide Closing Remarks at the Secretaryial Summit on Health
navigate to appropriate section
GET INVOLVED

How You Can
help us to help you

- what websites from today would you want to be able to consult in five, ten, twenty years’ time?
- have you told us what is important to capture?
End of Term 2012

The End of Term project has resumed for an End of Term 2012-2013 archive, and we need help to identify websites for collection, particularly those that might be most at-risk of change or deletion at the end of the presidential term. If you or your organization would like to participate by nominating important websites for capture to help us build this archive, please use our Nomination tool. Read more about the project here.

Who can help:

We're looking for government document experts, subject experts, and anyone interested in helping identify U.S. Federal Government websites for collection and preservation.

What you can do:

Nominations of any U.S. Federal Government domains are welcome. There are a few topic areas that we particularly need assistance identifying, including but not limited to:

- Judicial Branch websites
- Important content or subdomains on very large websites (such as NASA.gov) that might be related to current Presidential policies
- Government content on non-governmental domains (.com, .edu, etc.)

Volunteer nominators will be asked to contribute as much time and effort as they are able, whether it be a nomination of 1 website or 500 websites. Nominators will be given access to the Nomination Tool, which is an easy-to-use tool created by the University of North Texas to help with collaborative selection projects. Volunteers are asked to submit some simple metadata about the site that they are nominating. To participate, please use our Nomination tool.

Project Timeframe

2012

- January 2012: Project team will begin accepting early nominations of priority websites via the Nomination Tool.
USEFUL RESOURCES
CyberCemetery

What is the CyberCemetery?

The CyberCemetery is an archive of government websites that have ceased operation (usually websites of defunct government agencies and commissions that have issued a final report). This collection features a variety of topics indicative of the broad nature of government information. In particular, this collection features websites that cover topics supporting the university's curriculum and particular program strengths.

Who created and maintains the CyberCemetery?

The University of North Texas Libraries and the U.S. Government Printing Office, as part of the Federal Depository Library Program, created a partnership to provide permanent public access to the Web sites and publications of defunct U.S. government agencies and commissions. This collection was named the "CyberCemetery" by early users of the site.

This collection supports the mission of the UNT Libraries by acquiring, preserving, and providing access to recorded knowledge in the form of government websites.

This page is maintained by Starr Hoffman, and was last updated on February 02, 2009.
The Library of Congress Web Archives (LCWA) is composed of collections of archived websites selected by subject specialists to represent web-based information on a designated topic. It is part of a continuing effort by the Library to evaluate, select, collect, catalog, provide access to, and preserve digital materials for future generations of researchers. The early development project for Web archives was called MINERVA.
Project One Web Archives
**links**

- International Internet Preservation Consortium: [http://netpreserve.org/](http://netpreserve.org/)
questions?

webcapture@loc.gov