Using Wayback Machine for Research

Nicholas Taylor
Repository Development Group
What Is the **WAYBACK MACHINE**?
WABAC Machine?
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

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
Archive for the ‘Euro-2004’

EURO 2004 - A to Z (F)

Fernando Couto - O grande capitão. Entrou na prova como titular, mas não resistia à revolução de Scarioli, na sequência da derrota com a Grécia. A partir daí, Ricardo Carvalho mostrou ser dono incontestável da defesa portuguesa... e 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 a troféu que o distinguia como melhor jogador. De qualquer forma, decisivo na solidez da equipa portuguesa, formando um meio-campo de luxo com Coentrinho, Deco, Maniche e Cristiano Ronaldo.

França - Uma equipa nos limites. A sua solidez competitiva, a qualidade técnica dos seus protagonistas, mesmo em mãos de ma, e alguns momentos
Catalonian Web Archive
Catalonian Web Archive
California Digital Library Web Archiving Service

California Water Districts Web Archive

Title: BAWSCA
Archival URL: http://webarchives.cdlib.org/tw16690n353http://bawscsa.org/
Date captured: 04/23/12 11:00 PM

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

34th Silicon Valley Water Conservation Awards
The Silicon Valley Water Conservation Awards recognizes 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
TRY WHISTLING THIS

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 dr's 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. HE WILL NOT EAT. Nate's 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?!
Common

LIMITATIONS AND WORKAROUNDS
How Much Information? 2003

Summary

Information Flows

Stored Information

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

About this Report
About SIMS
Printable full report (PDF)

Limitation: banner displaces page elements
workaround: hide the banner
limitation: AJAX-enabled sites
limitation: AJAX-enabled sites
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


Wayback Machine URL  collection   date/timestamp  (YYYYMMDDHHMMSS)  URL of archived resource
URL-based access
URL-based access
### 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>0</td>
<td>0</td>
</tr>
</tbody>
</table>

- 2011: 4 pages
- 2012: 10 pages

* indicates pages available.
### Search Results for Jan 1, 2011 - Dec 31, 2011

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Pages</td>
<td>0 pages</td>
<td>0 pages</td>
<td>0 pages</td>
<td>1 page</td>
<td>1 page</td>
<td>1 page</td>
<td>1 page</td>
<td>2 pages</td>
<td>1 page</td>
<td>1 page</td>
<td>1 page</td>
<td>1 page</td>
</tr>
</tbody>
</table>

Date: *Apr 6, 2011* *May 5, 2011* *Jun 2, 2011* *Jul 7, 2011* *Aug 4, 2011* *Aug 16, 2011* *Sep 8, 2011* *Oct 7, 2011* *Nov 3, 2011* *Dec 7, 2011*
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
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
find archived content now at a new URL

- congressional committee hearings archive
- live site URL doesn’t work in archive
- 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>5</td>
</tr>
</tbody>
</table>

- Jan 8, 2011
- Jan 16, 2011
- Feb 2, 2011
- Feb 8, 2011
- Mar 3, 2011
- Apr 6, 2011
- May 4, 2011
- Jun 1, 2011
- Jul 7, 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
truncation archival access URL
snapshot from prior to site change
navigate to appropriate section
navigate to appropriate section
find archived content now at a new URL

- records currently stored in password-protected part of site may have previously been publicly-accessible
- conceptual site organization lasts longer than exact link construction
- 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
navigate to appropriate section
navigate to appropriate section
How You Can

GET INVOLVED
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?
for more information

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

webcapture@loc.gov