Memento: Time Travel for the Web

http://www.mementoweb.org

Herbert Van de Sompel – hvdsomp@gmail.com
Michael L. Nelson – mln@cs.odu.edu
Robert Sanderson – azaroth42@gmail.com

The Memento Experiment was partly funded by the Library of Congress

Based on the paper at http://arxiv.org/abs/0911.1112
Memento Team

• At the Los Alamos National Laboratory, Prototyping Team:
  o Lyudmilla Balakireva
  o Robert Sanderson
  o Harihar Shankar
  o Herbert Van de Sompel

• At Old Dominion University, Web Science and Digital Library Research Group:
  o Scott Ainsworth
  o Michael Nelson
Looking at the Past can be Fun …
Looking at the Past can be Fun

Feb 14 2006

Cheney prays for hunt victim
Looking at the Past can be Fun

Feb 14 2006

Press Attacks Cheney
And Memento wants to make it Easy
Recap of the Basics …
W3C Web Architecture: Resource – URI - Representation

Diagram:
- URI
- Resource
- Representation

Relationships:
- URI identifies Resource
- Resource represents Representation
- dereference
W3C Web Architecture: Resource – URI - Representation

URI

Identifies

Resource

Represents

Representation 1

Represents

Representation 2
Problem Statement …
Resources have Representations

resources have representations
Resources have Representations that Change over Time
Only the Current Representation is Available from a Resource
Old Representations are Lost Forever
Archived Resources Exist

the old representations are lost forever

well, that is not not entirely true
** Archived Resources **


---

**Memento: Time Travel for the Web**

Herbert Van de Sompel, Michael L. Nelson, Robert Sanderson

CNI Fall 2009 Task Force Meeting, Washington, DC, December 15 2009
Finding Archived Resources

Go to http://www.archive.org/ and search http://cnn.com

Finding Archived Resources

Go to http://en.wikipedia.org/wiki/September_11_attacks and click History

Memento: Time Travel for the Web
Herbert Van de Sompel, Michael L. Nelson, Robert Sanderson
CNI Fall 2009 Task Force Meeting, Washington, DC, December 15 2009
Navigating Archived Resources


Dec 20 2001, 4:51:00 UTC

Pentagon

Current

http://en.wikipedia.org/wiki/The_Pentagon
Navigating Archived Resources

archived resource for http://cnn.com


Memento: Time Travel for the Web
Herbert Van de Sompel, Michael L. Nelson, Robert Sanderson
CNI Fall 2009 Task Force Meeting, Washington, DC, December 15 2009
Current and Past Web are Not Integrated
The Memento Solution …
Navigate the Web of the Past

http://en.wikipedia.org/wiki/Web_Archiving

Oct 11 2009, 05:30:33 UTC

From Wikipedia History

Oct 01 2009, 16:30:00 UTC

Robots Exclusion Protocol

Oct 11 2009, 05:30:33 UTC
Navigate the Web of the Past


From Wikipedia History

Oct 11 2009, 05:30:33 UTC

Robots Exclusion

From Wikipedia, the free encyclopedia

This is an old revision of this page, as edited by 128.188.120.54 (Talk) at 20:49, 15 September 2009. It may differ significantly from the current revision.

(diff) ← Previous revision | Current revision | (compare) → (diff)

For restricting Wikipedia bots, see Template:Bots.

The Robots Exclusion Standard, also known as the Robots Exclusion Protocol or robots.txt protocol, is a convention for informing web crawlers about which parts of a website are not to be crawled. It is usually specified in the robot.txt file at the root of a website. It is often used by search engines to categorize and archive web sites, or by webmasters to protect read source codes. The standard is asymmetrical to, but can be used in conjunction with, sitemaps, a robot inclusion standard for websites.

Contents

1 History
2 About the standard
3 Disadvantages
4 Automated Content Access Protocol
5 Examples
6 Nonstandard extensions
7.1 Crawl-delay directive
7.2 Allow directive
7.3 Sitemap
7.4 Extended standard
8 See also
9 References
10 External links

History

robots.txt was popularized with the advent of AltaVista, the first popular search engine.

About the standard

If a site owner wishes to give instructions to web robots he must put a text file called robots.txt to the root of the web site hierarchy (e.g. www.example.com/robots.txt). This text file should contain the instructions in a specific format (see examples below). Robots that wish to follow the instructions try to fetch this file and read the instructions before fetching any other file from the site. If the file doesn’t exist web robots assume that the web owner wishes to provide no specific instructions.

A robots.txt file can be, for example, extra, be misleading or

For websites with multiple subdomains, each subdomain must have its own robots.txt file. If example.com had a robots.txt file but a.example.com did not, the rules that would apply for example.com would not apply to a.example.com.

Disadvantages

The protocol is purely advisory. It relies on the cooperation of the web robot, so that making an area of a site out of bounds with robots.txt does not guarantee privacy. Some web site administrators have tried to use the robots.txt to make private parts of a website invisible to
Navigate the Web of the Past

http://www.robotstxt.org/

Oct 11 2009, 05:30:33 UTC

From Internet Archive

Nov 09 2007, 06:21:04 UTC

Memento: Time Travel for the Web
Herbert Van de Sompel, Michael L. Nelson, Robert Sanderson
CNI Fall 2009 Task Force Meeting, Washington, DC, December 15 2009
How does Memento do This?

There are two components to the Memento Solution:

• **Component 1:** Navigation towards an archived resource via its original resource, by leveraging content negotiation.

• **Component 2:** A discovery API for archives that allows requesting a list of all archived versions it holds for a resource with a given URI.
How does Memento do This?

• **Component 1: Navigation towards an archived resource via its original resource, by leveraging content negotiation.**
The Web without a Time Dimension

Original resource

Archived resources

HTTP GET

URI-R

HTTP GET

URI-M₁

HTTP GET

URI-M₂

HTTP GET

URI-M₃
The Web with Time Dimension added by Memento
The Web with Time Dimension added by Memento
How does Memento do This?

In order to fully understand how Memento introduces a time dimension to the Web, we present a brief recap of Transparent Content Negotiation (conneg) in HTTP.

GET with conneg on URI T – Server Choice – 302 Found – Step 1

```
GET T HTTP/1.1
Accept: text/html, application/pdf;q=0.8
Accept-Language: en-US, fr;q=0.7, de;q=0.5

HTTP/1.1 302 Found
TCN: choice
Vary: negotiate, accept, accept-language
Location: A
Content-Type: text/html
Content-Language: en
Alternates: "A" 1.0 {type text/html}
           {language en}, "B" ..., "C" ...
```
GET with conneg on URI T – Server Choice – 302 Found – Step 2
GET with conneg on URI T – Server List – 406 Not Acceptable
How does Memento do This?

• **Component 1:** Navigation towards an archived resource via its original resource, by leveraging content negotiation.
Terminology Intermission

We introduce the term Memento to refer to an archived version of a resource.

A Memento for a resource URI-R (as it existed) at time $t_i$ is a resource URI-M$_i$ [URI-R@$t_i$] for which the representation at any moment past its creation time $t_c$ is the same as the representation that was available from URI-R at time $t_i$, with $t_c \leq t_i$. Implicit in this definition is the notion that, once created, a Memento always keeps the same representation.
DT-conneg: Content Negotiation in the datetime dimension

- RFC 2295 introduces conneg in the following dimensions: media type, language, compression, character set, e.g.:
  
  ```
  Accept-Language: en-US
  ```

- Memento introduces conneg in the datetime dimension:
  
  ```
  X-Accept-Datetime: {Mon, 12 Oct 2009 14:20:33 GMT}
  ```

- This means that *somewhere*, we will need transparently negotiable resources to get to appropriate Mementos.

- This will be discussed for 2 classes of servers.
Servers With Internal Archival Capabilities

- This type includes:
  - Content Management Systems
  - Version Control Systems
  - Servers that archive resource representations in the cloud and keep track of the URIs and datetimes of remotely archived resources.

- These servers have all the essential information (URI-Ms, and associated datetimes) to respond to a DT-conneg request.
Memento: Time Travel for the Web
Herbert Van de Sompel, Michael L. Nelson, Robert Sanderson
CNI Fall 2009 Task Force Meeting, Washington, DC, December 15 2009

Mementos

September 11 attacks

The September 11 attacks were a series of four coordinated terrorist attacks by al-Qaeda against the United States on the morning of Tuesday, September 11, 2001. The attacks are widely considered among the most significant events of the 21st century.

- The attacks were carried out by 19 hijackers who employed four commercial airplanes as weapons.
- Two planes crashed into the Twin Towers of the World Trade Center in New York City.
- A third plane crashed into the Pentagon in Arlington, Virginia.
- A fourth plane crashed in a remote area of rural Pennsylvania.

These attacks resulted in the deaths of nearly 3,000 people, including passengers, crew members, and first responders. The attacks also caused significant damage to infrastructure and property.

Following the attacks, the United States launched a military campaign against Afghanistan, which led to the removal of the Taliban government.

These events had a profound impact on the United States and the world, leading to increased security measures and changes in international relations.
DT-conneg with URI-R to get URI-M

original resource

Mementos

variant resources

transparently negotiable resource
We introduce the term **TimeGate** to refer to a transparently negotiable resource that supports the datetime dimension.

A TimeGate for an original resource URI-R is a transparently negotiable resource URI-G[URI-R] for which all variant resources are Mementos URI-M_i[URI-R@t_i] of the resource URI-R. Since multiple archives may host versions of URI-R, multiple TimeGates may exist for any given resource, i.e. one per archive.
Memento: Time Travel for the Web
Herbert Van de Sompel, Michael L. Nelson, Robert Sanderson
CNI Fall 2009 Task Force Meeting, Washington, DC, December 15 2009

DT-conneg with URI-G/URI-R to get URI-M

original resource

same

transparently negotiable resource

TimeGate

Mementos

variant resources
Servers With Internal Archival Capabilities: Successful Flow

GET http://en.wikipedia.org/wiki/Clock HTTP/1.1
X-Accept-Datetime: (Sat, Oct 10 2009 00:00:00 GMT)

HTTP/1.1 302 Found
TCN: choice
Vary: negotiate, X-Accept-Datetime
X-Archive-Interval: {Thu, Sep 20 2001 18:58:00 GMT} - {Mon, Oct 19 2009 21:27:00 GMT}
Alternates:
{"http://en.wikipedia.org/w/index.php?title=Clock&oldid=318050912" 1.0 {type text/html} {language en} {dt Fri, Oct 9 2009 03:30:00 GMT}},
{"http://en.wikipedia.org/w/index.php?title=Clock&oldid=319628197" 0.8 {type text/html} {language en} {dt Tue, Oct 13 2009 14:35:00 GMT}},
{"http://en.wikipedia.org/w/index.php?title=Clock&oldid=318683946 " 0.6 {type text/html} {language en} {dt Thu, Oct 8 2009 16:31:00 GMT}}

GET http://en.wikipedia.org/w/index.php?title=Clock&oldid=318050912 HTTP/1.1
X-Accept-Datetime: (Sat, Oct 10 2009 00:00:00 GMT)

HTTP/1.1 200 OK
X-Archive-Interval: {Thu, Sep 20 2001 18:58:00 GMT} - {Mon, Oct 19 2009 21:27:00 GMT}
X-Datetime-Validity: {Fri, Oct 9 2009 03:30:00 GMT} - {Tue, Oct 13 2009 14:35:00 GMT}
<table>
<thead>
<tr>
<th>Case</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Client issues request without datetime content negotiation against original resource URI-R that functions as its own TimeGate.</td>
</tr>
<tr>
<td></td>
<td>GET URI-G HTTP 1.1</td>
</tr>
<tr>
<td>2</td>
<td>Server that is the target of request [1] responds:</td>
</tr>
<tr>
<td>Case 1</td>
<td>URI-R exists.</td>
</tr>
<tr>
<td></td>
<td>HTTP/1.1 200 OK</td>
</tr>
<tr>
<td></td>
<td>Content-Location: URI-R</td>
</tr>
<tr>
<td>Case 2</td>
<td>URI-R does not exist.</td>
</tr>
<tr>
<td></td>
<td>HTTP/1.1 404 Not Found</td>
</tr>
<tr>
<td>3</td>
<td>Client issues datetime content negotiation request against Memento URI-M for original resource URI-R:</td>
</tr>
<tr>
<td></td>
<td>GET URI-M HTTP 1.1</td>
</tr>
<tr>
<td></td>
<td>X-Accept-Datetime: (datetime,)</td>
</tr>
<tr>
<td>4</td>
<td>Memento server (same as server of original resource URI-R) that is the target of request [3] responds:</td>
</tr>
<tr>
<td>Case 1</td>
<td>Server detects X-Accept-Datetime &amp; URI-M exists.</td>
</tr>
<tr>
<td></td>
<td>HTTP/1.1 200 OK</td>
</tr>
<tr>
<td></td>
<td>Content-Location: URI-M</td>
</tr>
<tr>
<td></td>
<td>X-Archive-Interval: (datetime_start) - (datetime_end)</td>
</tr>
<tr>
<td></td>
<td>X-Datetime-Validity: (datetime_first) - (datetime_last)</td>
</tr>
<tr>
<td>Case 2</td>
<td>Server detects X-Accept-Datetime &amp; URI-M does not exist.</td>
</tr>
<tr>
<td></td>
<td>HTTP/1.1 404 Not Found</td>
</tr>
<tr>
<td>Case 3</td>
<td>Server does not detect X-Accept-Datetime &amp; URI-M exists.</td>
</tr>
<tr>
<td></td>
<td>HTTP/1.1 200 OK</td>
</tr>
<tr>
<td></td>
<td>Content-Location: URI-M</td>
</tr>
<tr>
<td>Case 4</td>
<td>Server does not detect X-Accept-Datetime &amp; URI-M does not exist.</td>
</tr>
<tr>
<td></td>
<td>HTTP/1.1 404 Not Found</td>
</tr>
</tbody>
</table>

Servers Without Internal Archival Capabilities

• This type includes:
  o Servers that are crawled by a web archive
  o Servers with an associated transactional archive

• These servers do not have the essential information (URI-Ms, and associated datetimes) to respond to a DT-conneg request.

• But they can still be really constructive by redirecting (HTTP 302) a client to an archive that can respond to the DT-conneg request.
Memento: Time Travel for the Web
Herbert Van de Sompel, Michael L. Nelson, Robert Sanderson
CNI Fall 2009 Task Force Meeting, Washington, DC, December 15 2009

original resource

transiently negotiable resource

TimeGate

Mementos

variant resources

DT-conneg with URI-G to get URI-M
Memento: Time Travel for the Web
Herbert Van de Sompel, Michael L. Nelson, Robert Sanderson
CNI Fall 2009 Task Force Meeting, Washington, DC, December 15 2009

DT-conneg with URI-G to get URI-M

Mementos
variant resources

TimeGate
transparently negotiable resource

original resource
redirect

Memento Experiment

DT-conneg with URI-G to get URI-M
How to redirect from Original Resource to its (external) TimeGate

- Q1: Which archive to redirect to?
  - The archive with the best coverage for the server at hand.
  - There are quite a few nuances, here.
  - Always redirect to an Aggregator (see later)

- Q2: What is the TimeGate URI-G for URI-R on the chosen archive?
  - Convention for syntax of URI-G as function of URI-R.
  - Always redirect to an Aggregator (see later)
Servers Without Internal Archival Capabilities: Successful Flow

GET http://cnn.com/ HTTP/1.1
X-Accept-Datetime: {Tue, Sep 11 2001 20:35:00 GMT}

HTTP/1.1 302 Found

GET http://web.archive.org/web/timegate/http://cnn.com/ HTTP/1.1
X-Accept-Datetime: {Sat, Oct 10 2009 00:00:00 GMT}

HTTP/1.1 302 Found

TCN: choice
Vary: negotiate, X-Accept-Datetime
X-Archive-Interval: {Tue, Sep 15 2000 11:28:26 EDT} - {Tue, Jul 08 2008 09:34:33 GMT}
rel="aggregation"
Alternates:
1.0 {type text/html} {language en} {dt Tue Sep 11 2001 20:36:10 GMT}},
{http://web.archive.org/web/20010911203051/http://www2.cnn.com/"
0.8 {type text/html} {language en} {dt Tue Sep 11 2001 20:30:51 GMT}},
{http://web.archive.org/web/20010911204733/http://www2.cnn.com/"
0.6 {type text/html} {language en} {dt Tue Sep 11 2001 20:47:33 GMT}}

X-Accept-Datetime: {Tue, Sep 11 2001 20:35:00 GMT}

HTTP/1.1 200 OK
X-Archive-Interval: {Tue, Sep 15 2000 11:28:26 EDT} - {Tue, Jul 08 2008 09:34:33 GMT}
## Servers Without Internal Archival Capabilities: Other Scenarios

<table>
<thead>
<tr>
<th>Case 1</th>
<th>TimeGate server detects X-Accept-Datetime &amp; URI-G exists &amp; datetime can be parsed &amp; datetime is in range for which server has Mementos for original resource URI-R.</th>
</tr>
</thead>
</table>
|        | HTTP/1.1 302 Found  
|        | Location: URI-R  
|        | Vary: negotiate, X-Accept-Datetime  
|        | TCH: choice  
|        | Alternates: {URI-Mi {dt datetime_i}}, {URI-Mk {dt datetime_k}}  
|        | X-Archive-Interval: (datetime_start) - (datetime_end)  
|        | Link: URI-R; rel="aggregation"  

<table>
<thead>
<tr>
<th>Case 2</th>
<th>TimeGate server detects X-Accept-Datetime &amp; URI-G exists &amp; datetime can be parsed &amp; datetime is outside range for which server has Mementos for original resource URI-R.</th>
</tr>
</thead>
</table>
|        | HTTP/1.1 406 Not Acceptable  
|        | Vary: negotiate, X-Accept-Datetime  
|        | TCH: list  
|        | Alternates: {URI-Mi {dt datetime_i}}, {URI-Mk {dt datetime_k}}  
|        | X-Archive-Interval: (datetime_start) - (datetime_end)  
|        | Link: URI-R; rel="aggregation"  

<table>
<thead>
<tr>
<th>Case 3</th>
<th>TimeGate server detects X-Accept-Datetime &amp; URI-G exists &amp; datetime cannot be parsed.</th>
</tr>
</thead>
</table>
|        | HTTP/1.1 400 Bad Request  
|        | X-Archive-Interval: (datetime_start) - (datetime_end)  
|        | Link: URI-R; rel="aggregation"  

<table>
<thead>
<tr>
<th>Case 4</th>
<th>TimeGate server detects X-Accept-Datetime &amp; URI-G does not exist.</th>
</tr>
</thead>
</table>
|        | HTTP/1.1 404 Not Found  

| [3] | Client issues datetime content negotiation request against TimeGate URI-G for original resource URI-R:  
| GET URI-G HTTP 1.1  
| X-Accept-Datetime: (datetime_i)  

| [4] | TimeGate server that is the target of request [3] responds:  

---

See [http://www.mementoweb.org/guide/http/remote](http://www.mementoweb.org/guide/http/remote)
How does Memento do This?

There are two components to the Memento Solution:

- **Component 1:** Navigation towards an archived resource via its original resource, content negotiation.

- **Component 2:** A discovery API for archives that allows requesting a list of all archived versions it holds for a resource with a given URI.
How does Memento do This?

• **Component 2**: A discovery API for archives that allows requesting a list of all archived versions it holds for a resource with a given URI.
• Mementos for any given URI-R are distributed across archives.

• In order to get a correct perspective of available Mementos, different archives need to be consulted.

• Can do so in distributed consultation mode (sloooow), or by consulting an aggregator.
Terminology Intermission

We introduce the term TimeBundle to refer to a resource via which an overview of all Mementos for an original resource URI-R is available.

A TimeBundle for a resource URI-R, is a resource URI-B[URI-R] that is an aggregation of:

(a) All Mementos URI-Mi [URI-R@ti] available from an archive,
(b) The archive's TimeGate URI-G for URI-R,
(c) The original resource URI-R itself.
Memento: Time Travel for the Web
Herbert Van de Sompel, Michael L. Nelson, Robert Sanderson
CNI Fall 2009 Task Force Meeting, Washington, DC, December 15 2009
TimeBundle API: For Discovery, Cross-Archive Services

- Archive uses common approaches to make TimeBundles/TimeMaps discoverable:
  - SiteMaps,
  - Atom Feeds,
  - OAI-PMH.

- Aggregator harvests and merges TimeMaps. Based on this information, the Aggregator exposes its own TimeGates.
  - Cross-archive
  - Finer datetime granularity
  - Better chances of matching a client’s datetime preference.
  - Can become a shared target for redirection for many web servers.
How does Memento do This?

There are two components to the Memento Solution:

- **Component 1**: Navigation towards an archived resource via its original resource, by leveraging content negotiation.

- **Component 2**: A discovery API for archives that allows requesting a list of all archived versions it holds for a resource with a given URI.
The Memento Experiment…
The Memento Experiment
The Memento Movie
Discussion …
Discussion: Memento and Lost Causes (1)

- URI-R vanishes, but the server that used to serve it is still operational:
  - In this case, the server should still issue the redirect to a TimeGate. Irrespective of whether the client issues a DT-conneg request or not.
  - This allows seamless access to a Memento of URI-R, even if the server no longer hosts the original.
Discussion: Memento and Lost Causes (2)

- A domain vanishes:
  - The client is looking for a current representation of URI-R that was hosted by the domain, but fails.
  - The client resorts to interaction with archives (or with a TimeBundle aggregator) and arrives at the most recent Memento of the resource.
Discussion: Memento and Caching

• Caches do not take `X-Accept-Datetime` header into account.

• In order to avoid retrieving current representation of URI-R, caches between client and server (included) must be bypassed when doing datetime content negotiation:
  
  o `Cache-Control: maxage=0` (when in time travel mode)

• In order to avoid returning a Memento when the current representation is requested (no datetime content negotiation), Mementos must not be cached:
  
  o `Cache-Control: nocache` (when in time travel mode)

• Once caches support a datetime header, the above are no longer required.
Discussion: Memento and Web Archives

• Web Archives rewrite URLs in archived pages, in order to avoid:
  - Serving current representations of embedded resources;
  - Linking to current representations of resources

• The upside: Archived pages are self-contained.

• The downside: Cannot navigate beyond the archive’s content, even if other servers (archives or original) may have appropriate version of embedded or linked resource.

• Would be interesting to explore novel strategies with this regard.
Towards Acceptance …
Towards Acceptance: Hyping

**New Scientist**

Time-travelling browsers navigate the web's past

15:41 16 November 2009 by Paul Marks

Finding old versions of web pages could become far simpler thanks to a "time-travelling" web browsing technology being pioneered at the Los Alamos National Laboratory in New Mexico.

Bookmarking a page takes you to its current version – but earlier ones are harder to find (to see an award-winning 1990s incarnation of newscientist.com, see our gallery of web pages past, right). One option is to visit a resource like the Internet Archive's Wayback Machine. There, you key in the URL of the site you want and are confronted with a matrix of years and dates for old pages that have been cached. Or, if you want to check how a Wikipedia page has evolved, you can hit the "history" tab on a page of interest and scroll through in an attempt to find the version of the page on the day you're interested in.

It’s a lot of hassle. But it shouldn’t be, says Herbert Van de Sompel, a computer scientist at Los Alamos.

---

**China National News**

Soon, 'time-travelling' browsing technology to navigate web's past

London, Nov 17: A 'time-travelling' web browsing technology is being pioneered at the Los Alamos National Laboratory in New Mexico to make it simpler to find older versions of web pages.

Called Memento, the system was developed by Herbert Van de Sompel, a computer scientist at Los Alamos, in collaboration with colleagues from Old Dominion University in Norfolk, Virginia.

Called Memento, the system Herbert Van de Sompel, a computer scientist at Los Alamos, is developing alongside colleagues from Old Dominion University in Norfolk, Virginia, to make it simpler to find older versions of web pages.

**Memento: Time Travel for the Web**

Herbert Van de Sompel, Michael L. Nelson, Robert Sanderson

CNI Fall 2009 Task Force Meeting, Washington, DC, December 15 2009

---

**Memento: Time Travel for the Web**

Herbert Van de Sompel, Michael L. Nelson, Robert Sanderson

CNI Fall 2009 Task Force Meeting, Washington, DC, December 15 2009
Towards Acceptance: Discussing

Towards Acceptance: Thinking

Telstar
Integrating References and Citations into Learning Environments

The when of the web
Any discussion about citing or referencing web-based resources seems to inevitably turn to the issue that web pages, and other resources, are subject to change. This means that if I cite a web page today, there is no guarantee that when you look at it tomorrow the content will be the same. This is in contrast with our expectations of the physical world, where if I cite a book, you’d expect that book to be the same (bar physical damage) as when I cited it.

Because of this it is standard practice to include in a web page reference a date indicating when you accessed the page. For example:


This clearly says when I accessed the page. However, if you followed the URL I’ve included in the reference, you’d get the page as it looks today, not the page as I saw it on the 24th November 2009. Although this could be the same, clearly especially with fluid sites such as Wikipedia, the content is quite likely to have changed in some respect. In areas where new information becomes available, or there is disputed information, the page could change quite radically.

A couple of weeks ago Herbert Van de Sompel (@hvdsmop), Rob Sanderson (@azaroth42) and others published a paper on ‘Memento’ – a proposal to enable archived versions of webpages to be served instead of the current one. They have named this approach ‘Memento’ – the detailed paper is available at http://arxiv.org/abs/0911.1112, and some more information is available at the Memento website. You can also see a webinar Herbert did at OCLC recently.

Towards Acceptance: Ongoing

- Working towards acceptance of Memento solution for the MediaWiki platform on which Wikipedia (and many other wikis) operate. Software available; community acceptance process started.

- Internet Archive to implement Memento. Process will start early next year with joint brainstorm.

- JISC funds developer contest for the creation of Memento-based prototype demonstrators. End of February 2010.
Towards Acceptance: Working

- Install Apache rewrite rule that redirects when X-Accept-Datetime is present.
  - http://mementoweb.org/tools/apache
- Implement Memento natively for a CMS platform.
  - http://mementoweb.org/guide/http/local
- Join memento-dev Google Group
  - http://groups.google.com/group/memento-dev
- Use ModifyHeaders FireFox extension to test.
Towards Acceptance: Funding

• Significant effort required to make Memento happen for real:
  o Standardization.
  o Outreach.
  o Research.
  o Software development.
• We would love some funding to continue the Memento effort.
Memento wants to make Navigating the Past Easy

http://www.mementoweb.org