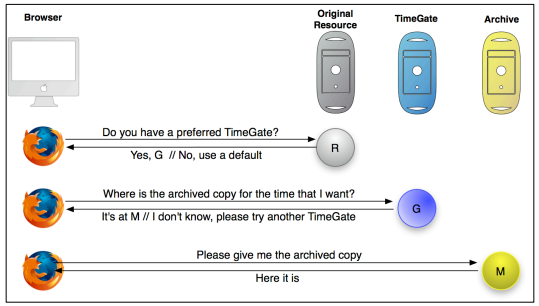
Memento: Time Travel for the Web

Memento extends content negotiation on the web with the time dimension, allowing both browsers and software agents to automatically discover previous versions of online resources, given only a desired datetime and a URI.

Interaction Summary



Original Resource:
Any resource on the web

TimeGate:
New resource that performs datetime content negotiation for a (URI, datetime) tuple

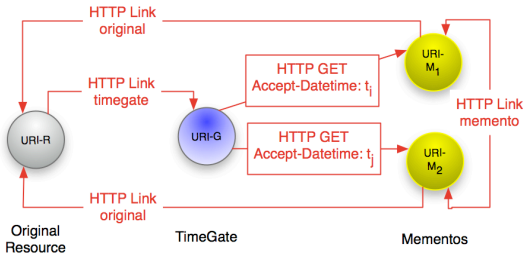
Memento:
A resource which represents the state of an Original Resource at a certain point in time, typically in an Archive or Content Management System

Technical Details

Link Header:
Resources use an HTTP response header to link to other resources. The type of link is specified in a "rel" attribute.
rel=original: Link back to the original resource
rel=timegate: Link from original to preferred TimeGate
rel=memento: Link to another Memento of same Original

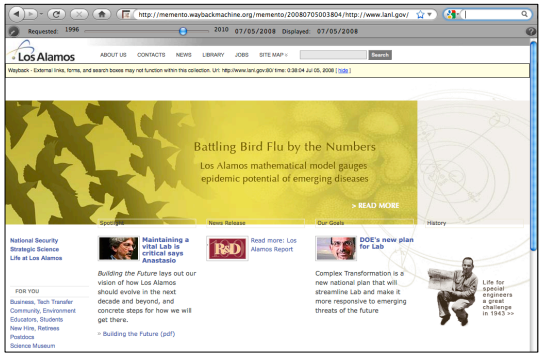
Accept-Datetime Header:
HTTP Request header sent to a TimeGate to initiate datetime content negotiation. The TimeGate responds with a 302 redirect to the appropriate Memento.

Memento-Datetime Header:
HTTP Response header sent by a Memento to give its actual datetime



Human Users: Surf the Past Web

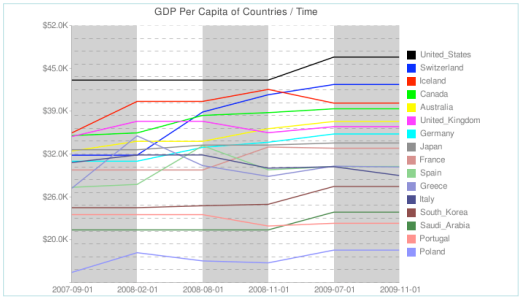
MementoFox:
Add-on for Firefox that implements the Memento protocol
Approved by Mozilla
2700 downloads, sustained 400 daily users



Time Traveling back to 2008 using MementoFox

Server Implementations:
Internet Archive, UK Web Archive implement natively
Extension module for MediaWiki: W3C Wiki + others
Proxy for all other known MediaWikis, 8 other archives
Transactional archive implemented that stores all HTTP responses from a willing web server

Software Agents: Linked Data Time Series Analysis

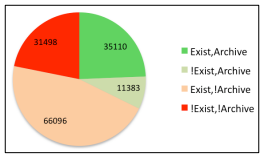


DBPedia (Linked Data version of Wikipedia) implements the Link header to a Memento archive at LANL.

Software agents can request descriptions of resources at desired point in time to see historical data, allowing time series analysis to be performed, generating graphs such as above.

Software Agents: Analysis of Reference Persistence in Scholarly Communication

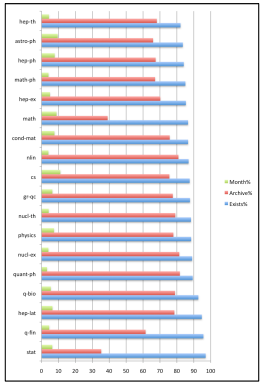
Study performed over the URIs extracted from the full text of all arXiv papers, up until January 2010: 144,087 unique URIs from 400,000 papers.



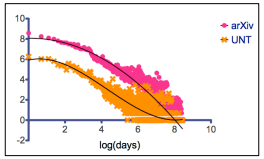
Each (URI, Paper Publication Date) combination was examined to determine if the resource (a) still existed, (b) was archived and (c) the difference between archiving date and publication date.

Overall availability was 78%, however 45% of referenced resources are not archived, and at risk of being lost.

This was also performed on a per-discipline basis, using the top level arXiv subjects as keys.



The same study was repeated against a smaller institutional repository at UNT, and the results are comparable. A log(count) vs log(days until archiving) shows a very similar curve.



A larger study involving further repositories is currently being undertaken.

Awards, References

- Winner of the 2010 Digital Preservation Award
- Ongoing $1M funding from Library of Congress

1. Van de Sompel, Nelson, Sanderson "*HTTP framework for time-based access to resources states – Memento*" Internet Draft, Nov 2010: http://www.ietf.org/id/draft-vandesompel-memento-00.txt
2. Van de Sompel, Nelson, Sanderson et al. *"Memento: Time Travel for the Web*" Technical Report, Nov 2009: http://arxiv.org/abs/0911.1112
3. Van de Sompel, Sanderson, Nelson et al. "*An HTTP-Based Versioning Mechanism for Linked Data*" LDOW Workshop, April 2010: http://arxiv.org/abs/1003.3661
4. Sanderson, Van de Sompel "*Making Web Annotations Persistent over Time*" JCDL, June 2010: http://arxiv.org/abs/1003.2643
5. Sanderson, Phillips, Van de Sompel "*Analyzing the Persistence of Referenced Web Resources with Memento*" Open Repositories, June 2011 (to appear)
6. Sanderson, Ainsworth, Adams et al. "*Implementing Time Travel for the Web*" Code4Lib Journal, April 2011 (to appear)