



Re-Finding and Keeping Information Found on the Web: Creating a Web Session Framework

Michael Adcock
MLIS Candidate
University of Washington



Motivation

- Current methods for re-finding web pages can be improved.^{4, 10, 11-13}
- Most successful methods for re-finding require no keeping activity up-front.²⁻⁵
- Search is a popular means of re-finding,^{1, 2-5} but the dynamic nature of the web can make searching difficult.⁹
- Finding and re-finding are two different activities.^{3, 6}
- Contextual information about web search and usage may be leveraged to enhance tools and aid the user.⁸
- Users primarily use hyperlinks and the back button when fact finding or information gathering while using the web.⁷
- Visualization of paths taken during web browsing may enhance re-findability.⁴

Current Strategies

Method	Likelihood of Use	Issues
Search and browse	Often	Constructing "good" search terms, taking "good" paths
Bookmarks/favorites	Sometimes	Organization and management
Browser history		
Send email to self/others	Rarely	Remembering the artifact and finding it.
Save web page to a file		
Print web page		

Adapted from Figure 3 in Adis et al.

Prototype Features

FUNCTIONAL DESCRIPTION

- Initiate session for information gathering on a topic
 - Browsing activity is tracked and saved in the background:
 - Descriptive metadata
 - page title
 - keywords
 - need to investigate ways to extract useful ones from page
 - Relational metadata
 - links taken (build a network of paths)
 - Contextual metadata
 - number of times page is visited
 - time spent on page
 - short times = user skimmed quickly (?) → content less useful
 - longer times = user read content → content more useful
 - Visualization of path is provided and can be manipulated
 - Show path linkages
 - Attempt to show "usefulness" of pages
 - based on the various metadata collected
 - Allow user to
 - set "importance" of page
 - delete/hide pages in path
 - Sort and filter display of pages based on
 - calculated "usefulness" and "importance"
 - access time and date
 - other interesting criteria
 - Search functionality
 - Icons represent picture of pages
 - Show descriptive keywords for pages (toggle)
- Resume session later
 - Enter saved path at any point
 - Track additional activity and update session

CONCEPTS TO EXPLORE

- Session as document
 - It's just a file on your computer
 - Stored in a self contained HTML page
 - Cross-platform capability
 - Can be shared with others
 - Visualization available
 - Session can be updated by further browsing
- Session as "personal search results"
 - Search within the session for a page
 - Possible to sort pages by PageRank?
 - perform search based on keywords
 - order pages based on matches in Google results
 - Suggest terms for a new Google search based on collected keywords
 - could lead to unexplored but relevant pages
 - new pages could be saved back into the session
 - a sort of search recursion process

FACILITATING TECHNOLOGIES

- JavaScript and HTML
 - Write to local file (session document)
 - Self-modifying functionality
 - enabling code and data are in the same file
 - Methods are available to circumvent normal browser security*
- Browser Add-ons
 - Firefox extension could be used for:
 - data collection
 - injection of session path into browser history (if possible)
 - management of session files

*TiddlyWiki is a working example. (<http://www.tiddlywiki.com/>)

Challenges

- Limit the scope to something manageable for a single quarter independent study.
- Handle the dynamic aspect of information on the web.
- Discover and overcome implementation and technical issues.

Goals

- Research literature and any similar tools.
- Create working prototype.
- Obtain feedback and evaluate prototype strengths and weaknesses. (Pilot testing?)

References

1. Adis, A., Jhaveri, N., and Kaki, M. (2005) Information Search Reaccess Strategies of Experienced Web Users. *Proc. of 4th WWW* (2005), 583-592.
2. Bruce, H., Jones, W., and Dumais, S. (2004). Keeping and Re-Finding Information on the Web: What Do People Do and What Do They Need? In *Proc. of ASIST 2004*, Providence, RI.
3. Capra, Robert and Pinsky, Mary and Perez-Quinones, Manuel A. (2005) Re-finding is Not Finding Again. Technical Report TR-05-40, Computer Science, Virginia Tech.
4. Hightower, R.R., King, L.T., Helfman, J.L., Bolewski, B.B., and Hollan, J.D. (1998) Graphical Multiscale Web Histories: A Study of PathPrints. In *Proceedings of ACM Conference on Hypertext 1998*.
5. Jones, W., Bruce, H., and Dumais, S. (2003) How do people get back to information on the web? How can they do it better? In *Proc. INTERACT 2003*, 789-796.
6. Kamanadina, A. (2005) Searching In Public versus Re-finding In Private: Are they the same? In *Proceedings of CHI 2005*.
7. Kellar, M., Watters, C., and Shepherd, M. (2006). The Impact of Task on the Usage of Web Browser Navigation Mechanisms. In *Proceedings of Graphics Interface 2006*, Quebec City, Canada, To Appear.
8. Lawrence, S. (2000). Content in Web search. *IEEE Data Engineering Bulletin*, 23(3): 23-32.
9. Levy, D. (1994). *Fixed or fluid? Document stability and new media*. In *Proceedings of European Conference on Hypertext*, pp. 24-31.
10. MacKay, B., Kellar, M., and Watters, C. (2005). An Evaluation of Landmarks for Re-Finding Information on the Web. In *Proc. of CHI 2005*, Portland, OR, 4009-4012.
11. Teruan, J. (2005). The Re-Search Engine: Helping people return to information on the Web. In *Proceedings of SIGIR '05 (International Consortium)*.
12. Teruan, J. (2005) Where'd It Go?: Re-Finding Information in the Changing Web. in *Student Organized Workshop*, (Gloucester, MA, 2005).
13. Teruan, J., Avraido, C., Ackerman, M. S., and Karger, D. R. (2004). The perfect search engine is not enough: A study of orienting behavior in directed search. In *Proceedings of CHI '04*, pp. 415-422.