

Supporting Recall of Web Browsing History by Generating User's Knowledge Space

MURAKAMI Harumi[†] and HIRATA Takashi[‡]

We have developed a system called Memory-Organizer, which helps users to construct externalized memory to support human memory. The externalized memory is a concept that externalizes and stores the contents of human working memory on computers. In this paper, we describe a system that generates the user's knowledge space from his/her Web-browsing history, and helps users to recall their history by using Memory-Organizer. Preliminary experimental results show that a new history, which extracts keywords from a line including a link, can become information sources for generating knowledge space. In particular, in the case of newspaper articles, most anchor texts (article headers) are prime candidates for generating knowledge space.

Keywords: Web browsing history, knowledge space, recall, externalized memory, Memory-Organizer

ユーザの知識空間作成による Web ブラウジング履歴の想起支援

村上 晴美[†] 平田 高志[‡]

我々はこれまでに人間の記憶を支援するために、コンピュータ上に外化記憶を構築するシステム Memory-Organizerを開発してきた。外化記憶はコンピュータ上に人間の作動記憶の内容を蓄積する概念である。本稿では、Memory-Organizerを用いて、Web ブラウジング履歴からユーザの知識空間を作成することにより、履歴の想起支援を行うシステムについて述べる。予備的な実験により、アンカーテキストを含む文字列からキーワードを抽出する本研究における履歴が知識空間作成のための情報源となりうること、特にニュース記事の場合はほとんどのアンカーテキストが記事の見出しであり、知識空間を生成するための候補となることがわかった。

キーワード: Web ブラウジング履歴, 知識空間, 想起, 外化記憶, Memory-Organizer

1 Introduction

We have developed a system called “Memory-Organizer,” which helps users to construct “externalized memory” to support human memory. The externalized memory is a concept that externalizes and stores the contents of human working memory on computers.

The system consists of (a) a thinking-space browser, which helps users to input and edit externalized memory and support creative thinking; (b) an overlay Web browser, which helps users to gather, clip, and create externalized memory while browsing the Web; (c) an interest-space browser, which helps users to recall externalized memory by arranging it on the user's interest space in chronological order; (d) a map browser, which helps users to recall externalized memory by arranging it based on position information; (e) a calendar browser, which helps users to recall

externalized memory onto a calendar view based on date; and (f) a knowledge-space browser, which helps users to recall externalized memory by generating the user's knowledge space. Figure 1 shows an overview of the Memory-Organizer. See [1, 2, 3] for the details on each browser.

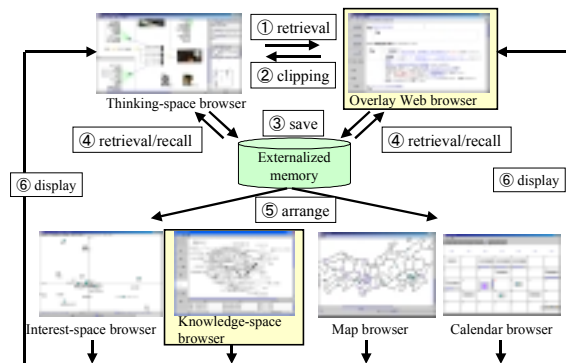


Figure 1: Overview of the Memory-Organizer.

In Section 2, we provide an overview of the externalized memory. In the following sections, we describe a system that helps users to recall their Web browsing history by generating knowledge space with Memory-Organizer.

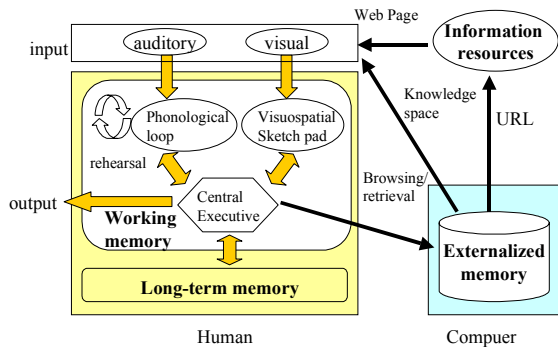
[†] Graduate School for Creative Cities and Media Center, Osaka City University
大阪市立大学大学院創造都市研究科 兼 学術情報総合センター

[‡] Japan Defense Agency
防衛庁

2 Overview of Model of Externalized Memory

Externalized memory is a concept that externalizes and stores the contents of human working memory on computers. We aim to capture information processed during human cognitive processes via working memory presented by Baddeley¹⁾ [4].

Figure 2 shows an overview of the externalized memory. The information processed by the user such as on Browsing the Web is accumulated into the user's externalized memory. The knowledge space is generated from the externalized memory under the user's direction, becoming the user's input information. When the user selects a URL from the Web browsing history, the system sends a URL through the externalized memory, the designated Web page is displayed on a browser, and this again becomes input information to the user. This is how circulative information processing is performed.



Note: The working memory model is modified from [4, 5].

Figure 2: Overview of the Externalized Memory Model.

3 Recall of Web Browsing History by Generating Knowledge Space

In this section we describe a system that helps users to recall their own Web browsing history by generating knowledge space for them from the history by using Memory-Organizer.

Figure 3 shows an overview of the system. When the user selects a link using an overlay Web browser, "date, URL designated by the link, terms extracted from a line including the link" are saved as a data set in a new history. We call this data set a "history set." The system displays the user's knowledge space generated from the history based on the user's selection of keywords, and displays history sets including selected keywords on to a knowledge-space browser. When the user selects a

history set that includes a URL from the history sets, the overlay Web browser connects to the Internet and displays the URL's Web page.

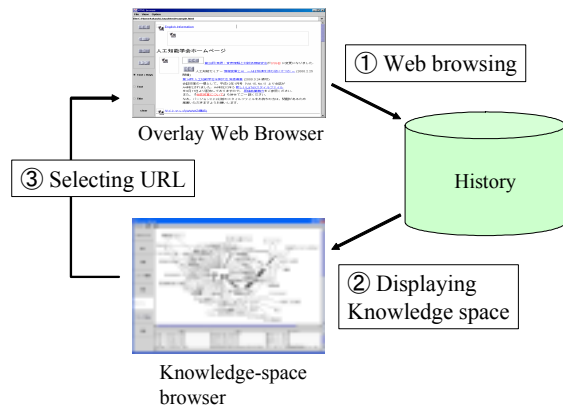
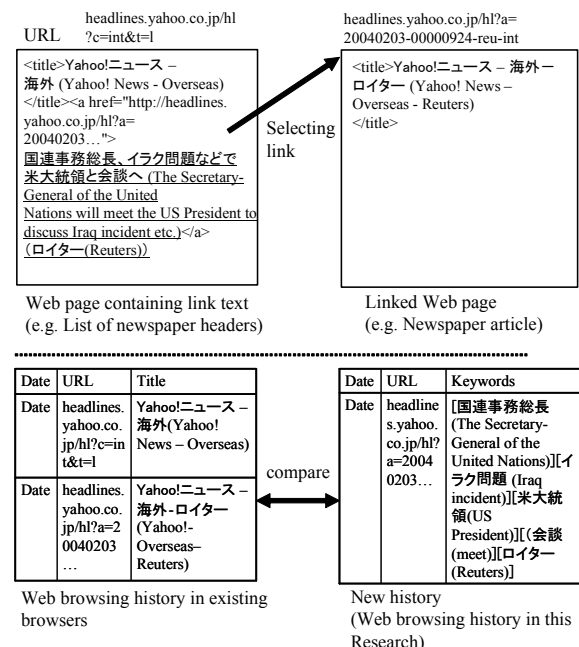


Figure 3: Overview of the system.

3.1 Overlay Web Browser

An overlay Web browser has functions that generate a new history different to existing histories, such as IE, when the user selects a link. Here, we describe our method of generating the new history.

The overlay Web browser links keywords extracted from a line including a link to a designated URL on Web browsing in order to recall the Web-browsing context. In addition, in this research, we do not save history when no keyword is extracted from the text.



Note: The text in this example was originally Japanese and has been translated into English.

Figure 4: Comparison of histories in existing Web browsers and in the overlay Web browser.

Figure 4 shows the difference between the

¹⁾ We think that it is impossible to accumulate the exact contents of the working memory currently, so we aim to accumulate it virtually.

history of existing Web browsers and the new history of the overlay Web Browser. The existing browsers save the date, the URL of a Web page, and text extracted from title elements as a data set in the history. In contrast, our overlay Web browser saves the date, a URL designated by an anchor element, and terms extracted from a line including the selected link as a data set in the new history.

For example, consider that the user selects a link “国連事務総長、イラク問題などで米大統領と会談へ (The Secretary-General of the United Nations will meet with the US President to discuss Iraq incident etc.)” from “Yahoo!ニュース -海外- (Yahoo! Japan News - Overseas).” The history of the existing browser saves “Yahoo!ニュース -海外-ロイター (Yahoo! News - Overseas - Reuters),” which is extracted from a title element of the Web page. Our new history saves “国連事務総長 (The Secretary-General of the United Nations), “イラク問題 (Iraq incident),” “米大統領 (US President),” “会談 (meet),” and “ロイター (Reuters),” which are extracted from a line including an anchor element of the previous Web page.

3.2 Knowledge-space Browser

A knowledge-space browser helps users to understand the user’s knowledge and recall an externalized memory by generating knowledge space that consists of nodes of keywords and links between them.

The browser comprises an “operation area” on which the user executes operations, a “knowledge-space display-area,” which displays the user’s knowledge space, and a “history display-area,” which displays history sets that includes keywords selected by the user.

The authors have been using this system since April 2001. Figure 5 shows a screenshot of the knowledge space for the first author when selecting “テロ (terrorism)” “アルカイダ (Al-Qaeda)” “自衛隊派遣 (sending Japan Self-Defense Forces)” and displaying the neighbors, respectively on 17th Dec 2003.

It is found that a keyword “テロ (terrorism)” is displayed in large type, and keywords related to the World Trade Center and Pentagon Attack, and Palestinian terrorism are also displayed. In this way, we know that one piece of knowledge is linked another. The user can subsequently browse the knowledge space by continually displaying neighbor keywords.

A list of history sets that includes URL and keywords is displayed in a history display-area to assist users in locating Web pages referred to by keywords. In an example from Figure 5, keywords such as “自衛隊派遣 (sending Japan’s

Self-Defense Forces),” “基本計画決定 (formulate a basic plan),” and “陸自 5 0 0 ~ 7 0 0 人 (500-700 people from Japan’s Ground Self-Defense Forces),” are displayed to help users to access a Web page browsed on 9th Dec 2003.

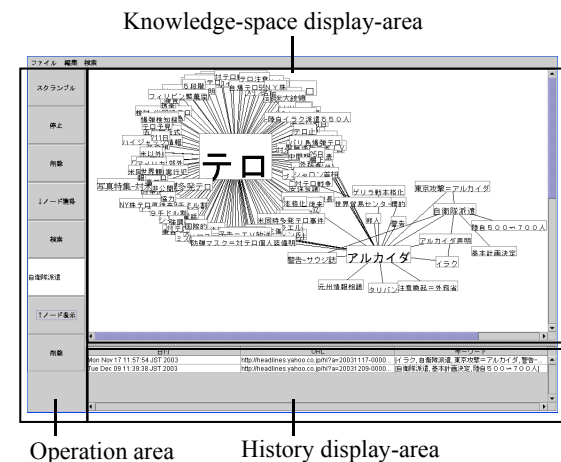


Figure 5: Screen image of the knowledge-space browser.

In regards to the knowledge-space interface, we refer to the theory of spreading activation [6], which is a model of human semantic memory. We assume that concepts and relations might represent the user’s knowledge space, thus when the user selects a keyword, keywords related to the original keyword are displayed. We change the size of the nodes to express the level of memory activation.

Currently, the semantics of the relations is not defined, and the distance between nodes does not correspond to the semantic distance in the theory of spreading activation.

4 Algorithm

The system comprises three stages: (1) Web browsing, (2) exploring knowledge space, and (3) selecting URL. The user performs (1) by using an overlay Web browser, and (2), (3) by using a knowledge-space browser. Figure 6 shows an outline of the algorithm.

4.1 Web Browsing

When the user selects a link using an overlay Web browser, “date, URL designated by the link, terms extracted from a line including the link” are saved as a data set (history set) in a new history. Details of the history-generating method are provided in [3]. In an example in Figure 6, when the user selects a link “A と B は、何日、C。(A and B do C on some day.)” on an overlay Web browser that displays a Web page of URL0, accessed date, URL1, keywords [A][B][C] are saved to the history as a history set.

4.2 Exploring Knowledge Space

According to the user's selected links, the system retrieves keywords from the history, displays the user's knowledge space (which connects keywords to a knowledge display-area), and displays a list of history sets that include keywords on a history display area.

Details of the system are as follows.

- Step 1: The user selects a keyword.
- Step 2: The user selects a keyword and displays it.
 - Step 2.1: The system retrieves keywords from the user's history and accumulates them to a "limited accumulative history (described later)."
 - Step 2.2: The system displays a knowledge space on to a knowledge-space display-area by employing the limited accumulative history.
 - Step 2.3: The system displays history sets, which include retrieved history sets, on a history display area.

The features of the knowledge space in this research are: (a) it does not pre-generate static knowledge space, but generates it according to the user's operation, (b) it is a simple method that connects correlated keywords in the history, and (c) it changes the size of the nodes that simulate the human cognitive process.

We assume that a cognitive process does occur as shown in Figure 2, and we consider that memory is activated according to the frequency of stimuli in working memory. Thus, we change the size of the nodes according to the activity level.

There can be twice of stimuli in working memory: memorizing and recall. We consider that changing the node size according to the frequency of keywords in a history corresponds to stimuli on memorizing. In this research, we devise a concept called limited accumulative history to calculate the level of stimuli of memorizing plus recall. This limited accumulative history is another type of history that adds history sets, which include keywords provided by the user when the system retrieves history.

In addition, in retrieving history, the system first attempts to execute an exact match; if this fails, however, it proceeds to attempting a partial match.

We explain the example in Figure 6 as follows.

- When the user inputs "A," "A" is displayed.
- When the user selects "A" and displays neighbors of "A," (hereafter "selects and displays") the system retrieves "A" from the history, URL1, which includes that "A" is added to the limited accumulative history, and it becomes (URL1). The system then displays "B" and "C," which correlate with "A," and connections from "A" to "B" and "C" are generated. Here, the frequency of "A," "B," and "C" are each "1," and these are displayed

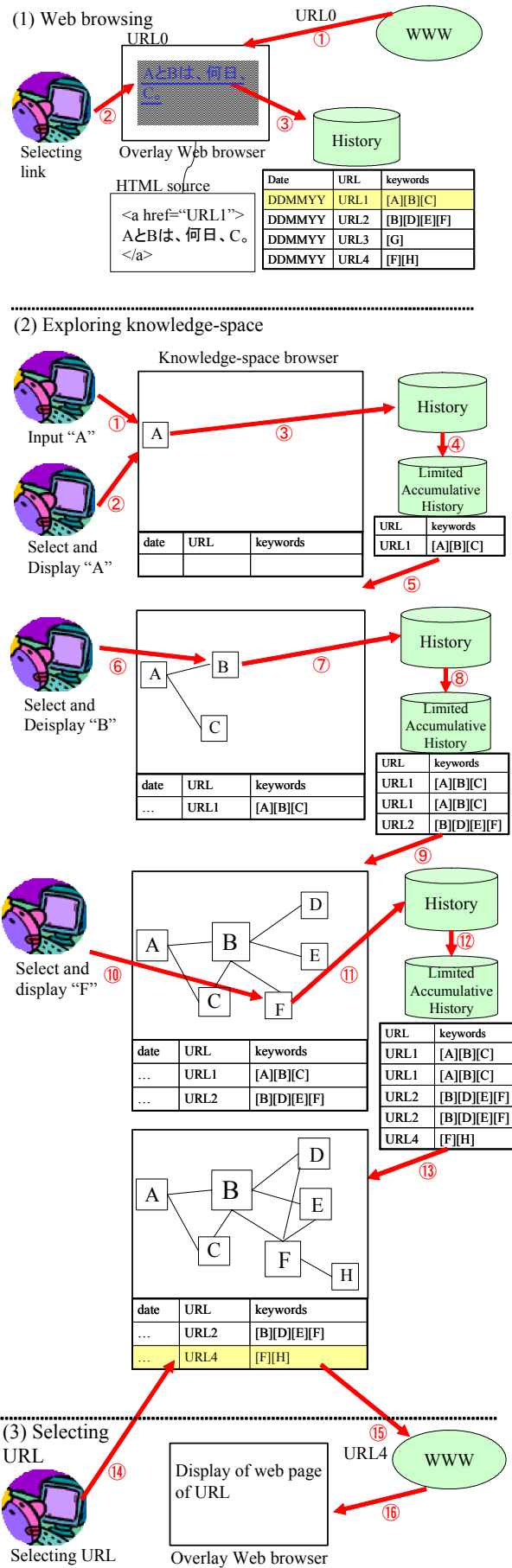


Figure 6: Overview of the algorithm.

in size according to the frequency.

- When the user selects and displays “B,” URLs 1 and 2, which include “B,” are added to the limited accumulative history, thus it becomes (URL1, 1, 2) as above. The system then displays “D,” “E,” and “F,” from “B,” and connects from “B” to “C,” “D,” “E,” and “F.” In addition, the respective frequencies of “A,” “B,” “C,” “D,” “E,” and “F” are “2,” “3,” “2,” “1,” “1,” and “1,” and these are also displayed in size according to the frequency.

- When the user selects and displays “F,” URLs 2 and 4, which include “F,” are added to the limited accumulative history, thus it becomes (URL1, 1, 2, 2, 4) as above. The system then displays “H,” which then connects via “F” to “D,” “E,” and “H.” In addition, the respective frequencies of “A,” “B,” “C,” “D,” “E,” “F,” and “H” are “2,” “4,” “2,” “2,” “2,” “3,” and “1,” and these are displayed in size according to the frequency.

In this way, a path such as “A→B→F→H” can be displayed according to the user’s operation.

In this study, we do not delete any nodes (this may correspond to forgetting) except for the user’s will. However, we think that there will be a need for functions that do not display keywords designated by the user and do not display certain keywords to be easy to see the screen.

4.3 Selecting URL

When the user clicks a history set that includes a URL, an overlay Web browser connects to the Internet and displays the designated Web page.

In the example in Figure 6, when the user selects URL4, the overlay Web browser displays the Web page of URL4.

5 Experiment

Whether a knowledge space generated by this method is appropriate depends on: (a) whether there is more than one keyword in a history set, (b) whether the same keyword exists in a different history set, and (c) whether the contents of the extracted keywords are appropriate. In this paper, we performed a preliminary experiment for (a) and (b).

5.1 Experiment 1

The subject was the first author. She performed a simulation of using an overlay Web browser in everyday life. The details of the conditions for use are described in [3]. Here we investigate her history from 9th Apr 2001 to 17th Dec 2003.

The number of selected links was 6,081, There were 4,364 selected links with no duplication, 20,158 extracted keywords, and 9,466 extracted keywords with no duplication. The average number of keywords per link was 3.3 (20,158/6,081). Figure

7 shows a graph of the relation of the number of keywords in a history set and the number of history sets. The proportion of histories with one keyword is 1,802/6,081 (30%). That is, in this history, 70% of the history sets can be candidates for generating knowledge space.

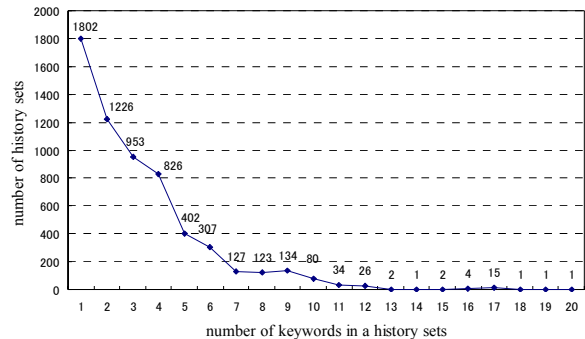


Figure 7: Relation between number of keywords in a history set and number of history sets.

5.2 Experiment 2

Four female subjects participated in this experiment. We downloaded newspaper articles (HTML files) over a three-month period from the newspaper Web site “Mainichi Interactive [7].” The subjects were asked to browse any article they wanted to see, at least one page per day. The details of the procedure are described in [3]. In the Mainichi Interactive site, a line including a selected link is the same as an anchor text and the article header.

The average number of selected links was 396, the average number of extracted keywords was 1,735, and the average number of average number of extracted keywords per link was 4.4 (1,735/396). The greatest number of keywords in the history is seven (See Figure 8).

Ninety-nine percent of the history sets contain more than one keyword. That is, most history sets are candidates for generating knowledge space.

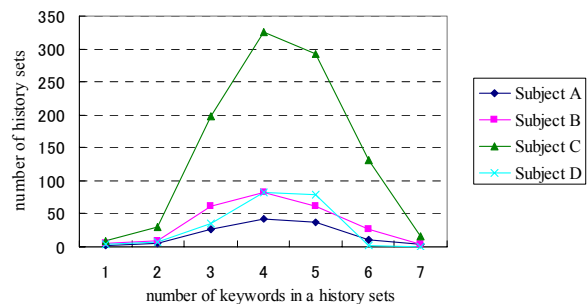


Figure 8: Relation between number of keywords in a history set and number of history sets.

The above results indicate that the new browsing

history, which extracts keywords from a line including the link, can become information sources for generating knowledge space. In particular, in the case of newspaper articles, most of the anchor texts (headers of the articles) are candidates for generating knowledge space.

6 Related Work

Much research concerning visualization of Web browsing history deals with existing history such as IE, and there is not much research on methods for generating new history. The typical approach is that “visualization of history helps users in navigating the Web,” in which Web pages or sites are expressed as nodes and the relation between nodes is displayed as a tree or graph structure. Examples of these can be seen in MosaicG[8] and PadPrints[9]. This approach is useful for “going back” to recently browsed Web pages, as in the case of existing browsers. However, it is unclear whether this approach is useful for displaying users' interests, and for helping people to access browsed pages they viewed more than one month ago.

The main difference between this research and related work is that (a) we generate a new history based on the user's link selecting activities, and (b) our method generates knowledge space from keywords extracted from anchor texts to recall the history.

7 Summary

We have developed a system called Memory-Organizer, which helps users to construct externalized memory to support human memory. In this paper, we described a system that helps users to recall their Web browsing history by generating the user's knowledge space with Memory-Organizer. Preliminary experimental results show that the new history, which extracts keywords from a line including a link, can be information sources for generating knowledge space. In particular, in the case of newspaper articles, most anchor texts (article headers) are prime candidates for generating knowledge space.

References

- [1] Murakami, H. and Hirata, T.: Memory-Organizer: A System for Constructing Externalized Memory, Proceedings of the 15th Annual Conference of JSAI, 3F1-03 (in Japanese) (2001).
- [2] Murakami, H. and Hirata, T.: Information Acquisition and Reorganization from the WWW by using Memory-Organizer, Bulletin of Osaka City University Media Center, Vol.3, pp.9-24 (2002).
- [3] Murakami, H. and Hirata, T.: A System for

Generating the User's Chronological Interest-space from the User's Link Selecting Activities on the Web, JCSS-TR-47, pp.1-12 (in Japanese) (2003).

[4] Baddeley, A. D., and Hitch, G.: Working memory, *The Psychology of Learning and Motivation*, Academic Press (1974).

[5] Logie, R. H.: *Visuo-spatial Working Memory*, Lawrence Erlbaum (1995).

[6] Collins, A. M., Loftus, E. F.: A Spreading-activation Theory of Semantic Processing, *Psychological Review*, 82, 407-428 (1975).

[7] <http://www.mainichi.co.jp/>

[8] Ayers, E. Z. and Stasko, J. T.: Using Graphic History in Browsing the World Wide Web, *Proceedings of WWW4* (1996).

[9] Hightower, R. R., Ring, L. T., Helfman, J. L., Bederson, B. B., and Hollan, J. D.: Graphical Multiscale Web Histories: A Study of Padprints, in *Proceedings of ACM Hypertext'98*, pp.58-65 (1998).