

# The Impact of Bookmarks and Annotations on Refinding Information

Ricardo Kawase, George Papadakis, Eelco Herder, Wolfgang Nejdl  
L3S Research Center / Leibniz Universität Hannover  
Appelstr. 4, 30167 Hannover, Germany  
+49. (0)511. 762-17715  
{kawase, papadakis, herder, nejdl}@L3S.de

## ABSTRACT

Refinding information has been interwoven with web activity since its early beginning. Even though all common web browsers were equipped with a history list and bookmarks early enough to facilitate this need, most users typically use search engines to refind information. However, both bookmarks and search based tools have significant limitations that impact their usability: the former are known to be hard to manage over the course of time, whereas the latter require the user to recall a specific combination of keywords or context. Most importantly, though, both are particularly inappropriate in cases where a piece of information is contained within an unstructured web page. In this paper, we present in-context annotation as a more efficient alternative to these methodologies. To verify this claim, we conducted a study in which we compare the performance of experienced users in all three approaches while revisiting specific pieces of information in the web after a long period of time. The outcomes suggest that in-context annotation clearly outperforms both traditional strategies.

## Categories and Subject Descriptors

H.5.4 [Information Interfaces and Presentation]  
Hypertext/Hypermedia - *Architectures, Navigation, User issues.*

H.5.3 [Information Interfaces and Presentation] - Group and Organization Interfaces - *Computer-supported cooperative work, Evaluation/methodology, Organizational design, Web-based interaction.*

## General Terms

Measurement, Performance, Design, Experimentation, Human Factors, Verification.

## Keywords

User study, evaluation, information refinding, Web annotation.

## 1. INTRODUCTION

When trying to refind information, users have shown to prefer search engines rather than dedicated history tools, such as the

back button, the history list and bookmarks. However, search engines usually suffer from the fact that a user has to *recall* a usually hard-to-remember combination of keywords in order to retrieve a specific web resource. Bookmarks, on the other hand, constitute the most popular refinding support method [17], as they optimize the searching stage in comparisons with search engines[13][1]. Their major drawback, though, is the fact that collections of bookmarks are *hard to manage*, due to their tendency to explode with the passage of time. This primarily stems from the inherent difficulty in structuring information [1], even in the context of alternative approaches, like hierarchical organization [6].

Most and for all, the aforementioned tools are inadequate in supporting *long-term refinding*, i.e. refinding that takes place some months after the initial visit [15], particularly when it involves unstructured resources. An approach that could well be employed for effectively dealing with these conditions is “*in-context*” annotations: these are annotations that are visible within the original resource, enhancing it with the observations and remarks of users. In fact, annotation tools allow users to mark just the desired part of a resource, and also provide them with some management and sharing facilities, similar to social bookmarking services. In this way, relocating a specific piece of information is alleviated to a great extent.

The motivation behind our preference to annotation stems from our latest work [12] and the conclusions drawn from it. In this previous work we analyzed paper-based annotation and found that its high utility is based on the facilitation of learning as well as of later reference. Therefore, all participants of the survey welcomed the opportunity to be able to act similarly in the digital environment, though not in a way that merely imitates paper-based annotation.

To the best of our knowledge, there is currently no study focusing on the actual impact of “in-context” annotation on re-visitation. In this paper we systematically investigate this effect through a user study that was conducted over a long period of time [11]. We asked a total of 30 people to relocate several pieces of information they had visited five months before, when the experiment started. The subjects were divided into three equivalent groups, with each group using exclusively one of the above tools (namely a search engine, a social bookmarking system and a web annotation tool). We then measured their efficiency in terms of the time taken to complete the browsing stages of all tasks. The outcomes of the experiments suggest that “in-context” annotation is significantly better than the other two approaches, reducing the required time to the half.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

HT'10, June 13–16, 2010, Toronto, Ontario, Canada.

Copyright 2010 ACM 978-1-4503-0041-4/10/06...\$10.00.

The rest of the paper is structured as follows. In section 2 we present similar studies and underline their difference from our own approach. Following that, in section 3 we present the settings of our study in detail, whereas in section 4 we analyze its outcomes. Finally, we draw our conclusions in section 5.

## 2. RELATED WORK

The early acknowledgment of the significance of refinding [18] in combination with the theory of information foraging [16] motivated a wide range of studies on analyzing the extent as well as the patterns of this phenomenon. Although the studies quantifying the so-called *recurrence rate* vary considerably in their estimates [7][18][5], they all advocate that refinding constitutes a prevalent user behavior, comprising more than half of the overall web activity. The analysis of this phenomenon further suggests that there is a great heterogeneity in the underlying incentives [2], and that chronologically it should be distinguished into *short-, medium- and long-term* refinding [15].

Several browser enhancements introduced over the years (e.g. history cache, URL auto-completion and bookmarks) together with search engines have so far been the only tools available for facilitating this behavior. Although their contribution to leveraging its efficiency has been documented in the literature [9], they are actually appropriate just for short- and medium-term refinding, that takes place in a period of time up to few weeks after the first encounter. Moreover, they all ignore the contextual information of the user's original search process, without taking advantage of any of the associated cues that have been reported to assist significantly in this process [14].

Additional, interesting aspects of this phenomenon have also been investigated in past user studies. Among them, Carpa et al. [3] focused on the use of search engines in the context of this behavior, and the effect of familiarity and task frequency in particular. Their study, however, focuses solely on the *short-term refinding* period of one week.

Obendorf et al.[15] present another interesting study that explores the evolution of users' refinding behavior over the last 15 years. To this end, the web activity of 25 people was analyzed. The outcomes suggested a significant deviation from the usage patterns that were presented in the early, related studies. More specifically, newly introduced browser features, like tabs, were found to have given rise to new navigation strategies, whereas dynamic web pages were recognized to act as obstacles to short-term refinding. The authors also identified opportunities for new browser tools that could cover users' medium-term refinding needs more effectively than available appliances, like RSS feeds and widgets.

Finally, MacKay et al. in [13] introduce another technique that extends traditional bookmarks with the aim of facilitating refinding, namely the Landmarks. In essence, this work investigates the efficiency of in-context clues in comparison with traditional bookmarks. The experiments conducted in this context entail 20 participants and consist of two sessions that were separated by a gap of three days. The outcomes suggest a clear *prevalence of in-context cues* over plain bookmarks. However, their experiments constitute a short-term refinding study that involves merely six tasks, which are actually not revealed.

## 3. EXPERIMENTS

In short, the goal of our study is to quantitatively estimate the efficiency and ease-of-use of three different refinding methodologies: web search, bookmarks and in-context annotation. Regarding the tools instantiating these methodologies, we selected the Del.icio.us<sup>1</sup> social bookmarking service while for in-context annotations we developed an experimental, straightforward annotation tool: SpreadCrumbs [10]. As for the web search, the participants were free to make use of the search engine they were more familiar with; they all turned out to prefer the Google<sup>2</sup> search engine.

### 3.1 Systems

*Del.icio.us* is a popular online social bookmarking system. The system combines bookmarking and tagging with social networking, features that turned it into one of the most successful social bookmarking services. With Del.icio.us a user can annotate bookmarked Web pages by associating tags and comments to them, while also having the option of sharing them with her contacts. A crucial difference with annotation tools, though, is that the added or tagged information can only be viewed and managed through their personal home page in the Del.icio.us' web site. Therefore, they are far from being in-context.

*SpreadCrumbs* is an in-context Web annotation system, which has been implemented as an extension of the Mozilla Firefox Web browser. The underlying assumption of SpreadCrumbs is that users can annotate Web resources with keywords or sentences and create hypertrails through a set of annotations. These annotations can then be used not only for her own reference, but also for sharing with her social network. The interface of SpreadCrumbs has deliberately been designed in a minimalistic way, so that users get easily acquainted with it: Post-its are quite popular in real life, and, thus, SpreadCrumbs offers a very familiar way of adding inscriptions and remarks to web content through a basic visual metaphor of Post-it notes (**Figure 1**). Moreover, it is transient in the sense that it is easily replaceable and not disruptive.

The following subsections elaborate on the two sessions of our experiments in greater detail.

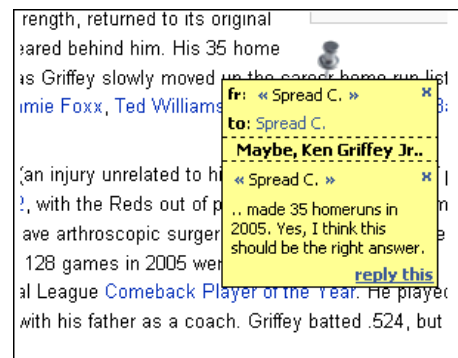


Figure 1. A SpreadCrumbs' annotation of one participant.

<sup>1</sup> <http://delicious.com>

<sup>2</sup> <http://www.google.com>

## 3.2 First Session

The purpose of this stage was merely to have the participants locate specific pieces of information that they will be asked to refind in the second phase of the experiments. In addition, the participants were asked to carry out a small task in order to evaluate the usability of the tools.

### 3.2.1 Participants and Settings

We designed our experiments to be comprised of two sessions; the first one includes all finding tasks, leaving all refinding tasks to the second session. It should be stressed here that the second session took place *five months* after the first one, thus allowing for the examination of long-term refinding behavior. The participants' pool consisted of 34 people in total (24 males and 10 females), who are aged 27 years old on average. All of them are quite proficient in on-line search and web technologies in general, due to their education background (the majority being graduate PhD students in the field of Computer Science).

In the context of the first session, we randomly divided the participants into two equivalent groups, with each one designated to a specific tool: Del.icio.us or SpreadCrumbs. To ensure familiarity with the application at hand, each individual was initially presented with a short tutorial of its features and functionalities.

### 3.2.2 Procedure

During the experiments, the participants were asked to locate the correct answers to 10 questions. All of them were specific information finding tasks that could be solved by a brief internet search with any popular search engine. They were in fact chosen at random from a set of 16 questions we had carefully prepared for the experiments; most of the answers were numerical in nature as an effort to make them sufficiently obscure and minimize the chance of participants knowing the answers themselves. One of the questions, for instance, was the following: "How many homeruns did Ken Griffey Jr. hit in 2005 playing for Cincinnati?".

After finding the required information, the participant was asked to annotate the web page that contained the answer for later reference. It should be clarified here, that there was no predefined set of acceptable web pages. Rather, subjects were free to mark any web resource they wanted. This first task was presented as 'just' an exercise in order to get used to the system. In reality, though, it was a preparational step for the second round of the experiment.

For the second task, we presented a scenario in the form of form of collaborative decision making: participants had to plan a day trip to London based on a collection of annotated pages recommended by a friend, which they were asked to evaluate. The goal was to obtain interaction data and user feedback on the bookmarking and annotation tools.

### 3.2.3 Primary Observations

Upon completion of the first task, we noticed that, far from exhibiting homogeneity, each participant followed her own approach in creating annotations: some included the answers in the annotation text; others added the questions, while the rest of them used just some keywords. Moreover, they followed different

strategies for positioning the annotations, as not every participant was concerned with placing them in a useful location; although most of them carefully posted them near the text, table, or paragraph containing the answer to a question, several individuals just added the annotation on the top of the page or over the margins. The latter are, thus, expected to experience some overhead in the course of the refinding task, especially in the cases of answers residing within web pages with a great deal of information and unstructured content, in general.

Furthermore, during the second task, we observed that most of Del.icio.us' users did not notice or read the additional comments on each bookmark. In contrast, all participants that performed the tasks using SpreadCrumbs did notice and read the comments. Some of them actually affirmed that their choices were influenced by those comments. This observation clearly suggests that in-context comments are more likely to be used during refinding.

## 3.3 Second Session

During the second session, five months after the initial session of the experiments, the participants were asked to relocate the answers they had previously found during the 'preparatory' task of the first round. This long time interval ensures that the participants remembered neither the answers they had provided nor the resources they had used in the course of the first session.

### 3.3.1 Participants and Settings

In total, 30 participants (21 males and 9 females) were involved in this phase of our study, out of the initial 34. They were asked to repeat the same tasks as the first time; in other words, each individual was given the initial set of 10 questions and had to refind the answers she had given in the first round. The participants were divided into three equivalent groups of 10 people, each one corresponding to a specific refinding methodology and tool.

The first group corresponds to the *search engine* approach and its members were allowed to employ solely search engines in their efforts to carry out their tasks. This group was formed by randomly choosing 5 participants from the bookmark group together with another 5 from the annotation group of the first session. They were not allowed to use the bookmarks or annotations they had already created, which implies that they could refind information only by searching and browsing from scratch. This group served as the baseline group during our analysis.

The second group represents the *bookmarks* approach and consisted of those subjects that used Del.icio.us both in the finding and the refinding sessions. The members of this group had the URLs of the visited resources at their disposal, saving in this way the burden of repeating the procedure of the first session. Additionally, some of them had added comments to their bookmarks, which invariably provided them with valuable clues for quickly relocating the answers.

Finally, the third group corresponds to the *in-context annotation* approach and was comprised of those participants that used SpreadCrumbs in both sessions. The URLs of the initially visited resources were thus available to them, similar to the bookmark group. Further, they were also assisted in their task by the annotations that they had composed during the first round of the experiments. However, as mentioned before, there was a great

diversity not only in the content of these annotations, but also in their positions. We expect these two factors to influence the performance of the participants.

### 3.3.2 Procedure

During the second phase, the participants were presented with one question at a time, chosen randomly, so that the order of questions is different from the one used in the first round. In this way, even the participants of the last two groups that were assisted by an application, had to devote some time to pinpoint the appropriate bookmark or annotation in their collection.

After the appropriate web resource was found, thus completing the searching stage, the participant had to locate the answer in the page and highlight it using the mouse (browsing stage). There were no instructions or restrictions as on how to proceed with this stage; the participants were allowed to perform this task the way they would in a non-controlled environment. It turned out that the vast majority of the participants took advantage of the browser's "find" functionality, which rapidly locates and highlights the given words in the page in view. This functionality was used not only in conjunction with some keywords taken from the question, but also with the whole answer (some subjects had it included in their bookmark comments).

Once the desired piece of information was highlighted, the participant was given the next question. Upon completion of all tasks, the subjects were asked to answer two questionnaires, one regarding the information refinding experience, and another one investigating their opinion on the tool they used.

The necessary data for estimating and evaluating the average and overall browsing time per individual were collected with the help of screen capture and data-logging software that recorded all participants' actions. Its analysis is discussed in the following section.

## 4. RESULTS

In this section we discuss the outcomes of our experiments, which are mainly concerned with *browsing time* - the time participants spent in the browsing phase while carrying out their task. In other words, our analysis focuses on the period of time that starts as soon as the page of interest finishes loading and ends the moment that the participant finds the required information. We begin with the analysis of the time measurements that were derived from the 297, in total, refinding activities. The corresponding tasks are evenly shared among the three groups mentioned above - the *Search Engine*, the *Bookmark* and the *Annotation* groups. That means that the performance of each group is represented by 99 time intervals expressing the duration of the tasks involved.

### 4.1 Browsing Time Measurements

The most appropriate metric for expressing the overall performance of each group is arguably the *average time* taken to complete the *browsing phase* - therewith ignoring the time it took participants to locate the page in the *searching phase*.

In our case, the available sample of 99 browsing times produces the following mean values: 46s for Search Engine, 38s for Bookmark, and 21s for Annotation. With an average mean of 21 seconds, the annotation group was significantly faster than the bookmarking group (38 seconds;  $t(98)=3.88$ ,  $p<0.01$ ,  $r=.36$ ) and the search engine group (46 seconds;  $t(98)=4.07$ ,  $p<0.01$ ,  $r=.38$ ).

The differences between the two latter groups were found to be non-significant. It turns out, therefore, that the performance of Annotation is substantially better, corresponding to a time that is almost the half of the other two groups. This suggests that in-context annotation boosts refinding to a great extent.

By contrast, when comparing the performance of the first two groups, the outcome does not match our initial expectation that Bookmark would outperform Search Engine due to the wealth of cues associated with them, i.e. the comments that were attached to bookmarks as well as the keywords of the tags that were drawn from the questions or even the answers. This can partially be explained by the *theory of context-dependency* [17], arguing that all context knowledge acquired in the refinding process serves as relevant cues for refinding information. This includes even the non-semantically related elements located within the target information that search engine users acquired while searching and browsing the search results. Bookmark users, on the other hand, had to acquire the context during the browsing stage itself. It should also be stressed that the performance of the search group would have been significantly worse if we also took into account the searching stage, which is minimized for bookmarks' users.

### 4.2 CTRL+F

Thus far we have focused on the effect of the diverse tools on users' efficiency, thus ignoring another important factor: the use of browsers' "find" functionality. This functionality plays a major role in relocating a specific piece of information within a web page. In order to quantify its degree of use, we measured the percentage of each group's tasks that were carried out with its help: 53.5% (Search Engine Group), 62.3% (Bookmark Group) and 17.2% (Annotation Group).

Apparently, CTRL+F has been extensively used by the subjects of Search Engine and Bookmark, whereas participants of Annotation resort to it less frequently. They actually use it solely in the cases of modified web pages that result in misplaced or orphan annotations; in these cases annotations are anyway of little help and the user has to resort to other means for pinpointing the desired information.

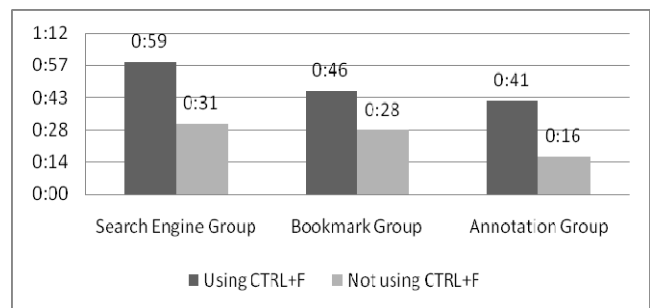


Figure 2. Average times of each group distinguishing tasks where the browser's "find" functionality was used.

Judging from the wide use of the CTRL+F strategy, it is reasonable to assume that "find" helps participants to perform better in refinding information. To verify this assumption, we estimated the average browsing time that corresponds to subjects using it and compared against the browsing time of those that did not use it. This comparisons were made in the context of all three

groups and their outcomes, presented in **Figure 2**, suggest the opposite: participants that took advantage of this functionality needed *significantly* more time in completing their tasks than those that did not. Hence, although this functionality is supposed to constitute a quite handy tool for locating information, in practice there is no evidence supporting its beneficial contribution to re-visitation efficiency.

## 5. DISCUSSION

In this paper we evaluated three different approaches to refinding information: web search engines, online bookmarks and online in-context annotations. The main focus of our study was on the reading and browsing phase that follows the searching stage of this process. The outcomes of our experiments suggest that bookmarks and annotations outperform search in terms of performance. Moreover, we observed a benefit of in-context annotation compared to bookmarks in terms of content recognition.

We also investigated the correlation of the browser's "find" functionality (CTRL+F) with refinding efficiency and observed that it does not actually account for any improvement in the browsing time. The questions that involved the "find" functionality typically entailed larger and unorganized pages; however, the expected benefits in terms of saving time could not be observed. For this reason, it would be beneficial for annotation tools to reduce this burden by minimizing users' cognitive load, interactions and wasted time.

Judging from our analysis, we expect annotation systems to be the next step in the evolution of hypermedia systems, just as personal/shared bookmark systems constituted a big step forward in the past. In our opinion, the main challenge will be on the user interface level: we see a need for more consistent, usable, integrated and interactive tools that will attract even the less engaged users. To do so, we plan to improve the management, browsability and searchability of collections of annotations. By structuring annotations and bookmarks in different dimensions and categories, and having multiple ways to reorganize it (e.g. alphabetic order, date of creation, date of re-visitation, date of shared access, etc) we aim to create hypertrails. These hypertrails are a personal/shared collection of hyperlinks that can be connected with each other, according to each user's preferences and needs, following the vision of Memex [3].

## 6. REFERENCES

- [1] Abrams, D., Baecker, R., and Chignell, M. Information archiving with bookmarks: personalWeb space construction and organization. Proceedings of the SIGCHI conference on Human factors in computing systems, pages 41–48, 1998.
- [2] Adar, E., J. Teevan, and S. T. Dumais. Large scale analysis of Web revisitation patterns. CHI '08, pages 1197-1206, 2008.
- [3] Bush, V. (1945) As we may think. The Atlantic Monthly, July 1945.
- [4] Capra, R. G. and Pérez-Quiñones, M. A. Using Web Search Engines to Find and Refind Information. Computer 38, 10 (Oct. 2005), 36-42.
- [5] Cockburn, A., and McKenzie, B. What do Web users do? An empirical analysis of Web use. Int. J. of Human-Computer Studies, 54(6): 903-922, 2001.
- [6] Guy, M., Tonkin, E. Folksonomies: Tidying up Tags? In D-Lib Magazine, 12(1), 2006.
- [7] Herder, E. Characterizations of user Web revisit behavior. In Proceedings of Workshop on Adaptivity and User Modeling in Interactive Systems, pages 32-37, 2005.
- [8] Herz, R. The effects of cue distinctiveness on odorbased context-dependent memory. Memory & Cognition, 25(3):375–80, 1997.
- [9] Jones, W., Bruce, H., Dumais, S. Keeping Found Things Found on the Web. In Proc. CIKM 2001, 119-126.
- [10] Kawase, R., Nejd, W. A Straightforward Approach for Online Annotations: SpreadCrumbs - Enhancing and Simplifying Online Collaboration. In: WEBISTINSTICC Press (2009), p. 407-410.
- [11] Kawase, R., Herder, E., Nejd, W. Annotations and Hypertrails With SpreadCrumbs - An Easy Way to Annotate, Refind and Share. In: WEBISTINSTICC Press (2010).
- [12] Kawase, R., Herder, E., Nejd, W. A Comparison of Paper-Based and Online Annotations in the Workplace. In Proceedings of the European Conference on Technology-Enhanced Learning (EC-TEL 09), 2009.
- [13] MacKay, B., Kellar, M., and Watters, C. An evaluation of landmarks for refinding information on the web. In CHI '05 Extended Abstracts on Human Factors in Computing Systems (Portland, OR, USA, April 02 - 07, 2005). CHI '05. ACM, New York, NY, 1609-1612.
- [14] Nishimoto, I. and Toda, M. Process-Recollective Refinding on the Web. In Proceedings of the 2006 IEEE/WIC/ACM international Conference on Web intelligence (December 18 - 22, 2006). Web Intelligence. IEEE Computer Society, Washington, DC, 883-892.
- [15] Obendorf, H., Weinreich, H., Herder, E., and Mayer, M. Web page revisitation revisited: Implications of a long-term click-stream study of browser usage. In Proceedings of CHI '07, pages 597-606, 2007.
- [16] Pirolli, P. 2007 Information Foraging Theory: Adaptive Interaction with Information. 1. Oxford University Press, Inc.
- [17] Pitkow, J. E. and Kehoe, C. M. 1996. Emerging trends in the WWW user population. Commun. ACM 39, 6 (Jun. 1996), 106-108.
- [18] Tauscher, L. and S. Greenberg. How people revisit Web pages: Empirical findings and implications for the design of history systems, International Journal of Human-Computer Studies, v.47 n.1, p.97-137, July 1997.