Annotating the Web: An Exploratory Study of Web Users' Needs for Personal Annotation Tools[*]

Xin Fu, Corresponding Author

School of Information and Library Science, University of North Carolina at Chapel Hill, CB# 3360, 100 Manning Hall, Chapel Hill, NC 27599-3360.fu@email.unc.edu

Tom Ciszek

School of Information and Library Science, University of North Carolina at Chapel Hill, CB# 3360, 100 Manning Hall, Chapel Hill, NC 27599-3360.tommy@email.unc.edu

Gary Marchionini,

School of Information and Library Science, University of North Carolina at Chapel Hill, CB# 3360, 100 Manning Hall, Chapel Hill, NC 27599-3360.march@ils.unc.edu

Paul Solomon

School of Information and Library Science, University of North Carolina at Chapel Hill, CB# 3360, 100 Manning Hall, Chapel Hill, NC 27599-3360.solomon@ils.unc.edu

This exploratory study investigates the needs Web users have to make annotations for their personal use when they view Web pages. Three forms of annotations observed on printed documents - text selection and emphasis, association building, and document re-segmentation - are examined in the Web environment. The study shows that

text selection and association building through notes or symbols remain the dominant forms of annotation on the Web, while structural annotation (re-segmentation) and layout annotation (change of font, color, etc.) are also prevalent. The study also investigates users' preferences for the tools designed to facilitate Web annotation practices. Findings suggest that usability is of utmost importance when developing Web annotation tools, and that under the current technical conditions, users welcome lightweight annotation functions, which are built into standard Web browsers.

Introduction

Annotation has been a popular topic at recent conferences on digital libraries, computer supported cooperative work (CSCW), knowledge sharing and management, and in other related fields. This is partly due to a consensus in the Web community that the current structure of the Web has not yet reached its full potential as 'an interactive information sharing device', the dream outlined half a century ago in Bush's landmark article 'As We May Think' (Bush, 1945). So far, the exchange of ideas on the Web is primarily one-way. Authors are given many tools for displaying whatever information they choose, while viewers are only given a limited means of interacting with that information, such as clicking on links and making bookmarks. It is believed that an interactive, two-way system where Web users could have an opportunity to mark-up Web pages in a similar way as they take notes when they read books would allow for a greater exchange of ideas on the Web (Heck, Luebke & Obermark, 1999).

Researchers typically refer to the notes taken by readers (of printed documents or Web pages) for later personal use or for the use of others as annotations. Formally, annotation can be defined as a value adding note or marking that is linked to an extant information object, representing a record of interaction between the reader and the information object (MacMullen, 2005). Marshall (1998) claimed that 'annotation is a fundamental aspect of hypertext' and 'ideally, such annotation increases not only the overall girth of the hypertext, but also its value'. She also noted that 'many of our systems and methods are (justifiably) geared toward the initial design of a hypertext, not toward the annotation of existing materials. While the role of the reader-as-navigator is widely acknowledged in Web development efforts, the reader-as-annotator is a far less common emphasis.' (p. 40)

The study presented here centers on Web users as annotators. It aims to understand Web users' needs for annotation tools to make 'the kinds of fluid annotations one sees on paper' (Marshall, 1998). We investigate how Web users utilize existing annotation tools

for personal annotations, their likes and dislikes of these tools, and what features would be desirable but are presently unavailable. We also discuss the design implications of these findings for the development of future personal annotation tools.

Related Studies

In this section, we first review studies of how people annotate printed documents, which sets a benchmark for studying annotation behavior in the Web environment. Then we briefly introduce a few general electronic annotation facilities before we review annotation tools that are available to annotate the Web. At the end of this section, we identify the gaps in previous studies and state the research questions that guided the study reported here and our ongoing efforts to investigate annotations in structured data environments.

Annotation of printed documents

Although the emergence of hypertext technology provides great convenience for the creation of links between two nodes, the practice of annotation had been in existence long before the Web and the hypertext era. Also, the term 'information object' is not confined to information in electronic form. A number of studies examined the practice of making annotations on paper as well as the use of annotations and their functionality in the context of hypertext. In a summarization task, O'Hara and Sellen (1997) outlined several general annotation types and functions from the usage perspective. Marshall (1998) conducted a long-term study on people's annotation behavior as they read textbooks and identified how five common hypertextual elements were realized in annotations on paper. Ovsiannikov, Arbib & McNeill (1999) carried out a survey to study how people make annotations in an academic research environment.

These investigations and observations reveal several forms that annotations on printed documents can take. One common form of annotation is *text selection and emphasis*: highlighting, underlining or circling a group of words, drawing stars, asterisks or other symbols in the margin adjacent to certain sections in the text, adding span notations to several sentences or paragraphs, folding a

page, or placing a bookmark between two pages. All these methods are often used to indicate that a particular part of the text is important. At the same time, they also serve as visually appealing shortcuts to the selections.

Another form of annotation is *building associations*, links or relations by making notes and drawing symbols. This can be further divided into four sub-types. The first is annotations at the 'collection' or composite level, which refers to many subparts of a single document. The second type consists of node-to-annotation links - annotations that do not visibly refer to any particular document element, but are localized within a document part (e.g., a longish note written orthogonally to the printing on a page). The third type is 'standard' hypertext associations - from anchored portions of the text to notes or commentaries. Such a note might be written in the margin, the top or the bottom of the page, or on a separate piece of paper. The last type is word-to-word associations usually written between lines of text (e.g., in foreign language texts, a student may translate a word into his or her native language).

A less frequently observed form of annotation is a *re-segmentation of the document*, which often happens when the physical structure of the document (e.g., for books, the arrangement of chapters, sections, etc.) does not suit the reader's purposes. Short of tearing out pages or making copies, cutting them up and reorganizing them, users have limited ways to make this type of annotation, one of which is switching marker colors (Marshall, 1998).

Research has also found that no matter how annotations are constructed, they provide benefits for the annotators and/or their communities. Ovsiannikov, Arbib and McNeill (1999) summarized the primary uses of annotations as 'to remember, to think, to clarify and to share'. Presumably, annotations made to share are less idiosyncratic and more standardized (e.g., proofreading marks as an extreme example). On the other hand, among these annotation methods, those for making content annotations (such as text selection, emphasis and adding notes) are well-supported given the variety of highlighters, pens, clips, post-its, and other technologies; however, structural annotations, such as creating a logical structure that is different from the physical structure or linking from one place to another place in a book, are inadequately supported due to the linear structure of printed documents. Another drawback of annotation on printed documents is the lack of support for collaborative annotation as a printed document may only be read by one person at a time unless it is duplicated at a cost. For example, as Heck, Luebke & Obermark (1999) commented, citing Ka-Ping Yee, the designer of the CritLink Mediator, 'an inaccuracy in a document that could have been corrected once by a reader must be noticed and re-corrected by each new reader; related supporting material to a document suggested by one reader must be re-found by each new reader; and so on, leading to incomplete information and wasted time'.

Annotation facilities in electronic environments

More and more computer applications now incorporate annotation functions into their design. Microsoft Word provides collaborative writing functions by allowing users to track changes when they edit a document. The latest version of Adobe PDF provides powerful 'comment' functions, from underlining and highlighting text to adding notes and time stamps.

There are also other annotation facilities which are designed to facilitate people's annotation practices in different situations. For example, Mu, Marchionini and Pattee (2003) introduced an application called ISEE that allows video annotation/discussion via a chat panel in sync with the video. Ramos & Balakrishnan (2003) presented a pen-based video annotation tool that supports frame or segment-based personal annotations.

Web Annotation Tools

The invention of hypertext technology and supplementary Web technologies provides incredible convenience for link making and path building between existing documents or document snippets. To harness the power of new technologies and give electronic documents some of the same note-taking possibilities as paper documents, people have developed various kinds of annotation tools and applications, from lightweight functions of adding a Web page to a 'Favorites' list or creating a short cut to the page on the toolbar, to a variety of more complex and specific-purpose annotation systems.

The functionality and implementation of existing Web based annotation systems varies greatly. Based on the method of implementation, we have identified four major types of annotation systems. The first type includes the annotation functionality *built into Web browsers* like Internet Explorer (IE), Netscape Communicator (NC), or Mozilla's Fire Fox (FF). These functions include saving a Web page (on a local drive and/or as a different format, such as text file), printing the Web page, temporarily selecting a section of the text, copying part of the Web page and pasting it somewhere else (such as in a text document), using a different color to

represent visited links, adding a Web page to the 'Favorites' list (IE) or the 'Bookmarks' (NC and FF), and creating a shortcut to the page on the toolbar (we will reference the last two functions as 'bookmarking' in the rest of the paper).

All these functions, except changing the color of visited links, text selection and copying-and-pasting a portion of text somewhere else, represent attempts to build links to the original Web page. Whether from the menu, the toolbar, a folder in the Explorer, or its hardcopy on paper, the links help a user in a future visit to the page. Changing the color of visited links is an emphasis function - it highlights the links that have been visited before. Selecting a section of the text by clicking and drawing the mouse over the text can also be used to emphasize, albeit temporarily. Copy-and-paste is another text selection function, but it stores a selected part of the page in a cache or a separate file so that the reader can concentrate on this part later in the session or in future visits.

A number of studies have been conducted to examine the use of these functions. For example, in the context of studying how people organize Web information for re-use (equivalent to annotations 'to remember'), Jones, Dumais and Bruce (2002) observed the use of 10 information keeping methods, including saving, printing, bookmarking, and using the 'Links' toolbar. Other studies focused on the use of one function (mostly the bookmarking function), such as Abrams, Baecker and Chignell (1998).

The second type of annotation system consists of *personalized Web information organization systems* provided by a few commercial Web portals for the public, such as My Yahoo! and My MSN, and by some organizations (i.e., libraries, businesses and universities, such as Morgan's MyLibrary [http://dewey.library.nd.edu/mylibrary]) for their user populations. The general aim of these systems is to help users to collect their favorite parts of the site and the Web in one place and display them in a favorable manner. When people log into one of these sites, they will be allowed to annotate a uniformly defined starting page and customize its content, layout and structure according to their interests and preferences.

The content is supplied by the home site and a variety of RSS (Really Simply Syndication) sources, sites that syndicate their contents. Most commercial systems only allow the users to select from the provided content, so it is not surprising that Chia and Garcia (2002) argued that 'those personalized pages are usually only additional features limited to the content offered on that particular site'. However, non-commercial systems often allow the users to enter the URL of their favorite sites and add these sites into their customized page. This is equivalent to the bookmarking function that standard browsers provide and becomes a form of hyperlink aggregation. Users can also rearrange the order of the content on their customized page, either through a logical sequence, which

makes more sense to them, or simply by a perceived order of priority. Finally, they can change the color and the layout of the page to make it more appealing.

The third type of annotation system on the Web is *interactive Web publication forums* such as blogs and wikis. Blogs are Web applications, which contain periodic, reverse chronologically ordered posts on common Web pages. They are the logical descendants of online forums and news groups. They can also be open and allow anyone to create new topics and start discussions. They turn a website into an active information distribution channel that could give rise to much more interactive websites for organizations, political groups, small companies, and so on (Ramos, 2004). We believe that blogs are the beginnings of an interactive and annotated Web. Given data from the Pew Internet and American Life Project (Rainie, 2005), over 8 million people (7% of U.S. adult Internet users) are creating blogs or web-based diaries and more than one-in-ten U.S. Internet users (12%) are posting on others' blogs. These statistics provide face-validity substantiating further research in the development of blogs as a tool for Web annotation.

Even more flexible is the wiki defined as 'a freely expandable collection of interlinked Web 'pages', a hypertext system for storing and modifying information - a database, where each page is easily editable by any user with a forms-capable Web browser client' (Leuf & Cunningham, 2001). Wikis go one step beyond blogs by not only allowing users to add content, but also allowing that content to be edited by anyone. (http://en.wikipedia.org/wiki/Wiki).

Both systems support easy interaction among Web page authors and readers. They provide crucial features that contribute to the decentralization of authority thus facilitating the correction of mistakes in Web content. Although we found no studies to date on the use of blogs and wikis from the explicit 'annotation' point of view, it is quite obvious that their support for annotation is limited - in blogs, readers can only publish their comments as new posts, but do not have the capability to attach annotations at specific parts of an existing post or edit the post itself. Wikis address this limitation, however Wikis and blogs rarely provide support for structural annotation - the arrangement of entries is mostly linear although certain flexibility is sometimes allowed in sorting the entries by date, author or other criterion. Even with these limitations, blogs and wikis represent a huge advance toward annotation of the Web by giving 'readers' channels for adding public annotations to the discourse.

The last type of online annotation system includes *annotation engines,* which help people mark-up documents on the World-Wide Web. These engines are referred to in some publications as 'Web annotation tools', but strictly speaking, they are only a subset of

Web annotation tools. They share the basic goal of allowing readers to add annotations to an arbitrary webpage and to reference them to any place within the page displayed in a browser. Additional common features that make them more flexible and useful than their paper counterparts include annotation search, indexes, shared annotations, typed links (e.g., query and issue), and reply threading (Chong & Sakauchi, 2001). Many authors, such as Garfunkel (1999), Chong and Sakauchi (2001), Denoue (1999) and Zohar (1999), have described and compared the implementation of annotation engines. Generally speaking, annotation engines are often implemented through either proxy-based intermediaries or client-based intermediaries. The pros and cons of each approach are discussed in Denoue (1999). For example, a proxy-based intermediary allows the annotation to be viewed in any browser, on any computer; but raises issues related to network traffic and privacy. Client-based intermediaries save the annotations locally; however, they usually require installation and configuration.

Although rarely discussed in the literature, there is a possibility of automatic annotation or recommendation by analyzing someone's Web visit log. The user selects which Web pages to annotate, then annotation (by means of recommendation) is done automatically every time the person visits these pages/sites based on his/her previous activities. For example, the system could display the most frequently visited news section at the top if the user selects to annotate a news page which is visited regularly. Another example is highlighting the most frequently visited links on the page (e.g., the links to library, gym hours and campus news) if a user elects to annotate his university homepage.

Taken together, the above four types of annotation tools provide almost all the functions that can be accomplished on paper. In addition, they also provide some functions, which are difficult to realize in the paper environment, such as easy re-segmentation of a document. However, there is not a single approach that supports all of the features that would work with any standard browser (Chong & Sakauchi, 2001). Also, the existing tools have little connection between each other. Some of them require installation. Some require users to log in before use. These factors may all become potential burdens for their users. Several researchers have noted these problems and called for improvement in annotation tools, especially the development of easy-to-use personal annotation tools. For instance, Zohar (1999) argued that 'there is nowadays no widespread annotation service'. Denoue (1999) claimed that 'there is no wide spread usage of' Web annotation systems. We strongly feel that a genuine understanding of the reasons for the lack of popularity of the existing annotation services and users' expectations for future tools will provide important guidance for the improvement of annotation system design and overall user productivity and satisfaction. Unfortunately, a survey of published research on Web annotation reveals that papers presenting prototype applications by far outnumber those trying to study

how and why Web users annotate. Although great effort is being directed towards the development of annotation tools, relatively few studies are dedicated to understanding what the users actually need. Furthermore, the existing literature shows an overwhelming bias towards annotation engines, while there are few studies addressing the other three types of annotation facilities.

The study presented here aims to address some of the above problems. Specifically, it focuses on Web users' explicit needs for personal annotation tools. By personal annotation, we refer to the type of annotation that people make primarily for their own use, in contrast with the annotation made with an intention to share with other people in a collaborative effort. As mentioned in Marshall & Bush (2004), this type of annotation constitutes the vast majority of the annotations that Web users currently make on the Web. Compared to public, collaborative annotations, they are closer in nature to annotations that people make on printed documents. However, as Marshall (1998) noted, 'there are on-line Web-based tools that support shared commentary, but most of these are not really intended for the lightweight personal annotations one encounters in the varying circumstances that constitute reading.' Meanwhile, Denoue and Vignollet (2000) pointed out that personal annotation systems required different design choices than annotation systems that were designed so that everyone could leave comments on the Web - 'When browsing the Internet, users should be able to quickly highlight texts, like they do on paper. This is in contrast to the remote server approach where delays can appear and where the system forces the browser to reload the entire document in order to display the new annotation. Furthermore, the server raises important privacy issues.'

In view of all of these comments, we feel the need to study personal annotation systems from the human point of view. In particular, we are interested in the following questions: What forms do annotations on the Web take? Are they the same as in the paper environment? Which annotation tools do Web users actually use now to make personal annotations? How do they like these tools? Do these tools meet their expectations? If not, what and where is this gap? What kinds of new annotation capabilities do they want?

A byproduct of this study is to assess whether people make different types of annotations when they access different types of Web pages, such as news pages, academic articles, personal homepages, institutional Web pages, and Web portals. Our hypothesis is that people make one form of annotation more than the other when they access certain type of pages. For example, they may do more structural annotations when visiting portal pages or institutional homepages, while they may do more content annotations when visiting a news page or a page displaying an academic article.

Methodology

The study relied on semi-structured interviews as its primary method. Participants were recruited through campus listservs. There is no evidence to show that Web users' needs for personal annotation varies according to their occupation or academic fields. However, experience with computers and the Web is expected to have an impact on use of annotation tools, that is, people with more exposure to the Web are more likely to have some experiences with at least some annotation tools, and people spending more time on the Web may have clearer needs for annotation tools. Thus, a sample biased towards people with medium to high levels of computer and Web experiences was drawn. The study was conducted in December 2004 and January 2005. Before the formal study, a pilot study was administered to test the interview protocol and get some preliminary understanding on how people utilize the available personal annotation tools.

Each interview took approximately one hour. It started with questions on participants' demographics and computer and Web experiences. The first part of questioning focused on the participants' Web annotation experiences. The interviewer mentioned the different forms of annotations observed in the paper environment (text selection and emphasis, association building through notes or symbols, and document re-segmentation) and asked the participants if they had made similar annotations on the Web. Next, the interviewer asked questions about the participants' experiences with four types of Web annotation tools reviewed in the last section: annotation facilities in the standard browsers, website supported annotation systems, interactive Web publication forums, and Web annotation engines. The participants were asked to describe their experience with each type of tool and, if used before, their likes and dislikes. If a participant had no experience with a particular type of tool, the functions of the tool were described (or demonstrated when necessary) by the interviewer and the participants were asked to assess the potential usefulness of that kind of tool.

Then, the participants were asked to reflect on their Web experiences and describe specific situations when they realized the need for making annotations but could not find any tool to support the need, either because they did not know that such a tool existed, or because they knew about some tools which could accomplish the task but decided not to use them due to their unfavorable features. Finally, the interviewer recapitulated several annotation scenarios that the participants mentioned and asked them to discuss if their annotation behavior was different in different scenarios as well as in some other scenarios that they had not

mentioned.

Results and Discussions
Participants

Twenty-one Web users participated in the study, among whom 13 were male and 8 were female. One of them was between 18 and 20 years old, eight were between 21 and 25, seven were between 26 and 30, three were between 31 and 35, and two were above 36. All of the participants identified their computer skills and Web use experiences as 'medium', 'medium to high' or 'high' on a five-point Likert scale. We are aware of the coarseness of these two measures, but considering the fact that we only use them as a rough check for the expertise of the sample, we feel that they are acceptable.

Forms of Annotation

Recall that three forms of annotations have been observed on printed documents: text selection and emphasis, association building through taking notes and drawing symbols, and re-segmentation of the document. In the Web environment, bookmarking, printing, saving, copying and pasting (copying part of the text or the URL and pasting it into a text file or an email), and temporary selection were the most frequently reported forms of annotations. As displayed in Table 1, all the participants in our study used bookmarking, printing, saving and copy/paste functions to annotate Web pages that they feel are important and likely to re-access. Eight (38.1%) participants regularly use temporary text selection as a place holder on the Web page that they are reading, either when they get distracted by another task or simply to help themselves focus. On the other hand, only one participant has added notes to and highlighted Web pages using an annotation engine.

Table 1. Mentions of different forms of personal annotations on the Web

Form of Annotation	Annotation Function	#	%
Content Annotation	Bookmarking	21	100
	Printing	21	100
	Saving	21	100
	Copy/Paste	21	100
	Temporary selection	8	38.1
	Adding notes	1	4.8
	Highlighting	1	4.8
Structural Annotation	Changing order of Web page content	13	61.9
Layout Annotation	Changing color, font, or size of the text	6	28.6

The use of structural annotation and layout annotation on the Web seems to be on the rise compared with the paper environment. Thirteen (61.9%) users reported experiences of changing the order of content on Web pages either by copy/pasting it to text editing software like Microsoft Word or using website supported annotation functions. Six (28.6%) users have changed the color, font, or size of the text on Web pages. The tools for making all of the above forms of annotations are discussed in the next section.

Generally speaking, the participants reported building links to a Web page from bookmark, toolbar shortcut, hardcopy, or aggregating them in other formats as the most common methods for annotating Web pages. The real 'content annotation', such as highlighting and note taking, is usually done on the printed or re-formatted copy of the page. We also noticed an apparent increase in structural and layout annotations on the Web relative to the paper environment due to better support of tools and the inherent features of the Web as a digital medium.

Use and Non Use

In the second part of the interviews, the participants were asked to talk about their experiences with each of the four types of Web annotation tools described above in the 'related studies' section. The number (and percent) of participants associated with each type of tool use experience are summarized in Table 2.

Table 2. Use of existing annotation tools

Type of Annotation Tool	+	(+)	-+	--
Annotation Facilities in Standard Browsers	21 (100)	0 (0)	0 (0)	0 (0)
Website Supported Annotation Systems	5 (23.8)	0 (0)	7 (33.3)	9 (42.9)
Interactive Web Publication Forums	2 (9.5)	5 (23.8)	3 (14.3)	11 (52.4)
Web Annotation Engines	1 (4.8)	0 (0)	9 (42.9)	11 (52.4)

Code: +: Have used before for personal annotation; (+): Have used before, but not for personal annotation; -+: Have not used before but feel potentially useful; --: have not used before and do not need the functions.

The results show that all the participants have employed the annotation facilities built into the browsers such as bookmarking functions. Five (23.8%) participants have tried to use website supported annotation functions. They gave My Yahoo! and/or the service provided by UNC (http://my.unc.edu) as examples of this type of system. Seven (33.3%) participants had some experience with blogs or Wikis, but only two (9.5%) of them used these forums for personal annotation purposes. Finally, only one (4.8%) participant has used Web annotation engines and for making personal annotation. Notably, around half of the participants have not

used and seemed not to need the last three types of annotation tools. The specific reasons for this apparent lack of interest will be explained in the next section, but the results summarized in Table 2 indicate that in general these participants prefer the tools that are ready to hand, such as those coming with standard Web browsers.

Likes and Dislikes

Participants' likes and dislikes for each type of tool were also investigated in the second part of the interview.

Annotation facilities in standard browsers

The biggest pros for annotation functions in standard browsers are their convenient availability and readiness to use (with the proviso that a standard Web browser application is being used), which enables wide spread usage as a lightweight personal annotation facility. However, browsers only provide limited support for annotation and the functions are not always satisfactory. Many participants complained about the bookmarking functions. A typical complaint was:

> I used to add a lot of pages that I came across and found useful into my 'Favorites' list, but in fact I seldom visited them again. When there are so many pages on the list, it's just difficult to find the page. Also, even if I do remember there is a link there and try to click it, it's not so easy to always click on the right one since each shortcut only takes a tiny space. If there is something I use a lot, I would rather drag them onto the toolbar. You know, that's a problem again. The toolbar doesn't hold too many icons either.

Another frequently mentioned defect of the bookmarking function is that the bookmarks as objects themselves are not conveniently accessible or not accessible at all across platforms and machines. They also easily get lost when one upgrades or changes computer. In fact, eight participants mentioned that they would rather make a Web page or a text file which contains links to their frequently visited pages, rather than bookmarking these pages in the browsers. Five participants already had this kind of personal

portal page or link file. They all felt this solution to be more convenient than the menu or toolbar shortcuts. Of course, designing and maintaining a personal portal page requires some Web authoring skills and access to Web space, which many Web users do not possess; saving links into a text document sometimes results in the loss of address information.

Although standard browsers support copy-and-paste of text on a Web page, several users felt that this approach was not convenient for selecting a portion of the text. They would prefer that the selection be made and stored on the page itself.

Website supported annotation functions

Website based annotation systems provide friendly interfaces to support users' annotation of Web page content, structure and layout. Although only five (23.8%) users had used them before, seven (33.3%) users indicated an interest in trying them after the interviewers described or demonstrated their functions. The idea was particularly appealing for those 'multi-task' participants who usually performed several tasks at the same time when they worked online.

However, a big reason that about half of the participants did not like these systems is that the website dictated what could be annotated and how. All the structure was predefined by the provided metadata and the users could only make adjustments within this frame. Personal comments or interpretations are not easily made. Additionally, the need to log in for access is also a factor that steers some users away. As one participant said,

> I have 'played with' a couple of sites like My Yahoo! but simply found them inconvenient. You had to type your username and password. Then you could only select from the contents that they provided. Rather than putting these contents on a quasi-personalized page, I would rather design such a page by myself.

Interactive Web publication forums

Only two (9.5%) participants had used interactive Web publication forums for personal annotations. They both kept diaries on

personal blogs, which they did not advertise to anyone. They liked the blogs' capacity of archiving previous posts, and one went on to express the desire that further content additions (such as pictures) be made available. Five (23.8%) other participants also reported on blog use, but not for personal annotation purposes. They liked the capability to shape and define the topic of a blog and the ease of receiving feedback from other people. However, privacy was a big concern when they added material to a blog.

Among the participants who have not used a blog before, three (14.3%) showed interest in its functions. One person would consider setting up a blog to save his pictures. He preferred a blog to an online album because he could add commentaries and stories to the pictures.

Five (23.8%) participants had used wikis before, but they all felt that wikis were more for collaborative annotation and research purposes rather than for personal use. They liked the freedom to edit and preserve in such a format. But although wikis provide a mechanism of saving all the previous posts, those interviewed also pointed out several features that they felt inconvenient, such as the need to log in, the use of non-standard (non-HTML) formatting tags (e.g., three quotes instead of <b> for bold face), the inability to save the posts locally, and the difficulty in permanently deleting a post. Given the increasing popularity of blogs and wikis, annotation capabilities are sure to improve and we will focus attention on experienced users in future studies.

Web annotation engines

Only one participant had used Web annotation engines. He said that he had tried several Web annotation services, but with only one success (Gibeo.net). For that one, he was excited with the functions that it provided (highlighting, adding comments, adding links from any part of text, performing a Google search for a word or phrase, etc.) and claimed that these functions were the closest to those he used to annotate printed documents. He also liked its feature of requiring no installation and being able to view the annotated pages without logging in (and thus be able to share the annotated page with a friend). His main complaints about the tool were two-fold. Firstly, it was sometimes slow due to the need to retrieve the annotated page from the annotation server. Secondly, the anchor had to be text in the current system. Therefore, it was inconvenient to add a comment to reference a whole paragraph. He also noticed that only strings longer than 4 characters could be used as anchors (e.g., 'at' or '100' are not qualified as anchor texts) and he was puzzled why it worked that way. Overall, he was quite happy with the engine, but meanwhile, he also emphasized

the frustration he experienced when he was trying other tools 'the installation and configuration looked intimidating'.

The participants who had not used annotation engines before had mixed feelings towards the devices. On the one hand, some participants liked the idea because of the possibilities they opened up for actually adding annotations on top of the Web pages. As one participant commented, it was very important to bring annotations close to the original information object so that people could easily see the associations. On the other hand, several participants voiced concerns towards both approaches of building the engines. For the engines implemented at the server side, several participants were very concerned about the privacy of their annotations when a company tracked not only what they visited, but also what they marked or wrote. Some participants expressed that they did not like the fact that they had to have Internet access to view their annotations. They were also concerned with the possibility of losing all their annotated pages if the service was closed or had server problems. One participant was worried that server-side annotation engines might charge the user repeatedly, so he would opt for a one-time payment to buy some software and run it on a local machine. Another participant raised the issue of copyright. He suspected that annotating other people's Web pages and putting the annotated pages online might be an infringement of copyright. The concerns for client side implementation were mostly related to the inaccessibility of annotated pages across computers and loss of the pages when computer problems occurred. But overall, the participants seemed to trust client based intermediaries more than those that are server based.

Expectations

Participants provided a variety of usage scenarios for Web annotations. Some of the functions that they expected may actually exist, although they were unaware of their availability or were reluctant to use them because of some annoying features or requirements. For example, one participant did a lot of reading on the Web and often came across good articles that she would like to recommend to her friends. Beyond emailing the entire page, she stressed that she wished to highlight the parts that she felt were particularly well written so that her friends would not miss them. This function was already provided by the annotation engine that another participant used.

Other expectations that were mentioned by the participants include:

- Seven people mentioned that they would prefer to have annotation functions built into their Web browser, instead of getting them from a website or a server-based service.
- Two participants mentioned that they would like to be able to search the annotations that they had made previously.
- Two people mentioned their wish for automatic annotation (or recommendation) by analyzing users' behavior. For example, when using the Google search function built into the right-click menu in Fire Fox, it would be nice if it would automatically refine the search by considering the context in which the search term appeared.
- Three people mentioned that all the current annotation functions were either associated with the entire Web page or a set of anchor text. A tool that would allow people to add annotations at an arbitrary place on the Web page and lay the annotations 'transparently' above the original page was highly desirable. The 'comment' functions provided in Adobe PDF are good examples of such flexibility.
- Two participants mentioned that they would like to have an annotation facility that allowed them to flag important passages within a Web page and automatically redirect them to that place at future visits.

Annotation in different situations

We tried to assess whether people made different types of annotations when they accessed different types of Web pages, but we realized that there were many other factors, besides the type of Web pages, that affected Web users' annotation behavior. The factors we discovered in this study included the type of task performed, the perceived importance of the page, the likelihood that the page will be revisited, their computer experience / Web experience, and their reading habits. For example, differences emerged between people according to their reading habits -- some participants really hated to read online and printed out almost everything that they could not finish reading within a couple of seconds, while some participants liked to do most of their reading online.

One pattern that we noticed was that people relied more on traditional methods such as printing or email when they wanted to annotate something really important, such as their work document, research or private information. This reflects a recurring theme with regard to the reliability and privacy of storing information on the Web.

Summary of Results

Generally speaking, these results suggest that users' annotation needs in the Web environment are similar to those in the paper environment. "To remember, to think, to clarify and to share" remain the major uses of personal annotations. Meanwhile, many people take advantage of the Web's hypertext nature and use annotations as an aid for navigation and organization. Content annotations, such as text selection, emphasis, note taking, link building, and page editing are still the most commonly used types of annotation and are extended by needs for structural annotation and layout annotation, which unsurprisingly increase in the Web environment.

Four types of Web annotation tools - annotation facilities in standard browsers, website supported annotation systems, interactive Web publication forums and Web annotation engines - support people's needs for personal annotations to various extents, each with different advantages and disadvantages. Annotation facilities in the standard browsers were favored by the participants because of the pervasiveness of the browsers. However, their support for text selection, emphasis, and note taking are negligible. Website supported annotation systems provide friendly interfaces and easy-to-use structural and layout annotation functions, but have rigid restrictions on the pages to be annotated. Their service is largely dependent on the availability of metadata on the Web pages. Interactive publication forums, such as blogs and wikis, are used more for collaborative annotation than for personal annotation. They also bring functional limitations and privacy issues.

Web annotation engines provide many functions that users like, but have severe usability defects. The technologies used by current annotation engines still need to address the bulk of the technical issues described in the literature, such as a question raised repeatedly by four participants: when users visit a page they have previously annotated, but the page gets updated in the interim, should they be given the original page with annotations, or the updated page without annotation, or rather a choice between the two? Several other issues are discussed in Ovsiannikov, Arbib, and McNeill (1999). Currently, the choice between the proxy-based and client-based approaches seems to be a dilemma since the merits of the two approaches are apparently complimentary and the users clearly desire both. As the participant who had used an annotation engine noted, he would prefer to have some functions similar to those of the annotation engine built into a standard Web browser so that they could be accessed more easily and

preferably without logging in.

Conclusions and Future Research

This study marks one of the first attempts to reach an intensive understanding on Web users' need for personal annotation tools. Although the sample size was relatively small, findings of this study provide important insights into Web users' annotation behavior.

The finding that annotation tools are not yet popular among Web users echoes the results of previous studies (e.g., O'Hara and Sellen, 1997). Without good support, Web users' needs for making annotations when they 'read' Web pages are largely latent or suppressed. Their immediate responses when asked if they made annotations on the Web were usually something like 'I would love to, but how can I?' As Marshall (1998) noted, personal annotation tools are on the horizon that will enable people to engage with documents and hyperdocuments in a way that aligns with observed practice. We also feel it is safe to claim that there are many opportunities for future research in this area and a potential market available for developments based on such research.

Preliminary findings from this study suggest that people's needs for making annotations in the Web environment do not differ significantly from their needs in the paper environment. With more and more reading being done online, most participants are enthusiastic about the possibility of making annotations on Web pages. All of them could enumerate several situations when they felt the needs for annotation capabilities. At the same time, however, we found that the participants were not entirely satisfied with the existing annotation tools. We noticed that usability was one of the chief concerns for the participants. For almost all of them, annotation functions are desirable, but not indispensable parts of their use of the Web; it seems to them that the Web offers 'clean' pages without scribbles. Lacking support for Web annotations, they have developed many workarounds such as printing out the pages and then marking the printed version or saving the text into a Word document and using the annotation functions there. Most of the participants noted that they were not sure if they still would like to use an annotation tool if they needed to log in every time they wanted to annotate something or they could not install the tools within a few clicks of mouse. Ideally, they want to have all the functions coming as a natural extension of the standard browsers that they use all the time. Reliability and privacy are also common concerns among the users. Thus, an easy-to-use, lightweight annotation facility built into today's standard Web browsers, supporting

basic functions of highlighting, marking, adding comments, and building links seems to be a promising next step based on the findings of this study.

The study has several limitations, which raise the need for further work. First of all, as an exploratory study, these results have limited generalizability. The participants that we interviewed are not representative of the Web user population and the four types of tool reviewed in this article and used in the study by no means cover all the latest annotation technologies on the Web. Moreover, there were only a few participants in our study who had used the last three types of annotation tools. Their judgment on the usefulness of the tools was based on our descriptions and demonstrations, instead of real experiences. Therefore, in the next phase of this study, we will aim to reach a larger user population with experience in each type of tools and conduct a more in-depth study. We hope to design a naturalistic study where we can observe people's annotation behavior when they perform different tasks or view different types of Web pages. We also aim to focus particular attention on expert bloggers and wiki users. Our annotation project is using a wiki for internal use and our own experiences will inform some of these investigations.

We fall short of answering which factors affect personal annotation behavior, but we do reveal several factors that need closer investigation, such as task performed, perceived importance of content and the likelihood of future need for identical content. Further studies are needed on each factor to reach a better understanding of Web annotation needs and behavior.

Acknowledgements

References

Abrams, D., Baecker, R., & Chignell, M. (1998). Information archiving with bookmarks: personal Web space construction and organization. *Proceedings of CHI '98*, (pp. 41-48). ACM Press.

Bush, V. (1945). As We May Think. Retrieved December 28, 2004 from http://www.ps.unisb.de/~duchier/pub/vbush/vbush.shtml

Chia, C. & Garcia, J. (2002). The personalization challenge in public libraries: perspectives and prospects. Retrieved January 19, 2005 from http://www.public-libraries.net/html/x_media/pdf/personalisation_engl.pdf

Chong, N. S. T., Sakauchi, M. (2001). Creating and sharing Web notes via a standard browser. *Proceedings of the 2001 ACM symposium on applied computing* (pp.99-104). ACM Press.

Denoue, L. (1999). Adding Metadata to Improve Retrieval: Yet Another Web Annotation System. Retrieved June 1, 2005 from http://www.fxpal.com/people/denoue/publications/TR1999-01.pdf

Denoue, L. & Vignollet, L. (2000). New ways of using Web annotations. Retrieved December 15, 2004 from http://www9.org/final-posters/poster46.html

Garfunkel, J. (1999). Web Annotation Technologies. Retrieved January 2, 2005 from http://look.boston.ma.us/garf/webdev/annote/software.html

Heck, R. M., Luebke, S. M., & Obermark, C. H. (1999). A Survey of Web Annotation Systems. Retrieved December 27, 2004 from: http://www.math.grin.edu/~rebelsky/Blazers/Annotations/Summer1999/Papers/survey_paper.html.

Jones, W. P., Dumais, S., & Bruce, H. (2002). Once Found, What Next?: A Study of "Keeping" Behaviors in the Personal Use of Web Information. *Proceedings of the Annual Conference of American Society for Information Science and Technology (ASIST 2002)*, 391-402.

Leuf, B., & Cunningham, W. (2001). The Wiki Way: Quick Collaboration on the Web. Addison-Wesley Longmann.

MacMullen, W. J. (2005). *Annotation as Process, Thing, and Knowledge: Multi-domain studies of structured data annotation*. SILS Technical Report TR-2005-02. UNC School of Information and Library Science.

Marshall, C. (1998). Toward an ecology of hypertext annotation. *Proceedings of ACM Hypertext '98*, Pittsburgh, PA (June 20-24,

1998). 40-49.

Marshall, C., Brush, A. J. (2004). Exploring the relationship between personal and public annotations. *JCDL '04*. 349-357.

Mu, X., Marchionini, G., & Pattee, A., (2003). The Interactive Shared Educational Environment: User interface, system architecture and field study. In *Proceedings of the third ACM/IEEE-CS joint conference on Digital libraries (JCDL)*, Houston, TX (May 27-31, 2003). 291-300.

O'Hara, K., Sellen A. (1997). A comparison of reading paper and on-line documents. *CHI '97</>. ACM Press. 335-343*

Ovsiannikov, I. A., Arbib, M. A., McNeill, T. H. (1999). Annotation Technology. International Journal of Human-Computer Studies, 50(4): 329-362.

Rainie, L. (2005). The State of Blogging. Pew Internet and American Life Project. Retrieved January 14, 2005 from http://www.pewinternet.org/pdfs/PIP_blogging_data.pdf

Ramos, A. (2004). FAQ: What Is a Blog? Retrieved January 2, 2005 from http://www.andreas.com/faq-blog.html.

Ramos, G. & Balakrishnan, R. (2003). Fluid interaction techniques for the control and annotation of digital video. In Proceedings of ACM UIST '03 (pp. 105-114). Vancouver, BC.

Zohar, R. (1999). Web Annotation – an Overview. Retrieved January 2, 2005 from http://dlp.cs.berkeley.edu/cgi-bin/pl_dochome?format=pdf&collection=Digital_Documents&id=55&show_doc=yes