

Peiwen Zhu. Web Annotation Systems: A Literature Review and Case Study. A Master's Paper for the M.S. in I.S. degree. April, 2008. 35 pages. Advisor: Bradley M.Hemminger.

Web annotation has been a popular research topic since the appearance of Internet and its supplementary technologies. This paper provides a literature review in the related research areas, and introduces some currently available web annotation systems with the comparison of these systems in seven aspects. In addition, this paper incorporates a description of the ongoing NeoNote project, and identifies limitations of this study for future work.

Headings:

Scholarly Communication

Web annotations

Information Systems/Design

WEB ANNOTATION SYSTEMS: A LITERATURE REVIEW AND CASE STUDY

by
Peiwen Zhu

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Information Science.

Chapel Hill, North Carolina

April, 2008

Approved by:

Bradley M. Hemminger

Table of Contents

Introduction.....	2
Literature Review.....	4
Current Systems.....	15
Case Study.....	26
Conclusion.....	30
References.....	32

Introduction

Web browsing plays an important role nowadays in people's daily life, study, and work. Since documents exist mostly in digital format on the web, people may spend a large part of their time on browsing or searching on the web to look for useful information. However, this used to be a one-way interaction with users having few options to mark texts or to highlight important sections in a web document; what's more, it is difficult to add extra information as reference on web pages, which is useful for further reference or sharing with friends.

Web annotation has been an ongoing research issue since the invention of hypertext and supplementary technologies such as HTML, XML, and Wiki. In this paper, web annotations will be defined as:

“Online annotations associated with web resources such as web pages, with which users can add, update or delete information from a web page without modifying the page itself”.

There are various areas of continuing research and implementation on web annotation systems. Researchers have taken different approaches to develop and implement web annotation systems. D. Grant Campbell (2002) tried to use Dublin Core in Web annotation programs, while Vasudevan, V. and Palmer, M (1999) discussed some missing elements of current Web infrastructure which made any implementation of annotation systems less than completely satisfactory, and potential changes to the Web architecture that might make the implementation of annotation systems more complete.

There are a lot of tools currently available on the web which could be used as web annotation system or at least have some features that could be adopted in a web annotation system. For example, Web Marker (<https://addons.mozilla.org/en-US/firefox/addon/2679>) allows people mark web pages and share them with the world; Diigo (<http://www.diigo.com>) goes one step further which lets users add persistent highlights and sticky notes to anywhere on any webpage; and Amaya, a fully-featured web browser, includes a collaborative annotation application based on Resource Description Framework (RDF), XLink, and XPointer, and can be used to create and update documents directly on the Web.

This paper will provide a literature review of the history and evolution of web annotation systems, and its current status and future as a service in social annotations and collaborations. The literature will be consisted of several sections each representing an area of research and application in web annotation. In addition to the literature review, this paper will also introduce some of the currently available web annotation systems and provide a comparison of their supported features. A case study is also included, which summarizes the NeoNote project performed at the University of North Carolina's School of Information and Library Science. The purpose of incorporating such a case study is to provide a source for further study and research for people with comparable requirements and experiences.

Literature Review

The literature review consists of several sections each representing an area of research and application in web annotation. The main areas of discussion on web annotation include: interaction between users and the web, web annotation and annotation systems, system requirements, limits of existing tools, necessary related technologies, and areas where web annotations would help. Each section will be discussed below.

2.1 Two-way interaction between users

Traditionally, people were only able to receive information passively when they are surfing on the Internet. The current structure of the World Wide Web has limited the ability of users to interact with the web as well as other individuals on the web. In the essay “As we may think” which was published in *The Atlantic Monthly* in July 1945, Vannevar Bush suggested a design for an interactive information sharing device and predicted a machine called the Memex that would allow people to surf from one information page to another (Bush, 1945). Bush’s essay has predicted many kinds of technology, such as hypertext, personal computers, the Internet, the World Wide Web, which have been realized in the last couple of decades. It is argued that Hypertext and the World Wide Web are based on or at least inspired by the Memex.

Many attempts have been made in order to involve users into the exchanging of information on the web. However, at the present time information on the web is still mostly exchanged in a one-way mode. The research of Heck et al.’s paper concluded that

“a two-way system where the viewer could have the opportunity to take notes on a web page for later use or the use of fellow readers would allow for a greater exchange of ideas” (Heck et al., 1999). What’s more, with the current structure of the web, a two-way information exchange could also help solve the problem of “incomplete information and wasted time” that was mentioned in Yee, 1998.

In the past two decades, many technologies and systems have been made available to enhance the ability of the web of allowing interaction and collaboration between users. A discussion board, for instance, is an online communication tool that allows an individual to post comments or questions. Additionally, users can share and discuss information and opinions through discussion board. However, “when feedback or questions need to be pointed to a specific part of the documents (e.g. portion of the text or an image), the discussion board... may not be effective” (Jung et al, 2006).

Blog is a type of website that is usually maintained personally, frequently updated and arranged in chronological order. It has the ability for readers to leave comments in an interactive way, which is very useful for users exchanging ideas with other people. Correspondingly, the appearance of Wiki has greatly influenced the way of online collaboration and sharing. Searchsoa.com states that “a wiki is a server program that allows users to collaborate in forming the content of a Web site”. Using wiki, people could directly edit the content of a document (i.e., a web page), including other users’ contributions.

Jung et al. (2006) stated that “communication and collaboration through more precise annotations may serve as an efficient collaboration tool...” What’s more, “A hybrid mechanism that allows annotations to be made to pages, and responses to those

annotations...may provide the benefits of standard annotation whilst allowing greater collaboration by the increased interaction between students” (Palme, 1999).

2.2 Web Annotation and Annotation Systems

2.2.1 Annotations

Annotation is typically defined as “extra information asserted with a particular point in a document or other piece of information” [Wikipedia]. Prior research has demonstrated that making annotation is an important accompanying activity to reading, with annotations used for diverse purposes.

Patrick et al (2004) discussed three attributes that are used to describe annotations, which are content, form, and functionality. Annotation content could be either very understandable to an occasional reader, or very personal in meaning. Annotation forms (types) include styles such as underlining and coloring, and different positions such as within the document and stand alone. Annotation functionalities include reading, editing, linking, and sharing.

Various types of annotations could be made on hardcopy documents, such as highlighting, commentary, link making, reading records, etc (Marshall 1998). Marshall also noted that annotations are a primary vehicle for supporting collaboration around documents.

Many purposes of making annotations have been identified. Marshall (1998) found that annotations were used to bookmark important sections, to make interpretive remarks, and to fine-grain highlight to aid memory. O’Hara and Sellen (1997) discovered that people use annotations to help them understand a text and to make the text more

useful for future tasks. Annotations are often helpful for other readers as well, even when they are not made with others in mind.

Glover, Xu, and Hardaker (2007) point out the two key advantages of inserting annotations into the web page, which are being able to share those notes with others and the ability to access the annotations from any web enabled computer. What's more, "annotations also provide third party, subjective metadata about the content of a web page that can be analyzed to provide additional information for use in web searching and dynamic link generation".

2.2.2 Web Annotation

Web annotation has been a popular research topic since the invention of hypertext technology and accompanying web technologies as well as the steady increase in web-based materials. As stated earlier, this paper will use the following definition of web annotations:

"A Web annotation is an online annotation associated with a web resource, typically a web page. With a Web annotation system, a user can add, modify or remove information from a Web resource without modifying the resource itself." (Wikipedia)

Heck et al. (1999) concluded that the solution to the "incomplete information and wasted time" problem would be the "instantiation of an annotation tool that can be used to make private, public, or shared annotations, or notes, on already existing web pages." Fu et al (2005) identified four types of annotation systems, which are annotation functionality built into web browsers, personalized web information organization systems, interactive web publication forums, as well as annotation engines. They concluded that

“the four types of annotation tools provide almost all the functions that can be accomplished on paper”, furthermore, these tools “also provide some functions which are difficult to realize in the paper environment”. However, they also mentioned that no single approach is available to support all of the features.

According to Denoue & Vignollet (2000), an annotation system usually consists of three modules: the first is used to view existing annotations, the second to create new annotations, and the third to store the annotations. Vasudevan and Palmer (1999) reviewed web annotation system’s architecture, and suggested that new technologies such as the Document Object Model (DOM) level 2 will be desirable to design high-quality annotation systems. Patrick et al. (2004) described a “Conceptual architecture of the individual mode of WATs (Web Annotation Tools)”, which is demonstrated in Figure 1 below.

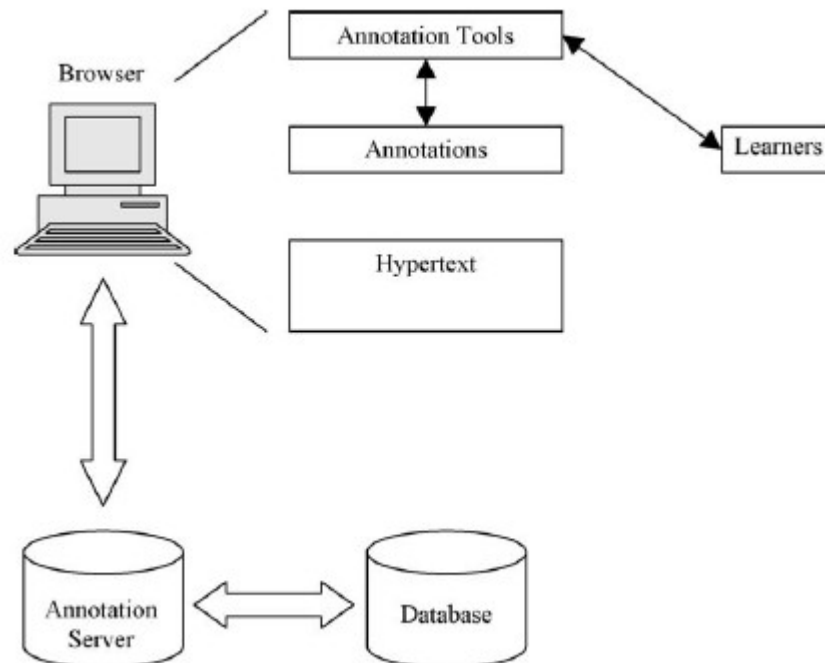


Figure 1. Conceptual architecture of the individual mode of WATs (Patrick et al, 2004)

2.3 Requirements

Patrick et al. (2004) stated basic requirements for the Web Annotation Tool that they suggested, for example, provide a usable interface that allows people to make and edit annotations, to insert annotations on hypertexts, to browse all the annotations, and to determine the logical relationships between annotations.

Fu et al. (2005) investigated the needs web users have to make annotations for their personal use when they view web pages, and examined three forms of annotations which are text selection and emphasis, association building, and document re-segmentation. They summarized that “content annotations, such as text selection, emphasis, note taking, link building, and page editing are still the most commonly used types of annotation”.

Glover, Xu, and Hardaker (2007) drew a number of requirements and desirable factors for web annotation systems, and broke them down into conceptual and technical, with conceptual factors being those that are intrinsic to annotation and the technical factors being those related to the implementation of the conceptual factors. Conceptual factors consist of basic text annotations, graphical annotations, shared annotations, etc.; while technical factors include “no additional software necessary”, accessible, open architecture, and so forth.

Ovsiannikov et al. (1999) conducted a survey of graduate and undergraduate students, professors, and professionals to obtain respondents’ expectations for features of good annotation software. Some essential features that were found are annotation of pictures, insensitivity to document format, keyword search, writing on margins, writing between lines, writing at the top, export of annotations, and multimedia comments. As

O'Hara and Sellen (1997) indicated, the smooth integration of annotating with reading must be one of the foremost priorities for an annotation system.

2.4 Limits of existing tools

Although a lot of attempts have been made in the research of web annotation systems, many problems, both technical and non-technical, still exist to keep these systems from successful and widely adopted.

Fu et al. (2005) argued that “relatively few researchers are dedicated to understanding what the users actually need”. Zohar (1999) argued that “there is nowadays no widespread annotation service”. Denoue (1999) claimed that “there is no wide spread use of” web annotation systems. No single approach is available now 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.

Denoue & Vignollet (2000) identified several limitations in the implementation of current web annotation systems. The most important of these is the lack of a standardized representation for web annotations, which means that current systems are divergent and proprietary, therefore limiting the possibility for third parties to create clients. Furthermore, they discussed some remaining challenges for annotation systems. The first one is users' privacy. With the current annotation systems' architecture and implementation, users' navigational habits as well as the words they like and dislike can be tracked. Another issue is interoperability. Current existing annotation systems adopt

different strategies to represent annotations, and use different ways to save these annotations. For example, XPointer have been proposed for XML documents and been adopted by Annotea, and Yawas (Denoue & Vignollet, 2000) adds the occurrence of the selected text. A detailed discussion appears in (Phelps, 2000) where the authors propose new ways to represent the annotation anchor. Their proposition not only applies to XML documents, but also to HTML, PostScript and PDF documents. Glover et al. stated that a major limitation of current solutions is encountered when attempting to annotate dynamic web pages.

2.5 Necessary Related technologies

2.5.1 Standardized Metadata

Because of the need for interoperability, identification and access rights, the Dublin Core plays a significant role in the annotation process (Campbell, 2002). Elements such as the title of the annotation, the name of the annotator, and the date created could be specified as Dublin Core elements, while other elements more specific to the annotation process, such as type, context and content, could either create or use another scheme. Annotea extended the using of Dublin Core with the adoption of more elements, which include language, format, publisher, identifier, etc.

However, some other elements could also be used for describing content of annotations as well as some of its important related information. For example, in the annotation process, “coverage” could be used to indicate the range of annotation and the related text that has been highlighted.

2.5.2 Basic ontological framework to facilitate inference (RDF)

In Annotea, annotations are described using Resource Description Framework (RDF). An RDF schema is used for describing properties of annotations (Kahan et al, 2002). When users post annotations in browsers like Internet Explorer or Firefox, the browser will generate an RDF description of the annotation which includes the metadata and the body, and send it to the annotation server using the HTTP POST method. Then the server assigns an URI to the annotation and the body, and replies with an RDF statement that includes these URIs.

RDF databases are implemented to store annotations in the server, which makes it possible to provide customized queries and limit the amount of data returned by the servers.

2.5.3 Locating annotations within content

The XML Pointer Language (XPointer) is defined as an address scheme for individual parts of an XML document. The XPointer addresses can be used by any application that needs to identify parts of or locations in an XML document. The reason for using XPointer is that a URL only points at a single, complete document, and it would be useful to be able to link to a particular element or group of elements on a page without having to change the document itself.

In an annotation process, a user may want to select a range of text and make annotation on that part. Here the range begins at one point and ends at another point. The start and end points are each identified by a location path. A sample XPointer address that specifies a range may look like this:

xpointer(start-point(string-range(/html[1]/body[1]/p[2], "", 38, 1))/range-to(end-point(string-range(/html[1]/body[1]/p[2], "", 51, 1))))

However, XPointer has its own limitations (Denoue and Vignollet, 2002). For example, it is hard to attach the annotation when the structure of the document changes, and the content pointed to by the XPointer is not human-readable.

2.6 Areas where web annotations would help

2.6.1 Collaboration

Previous research has shown the effectiveness of annotation systems as well as the potential of such systems to facilitate collaboration. Davis and Huttenlocher proposed a system called CoNote that enables a group of people to communicate via shared annotations on a set of electronic documents. They found that shared annotations of documents provide a richer communications forum. Marshall studied the effectiveness of annotations (Marshall, 1998), and noted that annotations are a primary vehicle for supporting collaborations around documents (Marshall, 1999). Jung et al. claimed that “when such annotations are implemented on the WWW...they can significantly contribute to enhancing both web-navigating experience and collaborative activities through the WWW” (Jung et al. 2006).

2.6.2 Information Retrieval

Denoue & Vignollet (2000) argued that highlighted texts can be used to augment the document representation. They conducted several experiments which tested how

annotations can be used to improve document access and document clustering. The results suggested three ways of using annotations in information retrieval:

- The highlighted texts can be used to build personalized document summaries, thus improving document access and retrieval.
- Automatic document clustering can also use them to generate user-directed document clusters.
- Automatic document classifiers can take advantage of the highlighted text to extract significant words from the documents without using the usual word frequency and inverse document frequency measures.

In the research of Golovchinsky, he showed how annotations improve information retrieval through automated relevance feedback. In his proposed approach, annotations attached to previously read documents can be used to automatically expand the query. (Golovchinsky, 1998)

Current Systems

Many attempts have been made to build web annotation systems. These systems take several different approaches to provide users with the option to annotate web pages, while many of them share some extent of similarity. This section will outline some existing tools (systems), including Annotea, Diigo, Fleck, and so on. Table 1 lists these systems with comparison of supported functions/features.

Software	Highlight	Annotate	Tag	Search	Bookmark	Share/ Collaborate	Page capture
Annozilla	X	X	-	-	-	-	-
Amaya	-	X	-	-	X	-	-
Co-ment	X	X	-	X	-	X	-
Diigo	X	X	X	X	X	X	X
Fleck	-	X	-	-	-	X	X
HyLighter	X	X	-	-	-	X	-
JumpKnowledge	-	X	-	-	X	X	X
SharedCopy	X	X	X	-	X	X	X
Shiftspace	X	X	-	-	-	-	-
Stickis	-	X	X	-	-	X	-
Trailfire	-	X	X	-	-	X	-

Table 1: Supported functions/features

3.1 Framework for comparing annotation systems

Seven typical functionalities/features that are supported by various annotation systems will be discussed below:

3.1.1 Highlighting

Highlighting is typical when users make annotations on a document, either traditional paper or web pages. With the support of highlighting, users can select a portion of web pages, for example, a range of texts, part of a paragraph, etc. The current version of Annozilla, Diigo, HyLighter, SharedCopy, and Shiftspace has the highlight feature.

3.1.2 Annotating

Annotating is the most basic function of a web annotation system, which allows users to post textual annotations to web pages. All of the web annotation systems (tools) that were talked in this paper support annotating.

3.1.3 Tagging

Some web annotation systems provide the functionality of tagging. Users can associate some keywords or terms with the annotations they made, or the pages they bookmarked. What's more, people can flexibly organize and share their own libraries of annotated web pages. Diigo has created a repository of quality content on almost every subject that people are interested in using the tags and annotations that were made by users.

3.1.4 Searching

The ability to search within the annotation repository is useful and may offer higher precision and faster response times than search using search engines.

3.1.5 Bookmarking

Bookmarking is an essential part of some web annotation systems, such as Diigo, JumpKnowledge, and SharedCopy. Users can save URL of web pages which they have highlighted or annotated for future reference, or for sharing with friends and colleagues.

3.1.6 Sharing/Collaborating

Sharing/collaborating in web annotation systems is the ability to let users share links, comments, annotations or the annotated web page with their friends or colleagues for collaborative research or study.

3.1.7 Page capturing

Different from bookmarking, page capturing allows users to save a copy of the web page. Some systems use page capturing to provide the functionalities of annotating and bookmarking, such as Fleck and SharedCopy.

3.2 Current Systems

3.2.1 Annotea

Annotea is a W3C project which was conducted to “enhance collaboration via shared metadata based Web annotations, bookmarks, and their combinations”

[<http://www.w3.org/2001/Annotea/>]. Kahan et al. describes Annotea as “a web-based shared annotation system based on a general-purpose open RDF infrastructure, where annotations are modeled as a class of metadata” (Kahan et al, 2002). Annotea employs a Resource Description Framework (RDF) based annotation schema for representing annotations and uses XPointer for locating texts that have been annotated in the document. Annotea currently requires the installation of either a new browser with integrated support (i.e. Amaya), or the installation of an extension for existing browsers (i.e. Annozilla for Firefox).

Some important requirements for designing Annotea are:

- Documents to be annotated should be well-formed, structured documents.
- Each annotation should be associated with a URI.
- Users can assign existing types to an annotation, create new types.
- Annotations should be represented with a RDF schema.
- Annotations are stored in RDF databases.

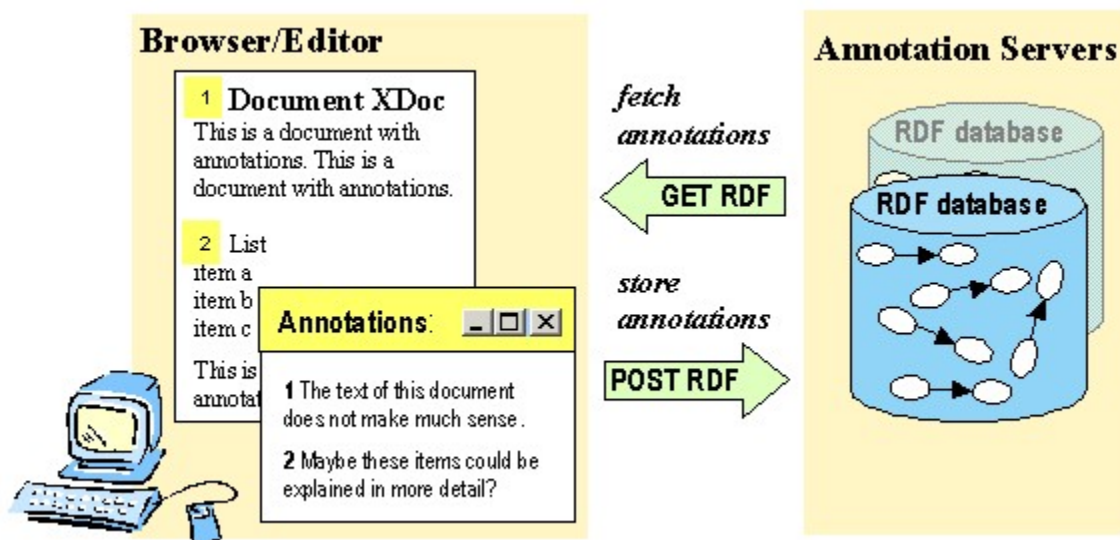


Figure 2: The basic architecture of Annotea (Kahan et al, 2002).

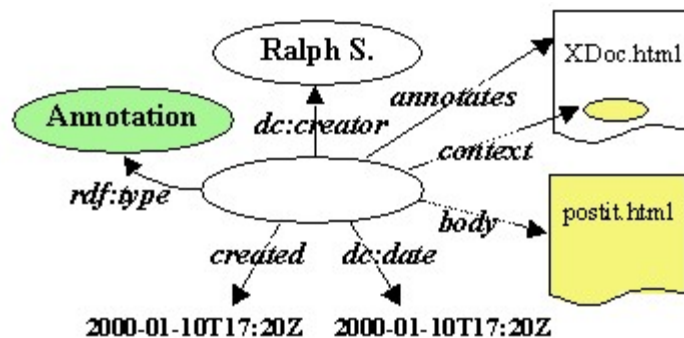


Figure 3: The RDF model of an annotation (Kahan et al, 2002).

3.2.2 Amaya

Amaya is a full-featured web browser and editor, which includes a collaborative annotation application. The current Amaya user interface for annotations is presented in the Amaya documentation [<http://www.w3.org/Amaya/Overview.html>].

Amaya adopts the same RDF schema as Annotea for describing annotations. Based on the documentation, an annotation has many properties including:

- Physical location: is the annotation stored in the local file system or in an annotation server
- Scope: is the annotation associated to a whole document or just to a fragment.
- Annotation type: 'Annotation', 'Comment', 'Query', etc.

Using Amaya, annotations can be saved in a local file system (local annotations), or be saved remotely on annotations servers (remote annotations).

3.2.3 Annozilla

The Annozilla project is designed to view and create annotations associated with a web page, as defined by the W3C Annotea project. Annozilla works as a Firefox extension, and further information available from [<http://annozilla.mozdev.org/>].

Like Amaya, Annozilla also stores annotations as RDF on a server, using XPointer (or at least XPointer-like constructs) to identify the position and range of the document being annotated. If configured, annotation icons will be inserted into the page at or near the annotated regions, which are presented as pencil marks at the start of selected texts. . The icons can be selected in order to load the annotation body. Annozilla can also be configured to automatically load annotations when a new page is loaded.

3.2.4 Co-ment

Co-ment [<http://www.co-ment.net/>] is an online collaborative text commenting system, which enables collaborative processes around texts that are more structured than on wikis. Users can upload documents and texts to be read and commented on by their friends. The user interface of co-ment is consisted of two side by side panel views, one containing the document, the other containing the comments. Phrases and words are highlighted within the document itself, indicating information related to various comments. Moreover, co-ment provides the search option that allows for filtering by date and keywords within the comments and within the text.

3.2.5 Diigo

Diigo [<http://www.diigo.com/>] is a Social bookmarking website which allows signed users to bookmark and tag web pages. The name "Diigo" is an abbreviation for

"Digest of Internet Information, Groups and Other stuff". Diigo allows users to highlight any part of a webpage and attach sticky notes to specific highlights or to a whole page. These annotations can be kept private, shared with a group within Diigo or a special link forwarded to someone else. Users can write on the transparency as they wish, as private notes or public comments. And they can read public comments on the transparency left by other readers of the same page, and hear their "two cents" and interact with them.

3.2.6 Fleck

Fleck is inspired on a story written in 1945 by Vannevar Bush and an article titled 'We Are the Web' by Kevin Kelly. Vannevar Bush predicted a machine called the Memex that would allow people to surf from one information page to another. One thing that the Memex had and the web doesn't is the ability to add new content to every page it contained. After reading the wired article by Kevin Kelly the authors decided to try to add a new level to the web by adding new tools that would allow its users to add information rather than just consuming it.

Fleck allows people to interact with pages on the web by adding notes and bullets to websites. Users can save the annotated page for themselves or send it to friends or colleagues.

3.2.7 HyLighter

HyLighter [<https://www.hylighter.com>] is a web highlighting and annotation research prototype which allows users of IE and Firefox to highlight passages of text and associate text annotations. What's more, users can compare what they have highlighted

with what others have highlighted, using colors to indicate who has highlighted what and shades to show how many users have highlighted a passage. According to the documentation, “HyLighter fills a gap in collaboration/social software by allowing you to tie conversations of any size group to specific sections of a text or image and rapidly distil the most relevant thoughts for a wide range of purposes”.

3.2.8 JumpKnowledge

JumpKnowledge is a new concept of annotate web pages that allows users to comment and annotate any part of a website. Annotated pages can be saved, linked and sent to friends. JKN works with https, framesets, supports multiple pages in a single annotation, and has an optional Firefox button that can annotate password-protected web pages and JavaScript-rendered pages. Some of the key features of JumpKnowledge that are different from other annotation systems include (adapted from <http://info.jkn.com/index.htm>):

- Multi-page: users can add comments to multiple web pages
- Auto displacement: users’ comments automatically displace the underlying text, so nothing of the original page is obscured and comments can be read in context.
- Permanent: even if the web page changes, added comments will be shown with the web page as it looked when you created the Annotation.
- Share: email the web page with the comments to friends and colleagues.

3.2.9 SharedCopy

SharedCopy is an AJAX based web annotation tool that allows users to mark-up, highlight, draw, annotate, cache, sticky-note and finally share any website. Annotation is delivered within the browser and without the use of JavaScript. Unlike Fleck, which loads a fresh copy of a marked web page, SharedCopy takes a snapshot of the page being noted and uses that page as the basis for annotations. Annotated pages can be saved for public viewing or for private reference. Moreover, SharedCopy supports sending annotated pages to friends, and bookmarking is built in and users are provided with profile/ bookmark list pages and user feeds for sharing.

3.2.10 Shiftspace

ShiftSpace is a web annotation application that allows the creation of sticky-notes, text highlighting or the alteration of the web page's source code. ShiftSpace works as an open source browser plugin (Firefox only). The plugin allows their users to annotate and remix a website saving it as a communally editable alternate version revealed in the browser by pressing Shift + Space. Modified pages are called “shifts”, and if made public, are shared on the ShiftSpace website. Users can subscribe to the shifts of users they like via RSS.

3.2.11 Stickis

Stickis is a web page annotation service that lets users subscribe to content “channels” from their friends and the community via a browser plugin. The service allows members to annotate web pages as they browse, and share conversation and commentary about their discoveries, both in the form of sticky notes left on top of the

very pages being annotated, and in the form of blog-like entries at a centralized, access-controlled repository.

3.2.12 Trailfire

Trailfire is a Firefox and Internet Explorer extension with a web-based frontend which combines social bookmarking, annotation, and trails (or a sequence of web pages). Users can post notes on top of a webpage and string them together with hyperlinks. The new version of the service will include the ability to make friends and share with them, follow all the trails made by a user, gather your friends into groups, and allow trails to be edited together by multiple users.

Case Study

NeoNote is an ongoing project that is performed at the School of Information and Library Science in the University of North Carolina at Chapel Hill, which was intended to address the problem that there is no comprehensive annotation system which can meet all users' needs. The basic idea of NeoNote is to develop a global shared annotation system as well as an easy-to-use user interface.

Some of the key requirements of the NeoNote project:

1. Be able to save annotations and grabbed web pages in a global, universally accessible server.
2. Be able to select portion of a web page, highlight the selected text, and make annotations on the selections.
3. Be able to grab the URL and content of web pages, and save out to the database.

Implementation:

Based on the system's requirements, we have built the NeoNote annotation systems based on W3C's Annotea infrastructure, which is built on top of open source technologies. The NeoNote annotation system consists of a user interface that is used to make annotations on web pages, a MySQL database that saves metadata information of annotations, and a Tomcat application server runs on Red Hat Linux. Information that is stored in the database about an annotation includes:

1) Range of the highlighted scope of text (the selection), represented using XPointer.

2) Content of the annotation that is represented in HTML

The basic structure of the NeoNote system is illustrated in Figure 4:

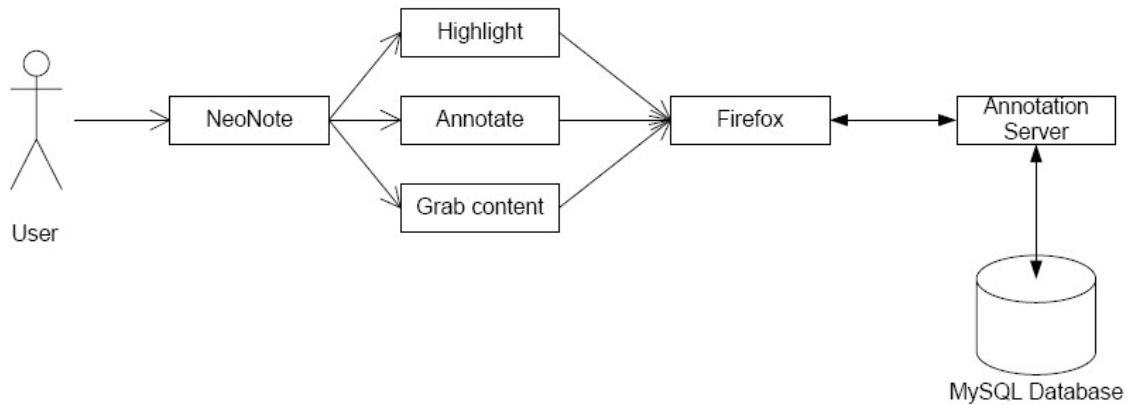


Figure 4: NeoNote architecture

The user interface of NeoNote was designed as a Firefox extension that allows users posting annotations to the preset remote annotation server. We have selected Mozilla Firefox as the standard web browser because it is widely used and easy to customize with extensions and plug-ins. Using extensions, we can add new functionality and additional features to Firefox.

Annozilla, as was described earlier, is an existing client side tool for posting annotations to Annotea-based annotation systems, which works as a Firefox extension. In version 0.6.8, Annozilla supports basic annotating functions, which can be used to annotate a whole page, or a portion of a page. However, it doesn't have the highlighting feature. Another Firefox extension, Web Marker, allows users to mark portions of web pages, and copy the marked-text to the clipboard. Our client tool, the NeoNote

highlighter and NeoNote grabber, is derived from Annozilla and Web Marker, and has combined the highlighting and annotating functions.

This NeoNote extension adds two buttons to the top menu bar of Firefox: a highlighter and a grabber. The highlighter button is modal, and when it was turned on, the selected text will be highlighted automatically. The grabber button is modeless and users can grab the URL and content of the current page by clicking this button.

Some features of this extension include:

- Selection of text means it is highlighted.
- Double clicking highlighted text means annotation note comes up, where you can add/edit text of annotation.
- Annotations could be rendered back when reloading an annotated page, highlighted areas will be re-highlighted.

Figure 5 shows NeoNote is being used to annotate the “web annotation” page of Wikipedia.

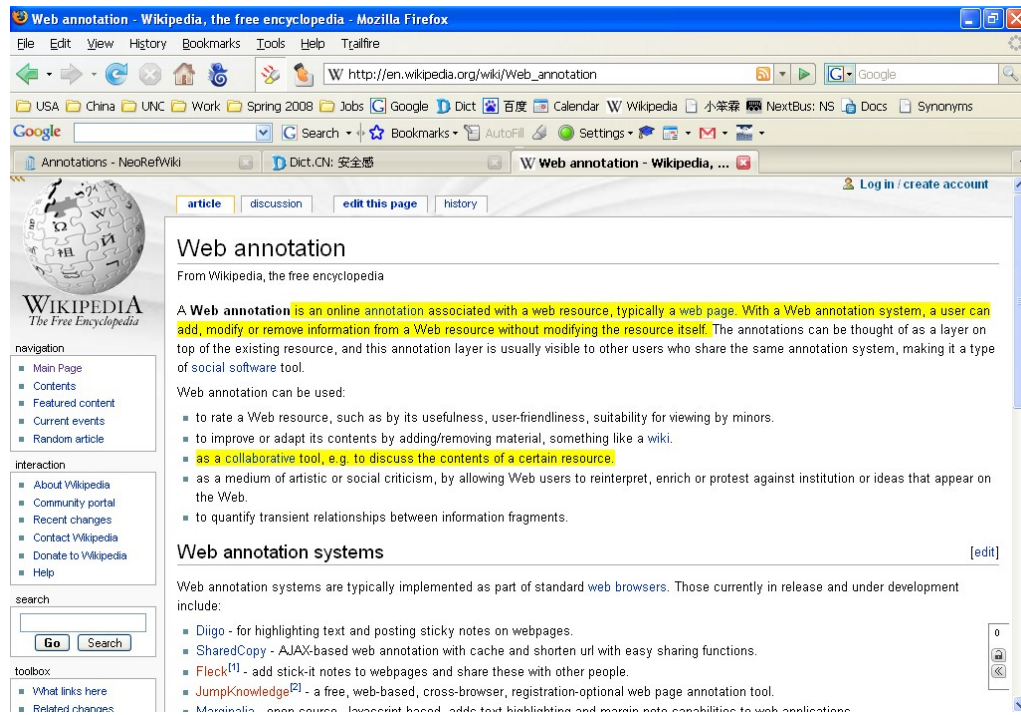


Figure 5 NeoNote in use

Conclusion

Web Annotation is emerging as an important field of current efforts to facilitate the collaboration and exchange of information on the World Wide Web. The main achievements of this paper have been the exploration of literature, which has covered various areas related to the research and development of web annotation systems, including users' interaction with other people on the web, annotation and web annotations, requirements of web annotation systems, limits of existing systems, some related technologies, and areas that web annotations could help the improvement.

In addition to the review of literature, this paper also presents some of the annotation systems that are currently available on the web. Seven typical functionalities of web annotation systems have been identified, such as highlighting, annotating, tagging, etc. After that, a description of the ongoing NeoNote project is followed, which has briefly introduced how we tried to address the limits of existing tools.

Nevertheless, the outcome of this paper has been limited by the inadequate information about the annotation systems that were discussed. Some of the systems such as Annozilla are open source; therefore it is possible to study its documentation and code to explore the structure. However, for many of the other systems, it is very difficult, if not impossible, to get to know their strategies of implementation. Therefore, in the next phase, I will aim to reach a thorough understanding of the implementation and structure of the

annotation systems. I am also planning to propose an improved database schema for describing and storing annotations.

References

- Brush, A., & David, B. (2001). Robust annotation positioning in digital documents. *Proceedings of the SIGCHI conference on Human factors in computing systems. Seattle, Washington, United States, ACM.*
- Bush, V. (1996). As we may think. *interactions*, 3(2), 35-46.
- Cadiz, J., & Anop, G. (2000). Using Web annotations for asynchronous collaboration around documents. *Proceedings of the 2000 ACM conference on Computer supported cooperative work. Philadelphia, Pennsylvania, United States, ACM1.*
- Campbell, D. (2002). The use of the Dublin Core in web annotation programs. *Proceedings of the 2002 international conference on Dublin core and metadata applications: Metadata for e-communities: supporting diversity and convergence*, 105-110.
- Chong, N., & Sakauchi, M. (2001). Creating and sharing web notes via a standard browser. *Proceedings of 2001 ACM symposium on Applied computing*, 99-104.
- Denoue, L., & Vignollet, L. (2002). Annotations in the wild. *Proceedings of the ECAI 2002 workshop on semantic authoring, annotation and knowledge markup.*
- Denoue, L., & Vignollet, L. (2000). An annotation tool for web browsers and its applications to information retrieval. *Proceedings of RIAO200, Apr. 2000.*
- Fu, X., Ciszek, T., Marchionini, G., & Solomon, P. (2005). Annotating the Web: An exploratory study of Web users' needs for personal annotation tools. *Proceedings of the American Society for Information Science and Technology*, 42(1).
- Glover, I., Xu, Z., & Hardaker, G. (2007). Online annotation - Research and practices. *Comput. Educ.*, 49(4), 1308-1320.
- Heck, R., Luebke, S., & Obermark, C. (n.d.). *A Survey of Web Annotation Systems*. Retrieved March 30, 2008, from http://www.math.grin.edu/~rebelsky/Blazers/Annotations/Summer1999/Papers/survey_paper.html.
- Hwang, W., Wang, C., & Sharples, M. (2007). A study of multimedia annotation of Web-based materials. *Computers & Education*, 48(4), 680-699.

- Jung, B., Yoon, I., Lim, H., Ramirez-Weber, F., & Petkovic, D. (2006). Annotizer: User-friendly WWW Annotation System for Collaboration in Research and Education Environments. *Web Technologies, Applications, and Services*, 524.
- Kahan, J., Koivunen., Prud'Hommeaux, E., & Swick, R. (2002). Annotea: an open RDF infrastructure for shared Web annotations. *Computer Networks*, 39(5), 589-608.
- Kim, S., & Slater, M. (2004). WebDAV-based hypertext annotation and trail system. *HYPERTEXT '04: Proceedings of the fifteenth ACM conference on Hypertext and hypermedia*, ACM.
- Marshall, C. (1998). Toward an ecology of hypertext annotation. *Proceedings of ACM Hypertext '98, PA*, 40-49.
- O'Hara, K., & Sellen, A. (1997). A comparison of reading paper and online documents. *CHI97 Conference Proceedings*, ACM Press, Atlanta, GA.
- Ovsiannikov, I., Arbib, M., & Mcneill, T. (1999). Annotation Technology. *International journal of human computer studies*, 50(4), 329-362.
- Palme, J. (1999). Talking back to the WWW. *Comput. Networks*, 31(21), 2281-2286.
- Patrick., Chen, S., & Chin, Y. (2004). Developing web annotation tools for learners and instructors. *Interacting with Computers*, 16(2), 163-181.
- Phelps, T., & Wilensky, R. (1997). Multivalent Annotations. *Proceedings of the First European Conference on Research and Advanced Technology for Digital Libraries*, Springer-Verlag.
- Takeshi, S., & Toshiyuki, A. (2001). A framework for sharing personal annotations on web resources using XML. *Proceedings of the workshop on Information technology for virtual enterprises*.
- Vasudevan, V., & Palmer., M. (1999). On Web Annotations: Promises and Pitfalls of Current Web Infrastructure. *Proceedings of the 32nd Hawaii International Conference on Systems Sciences*, maui, Hawaii.
- Web annotation - Wikipedia, the free encyclopedia. (n.d.). Retrieved March 30, 2008, from http://en.wikipedia.org/wiki/Web_annotation.
- Yee, K. (n.d.). CritLink: Better Hyperlinks for the WWW. Retrieved March 30, 2008, from <http://zesty.ca/crit/ht98.html>.
- Zohar, R. (1999). Web Annotation - an Overview. *Dept. of Electrical Engineering, Israel Institute of Technology*.

Web Resources

Amaya Home Page. Retrieved March 30, 2008, from <http://www.w3.org/Amaya/>.

Annotea project. Retrieved March 30, 2008, from <http://www.w3.org/2001/Annotea/>.

Annozilla. Retrieved March 30, 2008, from <http://annozilla.mozdev.org/>.

co-ment - Web-based text annotation. Retrieved March 30, 2008, from <http://www.co-ment.net/>

Diigo - Web Highlighter and Sticky Notes, Social Bookmarking and Annotation, Social Information Network!. Retrieved March 30, 2008, from <http://www.diigo.com>.

Fleck.com - Fleck the Web!. Retrieved March 30, 2008, from <http://www.fleck.com>.

HyLighter. Retrieved March 30, 2008, from <http://www.hylighter.com>.

JumpKnowledge. Retrieved March 30, 2008, from <http://info.jkn.com/>.

SharedCopy. Retrieved March 30, 2008, from <http://sharedcopy.com/>.

ShiftSpace | An Open Source layer above any webpage. Retrieved March 30, 2008, from <http://www.shiftspace.org/>.

Stickis. Retrieved March 30, 2008, from <http://stickis.com/>.

Trailfire. Retrieved March 30, 2008, from <http://trailfire.com/>.