CritLink: Advanced Hyperlinks Enable Public Annotation on the Web

Ka-Ping Yee
University of California, Berkeley
ping@zesty.ca

ABSTRACT
This paper describes a set of hypertext linking features that are essential for supporting critical discussion and document annotation, but are missing from the Web. An implementation of these features called "CritLink" enables users to attach annotations to any location on any public page, and to view the annotations on any page, without having to install any special client or server software. All browsers (including text-only browsers) and all operating systems are supported. Annotations can have types, can be public or private, and can annotate other annotations.

Keywords
Public annotation, critical discussion, hypertext, bi-directional linking, fine-grained links, extrinsic links, typed links, World-Wide Web.

INTRODUCTION
The World-Wide Web implements some of the ideas of hypertext [11], but the original hypertext vision has yet to be completely realized. Among the features missing are abilities that would enable critical discussion to take place on Web documents. These key abilities are *bi-directional linking*, *extrinsic linking*, *fine-grained linking*, *and link typing*. Together, they permit any reader to attach a comment to any existing document, where the comment can express a disposition (such as agreement or disagreement) with respect to a specific word or phrase in the document.

RELATED WORK
Many different Web annotation systems have been developed, including ComMentor [14], CoNote [4], GrAnT [15], HyperNews [12], HyperWave [8], Multivalent Documents (MVD) [13], Third Voice [16], WebVise [6], and YAWAS [5]. Software as early as NCSA Mosaic [10] included a Group Annotation feature. Most recently, the W3C has been developing annotation support for Amaya as part of their Annotea project [7].

CritLink, however, has some important properties that none of these other systems have. CritLink does not require the installation of a special browser or any add-on software; it works with any browser, even a text-only browser. Consequently, it is platform-independent; it is also accessible to the blind, unlike these other systems. CritLink can annotate any existing public Web document, not only a document within a closed authoring system. CritLink also supports annotations as first-class documents that can themselves be annotated. Unlike all of the above systems except HyperNews (which can only annotate specially instrumented Web pages), CritLink is based almost entirely on established Web standards, defined since version 2.0 of the HTML specification [1]. There are no new tags or file formats; there is only one simple extension to allow the specification of locations within the document at the phrase level. These properties are summarized in Table 1.

System	(a)	(b)	(c)	(d)	(e)	(f)
Annotea	no	no	yes	no	yes	no
ComMentor	no	no	yes	no	no	no
CoNote	yes	yes	yes	no	no	no
GrAnT	yes	yes	yes	no	no	no
HyperNews	yes	yes	no	no	no	yes
HyperWave	yes	yes	no	no	yes	no
MVD	no	no	no	no	yes	no
NCSA Mosaic	no	no	yes	no	yes	no
ThirdVoice	no	no	yes	no	no	no
WebVise	yes	no	yes	yes	yes	no
YAWAS	no	no	yes	no	no	no
CritLink	yes	yes	yes	yes	yes	yes

(a) view annotations in any Web browser
(b) create annotations in any Web browser
(c) annotate any public Web page
(d) serve annotations from any Web server
(e) create annotations on annotations
(f) all data stored in standard HTML

Table 1. Properties of various Web annotation systems.

CritLink is the only Web annotation system where all of the information is stored in standard HTML documents that can be created and accessed in all of the normal ways.

Also, CritLink's approach to the annotation problem is interesting and uncommon: rather than just providing annotations, it augments the hyperlink primitive to yield many new possibilities within the Web's current hypertext model, of which annotations are just one application.

ADVANCED HYPERLINKS

This section describes in more detail the four hyperlinking features, first highlighted in [9], that are enabled by CritLink and their significance to online collaboration.

Bi-directional Links

The Web currently supports links joining two anchors: an origin (the location containing the `<A>` tag) and a target (the location identified in the `HREF` attribute). Hyperlinks on the Web can usually be followed only from the origin to the target; in general, only links with their origin in the current document are visible, and links targeting the current document are not shown. Bi-directional links are hyperlinks that can be followed in either direction. The ability to find other works referring to a given document is important for synthesizing information from various documents.

Extrinsic Links

HTML links are *intrinsic*, meaning that they must be embedded in the linked document itself. Consequently, only the author of a document can link the document to anything else. To support better collaboration, it should be possible for other parties to contribute links relating a document to other documents. Such links must be stored outside the document, so they are called *extrinsic*.

Typed Links

When a link joins two anchors, it can be useful to know what kind of relationship it represents between the two anchors. This information can help user agents to provide appropriate means for displaying and interacting with the link. HTML [1] already provides the `REL` and `REV` attributes on hyperlinks for this purpose, though unfortunately they are ignored by all common browsers.

Fine-Grained Links

A coarse-grained anchor addresses an entire document, whereas a fine-grained anchor addresses a specific part of the document. The specification for HTML allows for both kinds of origin anchors (using the `<LINK>` and `<A>` tags respectively); however, most popular Web browsers ignore `<LINK>` tags. URLs [2] support both kinds of anchors; however, a fine-grained target anchor can only identify a fragment that has already been marked up in the target document. Even when the author of the target document has co-operated by marking a fragment, most Web browsers neglect to indicate the position of the fragment. They only scroll the user's view to an approximate location nearby, leaving the user to guess where the fragment actually lies. The requirement for co-ordination between the link creator and the target document author, together with the failure to display fragments, renders fine-grained targets largely useless for collaboration purposes.

In summary, current Web software properly supports only untyped, intrinsic, one-way links with fine-grained origin anchors and coarse-grained target anchors. As a result, readers cannot interact with or contribute to most documents. 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.

APPLICATIONS

The ability to follow links backward is useful in itself when applied to the vast collection of existing links on the Web. This can make it easier to find related work, which is an essential part of collaborating with others.

Typed, fine-grained links also permit the use of a Web document as a variation on an existing document. Each link could serve as a change marker to indicate the delta from an existing version of the document to a suggested revision.

All four linking features, taken together, can enable public annotation and critical discussion. If links can be followed backwards, then an annotation can simply be an ordinary Web document with a link to the annotated document. Extrinsic links permit others to annotate a document without requiring co-operation from the author. A fine-grained target anchor lets the annotation refer to a particular phrase in the document. Typed links let the annotation specify a relationship or disposition with respect to the target document. Because annotations are just ordinary documents, they can themselves be annotated, and they can be protected using all the same security mechanisms as ordinary Web documents.

USER EXPERIENCE

CritLink is a tool for both viewing and authoring annotations. It augments the browsing experience by *mediating* [18] all transactions between the browser and the server, thus requiring no extra software to be installed on either end. As pages are downloaded from a server, CritLink adjusts the hyperlinks from the page to request further pages via the CritLink mediator.

To begin a session, a user first directs any Web browser to http://crit.org/. From there, the user can browse the Web by following links normally and entering new URLs into the secondary location bar provided by CritLink. All pages are augmented with this extra toolbar at the top, annotation markers inserted into the content, and metadata and backward links at the end of the document.

EXAMPLE

Figure 1 shows how the ACM home page appears when viewed using CritLink. Selecting some target text (in this case "pioneering conferences") and pressing the "comment" button brings up the composition window in Figure 2. After a comment has been entered, indicator icons appear on the original page bracketing the target text; hovering over an icon pops up the title of the annotation, as shown in Figure 3. Selecting the icon displays the annotation, shown in Figure 4. Document metadata and lists of fine-grained and coarse-grained backlinks appear at the bottom of the page, shown in Figure 5.

In particular, this example demonstrates how a reader can construct a *trail* through the Web [3] by adding public links between documents that the reader does not own.

Figure 1 (above). ACM home page, about to be annotated.

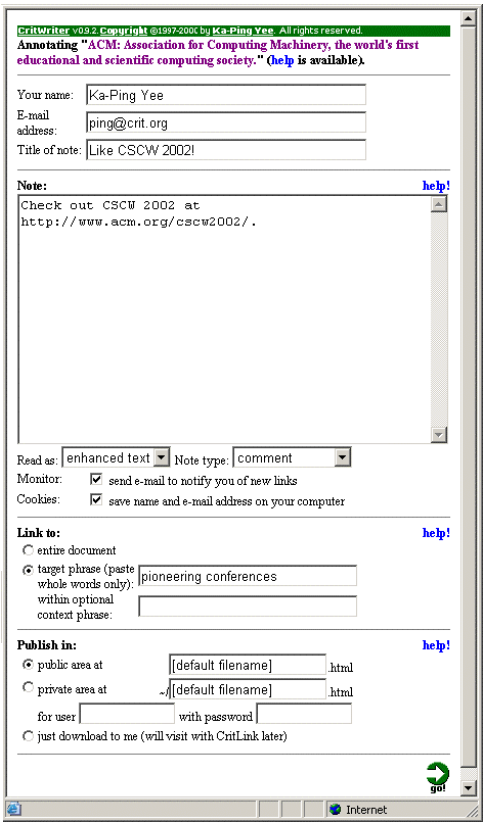

Figure 2 (above). Annotation entry form.

Figure 3 (above). ACM home page, with new annotation.

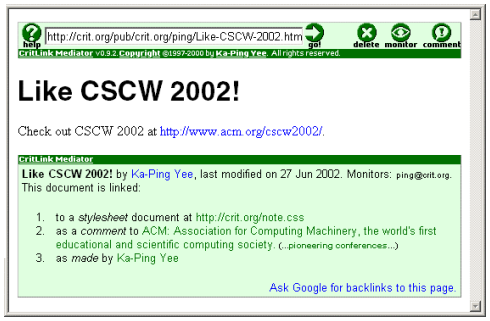

Figure 4 (above). New annotation document created by the form in Figure 2, reachable by clicking on the annotation marker in Figure 3.

Figure 5 (left). Metadata and backlink information appended to ACM home page document.

The CritLink toolbar also provides a "monitor" button, visible in Figures 1, 3, and 4. This button allows the user to register for notifications when a new link to a document is discovered. This would let the author of a document know when the document is being annotated. When a user creates a new annotation, CritLink automatically registers the user to monitor the annotation, so that the user is notified when someone replies to the user's comment.

IMPLEMENTATION

CritLink provides extrinsic linking by maintaining an external database of hyperlinks, and indexes the links by their target location to support bi-directional traversal. It updates this database as it finds new links on browsed pages. Typed links are supported by respecting the `REL` and `REV` attributes on links and providing assistance for setting those attributes when new annotations are created.

A single extension to standard syntax enables the location of fine-grained anchors within a target document. Standard URLs permit the name of an anchor to be given after a "#" character at the end of the URL, when such a named anchor has been marked up by the author of the target document. CritLink additionally understands URLs that specify a sequence of words to search for in the target document. The label "`:words:`" signals this extension. For example:

```
http://example.com/doc.html#:words:the-(blue)-dog
```

identifies, as an anchor, the position where "blue" appears as part of the phrase "the blue dog" in the text of the document located at http://example.com/doc.html. A detailed specification of this extension is given at [17].

The current prototype has the limitation that documents cannot be annotated while being viewed with in a frameset.

DEMONSTRATION

The presenter is the author of CritLink. A hands-on demonstration will be given to show how CritLink can be used to annotate public pages and engage in discussions about Web pages with other readers. The only necessary equipment is a computer with any Web browser and a connection to the Internet.

ACKNOWLEDGEMENTS

CritLink owes its design in large part to contributions from Mark S. Miller, Terry Stanley, Wayne Gramlich, Peter McCluskey, Chris Peterson, K. Eric Drexler, Dean Tribble, Marc Stiegler, Norm Hardy, Chris Hibbert, and Bill Frantz. This work was supported by the Foresight Institute.

REFERENCES

1. T. Berners-Lee, D. Connolly. Hypertext Markup Language 2.0. IETF RFC 1866 (Proposed Standard).
2. T. Berners-Lee, L. Masinter, M. McCahill, eds. Uniform Resource Locators. IETF RFC 1738 (Proposed Standard).
3. V. Bush. As we may think. *The Atlantic Monthly*, July 1945.
4. J. R. Davis, D. P. Huttenlocher. Shared Annotation for Cooperative Learning. *Proceedings of CSCL 1995*, p. 84–88.
5. L. Denoue, L. Vignollet. An annotation tool for Web browsers and its applications to information retrieval. *Proceedings of RIAO 2000.*
6. K. Grønbæk, L. Sloth, P. Ørbæk. WebVise: Browser and Proxy Support for Open Hypermedia Structuring Mechanisms on the WWW. *Proceedings of the 8th International WWW Conference*, 1999, p. 253–268.
7. J. Kahan, M.-R. Koivunen. Annotea: an open RDF infrastructure for shared Web annotations. *Proceedings of the 10th International WWW Conference*, 2001, p. 623–632.
8. H. Maurer, ed. Hyperwave: The Next Generation Web Solution. Addison-Wesley, 1996.
9. M. S. Miller, E. D. Tribble, R. Pandya, M. Stiegler. The Open Society and its Media. *Prospects in Nanotechnology*, John Wiley & Sons, 1995.
10. NCSA. NCSA Mosaic Group Annotations. http://www.ncsa.uiuc.edu/SDG/Software/Mosaic/Docs/group-annotations.html .
11. T. H. Nelson. The Hypertext. *Proceedings of the World Documentation Federation*, 1965.
12. D. LaLiberte. http://www.hypernews.org/ (1996).
13. T. A. Phelps, R. Wilensky. Multivalent Annotations. *Proceedings of the European Conference on Research and Advanced Technology for Digital Libraries*, 1997, p. 287–303.
14. M. Röscheisen, C. Mogensen, T. Winograd. Shared Web Annotation as a Platform for Third-Party Value-Added Information Providers. Technical Report STAN-CS-TR-97-1582, Stanford University, 1994.
15. M. A. Schickler, M. S. Mazer, C. Brooks. Pan-browser support for annotations and other meta-information on the World-Wide Web. *Proceedings of the 5th International WWW Conference*, 1996, p. 1063–1074.
16. ThirdVoice. http://www.thirdvoice.com/ (1999).
17. K.-P. Yee. Text-Search Fragment Identifiers. http://crit.org/draft-yee-url-textsearch-00.txt
18. K.-P. Yee. What's a mediator? http://zesty.ca/mediator.html