Wiki Wide Web: A Platform for Collaborative Editing of the Internet

Jacob I. Torrey and Jeanna N. Matthews
Clarkson University
Potsdam, NY USA
{torreyji, jnm}@clarkson.edu

ABSTRACT

This paper provides an overview of the Wiki Wide Web[1] system, its features, usage and implementation. Wiki Wide Web is a system that allows the collaborative editing of standard non-wiki web pages through the use of a client-side plugin and central diff server. In this paper, we describe the Wiki Wide Web prototype and the design decisions we made to greatly reduce the bandwidth requirements for the central server. We also discuss mechanisms included in the system to deal with disruptive edits.

Categories and Subject Descriptors

H.5.3 [**HCI: Group and Organization Interfaces**]: Theory and models, Evaluation/ methodology, Web-based interaction; K.4.3 [**Computers and Society: Organizational Impacts**]: Computer-supported collaborative work

General Terms

Design, Human Factors

Keywords

collaborative editing, wiki

1. INTRODUCTION

Wikis are becoming common in more places on the Internet as a method to store and collaborate on various documents by editing and changing them. However, it is not currently possible to do this on the majority of the 'plain web'. Wiki Wide Web is a method to combine the massive amount of content on the existing web with the editing and version tracking tools that a wiki can provide.

Wiki Wide Web is like a conventional wiki in many ways. Users can see the edits that others have made and can add their own changes and updates. However, one key difference

---

[1] http://www.wikiwideweb.com

is that with with Wiki Wide Web, each unmodified base web page represents an 'absolute truth' upon which every change is based and to which all changes can be compared. This differs from a more traditional wiki in which collaborative edits form the final product and there is no particular standard to which they should be compared.

Section 2 presents a description of the Wiki Wide Web prototype and information on early usage of the system. In Section 3, we present estimates of data storage and bandwidth requirements for Wiki Wide Web and discuss some alternative design points. In Section 4, we discuss how we deal with disruptive edits. In Section 5, we compare Wiki Wide Web to several similar systems. Finally, future work is discussed in Section 6.

2. WIKI WIDE WEB PROTOTYPE

Wiki Wide Web is implemented as a client-side browser plugin and a central server. The central server stores all the edits made by users. When a user browses to a web page for which changes have been made, the client-side browser plugin downloads the changes from the central server and applies them to the base page before displaying the merged content for the user. The browser plugin also enabled a user to edit any page. The edited content is sent to the central server for storage.

In our prototype, the client-side browser plugin is implemented as a Firefox extension. Changes or diffs to a web page are represented as an XML document that is communicated between the central server and browser plugin using Ajax. The browser plugin uses JavaScript to rewrite the content pane of the web browser dynamically. An advantage of this is that it provides real-time interaction with the diffs, allowing users to skip between then and see how the page evolved without having to wait for page loads. We have seen occasional problems with sites that dynamically modify pages after they are first rendered by the browser (e.g. example, a website that makes use of JavaScript to generate a mouse-over menu).

Figure 1 illustrates the process of editing the WikiSym web page to add a link to past proceedings. We have labeled three main areas of interest in the screenshot for clarity. Area 1 shows the underlying web page that is being edited. Area 2 shows the source code of the page, and has been edited to add some additional information. Area 3 shows a comment box in which the user can provide a reason for their changes.

Figure 2 displays the WikiSym page after the edit has been applied. The plugin italicizes any changed text. When
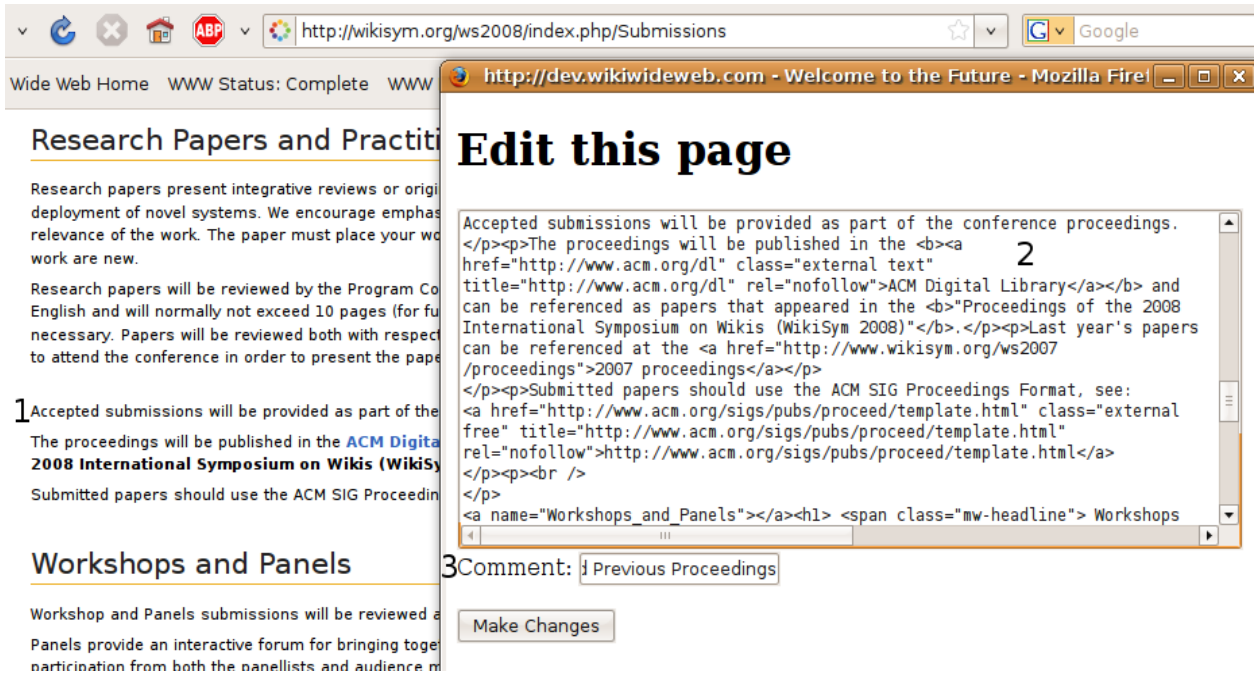
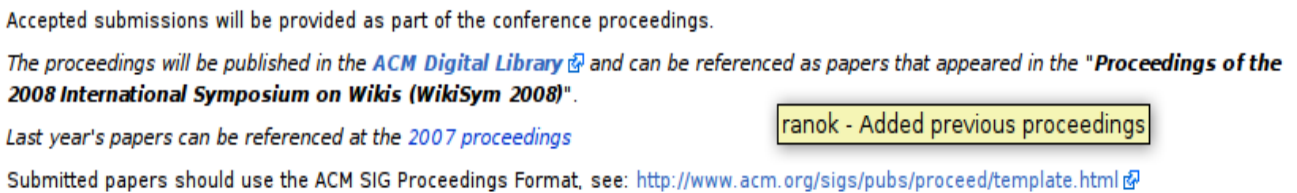Figure 1: Editing a page with Wiki Wide Web

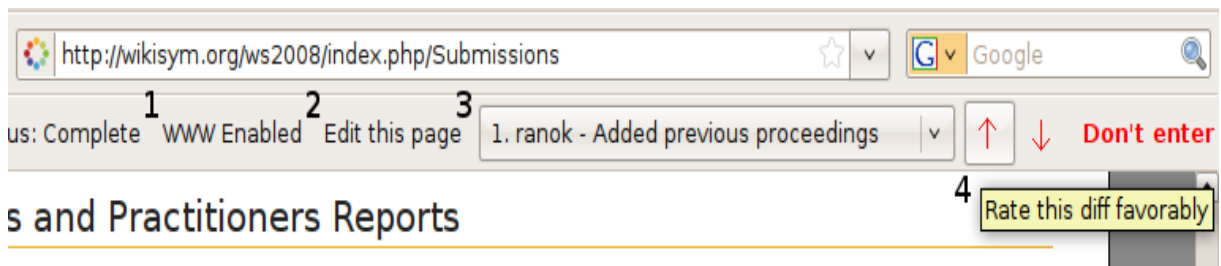Figure 2: Showing an edited page in Wiki Wide Web

Figure 3: Showing some of the controls on the Wiki Wide Web toolbar

the user holds their mouse over the changed text, a box will appear displaying the name of the person who made the change and any comment they entered about the change (from the text field on the edit window).

Figure 3 shows the Wiki Wide Web toolbar. Four areas have been numbered for clarity. Area 1 ('WWW Enabled') is a toggle button used to enable and disable the display of modified content. This allow the user to quickly switch between the unmodified base page and the modified page. Area 2 ('Edit this page') is used to launch an editor as shown in Figure 1.

Area 3 and 4 allow users to rate individual changes. Area 3 is the drop-down box of all the changes that have been made to the page. When a diff is chosen, the up and down arrows in area 4 allow users to quickly rate a diff up or down. This easy rate functionality allows users to participate in patrolling the system for inappropriate diffs.

The Wiki Wide Web prototype, while still under development, has a small user community. When Wiki Wide Web was first released as a private beta, 40 users signed up, editing 30 different sites making 40 total changes. Most of these changes were made in the week following the dispersal of beta invites. While these numbers are small, the feedback from the beta users was very thorough. While many of the users first edits were playfully testing the system by editing their own pages, after the novelty wore off, there were a good number of useful updates to tutorials, or adding more information to a site than was previously available.

3. DESIGN DECISIONS

When Wiki Wide Web was first conceived, it was designed to be a web proxy that would retrieve the requested site, apply any changes, and send the modified version of the page on to the user. After developing this to a point, it was found to be chronically slow, and doubled the bandwidth requirements of the Wiki Wide Web servers.

In Wiki Wide Web, there are two major issues that will arise with increased wiki usage: storage of diffs and bandwidth to quickly transmit those diffs to the users. In in our estimates, we used Amazon's price calculation tool to estimate the storage and bandwidth requirements for both the current design and the alternate design points we considered. We use some data from WikiMedia's Wikipedia statistics tool to estimate the size of edits to web pages over time.

Bandwidth is the first constraint for a web system's growth. It is possible to use S3 for both a storage facility and a method of transferring data to the users. The Wiki Wide Web servers would only have to act as a redirect for requests, telling the plugin which diffs to download from S3, and would compute the diffs from new edits. Because of the minimal role of the Wiki Wide Web servers, bandwidth in addition to the diff delivery would not amount to very much.

If the Wiki Wide Web system was designed to act as a proxy that would modify the page, then send the requested page with diffs applied to the client, the bandwidth requirements would be much larger. According to [4], the average size of an HTML web page is 25KB; the maximum size of a diff in Wiki Wide Web's database is 587B. There are an average of three diffs per page on Wiki Wide Web, making the total size of a pages diffs approximately 1.8KB, or $\frac{1}{14}$th the size. However, since any requested page must be brought into the server, and then sent back out, the total amount of

Usage Type	Diff-only	Whole-page	Improvement
Data Storage	4500 GB	4500 GB	-
Incoming Transfer	10 GB	14000 GB	1400×
Outgoing Transfer	1000 GB	14000 GB	14×
Total Monthly Cost	$856.00	$4,304.61	5×

Table 1: Usage and Transfer Monthly Cost

traffic for the server is double. Using those numbers, Amazon's calculator gives a monthly fee of $4,304.61. Using this proxy method wouldn't change the data storage too considerably so the 4500GB number used in the next paragraph is left the same. The storage usage would be the same as even in the proxy design, only the diffs are stored, the rest of the site is fetched for each request.

Looking at WikiMedia's Wikipedia statistics page[8] and the Wikipedia download page[7], the compressed (bzip) size of all the English articles (12.28 million), with a complete history in XML format is 147.7 GB. Without the history of each page (just the current version), the compressed size is 6.4 GB. This predicts that the storage usage for Wiki Wide Web would be large, even when storing just the diffs. However, even if the uncompressed size is 30 times that of the compressed size, that is still only 4.23 TB, which by today's standards is not an extreme amount.

With scalable options such as Amazon's S3 scalable storage solution, it would be quite feasible to store that much data relatively cheaply. To get a more exact estimate, Amazon's web service price calculator[1] can give an exact monthly charge. In Table 1, the approximate price for storage and the bandwidth for accessing that data is shown. For a service that is serving up that amount of data, $856.00 is quite reasonable for the amount of data. Also, since it is possible to serve diffs directly from S3, the bandwidth costs for serving those diffs would be integrated in with those for the storage. Using this method, the monthly costs are roughly a fifth of the proxy costs. The amount of data transfer in and out were estimates that there are far fewer diffs made in a month than there are requests for diffs. 10 GB is most likely an exaggeration since the Wikipedia database only grew by a maximum of 4.4GB from January 2001 to October 2006 (more recent data was not available). The outgoing transfer was set to one terabyte, as not all of the stored diffs would most likely be accessed in a given month.

While the numbers given earlier in this section are largely based on Wikipedia's data, it shows hope for Wiki Wide Web's scalability in the face of widespread adoption. While the $856.00 a month price for S3 seems expensive, looking at WikiMedia's operating costs[6], they have over $25,000 allotted for hosting and bandwidth. With a budget of that size, hosting Wiki Wide Web should be quite manageable, even if the estimates given were far off from the real figures.

4. DEALING WITH DISRUPTIVE EDITS

Much like any public wiki, there are users who prefer to vandalize pages, or add factual inaccuracies purposely. While these changes can be seen and otherwise fixed, it still takes a large body of motivated and caring individuals to discover and revert the offending changes. With Wiki Wide Web, the problems with disruptive edits can be even greater

for several reasons. First, users can edit web pages that represent the public web presence of individuals, organizations, companies, and governments, thus giving their edits wide visibility. Second, users could modify pages which collect personal information like user-name and password so as to capture that information. Third, users can modify the navigational content and other meta-content as well as textual content. Therefore, we have included a number of mechanisms to deal with disruptive content in the system.

The plugin will automatically disable password fields in all forms to prevent users from inadvertently entering private or personal data whiles the plugin is enabled. The enable/disable toggle switch allows the user to display the original page/disable the plugin at any time.

We also allow certain pages to be locked completely by central server administrators to prevent modification. Sites that are especially vulnerable to attacks, such as PayPal or other financial institutions, are especially appropriate to lock in order to prevent any malicious tampering.

As illustrated in Figure 3, users can rate any diff positively or negatively. This allows diffs with low ratings to be removed from the system. This is especially important because when edits can be applied to any page in the Internet it would be difficult for a team of administrators to patrol all the changes.

5. RELATED WORK

In this section, we briefly discuss several similar projects and how each differs from the Wiki Wide Web project.

5.1 ShiftSpace

The closest related project is be the ShiftSpace [5] project. Like Wiki Wide Web, it is a browser add-on that lets users leave notes and highlight web pages that can optionally be seen by other ShiftSpace users. ShiftSpace also provides the ability to swap images and edit the source of a website.

While this is similar in some regards to Wiki Wide Web, ShiftSpace has a different vision, it tries to be a "layer above any webpage", whereas Wiki Wide Web tries to appear as transparent as possible to the end user. ShiftSpace's 'Source-Shift', which lets users edit the source code of a web page is just one part of ShiftSpace and only lets users make a change and let others see those changes. Wiki Wide Web goes one step further and encourages users to build upon other users' changes.

ShiftSpace provides the features of a user editable web, but is less transparent to the users and requires active user participation. Wiki Wide Web attempts to be a seamless and easy to use collaboration system for the web, it also encourages editing of previous edits and rating them. The rating feature lets users decide which edits should stay, and which should be purged from the database, or hidden. This adds to the collaboration between users, compared to in ShiftSpace, in which users can't have a say in what other people decide to share.

5.2 Fleck

Fleck[3] is a system to annotate websites and share them with other users. It provides a social networking aspect to browsing the web, by letting its users share their annotations (much like Post-It^{TM} notes) and form groups to share their notes and chat about them. It also lets user keep their friends on other social networks such as Twitter^{TM} updated on their Fleck annotations.

Fleck is different from Wiki Wide Web as it provides a way for users to add clearly identifiable annotations to a site, whereas Wiki Wide Web merges those changes into the page directly. When a Fleck user visits a page that another use has added to, they see boxes that contain comments, or clearly noticeable highlights. This is more of a layer above the page, which remains static and unchanged throughout the annotation process. Wiki Wide Web integrates the user changes and comments right with the original site content. Wiki Wide Web also shares your changes and edits with the world, not just your collection of friends. You can have friends in Wiki Wide Web, but they are mostly a way to see last updates, rather than a method to limit change distribution.

5.3 Diigo

Diigo[2] is another web annotation system with a social networking aspect. It lets users highlight and annotate web pages and share those changes with their friends or other connections. However, it is first and foremost a personal bookmark and note manager. While you can share your bookmarks and annotations with others, this is again on another layer from the actual web content.

6. FUTURE WORK

While the Wiki Wide Web code-base and implementation still has work to be done, and more features to add, the most exciting future work will come once as it is adopted by a larger audience and statistics can be generated regarding the usage.

We are especially interested in exploring how well the page rating system will work to democratically keep the pages on topic and appropriate. It will also be very educational to see which types of pages get edited the most, and how in a more democratic system the biases and differing viewpoints will get resolved.

Another consideration that needs to be taken into account is how the system should handle changes to the base web site. Since all the diffs are built upon this base layer, when a change occurs on the base page, some if not all of the diffs become broken. There are a few different ways to remedy this problem, the diffs that get broken can be removed and put into a different area for people to fix, or the old page with the diffs that break can be saved as an immutable page that can get stored on Wiki Wide Web's servers as an extended history. Research still needs to be done to pick the best solution as there is nothing like this in currently implemented wikis as they have no mutable base layer.

We are also interested to explore the impact of a WYSIWYG editor. By enabling the editor for some users and not for others, we can explore whether the lower technical bar of entry will help the content of the web, or let in more vandals who were unable, or unwilling to disfigure sites.

As previously mentioned, the plugin should be ported to other browsers to help increase user adoption. With more browsers supported, it should be possible to get the user base needed to make Wiki Wide Web a valuable system to users, and to have a system to perform widespread social experiments on the democratic wiki system.

7. CONCLUSION

Wiki Wide Web shows great potential to provide a base for further study from its contributions. It can give very detailed statistics regarding its usage and can provide some insight in a truly democratic web. It's rating feature, and soon the WYSIWYG editor will help let more Internet users have a say in the content online, not just as content consumers. The collaborative editing that Wiki Wide Web provides lets users work together in a unique way, making the web their canvas.

8. ACKNOWLEDGMENTS

9. REFERENCES

[1] AMAZON. Simple Monthly Calculator, 2007. [Online `http://calculator.s3.amazonaws.com/calc5.html`; accessed 28-April-2008].

[2] DIIGO. Highlight and Share the Web!, 2008. [Online `http://www.diigo.com`; accessed 23-April-2008].

[3] INSTALINK. Fleck.com - Annote the web, 2008. [Online `http://www.fleck.com`; accessed 24-April-2008].

[4] KING, A. The Average Web Page, 2008. [Online `http://www.optimizationweek.com/reviews/average-web-page/`; access 5-May-2008].

[5] PHIFFER, D. ShiftSpace: What is ShiftSpace?, 2007. [Online `http://www.shiftspace.com/what-is-shiftspace`; accessed 23-April-2008].

[6] WIKIMEDIA. What we need the money for (June 2007), 2007. [Online `http://wikimediafoundation.org/wiki/What_we_need_the_money_for`; accessed 27-April-2008].

[7] WIKIMEDIA. Wikimedia Downloads, 2008. [Online `http://download.wikimedia.org/`; accessed 26-April-2008].

[8] ZACHTE, E. Wikipedia Statistics, 2008. [Online `http://stats.wikimedia.org/EN/Sitemap.htm`; accessed 26-April-2008].