

Lifting the Veil: Improving Accountability and Social Transparency in Wikipedia with WikiDashboard

Bongwon Suh, Ed H. Chi, Aniket Kittur, Bryan A. Pendleton*

Palo Alto Research Center Inc.

3333 Coyote Hill Road

Palo Alto, CA 94304 USA

{suh, echi, akittur}@parc.com, *bp@cmu.edu

ABSTRACT

Wikis are collaborative systems in which virtually anyone can edit anything. Although wikis have become highly popular in many domains, their mutable nature often leads them to be distrusted as a reliable source of information. Here we describe a social dynamic analysis tool called WikiDashboard which aims to improve social transparency and accountability on Wikipedia articles. Early reactions from users suggest that the increased transparency afforded by the tool can improve the interpretation, communication, and trustworthiness of Wikipedia articles.

Author Keywords

Social transparency, accountability, WikiDashboard, Wikipedia, wiki, collaboration, trust, visualization.

ACM Classification Keywords

H.5.3 [Information Interfaces]: Group and Organization Interfaces – Collaborative computing, Computer-supported cooperative work, Web-based interaction, H.3.5 [Information Storage and Retrieval]: Online Information Systems, K.4.3 [Computers and Society]: Organizational Impacts – Computer-supported collaborative work

INTRODUCTION

Collaborative knowledge building has become a highly successful paradigm for creating, finding, and consolidating content online, with popular examples such as Digg.com, del.icio.us, and Wikipedia. Such systems allow virtually anyone to add content and even to change content that others have added. Wikis exemplify this paradigm of user involvement, in which any page can be edited by anyone. The most visible success story of wikis is the Wikipedia project, an online encyclopedia in which any reader can also be a contributor, with their changes immediately visible to

subsequent visitors. This participation model has resulted in a highly popular site with a large amount of content (over two million articles in the English Wikipedia alone).

Despite their tremendous success, collaborative models of knowledge building are still viewed with skepticism. The quality, accountability, and trustworthiness of the articles in Wikipedia has been debated heavily in the press [4, 5, 10, 13]. Wikipedia itself keeps track of these issues and openly discusses them. Even Wales, a co-founder of Wikipedia, said that he would like to make it known that he does not recommend it to college students for serious research [6].

The opposite point of view, however, has not been debated or expressed nearly as much: Precisely because anyone can edit anything and that anyone can examine the edit history and see who has made them, it can (or has already) become a reliable source of information. Because the information is out there for anyone to examine and to question, incorrect information can be fixed and disputed points of view can be examined side-by-side. In fact, this is precisely the academic process for ascertaining the truth. Scholars publish papers so that theories can be put forth and debated, facts can be examined, and ideas challenged. Our research was motivated by the fact that *social transparency* and the attribution of ideas and facts to individual researchers is a crucial part of scientific progress.

ACCOUNTABILITY AND SOCIAL TRANSPARENCY

The Wikipedia community is well aware of these issues and tries to address the problem by requiring *attribution* [14], which is a key official Wikipedia policy. Wikipedia mandates that all presented information must be verifiable through established sources.

Furthermore, revealing the source of information, which is a form of social transparency, has been recognized as an important factor influencing trust in many online interactions [3, 7, 8] and it plays an increasingly important role in collaborative knowledge systems such as wikis [2]. Systems which store transaction logs can provide a rich data source for aggregate information that can provide a meaningful history to users and help them judge the trustworthiness of other users or information [7]. Shneiderman discusses how (1) disclosing patterns of past performance and (2) providing

rich feedback about users and content are best practices for increasing trust online [8].

Indeed, wiki-like systems may be especially fruitful targets for increasing trust through aggregating and surfacing relevant accountability information. Wikipedia is a prime candidate for information aggregation and surfacing, as the history of every edit is stored in the system.

Recently, WikiScanner (wikiscanner.virgil.gr) brought the issue and idea of social transparency to the forefront. It uncovers the organizations where anonymous edits in Wikipedia are coming from. Another site, WikiRage (wikirage.com), helps identify the hottest trends in Wikipedia. The WikiRage idea itself was derived from the “Recent Changes” page on Wikipedia. IBM has developed a tool called History Flow [11] that visualizes the edits to article pages in Wikipedia, and the Content-driven trust coloring [1] demonstrates how trust can be visualized on a line-by-line basis. Erickson et al. [3] discussed the importance of making socially-relevant hidden information visible in interactive communities. On the other hand, provenance of information also has been used as an important factor for assessing the quality of knowledge [9, 15].

These are all examples of how improving social transparency can increase the ability of readers to use the histories of writers to judge the quality of the content, and also possibly encourage writers to be more responsible. In this paper we discuss how to increase accountability and social transparency by surfacing hidden editing information through an unobtrusive “dashboard” embedded in each Wikipedia page.

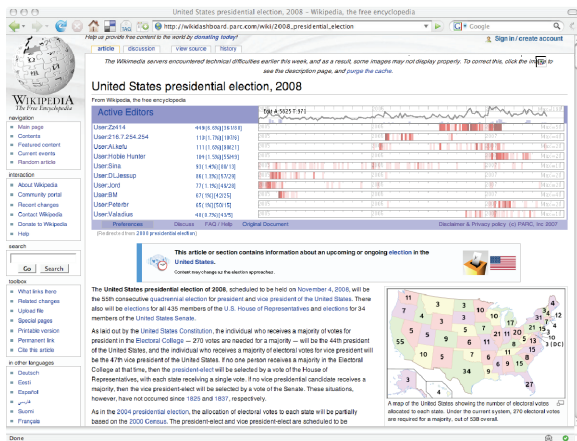


Figure 1. WikiDashboard is a visualization overlay for live Wikipedia pages. The dashboard provides a useful visual digest about who edits how many revisions on each Wikipedia page. It allows users to easily evaluate social activities and patterns around the page, which may be hard to detect otherwise. This figure shows an example of the tool applied to the Wikipedia article “United States presidential election, 2008”

WIKIDASHBOARD

The open participatory model of wiki systems – allowing anyone change anything – makes them inherently dynamic and encourages them to have a large number of editors with various points of view. Furthermore, although users can access past revisions of every page, it is difficult and time-consuming even for dedicated users to make sense of the history of a page, because many page histories run into the thousands of edits (as of June 16th of 2007, 10,453 Wikipedia pages had more than 1,000 edits).

Our goal here is to investigate how providing access to this type of accountability information, i.e. who edits how many revisions for an article, in a digestible form could affect users’ trust and interpretation of an article. If so, the approach can result in reducing the risks many perceive as inherent to a system [2] in which anyone can contribute or change anything.

To address this challenge, we designed WikiDashboard (<http://wikidashboard.parc.com>), a tool that helps users to identify interesting edit patterns in Wikipedia pages, patterns that may be very hard to detect otherwise. As shown in Figure 1, the site provides a dashboard embedded within each page in Wikipedia, while proxying the live content from Wikipedia. The dashboard provides a visualization overlay onto every live Wikipedia page, enabling users to be aware of social dynamics and context around the page they are about to read. The prototype can be used just as if users are on the Wikipedia site itself. All of the functions (such as the article search function, and the discuss and history tabs) work just as in Wikipedia.

Based on the type of Wikipedia pages, the prototype provides two types of dashboards: Article Dashboard and User Dashboard.

Article Dashboard

Each article has an associated article dashboard that displays an aggregate edit activity graph representing the weekly edit trend of the article, followed by a list of the top active editors for that page (Figure 1 and Figure 2).

The top summary graph shows two trends: a gray line graph representing the edits made on the article and a blue bar graph denoting the edits on the corresponding “Talk” discussion page. This graph can help users to easily identify any interesting incidents in the article history, e.g., a sudden burst of edits that might correspond with some heavy discussion.

Below the article edit summary, the active users of the article are ordered by the number of edits they made on the page and its talk page combined, allowing users to easily identify the top editors. Furthermore, the weekly edit activity graph of each editor on the right side of the dashboard enables users to investigate when the edits by that editor were made. A darker red bar denotes more activity in a particular week. With this graph, the tool can help users to examine patterns on how an article evolves over time between multiple editors,

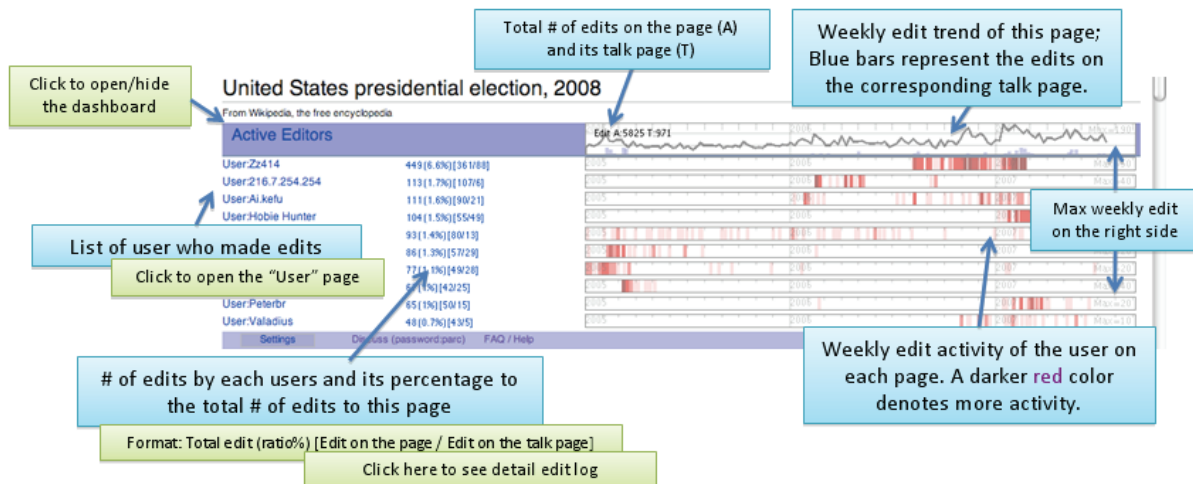


Figure 2 Article Dashboard. The top summary graph shows the weekly edit trend of this page. Below the summary graph, there is a list of active editors and their activities on the article. In this example, it is easy to identify that user Zz414 is the most active editor of the article. It is interesting to see that Zz414 suddenly stopped editing as of Feb. 2007.

e.g., co-editing relationship, recent editors, the burstiness of edits, etc.

User Dashboard

A user page is a special space like a home page to display information relating to a user. In WikiDashboard, each user page has a User Dashboard embedded, displaying the article contribution and editing patterns of that user (Figure 3).

Similar to the Article Dashboard, the top summary graph shows the editor's weekly edit activity (cf. Article Dashboard summarizes edit activities made on each *article*), which allows users to easily examine the editor's overall edit patterns in the past. The summary graph is followed by the list of Wikipedia pages where the editor has made edits. The list is ordered by the volume of contribution and includes the corresponding article-editor activity graphs on the right side.

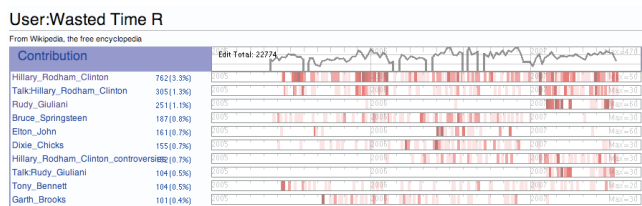


Figure 3 User Dashboard is embedded in each user page of Wikipedia. The dashboard displays weekly edit trend of an editor as well as the list of articles that the editor made revisions on. This example shows a user, "Wasted Time R" made significant edits on articles related to New York politicians and pop singers.

As shown in Figure 3, User Dashboard is designed to facilitate an easy inspection of topics of interest that the editor might have. Furthermore, the dashboard allows investigation of the evolution of editor's topics of interest. For example, the editor in Figure 3 recently developed interest in the Rudy Giuliani article.

Article titles and user names in both dashboards are clickable links, allowing users to browse through them for further exploration. For example, clicking on an article title brings up the corresponding article and article dashboard.

Clicking on the statistics summarizing the number of edits for a user launches the Detail Edit Log (Figure 4), which displays the list of every edit that a specific editor made on the article. This drill-down tool enables readers to examine each individual revision for validity, which is hard to accomplish when only provided with aggregate visual summaries.

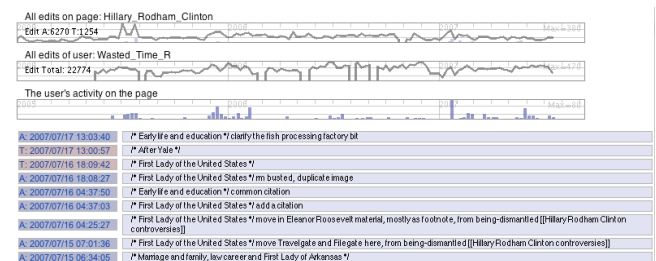


Figure 4 Detail Edit Log is provided for detail investigation of individual edits made by a particular editor on a selected page.

DISCUSSION AND FUTURE WORK

Theories of social translucence [3] state that three building blocks are necessary for effective communication and collaboration: making socially significant information visible and salient; supporting awareness of the rules and constraints governing the system; and supporting accountability for actions. The idea of social translucence suggests that WikiDashboard could benefit not only readers but also improve the effectiveness of active writers. Indeed, we are very interested in analyzing the impact of improved social transparency on both readers and editors. We plan to explore how readers process the given social context and use the information to evaluate user-generated contents as well

as potential behavioral change of editors due to improved social transparency.

Even though WikiDashboard has only been available to the public for a week, we have had already 2,388 visits and 12,157 page views. Thus, we have already been able to capture a number of insightful feedbacks from various users:

*“WikiDashboard appears to be a valuable tool that can provide some good insights into individual edit patterns and edit conflicts on specific articles. As a means of learning about the tool I have found it useful to use it on articles that I have an intimate understanding of development in order to get a feel of how it can be used and interpreted.”*¹

*“This is a revolutionary tool for us wiki-watchers.”*¹

*“This is very useful for getting a quick glance of the user's editing interests over time. ... I actually think a tool like WikiDashboard presents significantly more utility, and is the beginning of an interesting trend of repurposing metadata to create a trust heuristic.”*²

“WikiDashboard (is) a quick way to find the most active editors of an article. On (Hillary) Clinton's article, (user) Tvoz ranked third highest. On (Fred) Thompson's, (user) Ferryldodge was unmatched.” [10]

These feedback and testimonials are very encouraging and show that WikiDashboard has a great potential to change the way of readers' interpreting Wikipedia articles, influence editors to be more responsible, and ultimately improve trust in the content.

In future work, we are considering a number of improvements: (1) Providing richer social context on the dashboard would be an obvious next step. For example, showing the last few editors might be useful for readers to evaluate an article. (2) We plan to improve the measurement of editors' contribution in the system. Some edits clearly are more substantial than other minor fixes. Simple measures such as number of net words changed might be useful. (3) We are also interested in making the tool available as a collaborative visual analytic platform for Wikipedia, where users can discover noteworthy editing patterns and share them with other users, similar to the ManyEyes system [12].

One more important aspect of social transparency is to control what and how much to expose. It is crucial for some information to remain private. Furthermore, exposing particular types of information could have a negative impact on the system. Would we encourage editors to game the system to get to the top 10 editor list of many articles? It is an important question for future research as to which information is better left private vs. made transparent.

CONCLUSION

In this paper we introduced WikiDashboard, a social dashboard that aggregates and surfaces “under the hood” information in Wikipedia. WikiDashboard has great potential in impacting users' interpretation of Wikipedia. We hope the ideas presented in this paper may inform designers aiming to increase social transparency and user trust in other collaborative systems.

REFERENCES

1. Adler, B.T. and de Alfaro, L. A. “Content-Driven Reputation System for the Wikipedia”, *In Proc. Of WWW 2007*, ACM Press (2007).
2. Denning, P., Horning, J., Parnas, D., and Weinstein, L. “Wikipedia risks”, *Comm. ACM*, 48(12), ACM Press (2005), 152-152.
3. Erickson, T., Halverson, C., Kellogg, W.A., Laff, M., and Wolf, T. "Social translucence: designing social infrastructures that make collective activity visible", *Comm. ACM*, 45 (4), ACM Press (2002), 40-44.
4. Giles, G. “Internet encyclopaedias go head to head”, *Nature*, 438(7070), (2005), 900-901.
5. Hafner, K. “Seeing Corporate Fingerprints in Wikipedia”, *The New York Times*, Aug 17th, 2007
6. Lombardi, C. “Wikipedia: Stop citing our site”, On CNET news blog, June 16th, 2006, retrieved from http://www.news.com/8301-10784_3-6084715-7.html
7. Resnick, P., Zeckhauser, R., Friedman, E. and Kiwabara, K. “Reputation systems” *Comm. ACM*, 43(12), ACM Press (2000), 45-48.
8. Shneiderman, B. "Designing trust into online experiences", *Comm. ACM*, 43(12), ACM Press (2000), 57-59.
9. Stvilia, B., Twidale, M. B., Smith, L. C., and Gasser, L. Assessing information quality of a community-based encyclopedia. *In Proc. of ICIQ* (2005), 442-454.
10. Vargas, J.A. “On Wikipedia. Debating 2008 Hopefuls' Every Facet”, *The Washington Post*, Sept 17th 2007.
11. Viégas, F.B., Wattenberg M., Dave K., “Studying cooperation and conflict between authors with history flow visualizations”, *In Proc. of CHI 2004*, ACM Press (2004), 575-582.
12. Viégas, F.B., Wattenberg, M., van Ham, F., Kriss, J., and McKeon, M. "Many Eyes: A Site for Visualization at Internet Scale", *In Proc. in InfoVis '07*, (2007)
13. Waters, N.L. "Why You Can't Cite Wikipedia in My Class", *Comm. ACM*, 50(9), ACM Press (2007), 15-17.
14. Wikipedia.org, Wikipedia:Attribution, <http://en.wikipedia.org/wiki/WP:A> Retrieved Sep 15, 2007.
15. Wilkinson, D. M. and Huberman, B. A. “Cooperation and quality in Wikipedia”, *In Proc of WikiSym '07*, ACM Press (2007), 157-164.

¹ The Wikipedia Review, <http://www.wikipediaireview.com>

² Unit Structures, <http://chimprawk.blogspot.com/>