

Evaluating Reliability in Collaborative Information Products:

A Reputation and Rating System for Wikipedia

Bradley Herman

UC Berkeley School of Information

## **Evaluating Reliability in Collaborative Information Products:**

### **A Reputation and Rating System for Wikipedia**

In this paper, I propose a design for a reputation and rating system to increase the quality of articles on Wikipedia and its reliability as a source of information. A combination of reader feedback and distributed moderation will algorithmically generate reputations for contributors and ratings for articles. The ratings help readers evaluate reliability by providing needed cues that summarize the extent of and participants in the editing process for an article as well as past readers' opinions of it. Furthermore, individual reputations have been shown to increase quality and adherence to shared norms in virtual communities. These tools can help Wikipedia address some of the problems it faces while maintaining the openness of its operation.

Wikipedia is a free encyclopedia available on the Internet. The goal of the project is to make the world's knowledge available to any person in his or her own language. Articles are produced collaboratively and entirely by volunteers. It is open, meaning that anyone can create or edit an article at any time, with the system storing a copy of all previous versions. A community of contributors and enthusiasts has grown up around Wikipedia as the project has experienced rapid growth. It has surprised observers with the size and scope as well as the overall quality of its offerings.

However, Wikipedia faces a number of significant problems. While many are good, the quality of some of its articles is quite low.<sup>1</sup> Quality, in this context, is a measure of factual accuracy, comprehensiveness, and objectivity – what Wikipedia calls “neutral point of view.”<sup>2</sup> A more subtle problem is that readers, especially those unfamiliar with the project, may have difficulty evaluating the quality of articles. The nature of an encyclopedia article means that

<sup>1</sup> The John Seigenthaler Sr. incident is an example of low quality, where an anonymous contributor edited Mr. Seigenthaler's biographical article to claim that he was involved in the Kennedy assassinations. See Seelye (2005) for more on quality concerns in Wikipedia.

<sup>2</sup> See <http://en.wikipedia.org/wiki/NPOV> for more on neutral point of view and quality in Wikipedia

most readers will not be able to evaluate its claims. For Wikipedia articles, the editing history is obscure and the authors are usually pseudonymous or anonymous. As a result, it is often hard to tell what information is reliable. Moreover, the openness of the system means that articles are prone to vandalism and that controversial topics occasionally spark “editing wars” among opinionated contributors.

Reputation systems can be used to resolve some of these problems. Raub and Weesie (1990) note that a reputation develops when an individual’s history of behavior is communicated to the people with whom he or she interacts. It represents a prediction of that individual’s future actions (Kollock 1999b) and encourages good behavior (Resnick, Zeckhauser, Friedman, & Kuwabara, 2000). Reputation systems enable people to make better judgments about uncertain information. Epinions readers, for example, leverage ratings of reviewers in deciding which product reviews to rely on. In addition, eBay buyers and sellers complete successful transactions more often than outside observers would expect because partners' feedback affects their ongoing reputation (Kollock, 1999b).

There are a few interesting aspects of Wikipedia that this paper does not consider. First, Wikipedia exhibits the characteristics of a public good. While this has consequences for the motivation and coordination of contributions,<sup>3</sup> a discussion of those is beyond the scope of this paper. Thus far, there has been an abundance of participation in Wikipedia. Here, I take strong community interest and participation for granted.

Second, the structure and nature of the community that has grown around Wikipedia play a significant role in how people use the encyclopedia. It is strong enough that members even have a name for themselves, “Wikipedians.” As Wellman and Gulia (1997) note, “those who use

---

<sup>3</sup> For an examination of the challenges of public goods on the Internet generally, see Kollock (1999a). For discussion specific to Wikipedia, see Cifforilli (2003)

the Net seek not only information, but also companionship, social support, and a sense of belonging.” While the community affects the design and efficacy of any reputation and rating system, a thorough investigation of the Wikipedia community is best left for another author.

### **System Design Considerations**

To address the problems facing Wikipedia, the feedback system I propose has two interconnected parts: ratings of articles and reputations for individual contributors. Article ratings allow readers to better evaluate the reliability of that information. Individual contributors have reputations to help the community identify high and low quality edits as well as to encourage more and better contributions. In this context, reliance means readers trust that the authors have created a high quality article.

The first part is to rate the articles themselves on quality<sup>4</sup>. Readers have the opportunity to give a positive, neutral, or negative mark to each article they read. To discourage gaming behavior, a single user account (or IP address for anonymous readers) can only affect the score of a particular article once; future ratings of the same article erase the previous mark. The aggregate of all readers' marks contribute to that article's rating. Lih (2004) identifies the total number of edits to an article and the total number of unique editors as indicators of relative quality in Wikipedia articles. Those metrics are calculated automatically and also contribute to the article's rating. Finally, since the quality of an article is a reflection of its authors, their reputations factor into the rating of the article. The mechanics of these calculations are discussed below.

The second component to this system is establishing reputations for contributors. Since Wikipedia users usually operate under pseudonyms, there is no possibility of increasing their

---

<sup>4</sup> Here, many dimensions of quality are reduced to a single score for simplicity. For a more detailed discussion of the dimensions of quality in encyclopedic articles, see Stvilia, Twidale, Gasser, & Smith (2005)

trustworthiness by verifying credit card information or other real-world identifiers. Instead, they are judged on the quality of their contributions. When a reader marks an article, the system allocates a fraction of that feedback to each author based on the proportion that he or she is responsible for the current state of the article. The WikiRep framework provides an elegant example of this approach (Sabel, Garg, & Battiti, 2005). Additionally, community members themselves rate other users' contributions. At standard intervals, a few regular Wikipedia users will be selected randomly to distribute a small number of reputation points for particularly good edits they find, like gold stars on noteworthy schoolwork. This process is modeled after the distributed moderation system developed at Slashdot (Malda, 2006). These points are the second factor that determines a contributor's reputation.

This design assumes two things about the production of quality contributions. The first is that users with higher reputations are likely to make higher quality contributions than those with low reputations. The second is that a larger contribution has a greater affect on the quality of an article than a smaller contribution. These are simplifications of a complicated process, but I feel that, for the purposes of this system, these are descriptive and valid assumptions to make.

Resnick et al. (2000) identify three required properties for a reputation system to operate effectively. Identities in the system must persist for a long time, the system must collect and publicize feedback about interactions, and it must present that information in a way that informs users' trust decisions. Wikipedia presents interesting challenges and opportunities for implementing each of these properties.

As mentioned above, pseudonymous accounts and anonymous contributions are common among Wikipedia users. While some restrictions have recently been placed on anonymous contributions (Stross, 2006), there is strong support for continued anonymous access within the

community (Mayfield 2005). To balance this desire and the need for long-lived identities, the reputation system discourages users from changing their pseudonyms by forcing new accounts to “pay their dues.” Depending on the implementation and community support, this could be accomplished either through limiting access or mandatory quality reviews for contributions by new accounts. This is the most efficient way to maintain an effective reputation system in the face of easily acquired pseudonyms (Friedman and Resnick, 2001). Further, since an anonymous contribution is likely to be of lower quality than signed one (Lampe and Resnick, 2004), anonymous edits could be made to go through the review process as well. Since this imposes an additional burden, the specific mechanism should be determined by the community.

Capture and assignment of feedback is complicated in a collaboratively edited environment like Wikipedia. Articles change over time and it is not always easy to tell which author is responsible for which parts. Moreover, the nature of an encyclopedia article means that most readers are not in a position to evaluate the accuracy of its claims, though they can tell if it is well written. The WikiRep and Slashdot distributed moderation models can help overcome these hurdles.

The WikiRep system allows ratings of an article to be distributed to each of its authors' reputations in proportion to each author's contribution to the current version of the article. It also allows an article that has just been edited to inherit a rating from the previous version (Sabel et al., 2005). One can imagine extending this model to include adjusting an article's rating based on the reputation of the author who makes an edit and the size of the edit. If, for example, a highly rated user makes a large edit to a mediocre article, the system would provisionally rate the new version higher. Subsequent feedback from readers would reinforce this rating and the contributor's reputation, or lower the article's rating and contributor's reputation. Anonymous

contributions are treated skeptically and affect an article negatively until marked otherwise by a known user.

Such a scheme is not immune from unscrupulous users. A contributor might make a series of small but malicious edits to popular articles expecting that subsequent readers would still rate the overall article positively and, thus, raise his reputation. Here, empowering community members to rate individual edits by their peers can have a positive effect.

In the proposed Wikipedia reputation system, every thirty minutes a few individuals are chosen from the pool of currently active regular contributors to act as “moderators” and hand out a few reputation points. They can give a positive or negative point to any edit they come across during that time. Positive points add to the contributor's reputation as well as a fraction of the article's rating proportional to the size of the contribution. Edits marked down contribute negatively. Once all of a moderator's points are given out or the time expires, that individual goes back in the pool of users to be chosen in the future. Safeguards are built in to prevent a moderator from marking edits in an article he or she has edited recently. Additionally, there is a meta-moderation system, where signed-in users are asked to evaluate a specific recent moderator action, to weed out moderators who are consistently unfair or biased.<sup>5</sup> Analysis of the implementation on Slashdot indicates that distributed moderation can achieve ratings that accurately reflect the opinions of the community (Lampe and Resnick, 2004).

The third property this system needs to be effective is for Wikipedia readers and community members to use these ratings and reputation scores to guide their trust decisions. For articles, the rating is displayed as an image of one to four stars near the title of each article. There is also a link to a webpage explaining the rating system and what the score means for

---

<sup>5</sup> For a more detailed description of this type of distributed moderation system see Malda (2006)

readers trying to evaluate the reliability of that article. Reputation displays for contributors work in a similar manner, with a number of stars displayed on that user's personal page.

### **Efficacy of This System**

Will a reputation and rating system like the one described above effectively increase the quality and reliability of articles on Wikipedia? This question has two parts that need answering. Will article ratings help readers make better judgments about the reliability of articles? And, will reputations for contributors lead to higher quality contributions?

First, consider ratings for articles. Readers have very limited contextual information about the articles they read or the contributors who wrote them. This lack of cues makes it difficult to evaluate the information presented (Dellarocas, 2003). Ratings indicate how confident the community of contributors and moderators are in the quality of the article and, thus, how heavily the reader should rely on that information.

In fact, Wikipedia users have developed a feature related to article rating in the “featured article” designation. A bronze star image and the text “This is a Featured Article” appear on about 1000 of the best articles. The designation indicates to readers that the article is reliable because it has been reviewed for quality and provides an example of good quality to other contributors. While featured articles are a good start, readers can benefit even more from reliability indicators on all articles.

Vigilant community members often catch malicious contributions or mark poorly written articles for “cleanup,” but some make it through. While any reader can inspect the article's history of edits or its discussion page where contributors comment on the quality of contributions, these are often unknown or inscrutable to casual readers. Without a simple rating,

a reader might not know that a hypothetical article is unreliable because it was recently created by a contributor who has a reputation for inaccurate and inflammatory writing.

Ratings and reputation systems in other online environments are used effectively to aid user judgments about which people to trust and what information to rely on. eBay users, for example, often successfully base their trust of transaction partners on the opinions of other users (Resnick et al., 2000). While the marketplace aspect of eBay differs significantly from Wikipedia, users of sites that deal more exclusively in information benefit from ratings, too. Lampe and Resnick (2004) found that ratings given to comments on Slashdot offer “information of potential value to readers” (p. 8). It has also been effective at identifying high and low quality contributions (Rhiengold, 2002). Since encyclopedia readers seek information on a particular topic, they will not use ratings to filter out low-quality articles, but to choose between relying on Wikipedia or seeking another source.

Next, consider reputations for contributors. Well designed reputation systems tend to increase cooperation among participants in transactions (see, for instance, Raub and Wessie, 1990; Yamagishi and Matsuda, 2002). Cooperation among Wikipedia contributors means following community norms for high-quality contributions and respect for other contributors.

When the community's opinion of a contributor's actions is publicized in a simple way, it can affect how other members of that community will react to his or her next actions. In the Wikipedia community, this could manifest in a low-reputation user having more contributions re-edited or pulling less weight in discussions about community policy or article deletion. This “shadow of the future” will constrain his or her actions to remain more in line with the prevailing norms (Resnick et al., 2000), contributing in ways that improve article quality or leaving an insightful comment on an article's discussion page.

The presence of reputation information also constrains the way people interact with other participants. In the eBay community, where reputation affects the price reached in auction, users try hard to work out disagreements through communication before resorting to negative feedback (Kollock 1999b). Wikipedia struggles with rude contributors not respecting other users' contributions and with “edit wars,” where contributors with conflicting strong opinions on a topic will repeatedly edit an article to reflect their own point of view and undo the other person's edits. With a reputation system in place, there will be pressure to make a good impression on other members of the community and settle disputes in more amicable ways.

A formalized reputation system allows the quality control mechanisms used by Wikipedia to scale as the community grows. The site already has some activity around reputations, but it is informal and mostly dispersed over many discussion pages<sup>6</sup>. The “hard core” of the Wikipedia community is around 0.5% of users who account for half of the contributions (Mayfield, 2005). Members of this informal group have a significant amount of interaction and have built trust relationships over time. Because they have developed a rich sense of one another's reputations, these users can gauge the amount of oversight to give to a contribution by another member. As the number of less frequent contributors increases, these trust relationships will provide insight into a smaller fraction of contributions. This proposed system can help users evaluate contributions without requiring long term personal interactions. It can also expand the quality control process to include less frequent and newer users who have not or will not invest the time and energy cultivating trust relationships in that community.

By reporting contributors' reputations as a more tangible measure, this proposed

---

<sup>6</sup> Every article and user has a discussion or “talk” page where contributors comment on the edits they have made, want to make, or have reverted. They are used to explain actions, generate consensus, rebuke offenders, and resolve conflicts. See [http://en.wikipedia.org/wiki/Wikipedia:Talk\\_page](http://en.wikipedia.org/wiki/Wikipedia:Talk_page) for more information.

reputation system emphasizes recognition and community status as an additional motivation for high quality contributions. This can be seen on Slashdot, where other motivations to participate in a conversation are augmented by the promise of reputation points for quality comments. People who might otherwise not contribute to the conversation often post helpful contextual materials or text of the news-article under discussion because they know that such posts will boost their reputation. Detractors have labeled this practice “karma whoring”<sup>7</sup> because some such comments aren't especially original or insightful. In an encyclopedia context, however, these contributions are important even though they represent low-hanging fruit.

Finally, this proposed reputation system offers Wikipedia another way to catch and change poor contributions. There are committed members of the community that try to check every anonymous edit because these tend to be low quality or vandalism. Obviously, there are some logged-in users that contribute low quality edits as well. By knowing the reputations of contributors, community members can develop tools to highlight contributions from users with low reputations for attention by others. This way, consistently poor contributors or new users unfamiliar with the style conventions and norms of Wikipedia are less likely to reduce the quality of articles.

Despite the evidence that ratings and reputations help users make accurate reliability judgments and increase the quality of Wikipedia articles, there are some potential drawbacks to implementing this system that need to be addressed. The first is that, as with any reputation or rating system, contributors may try to game the system. A malicious user might build up a high reputation and then sneak an inaccurate or biased paragraph into an article in hopes that it will either be unnoticed or trusted. This proposal is meant to supplement Wikipedia's existing

---

<sup>7</sup> “Karma” is how Slashdot refers to reputation score; as one's reputation increases one accumulates more karma.

community ethic and quality control mechanisms, so we can expect that if users game the reputation system, they will still have a hard time gaming the whole community.

Another serious concern is that the external motivation of building reputation will crowd out users' current internal motivations for contributing. Changing the balance of motivations has implications for the sharing of knowledge (Osterloh and Frey, 2000) and can change the nature and quality of contributions. If that change is for the worse, it will reduce the effectiveness of the entire Wikipedia project. Time and empirical research will tell whether reputation does cause a significant change in motivation. The positive experience of Slashdot's reputation system with promoting a rather high quality conversation (Lampe and Resnick, 2004) bodes well for Wikipedia.

A final concern for this system is whether the diversity and nuance of a contributor's reputation can be effectively condensed into a single measure. This is highlighted by the system's failure to recognize different domains of knowledge. Somebody may have contributed a wealth of information about eighteenth century European classical music and built up a high reputation. He or she should not necessarily be trusted to make high quality contributions to articles on particle physics or the social history of the hot dog. Nonetheless, other components of quality, like good writing style and NPOV, do apply across all Wikipedia articles. The measurements of contributor reputation or article rating might be extended to capture more of the subtleties, but it must keep to a minimum the participant effort required to create and interpret those scores (Lampe and Resnick, 2004).

### **Conclusion**

The reputation and rating system proposed in this paper can improve readers' evaluations

of article reliability and improve article quality in Wikipedia. Feedback from readers will be combined with distributed moderation to calculate and present reliability scores for articles as well as reputation scores for contributors.

Article ratings convey important cues to readers about the level of confidence the Wikipedia community has in the quality of that article and, thus, its reliability as an encyclopedic source of information. By condensing scattered information about the number and reputation of an article's authors, the history and extent of their contributions, and the collective opinion of past readers, ratings give readers a simple metric to help evaluate what they read. Highly rated articles can be featured, raising the bar for all contributors, while poorly rated ones can be automatically targeted for cleanup.

Contributor reputation scores are useful for other contributors in the Wikipedia community. They serve to simplify and formalize the reputation that a contributor now acquires informally as he or she interacts with articles and other community members. Contributors interested in improving their own reputation scores will make their contributions of higher quality and closer in line with the Wikipedia norms. In addition, contributions by users with particularly low reputations will attract more critical review than those with high reputation.

While there are some pitfalls, rating and reputation systems have successfully increased quality and cooperation in other online collaborative and transactional communities like eBay and Slashdot. Implemented well with community support in Wikipedia, this system can help overcome some of the encyclopedia's reliability and quality problems as well as help the project scale while retaining the openness and participation that have made it successful.

### References

- Ciffolilli, A. (2003). Phantom Authority, Self-Selective Recruitment and Retention of Members in Virtual Communities: The Case of Wikipedia. *First Monday*, 8(12).
- Dellarocas, C. (2003). The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms. *Management Science* 49(10), 1407-1424.
- Friedman, E. J., & Resnick, P. (2001). The Social Cost of Cheap Pseudonyms. *Journal of Economics & Management Strategy* 10(2), 173-199.
- Kollock, P. (1999a). The Economies of Online Cooperation: Gifts and Public Goods in Cyberspace. In M. Smith & P. Kollock (Eds.), *Communities in Cyberspace* (pp. 220-239). New York, NY: Routledge.
- Kollock, P. (1999b). The Production of Trust in Online Markets. *Advances in Group Processes* 16.
- Lampe, C., & Resnick, P. (2004). Slash(dot) and Burn: Distributed Moderation in a Large Online Conversation Space. *Proceedings of the ACM Computer Human Interaction Conference 2004, Vienna Austria*. New York, NY: ACM Press.
- Lih, A. (2004). Wikipedia as Participatory Journalism: Reliable Sources? Metrics for Evaluating Collaborative Media As a News Resource. *5<sup>th</sup> International Symposium on Online Journalism*. University of Texas, Austin.
- Malda, R. (2006). Slashdot Moderation. Retrieved April 26, 2006, from <http://slashdot.org/moderation.shtml>.
- Mayfield, R. (2005). Governance, Scaling and Anonymity in Wikipedia. *Corante Many-to-Many*. Retrieved April 19, 2006, from [http://many.corante.com/archives/2005/08/12/governance\\_scaling\\_and\\_anonymity\\_in\\_wikipedia.php](http://many.corante.com/archives/2005/08/12/governance_scaling_and_anonymity_in_wikipedia.php).
- Osterloh, M. & Frey, B. S. (2000). Motivation, Knowledge Transfer, and Organizational Forms. *Organization Science* 11(5), 538-550.
- Raub, W. & Weesie, J. (1990). Reputation and Efficiency in Social Interactions: An Example of Network Effects. *The American Journal of Sociology* 96(3), 626-654.
- Resnick, P., Zeckhauser, R., Friedman, E., & Kuwabara, K. (2000). Reputation Systems. *Communications of the ACM* 43(12), 45-48.
- Rhiengold, H. (2002). The Evolution of Reputation. In *Smart Mobs* (113-132). Perseus Books Group.

- Sabel, M., Garag, A., & Battiti, R. (2005). WikiRep: Digital Reputation in Virtual Communities. *DIT Technical Report, DIT-05-050*. University of Trento, Italy. Retrieved April 10, 2006, from <http://eprints.biblio.unitn.it/archive/00000810/01/wikireptr.pdf>.
- Seelye, K. Q. (2005, December 4). Snared in the Web of a Wikipedia Liar. *The New York Times*. Retrieved May 4, 2006, from <http://www.nytimes.com/2005/12/04/weekinreview/04seelye.html>
- Stross, R. (2006, March 12). Anonymous Source Is Not the Same as Open Source. *The New York Times*. Retrieved April 26, 2006, from <http://www.nytimes.com/2006/03/12/business/yourmoney/12digi.html>.
- Stvilia, B., Twidale, M. B., Gasser, L., & Smith, L. C. (2005) Information Quality Discussions in Wikipedia. Technical Report ISRN UIUCLIS—2005/2+CSCW. Retrieved April 26, 2006 from <http://www.isrl.uiuc.edu/~stvilia/papers/qualWiki.pdf>.
- Wellman, B. & Gulia, M. (1999) Virtual Communities: Net Surfers Don't Ride Alone. In M. Smith & P. Kollock (Eds.), *Communities in Cyberspace* (pp. 167-194). New York, NY: Routledge.
- Wikipedia. (2006). Wikipedia: Featured Articles. Retrieved April 26, 2006, from [http://en.wikipedia.org/wiki/Wikipedia:Featured\\_articles](http://en.wikipedia.org/wiki/Wikipedia:Featured_articles).
- Yamagishi, T. & Matsuda, M. (2002) Improving the Lemons Market with a Reputation System: An Experimental Study of Internet Auctioning. Forthcoming.