Statistical Measure of the Effectiveness of the Open Editing Model of Wikipedia

Sara Javanmardi and **Yasser Ganjisaffar** and **Cristina Lopes** and **Pierre Baldi**
University of California, Irvine
{sjavanma, yganjisa, lopes, pfbaldi}@ics.uci.edu

Abstract

Wikipedia is commonly viewed as the main online encyclopedia. Its content quality, however, has often been questioned due to the open nature of its editing model. A high–quality contribution by an expert may be followed by a low–quality contribution made by an amateur or vandal; therefore the quality of each article may fluctuate over time as it goes through iterations of edits by different users. In this study, we model the evolution of content quality in Wikipedia articles in order to estimate the fraction of time during which articles retain high–quality status. The results show that articles tend to have high–quality content 74% of their lifetime and the average article quality increases as articles go through edits. To further analyze the open editing model of Wikipedia, we compare the behaviour of anonymous and registered users and show that there is a positive correlation between registration and quality of the contributed content. In addition, we compare the evolution of the content in Wikipedia known high–quality articles (aka. featured articles) and the rest of the articles in order to extract features affecting quality. The results show that the high turnover of the content caused by the open editing model of Wikipedia results in rapid elimination of low–quality content.These results not only suggest that the process underlying Wikipedia can be used for producing high–quality content, but also to question the viability of collaborative knowledge repositories that impose high barriers to user participation for the purpose of filtering poor quality contributions from the onset.

Introduction

Web 2.0 is the second generation of the web that emphasizes crowdsourcing, the process of outsourcing a task to a large group of people, in the form of an open call (Waldrop 2008). Using wiki technology, Wikipedia has become the largest crowdsourcing project and the main online encyclopedia (Zittrain 2008). It has been suggested that wiki technology can harness the Internet for science; "Wikinomics" is a recent term that denotes the art and science of peer production when masses of people collaborate to create innovative knowledge resources (Tapscott and Williams 2006). Because of its open editing model –allowing anyone to enter and edit content– Wikipidia's overall quality has often been under question. While it is difficult to measure Wikipedia's overall quality in a definitive way, two studies have tried to assess it manually by comparison of Wikipedia articles to their parallel articles in other reputable sources (Giles 2005; Chesney 2006). Nature magazine's comparative analysis of forty–two science articles in both Wikipedia and the Encyclopedia Britannica showed a surprisingly small difference; Britannica disputed this finding, saying that the errors in Wikipedia were more serious than the Britannica errors and that the source documents for the study included the junior versions of the encyclopedia as well as the Britannica year books[1].

The questions surrounding Wikipedia's open editing model have triggered a new generation of wikis like Citizendium[2] and Scholarpedia[3]. These online encyclopedias follow a much more traditional editing model, where a small number of experts produce most of the content, through a peer–reviewing process [4]. However, there is very little evidence that these traditional editing models are better than Wikipedia's model for the purpose of creating encyclopedic knowledge. To further address these issues, one must develop methods for automatically assessing Wikipedia's quality and the parameters that affect it.

Since Wikipedia is a highly dynamic system, the articles are changing very frequently. Therefore, the quality of articles is a time–dependent function and a single article may contain high– and low–quality content in different spans of its lifetime. The goal of our study is to analyze the evolution of content in Wikipedia articles over time and estimate the fraction of time that articles are in high–quality state.

This paper offers two main contributions to the state of the art. First, we develop an automated measure to estimate quality of article revisions throughout the entire English Wikipedia. Using this measure, we follow the evolution of content quality and show that the fraction of time that articles are in a high–quality state has an increasing trend over time. Then, we present an empirical study of Wikipedia statistics that may explain the results obtained in our study. We analyze the contributions of registered and anonymous users and show that there is a positive correla-

[1]http://bit.ly/cLDpXO
[2]http://en.citizendium.org/
[3]http://www.scholarpedia.org/
[4]http://bit.ly/9g4yzX

tion between user registration and quality of the contributed content. Furthermore, we compare the evolution of content in featured and non–featured articles over time. We show that featured articles are more closely followed and they benefit from higher content turnover (i.e. higher deletion and replacement of low–quality content).

Background & Related Work

In the open editing model of Wikipedia users can contribute anonymously or with untested credentials. As a consequence, the quality of Wikipedia articles has been a subject of widespread debate. For example, in late 2005, American journalist John Seigenthaler publicly criticized Wikipedia because of a collection of inaccuracies in his biography page, including an assertion that he was involved with the assassination of former U.S. President John F. Kennedy[5]. Apparently the inaccuracies remained in Wikipedia for 132 days. Because there is no single entity taking responsibility for the accuracy of Wikipedia content, and because users have no other way of differentiating accurate content from inaccurate content, it is commonly thought that Wikipedia content cannot be relied upon, even if inaccuracies are rare (Cross 2006).

To overcome this weakness, Wikipedia has developed several user–driven approaches for evaluating the quality of its articles. For example, some articles are marked as "featured articles". Featured articles are considered to be the best articles in Wikipedia, as determined by Wikipedia's editors. Before being listed here, articles are reviewed as "featured article candidates", according to a special criteria that takes into account: accuracy, neutrality, completeness and style[6]. In addition, Wikipedia users keep track of articles that have undergone repeated vandalism in order to eliminate it and report it [7]. However, these user–driven approaches cannot be scaled and only a small number of Wikipedia articles are evaluated in this way. For example, as of March 2010, only 2,825 articles (less than 0.1%) in English Wikipedia are marked as featured. Another difficulty of the user–driven evaluations is that Wikipedia content is, by its nature, highly dynamic and the evaluations often become obsolete rather quickly.

Due to these conditions, recent research work involves automatic quality analysis of Wikipedia (Javanmardi, Lopes, and Baldi 2010; Adler and de Alfaro 2007; Blumenstock 2008; Cross 2006; Dondio and Barrett 2007; Lih 2004; Liu et al. 2008; B. Stvilia and Gasser 2005; Zeng et al. 2006; Whner and Peters 2009). Cross (Cross 2006) proposes a system of text coloring according to the age of the assertions in a particular article; this enables Wikipedia users to see what assertions in an article have survived after several edits of the article and what assertions are relatively recent and thus, perhaps, less reliable. Adler *et al.* (Adler et al. 2008) quantify the reputation of users according to the survival of their edit actions; then they specify ownerships of different parts of the text. Finally, based on the reputation of the user, they estimate the trustworthiness of each word. Javanmardi *et al.* in (Javanmardi, Lopes, and Baldi 2010) present a robust reputation model for wiki users and show that it is not only simpler but also more precise compared to the previous work.

Other research methods try to assess the quality of a Wikipedia article in its entirety. Lih (Lih 2004) shows that there is a positive correlation between the quality of an article and the number of editors as well as the number of revisions. Liu et. al. (Liu et al. 2008) present three models for ranking Wikipedia articles according to their level of accuracy. The models are based on the length of the article, the total number of revisions and the reputation of the authors, who are further evaluated by their total number of previous edits. Zeng *et al.* (Zeng et al. 2006) compute the quality of a particular article revision with a Bayesian network from the reputation of its author, the number of words the author has changed and the quality score of the previous version. They categorize users into several groups and assign a static reputation value to each group, ignoring individual user behavior.

Stvilia et. al. (B. Stvilia and Gasser 2005) have constructed seven complex metrics using a combination of them for quality measurement. Dondio *et al.* (Dondio and Barrett 2007) have derived ten metrics from research related to collaboration in order to predict quality. Blumenstock (Blumenstock 2008) investigates over 100 partial simple metrics, for example the number of words, characters, sentences, internal and external links, etc. He evaluates the metrics by using them for classifications between featured and non–featured articles. Zeng *et al.*, Stvilia *et al.* and Dondio *et al.* used a similar method which enables the evaluation results to be compared. Blumenstock demonstrates, with an accuracy of classification of 97%, that the number of words is the best current metric for distinguishing between featured and non–featured articles. These works assume that featured articles are of much higher quality than non–featured articles, and recast the problem as a classification issue. Wohner and Peters (Whner and Peters 2009) suggest that, with improved evaluation methods, these metrics–based studies enable us to determine the accuracy of various submissions. Studying German Wikipedia, they believe that a significant number of non–featured articles are also highly accurate and reliable. However, this category includes a large number of short articles. Their study of German Wikipedia from January 2008 shows that about 50% of the articles contain less than 500 characters, and thereby they assume that some short non–featured articles are of high quality, since their subject matter can be briefly but precisely explained.

In addition, we and others (D. Wilkinson D 2007; Whner and Peters 2009) assume that when an article is marked as featured and is displayed on the respective pages, it attracts many more web users for contributions and demands more administrative maintenance. Wohners and Peters' investigation on German Wikipedia(Whner and Peters 2009) reveals this assumption to be true. For example, over 95% of all articles are edited with greater intensity, once they are marked as featured. Wilkinson and Huberman (**?**), in a similar study on English Wikipedia, show that featured articles gain an in-

[5] http://bit.ly/4Bmrhz

[6] http://en.wikipedia.org/wiki/Wikipedia:Featured_articles

[7] http://bit.ly/dy3t1Y

crease in the number of edits and editors after being marked as featured. According to these observations, the accuracy of the classification in the related work ((Zeng et al. 2006; B. Stvilia and Gasser 2005; Dondio and Barrett 2007)) will be valid only if featured articles are considered before they are marked as featured.

Data Set

Most of the content analysis research on the evolution of articles (like those enumerated in Section and our own work) require the full text of all revisions of articles. We have monitored the publicly available English Wikipedia dumps[8] since early 2006 with the last successful dump released in October 2007. Because of the exponential growth of Wikipedia, all of the history dumps have failed since then. Since the last dump data set is quite out–dated, we created a more recent data set which is now publicly available[9]. We used the Wikipedia API[10] to get the full text of all the submitted revisions in the history of Wikipedia. The API has a limit of 50 revisions per request and, since these types of requests are not frequent, the chance of having a cached version is slim which makes the process of fetching data expensive. On average, it takes more than one second for the server to send back the result for each request. In addition, we needed to compare the text of subsequent revisions in order to extract the edits made in a revision. This process is also computationally expensive. In order to maintain a reasonable processing speed and still remain polite to Wikipedia servers, we used a cluster of ten nodes which downloaded and processed the whole history of English Wikipedia from July through August 2009. A master node assigned articles to client nodes and waited for them to download and process the article history and send back the extracted statistics.

As of March 2010, English Wikipedia contains about 3.2M articles [11]. However, some portion of these articles are isolated stubs that are not referenced by any other article. In our analysis, we used crawler4j[12] to crawl the entire English Wikipedia and extract a list of articles accessible through links on the English Wikipedia home page[13]. We also ignored articles that were redirected to other articles. We ended up with a set of 2.2M articles. Then we downloaded the revisions of these articles through Wikipedia API which resulted in 130M revisions[14].

Measuring the Quality Evolution in Articles

Since Wikipedia is a dynamic system, the articles can change very frequently. Therefore, the quality of articles is a time–dependent function and a single content may contain high– and low–quality content in different periods of its lifetime. The goal of our study is to analyze the evolution of content in articles over time and estimate the fraction of time that articles are in high–quality state.

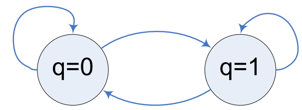

Figure 1: Transitions between high quality and low quality states

In our analysis of the evolution of the content quality in Wikipedia articles, we divide revisions to low– and high–quality revisions. Based on this assumption, an article can be in *low quality* ($q = 0$) or *high quality* ($q = 1$) states. In order to assess the quality q of a revision, we take into account two factors: the reputation of the author and whether this revision has been reverted in one of the subsequent revisions or not. The reputation of a contributor is a value between 0 and 1 and can be viewed as the probability that he produces a contribution of high–quality. This probability is computed based on the stability of the past contributions of the user using the methods developed in (Javanmardi and Lopes 2007; Javanmardi, Lopes, and Baldi 2010). The heuristic behind this reputation assessment is that high–quality contributions tend to survive longer in the articles as compared to low–quality contributions. This heuristic is also supported by other work (Hu et al. 2007; Adler and de Alfaro 2007).

As Figure 1 suggests, submission of a new revision can keep the state of the article or move it to the other state. If the revision is reverted later in the article history, we consider the new state of the article to be $q = 0$. Otherwise, if the reputation of the author of that revision is r, then with probability of r the new revision will be $q = 1$ and with probability of $1 - r$ the new revision will be $q = 0$.

With all these elements in place, we define $Q(T)$ as the ratio of high quality revisions submitted for the article up to time T:

$$Q(T) = \sum_{i=1}^{n} q(t_i)/n \tag{1}$$

where $q(t_i)$ is the quality of the revision submitted at time t_i and n is the total number of revisions up to time T. Figure 2 shows the distribution of $Q(T)$ for both all featured articles and a non–featured articles. While the average of $Q(T)$ is relatively high for both featured and non–featured articles, it is higher for featured articles –74% vs. 65%.

To estimate the proportion of time during which an article is in a high–quality state, we also define the duration $QD(T)$ by:

$$QD(T) = \frac{\sum_{i=1}^{n}(t_{i+1} - t_i)q(t_i)}{T - t_1} \tag{2}$$

The distribution of $QD(T)$ for both featured and non–feature articles are shown in Figure 3. Figure 4 also shows the average and standard deviation of $Q(T)$ and $QD(T)$ for

[8] http://download.wikimedia.org/enwiki/

[9] http://nile.ics.uci.edu/events-dataset-api/

[10] http://en.wikipedia.org/w/api.php

[11] http://en.wikipedia.org/wiki/Special:Statistics

[12] http://code.google.com/p/crawler4j/

[13] http://en.wikipedia.org/wiki/Main_Page

[14] The dataset is publicly available at http://nile.ics.uci.edu/events-dataset-api/

Figure 2: Distribution of Q(T) for featured and non–featured articles

Figure 3: Distribution of QD(T) for featured and non–featured articles

both featured and non–featured articles. Featured articles on average contain high–quality content 86% of the time. Interestingly, this value increases to 99% if we only consider the last 50 revisions of the articles. The same statistics for non–featured articles show that they have high–quality content 74% of the time. The difference between the averages of $Q(T)$ and $QD(T)$ suggests that typically low–quality content has short life span. This result is consistent with other studies reporting the rapid elimination of vandalism in Wikipedia (Viégas, Wattenberg, and Dave 2004; Kittur et al. 2007; Magnus 2008). For example, (Magnus 2008) reported that about one third to one half of the systematically inserted fictitious claims in Wikipedia are corrected within 48 hours.

Figure 5 shows the evolution of $QD(T)$ as a function of T for both featured and non–featured articles of the same age. Overall, $QD(T)$ tends to increase with T and its standard deviation decreases gradually.

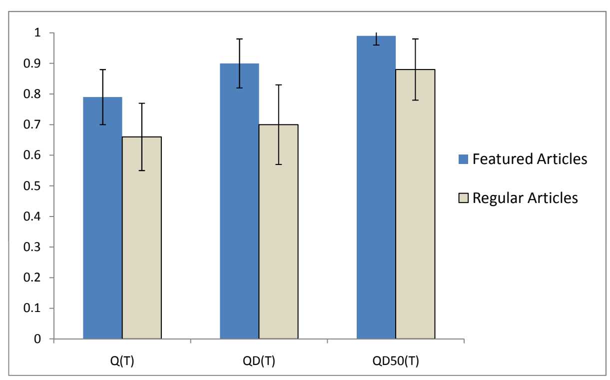

Figure 4: Average article quality for featured articles and non–featured articles. Quality is assessed by the average and the standard deviation of Q, QD, and $QD50$ for featured and non–featured articles. For each article, Q is the ratio of high–quality revisions. QD is the amount of time that an article spends in its high–quality state computed over its entire lifetime. $QD50$ is the value of QD when only considering the last 50 revisions of the article.

An Empirical Study of Wikipedia Statistics

In this section, we present an empirical study of Wikipedia statistics that may explain the results in Section . First we analyze user attribution in Wikipedia and compare behavior of anonymous and registered users. Second, we compare the evolution of content in featured and non–featured articles to see which parameters result in higher quality in featured articles.

Anonymous vs. Registered Users

Wikipedia users can contribute to wiki pages both anonymously or as registered users. Registered users are identified by their usernames, while anonymous users are tracked by

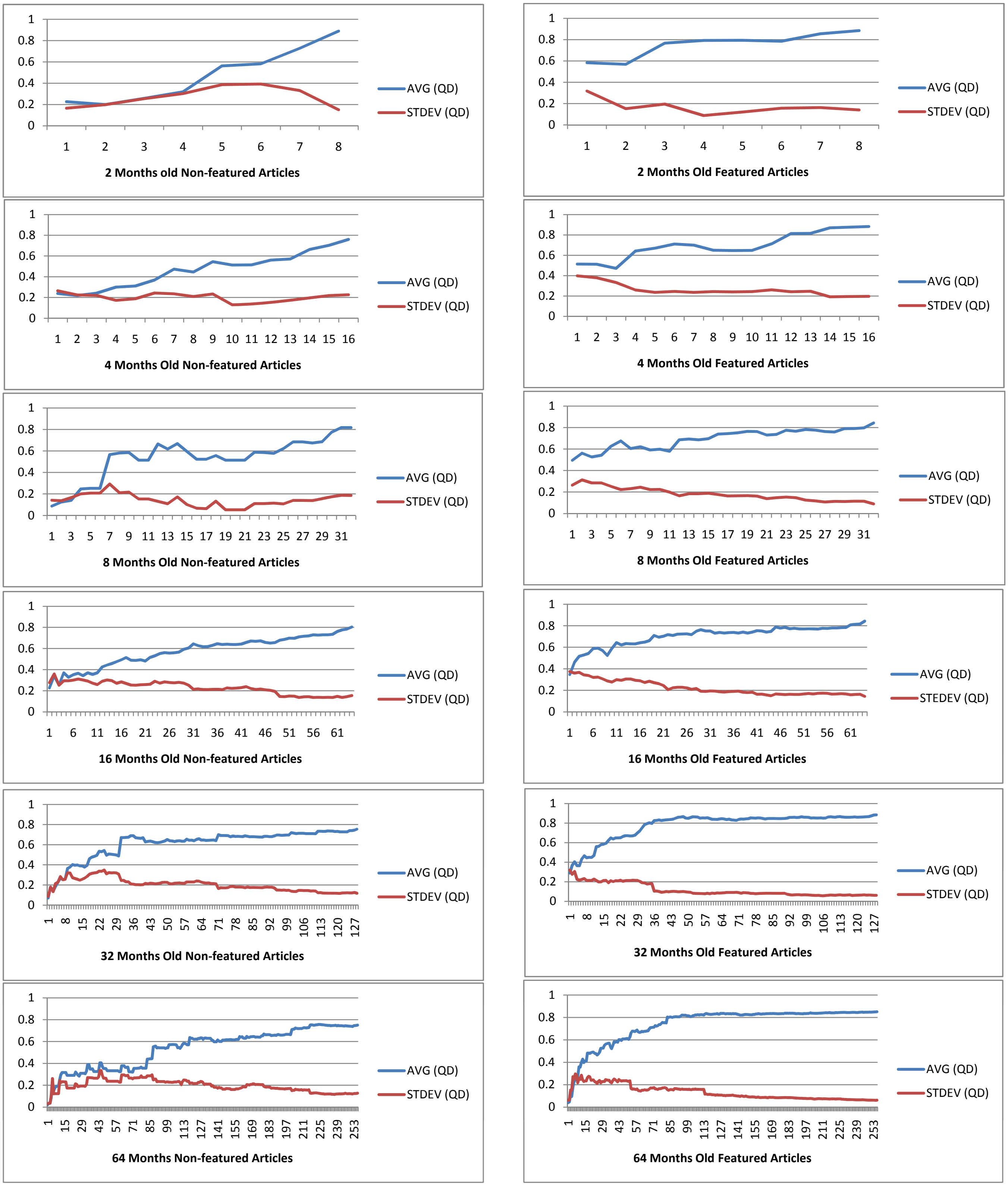

Figure 5: Evolution of article quality over time for same–age articles in Wikipedia.

their IP addresses[15]. Although there is no one–to–one correspondence between people and accounts or IP addresses, Wikipedia uses usernames or IP addresses to track user behavior for further promotions (*e.g.* admin assignment) or demotions (*e.g.* user block). To investigate the effect of open editing model of Wikipedia, we compare the behavior of anonymous and registered users to see if there is any correlation between registration and quality of the contributed content. We, as well as others (Zeng et al. 2006; Viégas, Wattenberg, and Dave 2004; Ekstrand and Riedl 2009), follow the same nomenclature as Wikipedia: a "user"in this study refers to a registered account or an IP address, and it does not refer to a real–world individual.

Wikipedia keeps the past revisions of articles and these revisions are accessible through history pages of articles. These history pages can be mined in order to analyze the behavior of registered and anonymous users in Wikipedia. Our first attempt to compare the behavior of anonymous and registered users was based on the revert actions done in Wikipedia articles. A revert is the action of undoing all changes made to an article, restoring it to what it was at a specific time in the past. According to the Wikipedia revert policy [16], reversion is used primarily to fight vandalism or similar activities such as spamming. Our study on all English Wikipedia reverts show that 96% of the reverts are done by registered users, while most of the reverted revisions are associated with anonymous users. Furthermore, in 73% of the time a revert restores the current revision of an article to a recent revision submitted by a registered user.

In order to have a more fine–grained analysis of the user behavior, we compared the text of consecutive revisions to extract the insertions and deletions made by each user in each revision. The granularity of inserts and deletes is measured in terms of single tokens (words). The results show that 60.6% of the total inserted content is contributed by registered users. We also followed the evolution of articles and extracted the contributions made to each article over time. Using this method, we were able to determine the contributor of each single token in the last revision of each article. The results show that 84% of the current content of Wikipedia articles (i.e. survived content in the latest revisions of the articles) has been contributed by registered users. Another interesting observation shows that 49.4% of the contributed content by registered users has been deleted over time while this value is 85.2% for anonymous users. These observations show the high dynamics in the evolution of content in Wikipedia and the higher stability of the registered contributions.

Comparison of the distribution of the reputation for anonymous and registered users clearly shows that registered users tend to have higher reputation. The average of reputation for registered users (as measured in (Javanmardi and Lopes 2007; Javanmardi, Lopes, and Baldi 2010)) is 59% while this is 49% for anonymous users (Javanmardi et al. 2009). Furthermore, 70% of the reverted revisions (vandalistic content) are associated with anonymous users. Together these results suggest that user registration has a positive affect on the quality of Wikipedia.

Featured vs. Non–featured Articles By comparing content evolution in both featured and non–featured articles, we aim to find what parameters affect content quality and how the open editing model of Wikipedia lets featured articles attain high–quality. In (D. Wilkinson D 2007; Ganjisaffar, Javanmardi, and Lopes 2009), authors have compared featured and non–featured articles and concluded that, on average, featured articles benefit from higher number of edits and distinct editors. In order to have a more detailed comparison between featured and non–featured articles, we examined the evolution of content in these articles and extracted the statistics summarized in Table 1.

Though 39.1% of the total inserted tokens in non–featured articles are contributed by anonymous authors, this figure drops to 15.2% when only the last revisions are considered. In the case of featured articles, the total percentage of inserted tokens by anonymous authors is 56.3%, with this figure dropping to 7.8% in last revisions. According to these statistics, most of the remaining content in both featured and non–featured articles belong to registered users, but this percentage is higher in featured articles. This observation, together with the result of Section provides strong evidence for why featured articles contain higher quality content throughout their lifetime.

Furthermore, the token survival ratio presented in Table 1 shows a much higher turnover of the content in featured articles over time; a higher ratio of tokens is deleted in featured articles compared to other articles. This might be counterintuitive as one might expect that the content inserted in the featured articles should be of higher quality and thereby more stable. However, it can be interpreted as higher dynamics in the evolution of the content in these articles that only allows very high quality content to survive. Note that in order to control the increased visibility and attention that articles might gain after being marked as featured, we have also reported the results of our analysis both before and after becoming featured.

Featured articles can also be distinguished from other articles in terms of proportion of reverted revisions. While on average, 9.9% of the revisions in non–featured articles are reverted, it becomes 25.4% after an article becomes featured. This significant increase in the ratio of reversions after articles are marked as featured is a matter of further study; this can be due to more vandalism as a consequence of higher visibility or it might be attributed to the fact that most of the featured articles have become mature and thus more resistant to change.

In summary, we conclude that (a) featured articles are more closely followed: although less than 0.08% of the articles are marked as featured, they comprise about 1.4% of the total number of revisions; (b) Wikipedia administrators contribute more actively to featured articles even before these articles are marked as featured; (c) the revert ratio in featured articles is about 1.8 times higher than the ratio for non–featured articles; (d) featured articles have a much higher

[15] http://en.wikipedia.org/wiki/Why_create_an_account

[16] http://en.wikipedia.org/wiki/Wikipedia:Revert_only_when_necessary

	Featured Articles	Non–featured Articles
Total Inserted Tokens		
Registered	43.7% (before: 50.9%, after: 36.7%)	60.9%
Anonymous	56.3% (before: 49.1%, after: 63.3%)	39.1%
Tokens in the last revisions		
Registered	92.2% (before: 93.1%, after: 88.2%)	83.9%
Anonymous	7.8% (before: 6.9%, after: 11.8%)	16.1%
Token Survival		
Registered	23.1% (before: 33.8%, after: 8.7%)	50.9%
Anonymous	1.5% (before: 2.6%, after: 0.7%)	15.2%
Ratio of Reverted Revisions	17.8% (before: 11.4%, after: 25.4%)	9.9%
Ratio of Revisions Submitted by Admins	17.4% (before: 20.7%, after: 14.2%)	10.9%

Table 1: Statistical Comparison of Featured and Non–featured Articles. The statistics for featured articles consider revisions submitted before and after the articles were marked as featured

turnover of content. This higher dynamic in the article's evolution allows very high quality content to survive. It is interesting to note that even at this lower survival rate, featured articles are on average longer than other articles (Blumenstock 2008). Overall, these statistics support the view that featured articles benefit from a higher degree of supervision as compared to other articles.

Conclusion & Future Work

Wikipedia is a highly dynamic environment and the quality of its articles can change over time as they go through iterations of edits by different users. In this study, we analyzed the evolution of content in English Wikipedia articles and showed that non–featured articles tend to have high–quality content 74% of their lifetime and this is 86% for featured articles. Furthermore, we showed that the average article quality increases as it go through edits while its standard deviation decreases. To analyze the parameters affecting quality, we compared the behavior of anonymous and registered users and showed that there is a positive correlation between user registration and quality of the contributed content. Furthermore, we examined the evolution of content in both featured and non–featured articles. Higher administrator submissions, higher reversions and higher content turnover are some of the parameters that result in the improvement of featured articles. According to these observations, we can conclude that the open editing model of Wikipedia is not a barrier to quality, its higher user participation rate providing a higher turnover of content, which is, after all, a major prerequisite for high–quality crowdsourcing systems.

In this work, we modeled the quality of an article revision using reputation of the editor and revert status of the revision. Although our study of featured and non–featured articles showed the effectiveness of the model, there are some limitations. First, for a significant number of users in Wikipedia, we don't have enough edit patterns to estimate user reputation accurately. Such data sparsity can degrade accuracy of reputation for these users. Another feature that the current model uses is the revert status of the revision. Reverts are extracted based on the history of the article and, for the purpose of predicting the revert status of a newly submitted revision, we need to estimate the likelihood of having vandalistic content in the current revision. In future work, we aim at extending the model in order to estimate the probability of being reverted for a newly submitted revision. To this aim, we use language model disagreement technique, which has been widely used for spam detection in blogs.

Acknowledgments

This work has been partially supported by NSF grant OCI-074806.

References

Adler, B. T., and de Alfaro, L. 2007. A content-driven reputation system for the wikipedia. In *WWW '07: Proceedings of the 16th international conference on World Wide Web*, 261–270. New York, NY, USA: ACM.

Adler, B.; K. Chatterjee, L. d. A.; Faella, M.; Pye, I.; and Raman, V. 2008. Assigning trust to wikipedia content. In *WikiSym '08: Proceedings of the 2008 international symposium on Wikis*.

B. Stvilia, M.B. Twidale, L. S., and Gasser, L. 2005. Assessing information quality of a community-based encyclopedia. In *In Proceedings of the International Conference on Information Quality*, 442–454.

Blumenstock, J. E. 2008. Size matters: word count as a measure of quality on wikipedia. In *WWW '08: Proceedings of the 17th international conference on World Wide Web*, 1095–1096. ACM.

Chesney, T. 2006. An empirical examination of wikipedia's credibility. *FirstMonday* 11(11).

Cross, T. 2006. Puppy smoothies: Improving the reliability of open, collaborative wikis. *First Monday* 11(9).

D. Wilkinson D, B. A. H. 2007. Assessing the value of cooperation in wikipedia. *First Monday* 12(4).

Dondio, P., and Barrett, S. 2007. Computational trust in web content quality: A comparative evaluation on the wikipedia project. *Informatica: An International Journal of Computing and Informatics* 31(2):151–160.

Ekstrand, M., and Riedl, J. 2009. rv you're dumb: Identifying discarded work in wiki article history. In *WikiSym '09 Proceedings of the 2009 International Symposium on Wikis*. ACM.

Ganjisaffar, Y.; Javanmardi, S.; and Lopes, C. 2009. Using user reviews to improve search in wikipedia. In *SIGIR 2009 Workshop on Search in Social Media*.

Giles, J. 2005. Internet encyclopaedias go head to head. *Nature* 438:900–901.

Hu, M.; Lim, E.; Sun, A.; Lauw, H.; and Vuong, B. 2007. Measuring article quality in wikipedia: models and evaluation. In *CIKM '07: Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, 243–252. ACM.

Javanmardi, S., and Lopes, C. 2007. Modeling trust in collaborative information systems. In *CollaborateCom '07: Proceedings of the 3rd International Conference on Collaborative computing: Networking, Applications and Worksharing*. New York, NY, USA: IEEE.

Javanmardi, S.; Ganjisaffar, Y.; Lopes, C.; and Baldi, P. 2009. User contribution and trust in wikipedia. In *Proceedings of the 5th International Conference on Collaborative computing: Networking, Applications and Worksharing*.

Javanmardi, S.; Lopes, C.; and Baldi, P. 2010. Mining wikipedia to extract user reputation. *Journal of Statistical Analysis and Data Mining (accepted for publicatoin)*.

Kittur, A.; Suh, B.; Pendleton, B. A.; and Chi, E. H. 2007. He says, she says: conflict and coordination in wikipedia. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, 453–462. ACM.

Lih, A. 2004. Wikipedia as participatory journalism: Reliable sources? metrics for evaluating collaborative media as a news resource. In *Proceedings of the 5th International Symposium on Online Journalism*.

Liu, H.; Lim, E.; Lauw, H.; Le, M.; Sun, A.; Srivastava, J.; and Kim, Y. A. 2008. Predicting trusts among users of online communities: an epinions case study. In *EC'08: Proceedings of the 9th ACM conference on Electronic commerce*, 310–319. ACM.

Magnus, P. D. 2008. Early response to false claims in wikipedia. *First Monday* 13(9).

Tapscott, D., and Williams, A. 2006. *Wikinomics: How Mass Collaboration Changes Everything*. Penguin Group. 70–77.

Viégas, F.; Wattenberg, M.; and Dave, K. 2004. Studying cooperation and conflict between authors with history flow visualizations. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, 575–582. ACM.

Waldrop, M. 2008. Editorial. big data: Wikiomics. *Nature News* (455):22–25.

Whner, T., and Peters, R. 2009. Assessing the quality of wikipedia articles with lifecycle. In *WikiSym '09 Proceedings of the 2009 International Symposium on Wikis*. ACM.

Zeng, H.; Alhossaini, M.; Ding, L.; Fikes, R.; and McGuinness, D. L. 2006. Computing trust from revision history. In *Proceedings of the 2006 International Conference on Privacy, Security and Trust*.

Zittrain, J. 2008. *The Future of the Internet–And How to Stop It*. Yale University Press.