Publiziert in: Alt et al. (Hrsg.), Tagungsband 15. Interuniversitäres Doktorandenseminar Wirtschaftsinformatik der Universitäten Chemnitz, Dresden, Freiberg, Halle-Wittenberg, Jena und Leipzig, Leipzig, 2011, S. 45-52.

Automatic Editing Rights Management in Wikipedia

Thomas Wöhner

Martin-Luther-Universität Halle-Wittenberg, Universitätsring 3, 06108 Halle

thomas.woehner@wiwi.uni-halle.de

Abstract: The free online encyclopedia Wikipedia is one of the most successful collaborative web projects. It is based on an open editing model, which allows everyone to edit the articles directly in the web browser. As a result of the open concept, undesirable contributions like vandalism cannot be ruled out. These contributions reduce the article quality temporarily, consume system resources and cause effort for correcting. To address these problems, this paper introduces an approach for automatic editing rights management in Wikipedia that assigns editing rights according to the reputation of the author and the quality of the article to be edited. The analysis shows that this approach reduces undesirable contributions significantly while valuable contributions are nearly unaffected.

1 Introduction

The free online encyclopedia Wikipedia is one of the most successful and famous collaborative web projects. It consists of more than 17 million encyclopedic articles in more than 270 languages overall [Wikipedia 2011]. The main characteristic of Wikipedia is the wiki concept, which allows everyone to edit the articles directly in the web browser [Cunningham/Leuf 2001]. Contributions are published without any control [Wikipedia 2011]. On the one hand, this open editing model attracts many voluntary web users who maintain and update the content of Wikipedia. On the other hand, undesirable edits like vandalism or spam cannot be ruled out [Denning et al. 2005]. Such edits harm Wikipedia in different ways: firstly, the article quality is reduced temporarily, secondly, system resources are consumed unnecessarily and thirdly, effort for correcting the articles is required. To avoid undesirable edits, Wikipedia administrators are able to ban malicious users or protect frequently vandalized articles from editing. However, these methods do not prevent undesirable edits effectively. A detailed analysis of the editing process in Wikipedia shows that about one third of the contributions in Wikipedia is short-lived and therefore can be judged as undesirable.

For that reason, many scientific papers have addressed problems with the quality of Wikipedia's content over the past few years. Some of these publications are involved in automatic detection of vandalism in Wikipedia (e.g. [Potthast et al. 2008], [Priedhorsky et al. 2007], [Smets et al. 2008], [Viégas et al. 2004]) These approaches can be used by Wikipedia bots to revert vandalism automatically. Further publications focus on the development of metrics for an automatic quality assessment to improve the quality transparency in Wikipedia (e.g. [Blumenstock 2008], [Dondio/Barrett 2007], [Lih 2004], [Lim 2006], [Wöhner/Peters 2009], [Zeng et al 2006]). [Javanmardi et al. 2010], [Adler et al. 2007] as well as [Wöhner et al. 2011] analyze approaches for an automatic reputation assessment in order to evaluate the editing behavior of Wikipedia authors. However, the main focus of currently known approaches is the detection and the revision of vandalized articles, but they are not suitable to prevent undesirable edits.

Therefore, in this paper, a models for an automatic editing rights management (ERM) to prevent undesirable contributions is introduced. The editing rights are assigned according to the reputation of authors and the quality of articles to be edited. The approach is evaluated by means of a simulation on the basis of the Wikipedia dataset.

The rest of the paper is organized as follows. Section 2 introduces the formal notation that is used to model Wikipedia and describes an approach to detect undesirable contributions. Section 3 provides the models for ERM and Section 4 evaluates this approach. The evaluation includes a description of the evaluation method and a discussion of the results. Finally, Section 5 presents a conclusion of the findings and shortly discusses the future work.

2 Transient and Persistent Contributions

In this paper undesirable contributions to Wikipedia articles are identified by a new approach that is based on the discrimination between persistent and transient contributions. This approach has been already discussed in a previous study in more detail [Wöhner et al. 2011]. Persistent contributions outlast a significant period of time Δt without being reverted. Therefore, it can be assumed that these contributions are accepted by the Wikipedia Community and improve the article quality. In contrast, transient contributions are rejected by the Community within the same period of time Δt. Hence, the contributions can be assessed as undesirable and do not contribute to the advancement of Wikipedia. Since contributions such as vandalism or spam are generally reverted within a very short period of time (three minutes) [Viégas et al. 2004], these edits are captured by the transient contributions.

The calculation method for persistent and transient contributions is exemplified in Figure 1.

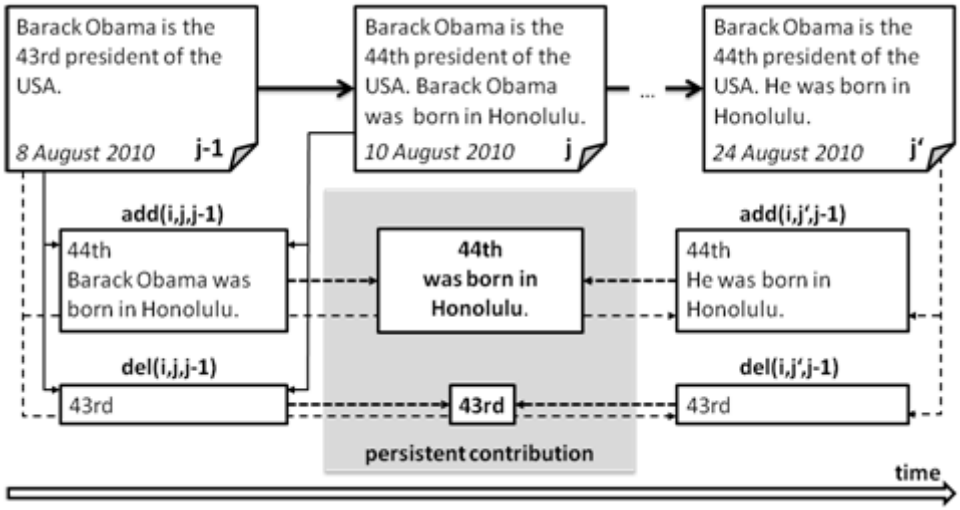

Figure 1: Determination of persistent contributions

The computation is based on different text comparisons. In order to calculate text differences the common Hunt and McIlroy algorithm [Hunt/McIlroy 1975] is used on word level. In this context a word denotes the sequence of characters between two whitespaces. At first, for a given version *j* of article *i* both the deleted text *del(i,j,j-1)* and the added text *add(i,j,j-1)* in comparison to the predecessor version *j-1* is calculated. Subsequently, the difference texts *del(i,j',j-1)* and *add(i,j',j-1)* are determined, where *j'*

is the most recent successor that exists after the time interval Δt. These two difference texts describe the modification of the article within Δt.

Afterwards, the common texts between *add(i,j,j-1)* and *add(i,j',j-1)* as well as between *del(i,j,j-1)* and *del(i,j',j-1)* is determined. This comparison calculates the portion of the contribution which is contained in the total article modifications done within Δt. This portion is classified as persistent and the remaining portion as transient. The number of characters of the persistent contribution is referred to as *pers(i,j)*, while *trans(i,j)* denotes the number of characters of the transient contribution. The efficiency *eff(i,j)* describes the percentage of persistently changed characters in version *j* of article *i*.

In the experiments the time interval Δt of two weeks is used. An empirical analysis described in a previous publication shows that the time interval of two weeks is the most suitable one [Wöhner et al. 2011].

3 Editing Rights Management

The goal of ERM is to decide automatically whether or not a given author is permitted to edit a given article. The intention is to block transient contributions while persistent contributions are permitted. However, both sub-goals cannot be simultaneously achieved perfectly, since most of the edits in Wikipedia comprise persistent as well as transient contributions. A tradeoff between blocked transient contributions and permitted persistent contributions can be achieved by parameterizing the ERM. The suggested approach for ERM combines two models: the reputation based ERM and the quality based ERM. Both models are deduced from findings of previous studies ([Wöhner/Peters 2009] and [Wöhner et al. 2011])

3.1 Reputation based Editing Rights Management

Reputation based ERM assigns editing rights based on the reputation of the author. Hence, according to the knowledge about the author's reputation it can be distinguished between the informed reputation based ERM and the uninformed reputation based ERM.

3.1.1 Informed Reputation based Editing Rights Management

The informed reputation based ERM is employed for all authors who have already contributed to Wikipedia before and thus their reputation can be assessed by means of their previous edits. A significant metric for reputation assessment is the average efficiency avg^{eff}_{a} of the contributions done by the given author *a* [Wöhner et al. 2011]. Therefore, within the informed reputation based ERM an author *a* is blocked if avg^{eff}_{a} is less than a given threshold efficiency eff^{t}. The parameter eff^{t} defines the strictness of the ERM and determines the relation between the permitted persistent contributions and the blocked transient contributions.

3.1.2 Uninformed Reputation based Editing Rights Management

Numerous edits in Wikipedia are performed by authors whose reputation is unknown. This includes anonymous edits as well as the first edits of registered users. For these contributions, the editing rights are deduced from the number of characters to be deleted $|del(i,j)|$ and the number of characters to be added $|add(i,j)|$.

Rules for the uninformed reputation based ERM can be identified by using machine learning algorithms on the basis of edits with an already known efficiency. In this work, the rule-based classification algorithm Repeated Incremental Pruning to Produce Error Reduction (RIPPER) [Cohen 1995] is applied to classify low-efficient *(eff<eff$^{\tau}$)* and high-efficient edits *(eff >=eff$^{\tau}$)*. In this way, typical patterns of |*del(i,j)*| and |*add(i,j)*| of low efficient edits are identified. Based on these patterns the editing rights are assigned.

However, the obtained rules depend on the defined strictness of the ERM (*eff$^{\tau}$*) and cannot be defined generally.

3.2 Quality based Editing Rights Management

An isolated use of the reputation based ERM results in a very strict assignment of editing rights, since authors whose first edit is of low efficiency are excluded from further contributions. To overcome this problem the reputation based ERM is combined with the quality based ERM. The quality based ERM defines editing rights with respect to the quality of the article. While low-quality articles can be edited without any restrictions, the editing rights for high-quality articles are assigned by means of the reputation based ERM.

The advantage of this approach is that low-reputation authors can improve their reputation by contributing to low-quality articles. In that case, potential undesirable edits cause only less harm since the articles are already of low quality. Furthermore, a previous study shows that with increasing maturity and quality of articles, transient contributions become more frequent since the acceptance for new contributions within the Wikipedia community declines [Wöhner/Peters 2009]. Hence, the editing rights for this relevant part of Wikipedia are controlled by the quality based ERM.

A previous study on quality assessment shows that the sum of the persistent contributions $sum^{pers}{}_i$ of a given article *i* is a suitable quality indicator [Wöhner/Peters 2009] that can be used for ERM. Thus, using the quality based ERM, an article *i* can be edited without restrictions if $sum^{pers}{}_i$ is less than a given threshold $sum^{\tau,\ pers}{}_i$; otherwise the editing rights are assigned by the reputation based ERM.

Figure 2 summarizes the complete set of rules that combines the quality based and the reputation based ERM.

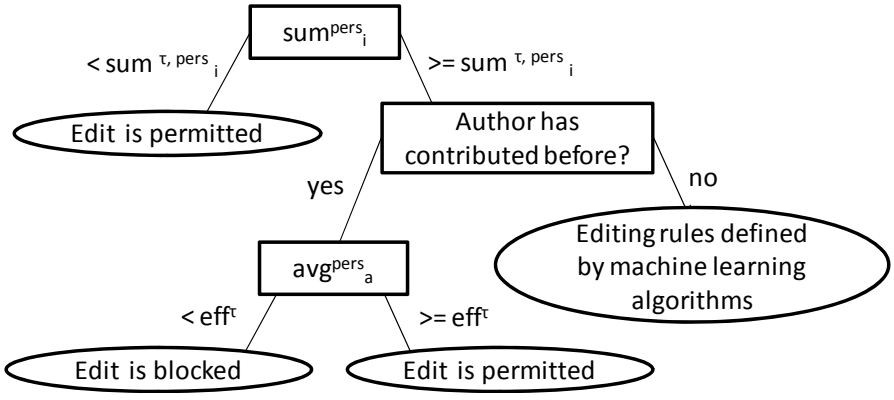

Figure 2: Decision tree for ERM

4 Evaluation

In this section, the evaluation method is described first and subsequently the results of the evaluation of the described approach for ERM are discussed.

4.1 Evaluation Method

The proposed ERM has been evaluated by means of a simulation on the basis of the data of the German Wikipedia from *21st January 2008*. To reduce the complexity of the simulation a small representative data sample of Wikipedia has been selected. The complete E-Business category including all sub-categories has been choosen. This data sample contains *258* articles and *22,040* article versions in total. *9,610* edits are performed by anonymous users; the remaining edits are done by *2,889* registered users. Since a complete category is used, it can be assumed that articles are edited by a quite closed user community so that the number of authors with unknown reputation is reduced.

The evaluation is performed in several steps. Firstly, for all article versions *i,j* of the data sample *pers(i,j)* and *trans(i,j)* is calculated as described above. Subsequently, the development of the articles within the data sample is simulated under the assumption that ERM is employed . For this purpose, all article versions are sorted by the creation date and it is decided sequentially whether or not a given edit is permitted according to the ERM. While doing so, the respective average efficiency of the authors $avg^{pers}{}_{a}$ and the sum of the persistent contributions of the articles $sum^{pers}{}_{i}$ is calculated continuously to determine both the current author reputation and the article quality. This calculation does not consider edits that were blocked by the ERM. According to the amount of the permitted persistent contribution sum^{pers} and the permitted transient contribution sum^{trans} within the simulation the effectiveness of the ERM can be assessed.

For the evaluation the ERM is parametrized with the aim to block edits with an efficiency *eff(i,j)* of less than *50%*. The remaining edits should be permitted. Accordingly, the threshold efficiency is defined as $eff^{*} = 0.5$.

The computation of the rule set for the uninformed reputation based ERM should not be based on the same data sample that is used for the evaluation of the ERM. Otherwise, there is the risk of overfitting. Therefore, the total available dataset is used to randomly select *10.000* edits of low efficiency and *10.000* edits of high efficiency done by anonymous authors or registered users performing their first edit. The classification rules obtained by the RIPPER algorithm, presented in Table 1, are applied for the uninformed reputation based ERM.

	Condition	Decision
1	$\|del(i,j)\|<10$ & $\|add(i,j)\|>13$	Blocking
2	$\|del(i,j)\| > 481$ & $\|add(i,j)\| < 305$	Blocking
3	Else	Permitting

Table 2: Rules for uninformed reputation based ERM

Within the quality based ERM, a threshold quality of $sum^{\tau,pers}{}_{i} = 5,000$ is used for each article. This parameter has been determined experimentally as follows: the simulation described above is rerun, ceteris paribus; and $sum^{pers}{}_{i}$ is increased by *500* in each run,

starting at $sum^{\tau,pers}{}_i = 500$ to $sum^{\tau,pers}{}_i = 50,000$. We obtained the maximum difference of sum^{trans} and sum^{pers} at $sum^{\tau,pers}{}_i = 5,000$.

In a practical implementation in Wikipedia this parameter would be defined in advance and could be estimated on the basis of past experience. Another more precise option is to apply a user driven approach. Thus, for example, administrators could decide for each article separately when it should be controlled by ERM. However, such an approach can be evaluated in the live system of Wikipedia only.

4.2 Results and Discussion

Table 2 presents the results of the simulation and compares the original development of the articles in Wikipedia with the development applying ERM.

	Original	ERM
sum^{pers}	6,250,896	5,475,196 (87.6%)
sum^{trans}	3,288,820	999,828 (30.3%)

Table 3: Persistent and transient contributions using ERM

The analysis shows that the ERM effectively prevents undesirable edits. *70%* of the transient contribution is prevented, whereas only *13%* of the persistent contribution is blocked. The results indicate that articles evolve in a similar way applying ERM, whereas undesirable edits are reduced considerably and therefore the damage for Wikipedia is limited.

Table 3 illustrates the accuracy of the ERM and shows how low efficient and high efficient edits are assessed by the ERM. In total, *77%* of the edits are assessed correctly by the ERM. Especially high efficient edits are predicted correctly. The true positive rate amounts to *88.3%*. In contrast, the weakness of the suggested ERM is the prediction of low efficient edits, since *63.8%* are permitted falsely (false positive rate).

	Blocked	Permitted
Low efficient	1,719	3,032
High efficient	2,027	15,262

Table 4: Accuracy of ERM

However, a detailed analysis shows that especially short contributions with an efficiency of just under *50%* are misclassified. Large contributions with a particularly low efficiency are classified correctly with a higher probability. Thus, *311* of the *471* edits that consist of more than *500* characters and having an efficiency of less than *20%* are blocked correctly.

In the final analysis, the efficiency of blocked and permitted edits is compared. The average efficiency differs clearly. While blocked edits have an average efficiency of about *54.4%,* the efficiency of permitted edits amounts to *83.1%* averagely. This evaluation finally proves the effectiveness of the provided approach for ERM.

5 Conclusion

This paper first introduced two models for an automatic editing rights management in order to restrict undesirable edits in Wikipedia. The reputation based editing rights management assigns editing rights depending on the author's reputation which is assessed by the average efficiency of his previous contributions. The quality based editing rights management restricts the editing of high quality articles whereas low quality articles can be edited without any control. In this context, the quality of articles is assessed by the sum of persistent contributions to the article.

The proposed approach for editing rights management is evaluated by means of a simulation based on the dataset of Wikipedia. The evaluation shows the effectiveness of the approach for ERM; *70%* of the transient undesirable contributions are blocked, whereas only *13%* of the persistent valuable contributions are rejected falsely.

For future work, the proposed approach will be integrated into the MediaWiki Software used by Wikipedia. Using this software prototype, the effectiveness of the approach will be validated and the set of rules will be adjusted experimentally.

References

Adler, B.T. and Alfaro, L. A Content-Driven Reputation System for the Wikipedia, In Proc. WWW 2007, ACM Press (2007), 261-270.

Blumenstock, J.E. Size Matters: Word Count as a Measure of Quality on Wikipedia. In Proc. WWW 2008, ACM Press, 2008, 1095-1096.

Cohen, W. Fast Effective Rule Induction. In Proc. ICML, 1995, 115-123.

Cunningham, W. and Leuf, B., The Wiki Way. Quick Collaboration on the Web, 2001, Addison-Wesley, Boston.

Denning, P., Horning, J., Parnas, D. and Weinstein, L., Wikipedia risks", Communications of the ACM, (48:12), 2005, pp. 152-152.

Dondio, P. and Barrett, S., Computational Trust in Web Content Quality: A Comparative Evalutation on the Wikipedia Project, Informatica – An International Journal of Computing and Informatics, 2007 (31:2), pp.151-160.

Hunt, J. and McIlroy, M. An algorithm for differential file comparison. Computer Science Technical Report, 41, Bell Laboratories, 1975.

Javanmardi, S., Lopes, C. and Baldi, P. Modeling User Reputation in Wikipedia. Journal of Statistical Analysis and Data Mining, 3, 2, 2010, 126-139.

Lih, A. 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, Austin, TX, 2004.

Lim, E.P., Vuong, B.Q., Lauw, H.W. and Sun, A., Measuring Qualities of Articles Contributed by Online Communities, in Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence,j. liu and B. Wah (eds.), Washington, DC, 2006, pp. 81-87.

Potthast, M., Stein, B. and Gerling, R., Automatic Vandalism Detection in Wikipedia, in Proceedings of the Advances in Information Retrieval - 30th European Con-

ference on IR Research, C. Macdonald, I. Ounis, V. Plachouras, I. Rutheven and Ryan White (eds.), 2008, pp. 663-668.

Priedhorsky, R., Chen, J., Lam, S.K., Panciera, K., Terveen, L. and Riedl, J., Creating, Destroying, and Restoring Value in Wikipedia", in Proceedings of the 2007 international ACM conference on Supporting group work, T. Gross and K. Inkpen (eds.), New York, NY, 2007, pp. 259-268.

Smets, K., Goethals, B. and Verdonk, B., Automatic Vandalism Detection in Wikipedia: Towards a Machine Learning Approach", in Proceedings of the AAAI Workshop, Wikipedia and Artificial Intelligence: An Evolving Synergy, R. Bunescu, E. Gabrilovich and R. Mihalcea (eds.), 2008, pp. 43-48.

Viégas, F., Wattenberg, M. and Dave, K. Studying cooperation and conflict between authors with history flow visualizations. In Proc. CHI 2004, ACM Press, 2004, 575-582.

Wikipedia (eds.), Wikipedia: About. http://en.wikipedia.org/wiki/Wikipedia:About.

Wöhner, T. and Peters, R. X, Assessing the quality of Wikipedia articles with lifecycle based metrics. In Proc. WikiSym 2009, ACM Press, 2009.

Wöhner, T., Köhler, S., Peters, R., Automatic Reputation Assessment in Wikipedia. International Conference on Information Systems 2011, Shanghai, China, 2011

Zeng, H., Alhoussaini, M., Ding, L., Fikes R. and McGuinness, D., Computing trust from revision history, in Proceedings of the 2006 International Conference on Privacy, Security and Trust, New York, NY, 2006.