VANDALISM AND CONFLICT RESOLUTION IN WIKIPEDIA. AN EMPIRICAL ANALYSIS ON HOW A LARGE-SCALE WEB-BASED COMMUNITY DEALS WITH BREACHES OF THE ONLINE PEACE

Paper for the
IADIS International Conference
"Web-Based Communities and Social Media 2011"

Rome, Italy

22 - 24 July 2011

Dr. Thomas Roessing

Assistant Professor
Department of Communication Studies
University of Mainz
55128 Mainz
GERMANY
++49 6131 39-24290
roessing@uni-mainz.de

ABSTRACT

The paper discusses the proceedings on the anti-vandalism page of the German language version of the online encyclopedia Wikipedia. Research questions address the structure of vandalism reports, the distribution over time of day and the relationship between conflict potential and conflict resolution. A quantitative analysis of 500 vandalism reports reveals that the anti-vandalism page is a good indicator for conflicts within the community and its deficits in dealing with them.

KEYWORDS

Wikipedia, web-based community, vandalism, conflict, conflict resolution

VANDALISM AND CONFLICT RESOLUTION IN WIKIPEDIA. AN EMPIRICAL ANALYSIS ON HOW A LARGE-SCALE WEB-BASED COMMUNITY DEALS WITH BREACHES OF THE ONLINE PEACE

1. RESEARCH TOPIC

Wikipedia is an online project for the development of a freely accessible encyclopedia (O'Sullivan, 2009). The encyclopedic content is provided by a web-based community of volunteers. The two largest language versions of Wikipedia are English and German. The English version (http://en.wikipedia.org) has about 3.5 Million articles and an active (at least one edit within the last 30 days) community of 148,715 users. The German version features about 1.2 million articles and 27,975 active users.[1] Wikipedia is one of the most attractive platforms on the Internet regarding the number of visitors. It is ranked by the alexa.com traffic ranking system 8th worldwide and 6th in the US as well as in Germany.[2]

The user-generated encyclopedic content covers a wide range of topics, including many that deal with politics, religion, and other issues where different world views collide. It has been suggested that social-psychological mechanisms have an influence on how conflicts within Wikipedia's community are dealt with (Roessing, 2008). Conflicts regularly require administrative attention to prevent damage to the project and the quality of the encyclopedic content. Dispute must not be allowed to poison the teamwork of the community and extremists must be prevented from injecting biased content into articles on relevant issues. Wikipedia's elected administrators (also referred to as 'sysops'; for their role within the community see Konieczny, 2009) have thus to deal with a considerable amount of conflict regarding different interests in the structure and the content of the online encyclopedia. This task is an addition to one of the administrators' most important areas of work, the fight against vandalism (Potthast, Stein, & Gerling, 2008). Open projects, where everyone can add, remove, or alter content at any time, even without registering, naturally attracts villains and fools. As a result, there is a lot of misbehavior, comparable to what is called "breach of the peace" in the offline world. Wikipedia deals with conflicts and vandalism on several different subpages of the project. The central page for the mentioned problems is *Wikipedia:Administrator intervention against vandalism*[3] (*Wikipedia:Vandalensperrung* in the German version[4]). Vandals and conflicts that are getting out of control are reported on these pages and administrators evaluate the situation and take action to prevent damage to Wikipedia.

This paper analyzes proceedings of the German language version of Wikipedia with the aim to gain information on the community by looking at the way it deals with conflicts and deviant behavior. Wikipedia's communities work differently in different languages and cultures, but the overall structure, proceedings, and outcome are similar at least for the larger communities (Pfeil, Zaphiro, & Ang, 2006). It is assumed that the results cannot be *directly* transferred to other language versions but that it is likely to find similar results at least among the larger communities (e. g. English, French, Japanese, and Italian).

---

[1] http://[en|de].wikipedia.org/wiki/Special:Statistics [2011-02-16]
[2] http://www.alexa.com/siteinfo/wikipedia.org [2011-02-16]
[3] http://en.wikipedia.org/wiki/Wikipedia:Administrator_intervention_against_vandalism [2011-02-22]
[4] http://de.wikipedia.org/wiki/Wikipedia:Vandalensperrung [2011-02-22]

It is useful to take a look into the structure of Wikipedia's community to understand the way it works. Therefore, the next paragraphs explain the hierarchy of power (this more or less applies to all language versions) and the structure of vandalism and conflicts in the German Wikipedia.

Wikipedia's articles can be edited by any Internet user. If not logged in, the edits of occasional editors are associated with their IP address (e. g. 134.93.160.121). Those contributors who registered and created an account are given different privileges depending on the number of edits or community elections.[5]

- *New Users* are not entitled to move articles from one lemma to another and can be excluded from editing certain pages along with anonymous users ("IPs").
- Four days after registration, new users are automatically promoted to be *autoconfirmed users*. This is the regular user status of Wikipedia.
- In some language versions of Wikipedia (the first of which was the German version) there exists an additional step in the hierarchy: the *editor* status. Editors are entitled to approve edits of IPs, new users, and users without editor-status. Without approval, edits are not visible for the general public. The "editor"-status is not to be confused with the colloquial term "editor" for someone who "edits", i.e. contributes to, Wikipedia.
- *Administrators* (*Sysops*) are users who have earned a reputation as authors and project organizers and who have been elected by the community to make use of additional functions of the Mediawiki-Software. They decide which users be locked out for misbehavior, which articles be deleted or protected against anonymous, new, or all users editing to prevent conflicts. Administrators are confronted with conflicts every day – and are the only members of the community with the power to put an end to them.
- *Bureaucrats* have technical and organizational duties. They provide elected administrators with their technical features.
- *Stewards* are appointed globally across all Wikimedia Foundation wikis. They can withdraw administrator's privileges, e. g. after a successful request for de-adminship.[6]
- Additional user groups that usually require that the applicant has been elected as an administrator are *Oversighters* (who can hide dangerous or illegal information even from administrators), *Check-Users* (who can read log files to discover foul play using multiple accounts), and *members of the arbitration committee*.[7]

A closer look at the German version of the page where *Administrator intervention against vandalism* is discussed reveals four major types of problems administrators have to deal with:

Edit War. Edit wars are a common phenomenon in the area of political conflict in Wikipedia, especially in disputed areas like politics, religion, or other morally loaded issues. Edit wars usually follow the pattern of *instrumental editing*. That is a way to avoid the restrictions of the neutrality-commandment ("NPOV", "Neutral Point Of View" in Wikipedia's terminology) by using facts (Kepplinger, 1992), wording, structure, or references to boost one's own, or to weaken one's opponent's position (Roessing & Podschuweit, 2011). Edit wars can be resolved by administrators either by blocking an article or a discussion page from being edited or by temporarily revoking the right to edit Wikipedia for one or both edit warriors. Repeated edit wars can result in the definite exclusion of edit warriors from the project.

---

[5] http://en.wikipedia.org/wiki/Wikipedia:User_access_levels; http://de.wikipedia.org/wiki/Wikipedia:Benutzer
[6] http://en.wikipedia.org/wiki/Wikipedia:Requests_for_de-adminship
[7] http://en.wikipedia.org/wiki/Wikipedia:Arbcom; http://de.wikipedia.org/wiki/Wikipedia:Schiedsgericht

Personal attacks. Personal attacks are not allowed within the Wikipedia community. This rule is meant to protect the spirit of teamwork that is inherent to a Wiki project. However, fierce disputes regularly escalate to intense argument and sometimes members of the community lose their temper. Personal attacks are stopped and at the same time punished by administrators with temporal or, in severe cases, definite blocking of the offenders' accounts.

General conflict. Some users try to use the page for administrator action against vandalism to get rid of opponents by calling for action against their alleged 'vandalism'. In most cases opponents are denounced as edit warriors or contributions in a discussion are taken as personal attacks. Due to social-psychological processes in the community, such conflicts often attract supporters of either side and can lead to fierce arguments, causing follow-up allegations of vandalism or personal attacks.

Simple Vandalism. The largest part of the site for administrator action against vandalism is occupied by the kind of behavior that the page originally was created for: deliberate damaging of Wikipedia's articles. Vandalism usually leads to temporal blockings of dynamic IP addresses, or infinite blockings of (in most cases newly registered) accounts.

2. RESEARCH QUESTIONS

The overall research question of this paper is: How does the community deal with problematic behavior such as edit wars, personal attacks, general conflict, and vandalism? It is assumed that the answer to this question allows inferences to the state of Wikipedia's community. This seems important with regard to the importance of Wikipedia as a source of information for the general public as well as for journalists of the traditional mass media (Roessing & Podschuweit, 2011).

In greater detail, there are four research questions:

RQ1:   What is the structure of the problems that are reported on the anti-vandalism page?

RQ2:   Does the distribution of the different types of problems alter with the time of the day (e. g. more vandalism after school is out, more general conflict later at night)?

RQ3:   Which resolutions are chosen by administrators to solve problems that are reported on the anti-vandalism page?

RQ4:   Is there a relationship between the extent of conflict inherent to a discussion on the anti-vandalism page and the actions taken by administrators?

3. HYPOTHESES

Vandalism is a common phenomenon whereas conflicts center around societal or internal hotspots (Roessing, 2008). This observation leads to Hypothesis 1:

H1:   The structure of the anti-vandalism page is dominated by many short issues of simple vandalism and fewer but longer discussions of conflicts.

It is known from analysis of the IP-Ranges used by vandals that a lot of mischief originates from school computers.[8] It is also known that the average member of the community is male, in his thirties and prone to technology (Schroer & Hertel, 2009). These two facts allow for Hypothesis 2:

H2:   Common vandalism occurs mostly during forenoon and noon. Community conflicts are solved when most of the community enjoys their free time, i. e. in the afternoon and the evening.

Administrators are elected to protect the project. If they comply with their duties, there should be very few unsolved cases. Additionally, the hardest consequence, the definite block of a registered user, should be rather rare since this usually leaves someone really angered. Hypothesis 3 thus consists of two sub-hypotheses:

H3a: Most cases end with short IP blockings, blockings of articles and temporarily blockings of registered members of the community. It is very rare that a case is left unresolved.

H3b: Definite blockings of registered users are rather rare.

It is every administrator's obligation to protect the project by settling conflicts. This leads to the fourth and last hypothesis:

H4:   The more conflict there is in an issue on the anti-vandalism page, the less likely is that the case is left unresolved.

4. METHOD

To determine the nature and extent of conflicts and vandalism the administrators of the German language version of Wikipedia usually have to deal with, a content analysis of 500 vandalism reports during the first ten days of November 2009 was conducted.[9] Unit of analysis was the single report, usually consisting of the reporting itself, a settled notice and in some cases additions and comments by other users. Coding units (categories of analysis) included the reporting account, the accused account (or IP) and the size (number of comments). The nature of vandalism reports was addressed by type, relation to Wikipedia's sections and content, potential for conflict, and the resolution of the report. Not every category was developed for the present paper; therefore only three constructs need further explanation:

Type of report: It was coded whether a report addressed blatant vandalism, improper behavior in connection with conflicts concerning the content of an article, an edit war, or other types of conflict.

Potential for conflict: A five-point scale was used to measure the amount of conflict that may rise from a vandalism report. Simple vandalism has little potential for conflict among users, while a vandalism report concerning the reference section in an article about political extremism carries a huge potential for conflict and dispute.

---

[8] http://de.wikipedia.org/wiki/Benutzer:Gardini/Vandalismusmeldung#Gelegenheitsvandalen.2C_Schulen_und_Unis [2011-02-22]
[9] Due to economic restrictions there was only one coder. In another analysis of similar complexity she reached an intra-coder reliability of 0.69 and 0.63.

Resolution: There are eight common resolutions for vandalism reports: (1) temporal blocking of IP addresses, (2) temporal blocking of user accounts, (3) infinite blocking of user accounts, (4) semi-blocking of articles, (5) complete blocking of articles, (6) EOD (End of Discussion), (7) reprimand, (8) archiving without action.

5. RESULTS

Sixty-four percent of the analyzed reports deal with simple vandalism. Fifty-two percent of the 500 analyzed cases ended with a temporal IP-Block. Eighty-eight percent of these reports contained only two entries: The report and the reaction of an administrator. Other issues, mostly regarding content or community life, were longer, containing up to 45 statements. Hypothesis one is thus supported.

Figure 1 shows the distribution of vandalism over the time of day for all types of vandalism and separately for simple vandalism.
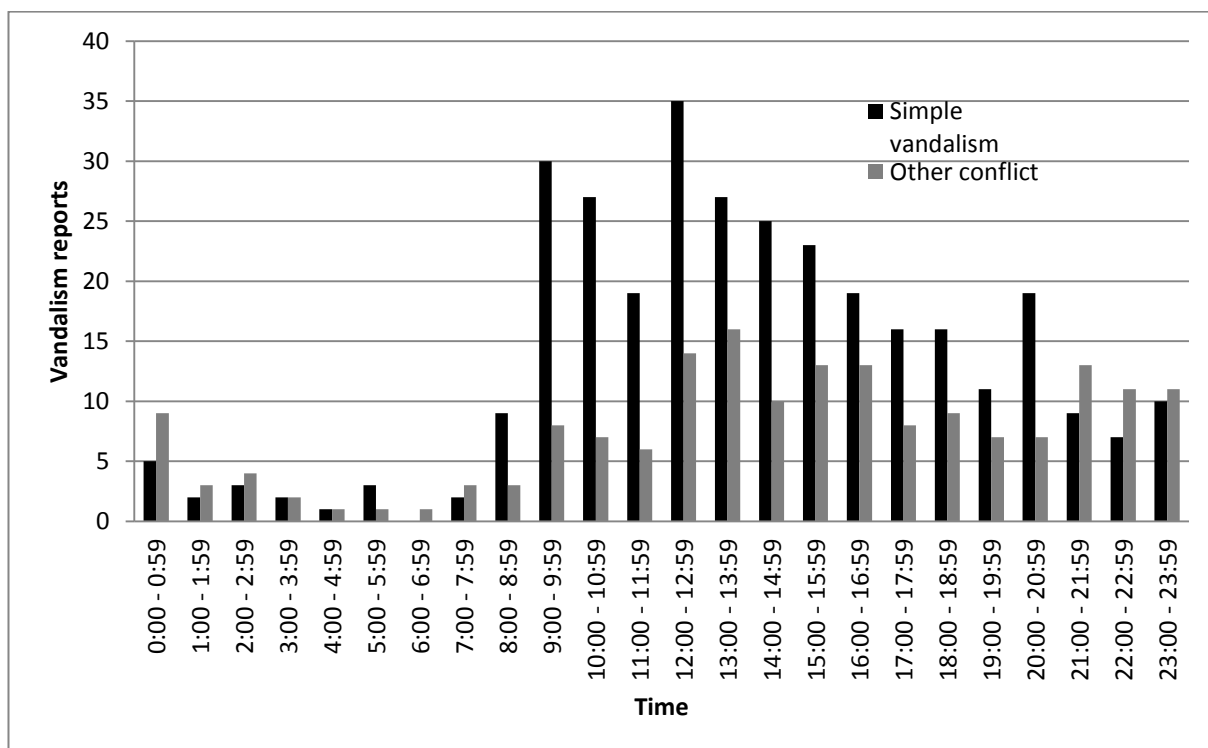

Figure 1: Distribution of vandalism over the day. n= 500

Hypothesis 2 receives partial support: There is an obvious relation between simple vandalism and school-hours. The suspicion that a considerable part of the vandalism on weekdays originates from schools is supported. But the evening is not dominated by other conflicts, because simple vandalism comes not exclusively from school computers.

The third hypothesis deals with the outcome of a discussion on the anti-vandalism page. It has already been revealed that the majority of the issues are short reports of vandalism ending in a temporary IP blocking. Figure 2 shows the complete distribution of conflict resolution.
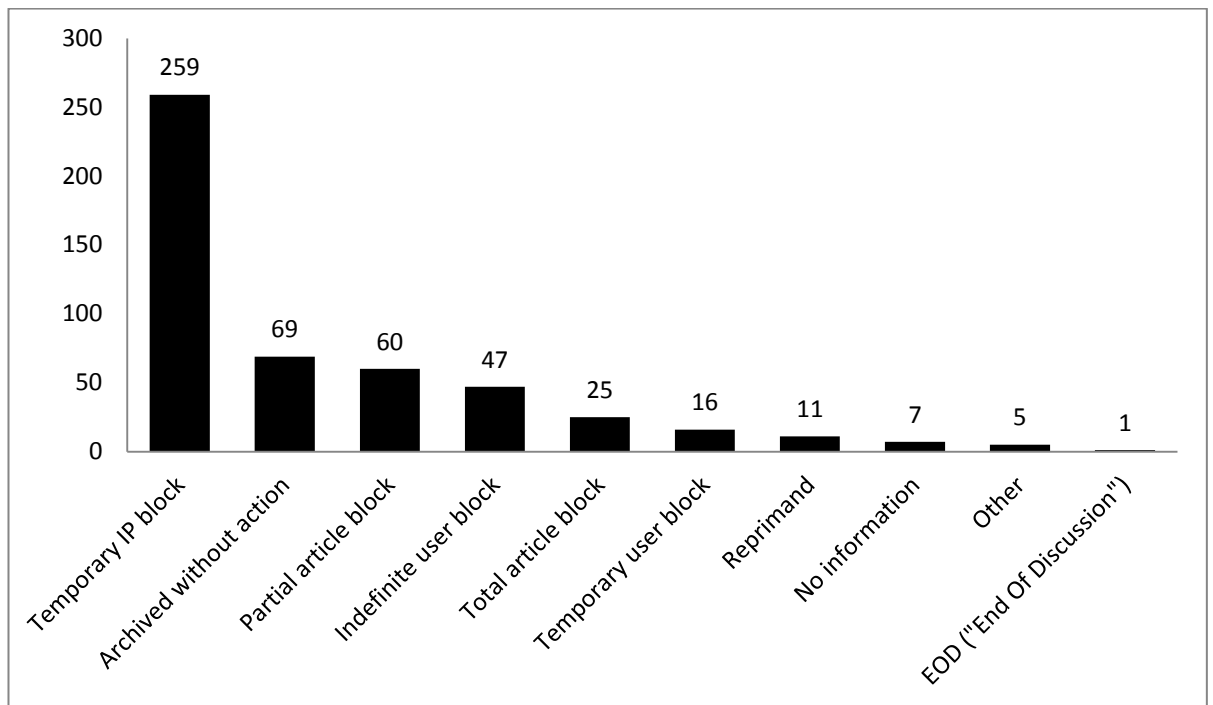
Figure 2: Conflict resolutions n=500

Hypothesis 3a is not supported by the data. Doing nothing and waiting for the archiving system is the most common outcome of issues on the anti-vandalism page following the majority of temporary IP blocks. As suggested by Hypothesis 3b, indefinite user blocks are rather rare.

Which kind of problem makes administrators refrain from their duty to resolve problems that were reported on the anti-vandalism site? Since it is the obligation of administrators to resolve conflicts, there should be very few of the issues that are not simple vandalism left to the automatic archiving system (Hypothesis 4). A cross-tabulation with the potential for conflict reveals that this is not the case. Of the 69 cases that were archived without a decision by an administrator or another resolution, 14 were consensual discussions, 9 somewhat disputed and 46 heavily disputed discussions. It is plausible that administrators fear to anger other users by deciding a disputed argument and thus refrain from doing so. To test this assumption further research seems advisable.

6. DISCUSSION

The findings of this study can be summarized in four statements:

1. The anti-vandalism page is mainly used for reporting and quick blocking of vandals. However, there are many other conflicts regarding Wikipedia's content or the community. Those issues tend to trigger longer discussions than simple vandalism does.
2. Mid-day hours are dominated by simple vandalism, probably originating from pupils making use of schools' computer equipment.
3. Most cases of simple vandalism are resolved by temporary IP blockings. Surprisingly, the second most common outcome of issues on the anti-vandalism page is 'no decision'.
4. 'No decision' is most likely in heavily disputed discussions.

The findings indicate that there is – leaving simple vandalism aside – a considerable potential for conflict within Wikipedia's community. Administrators play a key role in Wikipedia's system for conflict resolution. They do their work properly in simple and undisputed cases but tend to leave a discussion alone when it comes to heavy conflicts among users. Further research will be necessary to determine the connections of the anti-vandalism page with other places in Wikipedia where conflicts are discussed. It will also be necessary to study the role of administrators in community conflicts in greater depth.

REFERENCES

Kepplinger, H. M. (1992). Put in the public spotlight – Instrumental actualization of actors, events, and aspects in the coverage on Nicaragua. In: S. Rothman (Ed.), *The Mass Media In Liberal Democratic Societies* (pp. 201–219). New York: Paragon House.

Konieczny, P. (2009). Governance, Organization, and Democracy on the Internet: The Iron Law and the Evolution of Wikipedia. *Sociological Forum*, *24*(1), 162–192.

O'Sullivan, D. (2009). *Wikipedia: A new community of practice?* Farnham: Ashgate.

Pfeil, U., Zaphiris, P., & Ang, C. S. (2006). Cultural Differences in Collaborative Authoring of Wikipedia. *Journal of Computer-Mediated Communication*, *12*(1). Retrieved from http://jcmc.indiana.edu/vol12/issue1/pfeil.html.

Potthast, M., Stein, B., & Gerling, R. (2008). Automatic Vandalism Detection in Wikipedia. *Lecture Notes in Computer Science, 4956*, 663-668

Roessing, T. (2008). Opinion Formation in Wikipedia - Theory, Measurement, and Findings. *Masaryk University Journal of Law and Technology, 2,* 163-176.

Roessing, T. & Podschuweit, N. (2011). Wikipedia im Wahlkampf: Politiker, Journalisten und engagierte Wikipedianer. In: Schweitzer, E. & Albrecht, S. (Eds.): *Das Internet im Wahlkampf: Analysen zur Bundestagswahl 2009.* (pp. 297-314). Wiesbaden: VS Verlag.

Schroer, J. & Hertel, G. (2009). Voluntary Engagement in an Open Web-Based Encyclopedia: Wikipedians and Why They Do It. *Media Psychology*, *12*, 96 – 120.