

Wikipedia and Public Opinion

Paper for the 60th Annual Conference of the
World Association of Public Opinion Research (WAPOR)

Berlin, Germany
September 19 - 21, 2007

Thomas Roessing, University of Mainz, Germany

1. Introduction

The present paper deals with Wikipedia as an online community. It analyses the role of public opinion in discussions, disputes and the solution of conflicts among the users of the online encyclopaedia. A theoretical framework will be made up and methodological challenges and advantages of wiki-technology for an empirical approach to Wikipedia and public opinion will be discussed.

1.1 Wikipedia

When Wikipedia was founded in 2001, it seemed like just another online project. But it was not. It was different in many ways, especially with regard to the community behind the free online encyclopaedia and the scope of the whole project. The project soon started to grow rapidly from 2002 on, when more and more contributors added more and more articles to the online-encyclopaedia (Voß 2005). Today (2007), Wikipedia is one of the most popular web sites in terms of the number of visitors and page requests,¹ more than 100 servers handle 10,000 to 30,000 page requests *per second*.² There are about 250 language versions of Wikipedia, with English (en.wikipedia.org) as the largest section (about 1.65 Million articles) followed by German (de.wikipedia.org), containing about 600,000 articles.

Wikipedia is an online community (Preece/Maloney-Krichmar 2005) which can be described as an *Open Community Contribution System*. It aggregates efforts of many persons from different places toward a common task (Bos et. Al 2007) – the development of a free encyclopedia. Independently from certain language versions, exist three major user-groups: So called *IPs*, ordinary *Users* and *Sysops* (also called *Administrators*). When an unregistered or not logged in user contributes to Wikipedia, his contribution is registered with the IP-

¹ http://www.alexa.com/site/ds/top_sites?ts_mode=global&lang=none 02-21-2007

² <http://hemlock.knams.wikimedia.org/~leon/stats/reqstats/reqstats-monthly.png> 02-21-2007

address of his computer. Ordinary users are registered, logged in users; most of them write under a pseudonym, some natural persons use more than one pseudonym for their work on Wikipedia (so called *Sockpuppets*). Sysops are elected by the community and have additional technical measures, e. g. to delete an article or to lock users who have been found guilty of the breach of the peace or of vandalising articles. Therefore sysops are in charge not only of technical aspects of the online encyclopaedia, they play also an important role in the social structure of its users.

Wikipedia is organized in namespaces by the wiki-software (Voß 2005: 4 f). The most used namespaces are the article-namespace that contains the text of the encyclopaedia, the user-namespace that contains the personal profiles of the users and the Wikipedia-namespace with all the organisational things and meta-discussions. Every page, independent of its namespace, has a discussion page for the exchange of remarks and additional information with the aim to improve the content of the according page (usually an article). Consequently, the discussion pages are the part of Wikipedia, where conflicts between users or user groups become manifest.

1.2 Public opinion

The present paper uses the term public opinion in a way, that is oriented towards Noelle-Neumann's concept of public opinion (Noelle-Neumann 1984). To analyse the role of social control, of minorities and majorities and of decision-making within Wikipedia, this concept of public opinion is very useful, at least more useful than the understanding of public opinion as elite-opinion or as mere opinion distributions or other concepts from the wide area of public opinion research (Glynn et. al 1999; Childs 1965). Noelle-Neumann's concept of public opinion is especially suitable for the subject of the present paper, because the Wikipedia community fits some of the requirements for the functioning of public opinion as social control:

1. Wikipedia contains articles about subjects with a strong emotional potential (in terms of Noelle-Neumann's (1991): They are 'morally loaded'), e. g.: Abortion, religion, war, terrorism and other well known hot spots of social and online conflict such as political extremism (Roessing 2006; Roessing/Siebert 2006).
2. Most users of Wikipedia are either anonymously, or at least pseudonymously contributing to articles and discussions. They do not form a group of individuals that know each other, but they act in face of an anonymous public of other users and, of course the general public of read-only-users of the online encyclopaedia. With

publicness and anonymity, two more conditions for processes of public opinion are fulfilled.

3. There are plenty of conflicts, not only between single users or single users and the rest of the community but between different camps. Examples from the German Wikipedia are the continuing conflict between right- and left-wing users or between users with different opinions about the development of the Wikipedia project itself (*inclusionists* versus *deletionists*).

With conflicts and opinion camps, publicness and anonymity and with the emotional potential of many issues discussed in Wikipedia, Noelle-Neumann's theory of public opinion (especially its most prominent part, the Spiral of Silence) seems very appropriate for an analysis of social processes among *Wikipedians*.

2. Theoretical background

There are three levels of public opinion in Wikipedia.

2.1 The meta level of public opinion

The meta-level of public opinion in Wikipedia covers discussions among users that concern the structure, organisation, and the distribution of power in the Wikipedia project. These discussions exist in all language versions of Wikipedia, but the language versions differ in many ways. Cultural Differences precipitate in the way, collaborative authoring of Wikipedia works (Pfeil/Zaphiris/Ang 2006). The following analyses are based on the German language version of Wikipedia; in some cases links to English counterparts of German pages are additionally given. The applicability of the conclusions for other versions will be discussed in the discussion-section of the present paper.

An old, long-term and nevertheless fierce conflict is the one between *inclusionists* and *deletionists* (or *exclusionists*). The two camps foster different views on the question which articles are suitable for an encyclopaedia and which are not – and therefore are object to deletion by sysops.³ Central meeting ground for both parties is the candidates for deletion-page⁴, the list of criteria for relevant subjects⁵, some related areas⁶ and the according

³ <http://meta.wikimedia.org/wiki/Inclusionists> [retr. 2007-06-17]; – <http://meta.wikimedia.org/wiki/Exclusionist> [retr. 2007-06-17]

⁴ <http://de.wikipedia.org/wiki/Wikipedia:L%C3%B6schkandidaten> [retr. 2007-06-17]; - English: http://en.wikipedia.org/wiki/Wikipedia:Articles_for_deletion [retr. 2007-06-17].

⁵ <http://de.wikipedia.org/wiki/Wikipedia:Relevanzkriterien> [retr. 2007-06-17]; – English: <http://en.wikipedia.org/wiki/Wikipedia:Notability> [retr. 2007-06-17].

⁶ E. g. the quality improvement pages (<http://de.wikipedia.org/wiki/Wikipedia:Qualit%C3%A4tssicherung> [retr. 2007-06-17]) and the deletion revision (<http://de.wikipedia.org/wiki/Wikipedia:L%C3%B6schpr%C3%BCfung> [retr. 2007-06-17]).

discussion pages. The emotional potential of the conflict becomes evident by mutual verbal insults: Some deletionists accuse inclusionists of supporting garbage-articles in the Wikipedia⁷, deletionists are sometimes called “deletion-fetishists”.⁸ Both camps of public Wikipedia opinion maintain propaganda-sites: An inclusionist petition *against a strict deletion-policy*⁹ and its deletionist counterpart¹⁰, as well as a tongue-in-cheek *deletionists’ club*¹¹ that despite its ironical nature stirred much irritation among inclusionist wikipedians.¹²

Another meta-discussion that heated the minds of many Wikipedia-contributors was called *the great babel-war*. *Babels* are small text-boxes with a logo that are used by many wikipedians to indicate their preferences and opinions on their user-pages. Some users held the opinion that these babels were too much fun and nonsense and therefore not compatible with the idea of writing a serious encyclopaedia. Some sysops shared this opinion and many babels were deleted in a number of deletion-waves. Despite lengthy discussions in which a majority voted in favour of a moderate use of babels¹³, a minority still fights them, backed by the technical power and the social support of some sysops.

The most prominent effect of conflicts between parts of the users (and it should be kept in mind that Wikipedia-users are in most cases also authors and contributors of articles) can be found in the elections for the sysop-privilege.¹⁴ Sometime a candidate receives a fair amount of votes for or against him, because he is an inclusionist or deletionist or because he collects babels on his user-page or rejects doing so. Additionally, there are effects on the collaboration on the actual articles of Wikipedia, for example, when authors try to save an article from deletion and therefore try to make it deletionist-proof. This effect is part of the next level of public opinion in Wikipedia, the article discussion level.

⁷ E. g. http://de.wikipedia.org/wiki/Wikipedia:L%C3%B6schkandidaten/5._August_2005 [retr. 2007-06-17]; – http://de.wikipedia.org/wiki/Wikipedia:L%C3%B6schkandidaten/15._Juli_2006 [retr. 2007-06-17].

⁸ http://de.wikipedia.org/wiki/Wikipedia:L%C3%B6schkandidaten/28._Mai_2005 [retr. 2007-06-17].

⁹ http://de.wikipedia.org/wiki/Wikipedia:Unterschriftenliste_f%C3%BCr_eine_liberale_L%C3%B6schpraxis [retr. 2007-06-17].

¹⁰ http://de.wikipedia.org/wiki/Wikipedia:Unterschriftenliste_gegen_eine_liberale_L%C3%B6schpraxis [retr. 2007-06-17].

¹¹

http://de.wikipedia.org/wiki/Benutzer:Gardini/Verschw%C3%B6rung_f%C3%BCr_voreiliges_L%C3%B6schen [retr. 2007-06-17].

¹²

http://de.wikipedia.org/wiki/Benutzer_Diskussion:Gardini/Verschw%C3%B6rung_f%C3%BCr_voreiliges_L%C3%B6schen [retr. 2007-06-17].

¹³ http://de.wikipedia.org/wiki/Wikipedia:Meinungsbilder/Zul%C3%A4ssigkeit_von_Babelvorlagen and http://de.wikipedia.org/wiki/Wikipedia_Diskussion:Meinungsbilder/Zul%C3%A4ssigkeit_von_Babelvorlagen [retr. 2007-06-18].

¹⁴ <http://de.wikipedia.org/wiki/Wikipedia:Adminkandidaturen> [retr. 2007-06-18]; – English: http://en.wikipedia.org/wiki/Wikipedia:Requests_for_adminship [retr. 2007-06-18].

2.2 The article discussion level of public opinion

Articles are usually debated on their according discussion pages. The discussion pages of many articles are empty; this indicates that there has not yet been demand for discussions about the content and structure of these articles. Some articles have relatively small discussion pages with short remarks concerning particular pieces of information or the formulation of sentences or sections. On the other hand, some articles have huge discussion pages with hundreds or thousands of contributions. Many of these heavily discussed articles belong to the already mentioned area of value-laden subjects. Both the vehemence of the discussion and the subject of the article, and additionally the content of the contributions of the discussion play an important role for the conception of a content analysis of Wikipedia and will be discussed in greater detail in chapter 3 of the present paper.

Sometimes the most vehement discussions about articles grow into other areas of Wikipedia, e.g. when the opponents calumniate each other as vandals on the page for reports of vandalism.¹⁵

2.3 The article level of public opinion in Wikipedia

It is very probable that in many cases the article discussion – and indirectly the meta-level too – have an impact on the formulation, structure and content of the actual Wikipedia articles, which are read by many more people than the Wikipedia users who are involved in the development of the online encyclopaedia (“*wikipedians*”). This is because the particulate vehement discussions in most cases do not come from mere desire for discussion but from the feeling of some users that their point of view (POV) is correct and opposite opinions are not only wrong, but morally reprehensible. When one camp prevails over the other, its POV will govern the overall message of an article. As a matter of course, many articles are difficult to bias, therefore the number of users trying to write their POV into articles about guinea fowls¹⁶ is very small and many articles as unbiased as articles in printed, professional encyclopaedias. But a fair number of articles – of course again the same which were earlier mentioned as laden with emotions and well-equipped with lengthy and vehement discussions – can be expected to contain some bias. The following section of this paper will discuss how those biases, if they exist, can be detected and measured by means of content analysis.

Subsequently, methods have to be developed which can be used to gain some data about the

¹⁵ <http://de.wikipedia.org/wiki/Wikipedia:Vandalismmeldung> – Engl.:

http://en.wikipedia.org/wiki/Wikipedia:Administrator_intervention_against_vandalism

¹⁶ <http://de.wikipedia.org/wiki/Perlhuhn>

sources of bias, which are expected to be found in the other two levels of public opinion in Wikipedia.

3. Methodological challenges and advantages

The internet is a difficult terrain for content analyses (Roessing 2005). This is especially true for dynamic content management systems. Additionally in the case of Wikipedia, the wiki-system that allows anyone to add or alter pages at any time, constricts any stable measurement of content. Nevertheless, some aspects of public opinion can be measured by classical means of content analysis. These include issue, size, and bias of articles and discussions.

The history-function of the mediawiki-software provides the researcher with additional tools for the analysis of Wikipedia's content (Pfeil/Zaphiris/Ang 2006), but there is little experience with the use of these tools for public opinion studies. On that score the new possibilities add to the challenges of internet content analysis.

3.1 Variability of the content

Variability of the content is a popular argument when content analysis of web sites is discussed among communication researchers. In the case of Wikipedia it is a small issue anyhow. Nearly every page of Wikipedia has a link to a permanent version and through the "History"-button every previous version of every article or discussion can be retrieved. It would be no problem to define e. g. "All articles and discussions belonging to the category "Nuclear Power"¹⁷ in the version of June 15th 2007, 3pm as units of analysis.

3.2 Conflict and emotional potential

As mentioned before, conflict and emotional potential are prerequisites of processes of public opinion. Their existence must be proved empirically, not just assumed. This is difficult even for classic printed or broadcast media. And so it is in Wikipedia, but there the scientist is confronted with an additional problem. As explained in the second chapter of the present paper, public opinion manifests itself on different levels of Wikipedia. To determine the vigour of a discussion or the emotional involvement of arguments and discussants, the coders have to track an issue over a sometimes large number of discussion- and metadiscussion-pages. Serendipitously the wiki-software provides means and information to cope with the problem:

¹⁷ http://en.wikipedia.org/wiki/Category:Nuclear_power (retr. 2007-06-23)

- The history-page lists time, author, scale and an optional comment to every edit that changes the content of an article- or discussion-page.
- A difference-page can be used to compare any versions of a Wikipedia-page.
- System-links in the navigation bar and wiki-links implemented by users connect related articles, article-discussions and meta-discussions.

For example, to determine the vigour of a discussion or a conflict about an article, the history-page provides support for measurement. The vigour of a discussion can be measured as edits per time or as the median time between edits. The article discussion of the article about the German leftist party “Die Linke”¹⁸ received 25 edits in the 48 hours between 10:17, June 23rd and 10:17, June 25th (1.92 edits/hour) from seven users (two IPs and five registered users). The median time between the 25 edits is 11.5 minutes (mean 12,38 minutes). Especially the last measures can be used to compare discussions with different potential for conflict. The median time between the last ten edits of the article discussion of “Wasser”¹⁹ (Water) is 2051 minutes (mean 15,980.89), indicating the lesser potential for discussion of the water issue. Further indicators for the intensity of a discussion can be derived from these data by calculating the average number of edits per user or vice versa and the standard deviation of the edit-distance to investigate the homogeneity of the discussion process (the discussion about the article of the Leftist Party is with $s=283.3$ more homogenous than the discussion about “Water” ($s=27.129.5$). The number of verbal insults and the number of meta-discussions about misbehaviour of participants can be used as indicators of the emotional potential of an issue or a discussion between wikipedians. All the possible measurements mentioned before still await extensive testing and validation. The author of the current paper is already preparing an according pilot study.

3.3 Changes of article content

While the mere extent of changes is easily extractible from the history-pages, the change of the content is to be measured by coders looking at the difference-pages of each article. For example, there is a conflict among German wikipedians about the question if and where the assessment of a political organisation as ‘extremist’ by the authorities should be mentioned in the according articles. To measure the development of such a conflict, it is not sufficient to count the number of edits and the number of users involved. It is necessary to code the direction of the edits and the opinion that becomes visible in the edits, edit-comments and other contributions of the users involved.

¹⁸ http://de.wikipedia.org/wiki/Die_Linke (retr. 2007-06-23)

¹⁹ <http://de.wikipedia.org/w/index.php?title=Diskussion:Wasser&action=history> (retr. 2007-06-25)

4. Summary and discussion

Previous sections dealt with public opinion and Wikipedia. On the basis of the German language version of the online encyclopaedia features of the research object itself (articles, discussion pages, and users) and three levels of public opinion in Wikipedia were presented:

1. Meta level (discussions not directly belonging to certain articles)
2. Article discussion level (discussions belonging to certain articles)
3. Article level (possible biases in the content the ordinary reader of Wikipedia is confronted with).

The last chapter discussed the methodological challenges and advantages the mediawiki-software holds ready for the researcher. Further methodological, pure, and applied research is recommended and partially already under preparation by the author. Further research should include a comparison of different language versions of Wikipedia. While the phenomenon of public opinion can be found in all cultures of the world, the actual issues of public opinion and people's reactions to it are usually different in different nations and cultures (Noelle-Neumann 1984). It is a plausible hypothesis that this is also true for online communities.

References

- Bos, Nathan / Zimmerman, Ann / Olson, Judith / Yew, Jude / Yerkie, Jason / Dahl, Erik / Olson, Gary (2007): *From Shared Databases to communities of Practice: A Taxonomy of Collaboratories*. In: *Journal of Computer-Mediated Communication* 12, <http://jcmc.indiana.edu/vol12/issue2/bos.html> (retr. 2007-06-06).
- Childs, Harwood (1965): *Public Opinion: Nature, Formation, and Role*. Princeton, New Jersey u.a.
- Noelle-Neumann, Elisabeth (1984): *The Spiral of Silence. Public Opinion – Our Social Skin*. Chicago, London: University of Chicago Press.
- Noelle-Neumann, Elisabeth (1991b): *The Theory of Public Opinion: The Concept of the Spiral of Silence*. In: James A. Anderson (Hrsg.): *Communication Yearbook* 14. Newbury Park, London, New Delhi. Pp. 256-287.
- Pfeil, Ulrike / Zaphiris, Panayiotis / Ang, Chee Siang (2006): *Cultural differences in Collaborative Authoring of Wikipedia*. In: *Journal of Computer-Mediated Communication* 12, <http://jcmc.indiana.edu/vol12/issue1/pfeil.html> (retr. 2007-06-06).
- Preece, Jenny / Maloney-Krichmar (2005): *Online-Communities: Design, Theory, and Practice*. In: *Journal of Computer-Mediated Communication* 10, <http://jcmc.indiana.edu/vol10/issue4/preece.html> (retr. 2007-06-15).
- Roessing, Thomas (2005): *Einsatz einer Typenstichprobe für die Inhaltsanalyse politischer Internetinhalte [Use of a typology-sample for the content analysis of political internet*

contents]. In: Volker Gehrau / Benjamin Fretwurst / Gregor Daschmann (Eds.): *Auswahlverfahren in der Kommunikationswissenschaft [Sampling methodology in communication research]*. Cologne: Van Halem. S. 173-182.

- Thomas Roessing (2006): *Use of Electronic Guestbooks and Bulletin Board Systems on Right- and Left-Wing Extremists' Web Sites*. In: Radim Polcák / Martin Škop / David Šmahel (Hrsg.): *Cyberspace 2005*. Brno: Masaryk University. Pp. 147-154.
- Roessing, Thomas / Siebert, Sandra (2006): *Perception and Assessment of Left- and Right-Wing Extremism by Public Opinion. An Experimental Study*. Paper presented at the 59th WAPOR Annual Conference in Montreal, Canada, May 16-18, 2006.
- Voß, Jakob (2005): *Measuring Wikipedia*. Proceedings of the ISSI 2005 conference.