

Wikipedia vs. Academia: An Investigation into the Role of the Internet in Education, with a Special Focus on Wikipedia

Timo Staub*, Thomas Hodel

Department of Business, Health and Social Work, Bern University of Applied Sciences, Switzerland

Copyright © 2016 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract This paper considers Wikipedia and collaborative editing in general: what is Wikipedia, how does it work as a collaborative editing project? Who publishes there, how do these people collaborate, is there a hierarchy among them? And what about Wikipedia quality control: is it efficient, how good is the factual quality of the content? Can Wikipedia be re-used for academic work - and if so, where and how? How does Wikipedia cope with research findings; can they be found on the platform? What influence does Wikipedia have on research and education; how should universities cope with the fact that open knowledge can be found there within a matter of seconds? This paper addresses the issue in a rather hands-on and down-to-earth approach that will allow us to draw some interesting conclusions about the role of open Internet knowledge (such as that which can be found on Wikipedia) for learning and knowledge creation. We will be placing a special focus on academia: for instance, how should universities of applied sciences define “competency based learning” at a time when so many answers can be readily found on Wikipedia? Here the paper does not strive to come to generalized conclusions, but it does strive to find some modest, surprising and - last but not least - also practical answers. The current paper is based on library research, an online analysis of the current Wikipedia website, and interviews with Swiss Wikipedia activists.

Keywords Wikipedia, Collaborative Editing, Learning Goal Analysis, Competency Based Education, Online Learning, Connectivism, Applied Sciences, Higher Education

1. Introduction

Globalization has become a reality: many basic manufacturing tasks have been relocated from the Western world to China, and companies which could not cope with

the ever quicker pace of innovation have had to close their doors. Global information, innovation and knowledge exchange have become key factors for success, and the Internet has become a major driving force behind the economic, technological and social changes which we have seen in the past twenty years. In short, the Internet has changed the way economies work. And the same goes for people, too: workplaces look different nowadays, and knowledge workers such as engineers and scientists need both a computer and an open Internet access. This is relevant because the digital revolution has changed the way people work, live, and learn.

Numerous studies have shown that Internet content which can be found on collaborative platforms such as Wikipedia is not only abundant, but also trustworthy. As a result, accurate information about issues such as physics, chemistry, medicine, history and a whole lot of other issues can be found within a matter of seconds by anybody, anywhere and anytime. As a consequence, the notion of competency has changed: understanding and research skills have become more important, facts and figures have (probably?) become less important because they can be found on the Internet anyway. Universities of applied sciences should be struggling with this new kind of definition: if we really wanted to focus on competencies rather than factual knowledge, why do we forbid the use of the Internet during examinations? Even more important, should we not update all our curricula to include the teaching of basic Internet research skills and knowledge, such as relevance feedback or information evaluation? In short, how do we define competency at a time when information about ideas, more or less proven concepts and descriptive texts can be found and re-used all over the place?

The current research paper addresses this issue in a rather hands-on and down-to-earth approach: Wikipedia is a typical, well-known and well-established example of a collaborative editing community publishing open knowledge which can be re-used by anybody, be it within academia or outside

academia. So the paper looks at Wikipedia, and draws some conclusions about the role of open Internet information (such as that which can be found on Wikipedia) for learning and knowledge creation. It places a special focus on academia: for instance, how should universities define “competency” at a time when so many relevant pieces of information can be readily found on Wikipedia?

2. About This Paper

Research Methodology

Wikipedia has already been the subject of abundant research. As a result, a lot of papers have been written about the quality of Wikipedia articles, the intrinsic or extrinsic motivation of Wikipedia authors, and many other subjects such as the human community and the technical platform supporting the Wikipedia enterprise. Even more importantly, Wikipedia can be easily observed “in action” over the Internet, which allows many conclusions to be drawn. Wikipedia authors can also be interviewed, which provides further insights into the Wikipedia ecosystem. All three approaches (1. paper research, 2. analysis of the Wikipedia platform, and 3. interviews with Wikipedia authors living in Switzerland) have been used for the current paper, which is therefore based on three qualitative pillars:

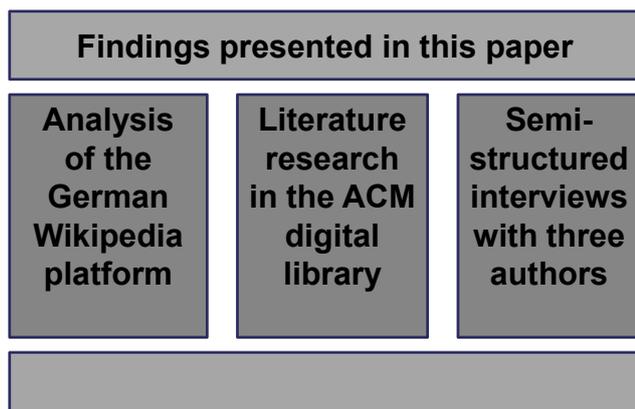


Illustration of qualitative research methodology used for this paper [1]

As a result of the three pillar approach, the Wikipedia content and ecosystem becomes much easier to describe - and the same also goes for the results which it produces.

Terminology

Collaborative editing can be defined by its attributes: *writing in a shared document* (different people write, edit and read the same document), *collaborative processes* (possibility to define workflows and user roles), *data lineage* (associations between same pieces of text in different documents), *distributed teams* (users work in virtual teams organized over the Internet), *placeless document philosophy* (users can store and find documents without having to specify a fixed location in a hierarchical structure), and finally *flexible handling of content and layout* (separation of

content and design, resulting in the automatic support of multiple layouts, if possible also cross media publishing). [2]

When looking at this definition, it becomes clear that Wikipedia does not fulfill all the attributes of a classical collaborative editing tool. E.g., Wikipedia does not work with separate documents, it works with *articles* – and these are integrated in a searchable knowledge base. However, most of the attributes apply: Wikipedia supports simultaneous writing, distributed teams, collaborative processes, and a separation of content and design (Wikipedia supports various text designs and screen sizes automatically; it is even possible to assemble online articles into printable paper books).

3. About Wikipedia

The basics of Wikipedia are well known: in its fundamental and original form, Wikipedia is nothing more than an *encyclopedia*. So it consists of *articles* which describe certain *words, terms or concepts*.

At the beginning, Wikipedia was edited by its readers – every reader could update the Wikipedia article which he or she was currently reading. In principle, this still applies today - but things have grown more complex. The following is a short description of our subject in its current form:

Community Interactions and Quality Control Take Place over the Internet

The software system behind Wikipedia is called *Mediawiki*. In addition to the obvious editing functions, it offers features which are directly relevant to quality control, such as *role support* (the Mediawiki software supports various organizational roles), *track changes and roll-back* (concurrent edits are tracked in detail, they can be quickly undone), various discussion and arbitration instruments (every Wikipedia article comes with a discussion page), and watchdog functions (alert e-mails can be sent out to original authors every time an article has been re-edited by somebody else).

The Mediawiki software is closely linked with the internal organization of the Wikipedia community; it supports the interactions of its members in a very efficient way. As a result, the vast majority of the interactions among Wikipedia members take place over the Internet, and the Mediawiki software has become a rather complex and interlinked system environment.

The Wikipedia Community Is Structured and Self-organized

German (and Swiss) Wikipedia members call themselves *Wikipedians*. They form a community of practice which has developed an inner hierarchical structure based on meritocracy, which is described below.

Anonymous readers can still update Wikipedia articles, and in many cases their *edits* will even go online immediately. But these people will only be registered with

their IP addresses; they, as anonymous Wikipedia contributors, are at the bottom of the Wikipedia hierarchy.

Registered Wikipedia users already have a Wikipedia name. They can most of the functions of the Mediawiki software. Usually, the Wikipedia name does not correspond with the real name, so a great deal of anonymity can remain. If a registered Wikipedia user produces a certain amount of work, then he or she gets the right to vote, thus participating in the election of those appointed to the higher echelons of the Wikipedia hierarchy, such as *administrators*.

Authors are the Wikipedians who actually write encyclopedic articles. Usually, they work within a very specialized domain of interest (such as the history of ancient art in Switzerland).

Visualizers work as illustrators, making drawings and animations which can afterwards be used to illustrate Wikipedia articles.

Cleaning staff are occupied with the maintenance of articles which have previously been written by other people: these people verify, write their own references, check the references of others, and they also mark dubious content so that it can be re-considered by the authors who wrote it.

Troll hunters identify Wikipedians who do not behave in a cooperative manner, such as political activists entering propaganda content. If necessary, the trolls' access to Wikipedia will be blocked.

If Wikipedians cannot agree among themselves (e.g. on the neutrality or the length of a certain article), then *arbitrators* will try to resolve the dispute.

Helpers explain the rules and processes to Wikipedia newbies, either in direct contact or via the specially written manual pages in the *author's portal*, which is explained below.

Most of the articles are now monitored by *reviewers/classifiers* (German: *Sichters*). For example, if a change has been made to an important article such as the one about the current pope, then this change will have to be verified and approved by a viewer/classifier prior to being published online. Viewers/classifiers are usually venerated Wikipedia authors who have risen through the ranks of the Wikipedia hierarchy because they have published a great deal of "valuable" text (valuable text is text which is not corrected by other authors afterwards - the Mediawiki software rates the authors accordingly).

Administrators have special rights assigned to them with regard to the use of the Mediawiki software. For example, they can protect certain articles in such a way that they cannot be edited at all. They can also lock out other Wikipedians. Administrators have a lot of power, and they occupy a high position in the Wikipedia hierarchy; this is the reason why administrators have to be elected by online voting. Usually only trusted authors will become administrators.

Bureaucrats have an even higher position than administrators. They can revoke special rights. Only bureaucrats can change an existing Wikipedia user name.

At the top of the hierarchy are *stewards*, who are something like supreme Wikipedia judges, simultaneously presiding over several national Wikipedia versions.

There is a *back office* which takes care of all tasks that do not directly concern the encyclopedia - such as online support, communication and public relations.

Finally, *software developers* write the Mediawiki software, which is an open source software system.

Wikipedia is supported and supervised by the *Wikipedia foundation*. The foundation collects funds, and these funds are used for the hosting of the encyclopedia and a number of other websites such as Wikibooks, Wikiversity and the Wikimedia picture website. The foundation is one of the most important Wikipedia stakeholders. Another is *Jimmy Donal "Jimbo" Wales*, who is one of the Wikipedia founders - he constitutes the public face of Wikipedia.

Updates are quick and self-explanatory

Wikipedia is quite self-explanatory. Therefore, the instructions for editing are straightforward. Go to Wikipedia (www.de.wikipedia.org), search for the article that you want to edit. Then click the *edit tab*, and edit the article using the typical Wikipedia syntax. It is recommended that you also click on the *discussion tab* - here, you should explain why you changed the article, and what you edited. The *version history* will show previous versions of the article, and it offers the roll-back function which allows the reviewers to undo previous changes.

Authors get help in the Author's Portal

Apart from the encyclopedia itself, authors will find the *author's portal* (German: *Autorenportal*) which offers help and instructions, such as technical documentation, authoring guidelines, information about copyright issues, and a great many templates. Even more importantly, the author's portal offers interactions with other Wikipedia members: some Wikipedians offer help as *mentors* (German: *Mentorenprogramm*), and processes such as *conflict resolutions* and the *election of administrators* also take place through the author's portal. Last but not least, the author's portal offers the opportunity to meet fellow Wikipedians at real-life offline events such as grill parties.

Wikipedians can have their own personal page

Much as in Facebook, active Wikipedia authors will have a *personal page* which allows them to present themselves to other Wikipedians. However, quite typically for Wikipedia, their personal pages will remain *centred around subject matter interest and encyclopaedic work* - and, unlike in Facebook, their personal pages will not contain a lot of bells and whistles.

4. About Quality Issues

As the reader might guess from the points mentioned above, Wikipedia is all about quality: within the author's

portal, much ink has been spilled on the discussion and detailed definition of quality assurance policies. Most of these (encyclopedia-related) policies have been implemented via the Mediawiki software. The whole Wikipedia meritocracy is based on quality, and only authors who have published a relatively large amount of quality content will get to superior hierarchical levels. As a result, the quality of Wikipedia articles is constantly being scrutinized and discussed by both readers and venerated Wikipedians alike. All these endeavors have a good effect on the content quality of Wikipedia articles:

Results of Library Research

Extensive research on the quality of Wikipedia articles has been undertaken. The results can be summarized as follows:

Wikipedia quality control mechanisms work and are highly effective: 42% of wrong or biased pieces of information are rectified by the first viewer, after ten viewers 70% of errors have been eliminated. [3]

Compared with peer-reviewed academic papers of the highest quality, Wikipedia will contain some errors. But these errors will not consist of wrong information, they will consist of “errors of omission”: not everything which should be said, has been said. [4]

Although it is much bigger than other encyclopedic collections, the quality of Wikipedia is on the same level as authoritative encyclopedias, such as Brockhaus and the Encyclopedia Britannica. In Wikipedia, errors are eliminated quicker than in standard encyclopedias. [5] [6]

Articles about prominent or important subjects get more attention than articles about subjects of lesser importance – the first ones are usually of a higher quality than the latter ones. [7]

There are two kinds of Wikipedians: the “Zealots” (who will first write an article, and then defend it) and the “good Samaritans” (who will maintain, add footnotes, and correct visible errors quickly and without much ado). [8]

Experience does not have a big influence on quality: articles written by Wikipedia newbies are not worse than articles written by Wikipedia veterans. [9]

People who use Wikipedia very often as readers will start to correct errors and eventually become authors. [10]

A lot of time has been spent on measuring the quality of Wikipedia articles in a quantitative manner. In the meantime, Wikipedia articles are increasingly being used as references in other contexts. [11] [12] [13] [14]

Results of the Interviews with Swiss Wikipedia Activists

We interviewed three Swiss Wikipedia authors (who were selected on criteria such as the number of written Wikipedia articles and nearby location). The first two of the three authors had written Wikipedia articles about Eastern religions and art history, the third author had written various minor contributions in subjects such as philosophy and contemporary history. The interesting finding was the fact that the first two authors had not only written Wikipedia articles about their respective areas of interest, they had also

written peer-reviewed academic papers and books about the very same areas of interest – in short, these were academic authors, and they knew very well what they were writing about! The third author wrote mostly about issues which he had previously studied at the university, so most of his edits could be backed up by book references – as a result, his edits were also of good quality.

So, interestingly enough, Wikipedia authors do not only adhere to professional quality standards, in our research project they also *are* professionals who have a very deep knowledge of the subject matter they are writing about.

5. Conclusions

The conclusions below are based on the three “pillars” illustrated in the picture above: 1) Library research in the ACM digital library (which can be summarized by saying that Wikipedia has “good” quality) 2) Interviews with Swiss Wikipedia authors (two of the three authors were senior authors with scientific publications about their respective areas of interest, and the third one had university studies about his respective area), and 3) An understanding of the Wikipedia ecosystem and quality control mechanisms (as observed on the German Wikipedia site). Our conclusions are stated in two parts – general conclusions and recommendations:

5.1. General Conclusions

Wikipedia can be seen as a research community: findings are constantly being published, scrutinized, verified, and the reputation of an author grows with the reputation of his or her contributions. In the meantime, quality control is almost organic and is automatically supported by the Mediawiki software. In short, Wikipedia is probably the perfect example of a community of practice which is supported by knowledge management software.

When looking at the humanities, Wikipedia can be seen as a valuable first entry point for research. It is possible to find well-known and generally accepted findings on Wikipedia. When looking at the humanities, older and bigger Wikipedia articles will give an accurate account of the state of knowledge within a given area of interest. Very often, the articles in question provide references, much as in the case of academic literature. As a result, Wikipedia also offers a convenient entry point for research, it gives a valuable introduction to a given subject matter area. The reader will find factual information and current terminology. Even more importantly, the discussion pages will give him or her initial insight into the current debates concerning the issue he or she is reading about.

Wikipedia is an example of a successful learning community. Whenever somebody wants to build a new knowledge management system, the lessons learned from Wikipedia should be taken into consideration: motivational aspects such as the public visibility of all contributions, a

quality control system which is both automatic and generalized, a clear focus on common terminologies and common processes.

6. Recommendations

Professors should know what Wikipedia says about their respective subject matter domain because sooner or later they will be confronted with it anyway (mostly by their students). They should look up their subject matter in Wikipedia, and briefly address their findings during their lectures: can the Wikipedia content be used as an entry point for further research or should it be considered off limits – and if so, why?

Wikipedia can be used for serious student projects: students need to learn about academic methodology and academic writing style. Much of the latter can be observed in Wikipedia. This demands a lot of work - but it can be done: e.g., a student could be asked to look up the Wikipedia article about a certain subject matter, and then enlarge the article using information derived from the lectures. In the meantime, the article should be improved by adding more references to academic literature. Students should then justify their edits on the Wikipedia discussion page, and consider the reactions of other Wikipedians. Afterwards, the students should print out their contributions, and hand them in for grading by the teacher.

Such a student project represents real work: it can be considered as a contribution to a public good, performed in front of an international audience. Even more importantly, the students have to adhere to a given quality standard which will be scrutinized both by the grading professor, and by the Wikipedians. The whole project should be well organized: e.g., there should be a preliminary list of the articles in question to avoid the possibility that all students work on the same article in the meantime (which might upset other Wikipedians).

Competency-based education should take into account the Internet: currently many schools put an emphasis on competency-based education, thus putting the focus on learning outcomes and the higher echelons of the learning goal pyramid proposed by Benjamin Bloom. This goal is both valid and important. But competency-based education should take into account the important role of the Internet in professional life:

- *The use of the Internet should be permitted during learning, research *and* examinations* because it will be employed during day-to-day professional work anyway. E.g., as examinations are increasingly being conducted using computers, students should be permitted to have open access to the Internet, and they should be allowed to retain this access during their examinations as well. Thus, the conditions which prevail during the examinations of students will be comparable with the conditions which they will encounter later on at their workplaces. In addition, one

or two open text questions should be asked so that students can show their research skills and subject matter mastery. Cheating can be prevented by strict examination time limits (no time for asking an Internet buddy via the chat function), by monitoring the Internet connection, or via the use of standardized plagiarism detection systems such as Ephorus.

- *Online research skills should become a part of standard student curricula, be it separately or within a certain subject matter.* This is because nowadays research is being undertaken by everybody and all the time; the material is available through the Internet at the click of a button, be it for leisure purposes, tangible work goals or academic verification. Therefore research techniques (such as the qualitative evaluation of a given information source or the triangulation of sources) have become increasingly important. Here, Wikipedia can serve as a tangible business case, as outlined above.

REFERENCES

- [1] T. Saub and T. Hodel. "Wikipedia vs. Academia". Proceedings of the 11th International Scientific Conference for eLearning and Software for Education, Bucharest, April 23-24, 2015 (these proceedings were accepted for the present journal publication).
- [2] T. B. Hodel and K. R. Dittrich, 2005. "Concept and prototype of a collaborative business process environment for document processing," *Data & Knowledge Engineering*, vol. 52, Special Issue: Collaborative Business Process Technologies, pp. 61-120.
- [3] Aaron Halfaker, Aniket Kittur, Robert Kraut, John Riedl, October 2009. "A jury of your peers: quality, experience and ownership in Wikipedia". Proceedings of the 5th International Symposium on Wikis and Open Collaboration
- [4] Gabriel De la Calzada, Alex Dekhtyar, April 2010. "On measuring the quality of Wikipedia articles". Proceedings of the 4th workshop on Information credibility
- [5] Giles, 2005. "Internet encyclopedias go head to head". *Nature*, Page 438, 7070, 900-901.
- [6] Aniket Kittur, Robert E. Kraut, November 2008. "Harnessing the wisdom of crowds in Wikipedia: quality through coordination". Proceedings of the 2008 ACM conference on Computer supported co-operative work
- [7] Thomas Wöhner, Ralf Peters, October 2009. "Assessing the quality of Wikipedia articles with lifecycle based metrics". Proceedings of the 5th International Symposium on Wikis and Open Collaboration
- [8] Denise Anthony, Sean W. Smith und Tim Williamson, April 2007. "Explaining Quality in Internet Collective Goods: Zealots and Good Samaritans in the Case of Wikipedia". Technical Report TR2007-606, Department of Computer Science, Dartmouth College
- [9] Aaron Halfaker, Aniket Kit-tur, Robert Kraut, John Riedl, October 2009. "A jury of your peers: quality, experience and

- ownership in Wikipedia”. Proceedings of the 5th International Symposium on Wikis and Open Collaboration
- [10] Stacey Kuznetsov, June 2006. “Motivations of contributors to Wikipedia”. SIGCAS Computers and Society , Volume 36 Issue 2
- [11] Sheizaf Rafaeli und Yaron Ariel, 2008. “Online Motivational Factors: Incentives for Participation and Contribution in Wikipedia”. In A. Barak (Ed.), Psychological aspects of cyberspace: Theory, research, applications (pp. 243-267). Cambridge, UK: Cambridge University Press.
- [12] Meiqun Hu, Ee-Peng Lim, Aixin Sun, Hady Wirawan Lauw, Ba-Quy Vuong, November 2007. “Measuring article quality in Wikipedia: models and evaluation”. Proceedings of the sixteenth ACM Conference on information and knowledge management
- [13] Joshua E. Blumenstock, April 2008. “Size matters: word count as a measure of quality on Wikipedia”. Proceeding of the 17th International Conference on World Wide Web
- [14] Sara Javanmardi, Cristina Lopes, July 2010. “Statistical measure of quality in Wikipedia”, Proceedings of the First Workshop on Social Media Analytics