A Brief Review of Studies of Wikipedia in Peer-Reviewed Journals

Chitu Okoli
John Molson School of Business
Concordia University, Montréal, Canada
Chitu.Okoli@concordia.ca

Abstract

Since its establishment in 2001, Wikipedia, “the free encyclopedia that anyone can edit” has become a cultural icon of the unlimited possibilities of the World Wide Web. Thus, it has become a serious subject of scholarly study to objectively and rigorously understand it as a phenomenon. This paper reviews studies of Wikipedia that have been published in peer-reviewed journals. Among the wealth of studies reviewed, major sub-streams of research covered include: how and why Wikipedia works; assessments of the reliability of its content; using it as a data source for various studies; and applications of Wikipedia in different domains of endeavour.

1. Introduction

Although the open source approach has traditionally been applied only to software products, recent years have demonstrated its applicability to the creation of other information products, most notably to the open content Wikipedia, “the free encyclopedia that anyone can edit” (www.wikipedia.org). In just seven years since its establishment in 2001, this comprehensive general encyclopedia has compiled over ten million articles in 253 languages (with over 2.5 million in English). Other open content encyclopedias, though not nearly as well developed, include the Association for Information Systems’ ISPedia (http://ispedia.terry.uga.edu), Enciclopedia Libre Universal en Español (http://enciclopedia.us.es), Wikinfo (http://www.wikinfo.org), and Citizendium (http://www.citizendium.org).

Wikipedia is based on wiki technology [1], a social collaboration Web application that allows viewers to add content to Web pages using minimal technical skills. The submitted content is published under the Free Software Foundation’s Free Documentation License (FDL)—the textual complement to the General Public License for open source software. This license permits anyone to freely copy, modify, and distribute the encyclopedia’s content under two conditions: the original source must be cited, and everyone else must be accorded the same right of free modification and distribution of all derivative works via the FDL.

Similar to open source software, Wikipedia aims to create high-quality digital information products through the participation of large numbers of contributors, mostly volunteers, though some might be paid. It is structured to eliminate or minimize individual agendas and strive towards public or industry welfare in setting policies for the development of the encyclopedia. Their organizational structures are generally rather loose, yet there is some central administration that permits the project to survive and flourish. All along its history, anecdotal evidence has abounded as to Wikipedia’s quality, mainly by referring to its popularity (it is regularly ranked by Alexa as a global top-ten website), and by noting the status of some of the organizations that regular cite it, such as the British Broadcasting Corporation.

However, more recently, Wikipedia has begun to attract the attention of scholars who have attempted to more rigorously study the Wikipedia phenomenon from many different angles. This study attempts to build a base for scholarly research on the subject of Wikipedia by reviewing most of the important scholarly work that has been done thus far. Because the focus here is on high-quality scholarly work, the studies reviewed here are limited to those that have been published in peer-reviewed journals. Thus, work that has been published in conference proceedings is not included, nor has other work that has not passed through scholarly peer review.

I begin by reviewing studies concerning how Wikipedia works, and why it works successfully. Then I examine a large body of research that uses various approaches, both comparative and epistemological, to assess the reliability of Wikipedia. Then there are many studies that refer to, depend on, or focus on Wikipedia as a source for information or data in the study. Finally, a few studies propose and examine
applications of Wikipedia and the Wikipedia concept in education and economics.

2. How and why Wikipedia works

The earliest peer-reviewed articles on Wikipedia were more or less general introductory reviews, introducing readers to the novel idea of an encyclopedia that anyone can edit. More advanced studies began to dig deeper by investigating the motivations of Wikipedians to contribute, and studying the details of the process of developing articles.

2.1. Reviews of Wikipedia

Remy [2] provided perhaps the first review of Wikipedia in a peer-reviewed journal, just a year after its establishment. She considered the free-editing concept dubious, and referred readers rather to Nupedia, the expert-written online encyclopedia that preceded Wikipedia [3].

Hall [4] and Strategic Finance [5] provided reviews that were generally neutral, mainly stressing the novelty of the concept. Krause [6] briefly discussed Wikipedia in the context of online social collaboration, as a prime example of wiki collaboration. McFedries [7] discussed wikis, referring to Wikipedia as the prime example. Morse [8] interviewed Jimmy Wales, the founder of Wikipedia. Wales mainly discussed the workings and value of wikis for companies, and when they were appropriate or inappropriate. Arter [9] reviewed Wikipedia from the point of view of its peer-review mechanism as a means of ensuring quality.

2.2. Motivations for contributing

In the study of open source software, the question of why people code for free has always been a fascinating subject of study. Similarly, a few studies have explored why Wikipedians contribute their free time to the encyclopedia, especially considering that, unlike in the case of open source software, it is extremely rare for anyone to be paid for their efforts.

In their study of public contribution to Amazon book reviews and to Wikipedia articles, Peddibhotla and Subramani [10] found multiple motives for contribution to Wikipedia, both self-oriented and other-oriented. Extending critical mass theory, they found that the higher the quality of contributions, the lower their quantity, and vice versa. Moreover, the motivations for contribution high-quality articles were different from those for contributing a high quantity of articles. Nov [11] found that Wikipedians are motivated to contribute primarily for the fun of it, and for ideological commitment to the project. However, other hypothesized motivation categories such as social reasons, career advancement, and protection, were not found to be very relevant.

2.3. The editorial process

A number of studies have gone further to try to understand the details of how the editorial process of Wikipedia operates to produce high-quality encyclopedia articles. Brandes and Lerner [12] specifically examined the historical progress of edit wars, where controversies lead to mutual revisions of contributions. They presented visual analysis tools to study the progress of such conflicts. Okoli and Oh [13] studied how Wikipedians’ interactions in both direct article creation and in discussions with each other affected their election to the status of “administrator,” which gives greater editorial privileges. Stvilia et al [14] examined the article-creation process in Wikipedia to discover how information quality is assured. They found that Wikipedia’s discussion pages provide an extensive resource of documentation on how the processes of error detection and correction occur. Den Besten and Dalle [15] studied the Simple English Wikipedia, a Wikipedia distinct from the English one that limits its vocabulary sense and grammatical structure to facilitate reading by children and learners of English. They investigated the editorial process that implemented the rules instated to keep articles “simple.”

3. The reliability of Wikipedia

So far, perhaps the most abundant body of scholarly work conducted on Wikipedia has been formal studies evaluating the reliability of Wikipedia, variously expressed as trustworthiness, quality, or accuracy. In other words, why trust the contents of an encyclopedia that anyone can edit? The most famous scholarly assessment of Wikipedia is a comparison of selected science articles in Wikipedia and Encyclopaedia Britannica conducted by Nature journal [16]. The study found Wikipedia’s accuracy comparable to those of Britannica. Among 42 articles, Wikipedia had an average of four errors each, and Britannica had three.

In this section I will first present a large number of negative assessments of Wikipedia’s reliability. Then I will present a number of studies that take an epistemic approach of assessing the reliability of Wikipedia; this perspective unequivocally considers the radical encyclopedia in a very favourable light.
3.1. Criticisms

Denning et al [17] question whether Wikipedia’s collaborative editing process is capable of producing accurate and authoritative information on a thoroughly comprehensive scope of human knowledge. Gorman [18] contends that Wikipedia has no basis to call itself an encyclopedia, and that without regulation of its article-writing process, its information is unreliable.

Fiedler [19] discusses a well-publicized case of a prank entry of John Seigenthaler, Sr., a famous journalist, implicating him in the assassinations of John and Robert Kennedy [20]. This false information remained on Wikipedia for over four months before being discovered. It has become a strong warning against the weakness of a major public information source “that anyone can edit.”

Svoboda [21] discusses various criticisms that some scholars have levied against Wikipedia. They contend that in spite of the attempts of some Wikipedians to provide quality control, the lack of formal controls results in the lowest quality contributions prevailing, with unclear standards of accuracy or writing quality.

Waters [22] contends that Wikipedia articles are at best a sum of thousands of opinions; this process, he says, could not logically result in articles of quality.

3.2. Epistemological analyses

Ironically, whereas most of the general review articles and those that assessed Wikipedia’s reliability have expressed strong doubts about its reliability, all the studies that have scrutinized Wikipedia from an epistemological perspective have strongly validated its epistemic qualities as a valuable source of knowledge. Epistemology, the theory of knowledge, is “a branch of philosophy concerned with the nature and scope of knowledge” [23]. A definite sub-stream of research has begun that explores Wikipedia as an epistemological phenomenon, examining how Wikipedia and related phenomena affect and shape people’s consciousness of how they know what they believe they know. This contrast in evaluation between classical views of knowledge and an epistemological re-evaluation suggests that Wikipedia represents a significant shift in how knowledge is evaluated and received, a shift of “seismic” proportions [24].

Fallis [25] presented a compelling case for the reliability of Wikipedia. He highlighted the epistemic problems with comparing Wikipedia with some conceptually “absolute” sources of knowledge, such as direct evaluation by experts. He argued that it is rather more meaningful to judge the reliability of Wikipedia by comparing it with other encyclopedias such as Britannica. With such criteria, he argued that Wikipedia has been repeatedly shown to be quite reliable [16]. Moreover, when compared to its more likely alternate sources on the Web, Wikipedia is strikingly superior as a source of knowledge [26]. He argued that Wikipedia has important epistemological properties (“e.g., power, speed, and fecundity”) that offset its shortcomings, and thus is an important source of knowledge today.

Dede [24], Eijkman [27], and Matychak [28] took similar perspectives in considering Wikipedia as a prime example of Web 2.0, which characterizes “a shift from the presentation of material by Web site providers to the active co-construction of resources by communities of contributors” [24]. They contrasted the epistemologies of the classical knowledge model of knowledge creation by experts with that of knowledge created by consensus of a community of contributors, proposing that various Web communities present epistemologies between these two extremes.

4. Wikipedia as a data source

Despite the negative assessments of some scholars, the highest approval of the reliability of Wikipedia’s content might be considered to be those peer-reviewed journal articles that use Wikipedia as a source of data. Here I arrange these studies in increase order of their involvement of Wikipedia data: many studies simply cite Wikipedia for definitions and general information; some go further to rely on information from Wikipedia as a major source of data for the study; still others make Wikipedia itself or its content the focus of their study.


Some articles merely refer to Wikipedia as an example of a wiki or of Web 2.0 phenomena. Lamb and Johnson [29] referred to Wikipedia as an example of a wiki, and they also referred to its article, “List of Social Networking Web Sites” as a convenient way to refer readers to more examples. In his discussion of freedom of expression in non-textual media, Soraker [30] referred to Wikipedia as a target for censorship by some nondemocratic governments, due to its textual nature. Lamb and Johnson [31] refer to the article on “Open source software” in describing some attributes of open source software.

Other studies referred to Wikipedia in quoting or developing definitions for key concepts. Shih et al [32] referred to Wikipedia in drawing up a definition of “Web 2.0.” Crowley-Henry [33] cited Wikipedia to define the “First World,” referring to high-income
countries. Chander and Sunder [34] cited Wikipedia extensively and authoritatively as a source of definitions for their paper justifying a particular subgenre of fan fiction as fair dealing (fair use) under copyright law. Olsen [35] referred to the Wikipedia article on “Critical thinking” to define the concept, in a paper arguing that operations research and management science professionals are uniquely suited as teachers to meet the rising trend in leading business schools of beginning the MBA programme with a course on critical thinking.

Finally, some studies drew from Wikipedia articles or Wikipedia itself as a source for ideas. Locander and Luechauer [36] credited the Wikipedia article on Dr. Seuss’s book *The Sneetches and Other Stories* [37] as the partial source for their idea for a paper on the harm of creating needlessly sharp dichotomies between the sales and marketing functions in a business. Swartz [38] referred to Wikipedia as an inspiration for Wikileaks, a blog site that gives a platform for whistleblowers to anonymously release government, corporate, and religious documents that expose unethical, illegal, corrupt and otherwise harmful practices. Wilkinson and Crossfield [39] proposed the Business Genome Project, inspired by the Human Genome Project and Wikipedia, an attempt “to map out the memes and a variety of other characteristics relating to the 75 million or so businesses that exist in the world, with a view to making this information available to support further development in the field” [40].

4.2. Wikipedia as a data source

Some studies have used Wikipedia as a primary source of data, from two general perspectives. While these studies do not take for granted that the content of Wikipedia articles is reliable—encyclopedias by their very nature are not primary sources of data—the content from Wikipedia articles provides major portions of the data for these studies.

In Alton et al.’s [41] analysis of the delayed response in 2005 to Hurricane Katrina in the United States, Wikipedia was one of many sources of documentation on details of the response. Bar-Ilan [42] studied a case of “Google-bombing,” where the top results to the search keyword “Jew” yielded the Wikipedia article and an anti-Semitic website. In a short paper about the uncertainty of the accuracy of information from the Internet for controversial questions, Buck [43] refers to Wikipedia as one source of information on the question of whether or not Hitler had Jewish ancestry.

Kim et al [44] proposed an approach to merging and matching ontologies, using Wikipedia’s philosophy ontology as an input ontology, along with oriental, western, and the Yahoo philosophy ontologies.

4.3. Wikipedia as the primary data source

In addition to the preceding studies, which use Wikipedia as one significant source of data among others, a number of studies use Wikipedia as the sole data source and the entire data focus of their study. Although several scholars still question Wikipedia’s reliability and credibility, it is significant that many other scholars have come to regard Wikipedia as a phenomenon and data source worthy of research and publication in peer-reviewed journals.

Wagner [45] analyzed knowledge acquisition in Wikipedia. By tracing the historical development of 80 articles, he found that the wiki approach to building knowledge facilitates the knowledge acquisition goals of knowledge management. Abraham et al. [46] presented a fictionalized case study of a sociology student discovering Wikipedia while searching for Internet information for a paper on microcredit. Spinellis [47] analyzed Wikipedia’s MediaWiki wiki platform (rather than Wikipedia itself) and found many instances of successful reuse of software.

Brandes and Lerner [12] produced a visual analysis of edit wars, when contributors with opposing points of view repeatedly revise each other’s entries. Nikolaos et al [48] investigated the development of quality articles in Wikipedia by using social network analysis to determine the authoritativeness of articles. Okoli and Oh [13] calculated the social capital of Wikпедiа who participated in article creation, within the social networks of their co-collaborators on various articles. Spinellis and Louridas [49] studied the process of the development of Wikipedia articles. In their study of the nature of the collaboration that operates to assure information quality in Wikipedia, Stvilia et al [14] accessed Wikipedia’s extensive discussion pages to observe the nature and directions of the conversations between the contributors. They found Wikipedia particular valuable in the richness of its textual data in documenting the process of discovering and correcting errors in the information. The lessons learned from the study of Wikipedia are applicable to other textual databases.

5. Applications of Wikipedia in education and economics

Some studies focused on demonstrating how Wikipedia concepts could be applied in certain fields
of endeavour. The fields that have been researched on thus far are mainly education and economics.

For education, Achterman [50], McPherson [51], and McPherson [52] referred to Wikipedia as a model for assigning schoolchildren projects to create encyclopedia-like content using wikis, building their literacy, and as a tool for them to receive collaborative feedback for their writing projects. Focusing on professors rather than on students, Black [53] examined the traditional academic process of peer review, highlighting many of its shortcomings, particularly those that stifled the creation and dissemination of knowledge. He posited that a new open model of peer review, such as that afforded by a Wikipedia-like mechanism, could result in more knowledge disseminating to the world.

From an economics perspective, Gloor and Cooper [54] describe Wikipedia as an example of what they call “swarm businesses,” businesses that derive value from the collaborative efforts of large numbers of contributors, including internal employees and members as well as external customers and partners. Hemphill [55], in criticizing calls for regulating Internet access providers by forbidding them from charging application or content providers for delivering content to consumers, argued that although Wikipedia competes with fee-based encyclopedias, the argument for regulation in its case is nonetheless narrow.

6. Conclusion

This paper has presented some important works in the rapidly growing body of research that has focused on the phenomenon of Wikipedia, the free encyclopedia. The broad range and high quality of work devoted to this topic indicate that many scholars consider it a significant area of inquiry. By further understanding how Wikipedia works, many insights will be obtained that will help to maximally benefit from not only Wikipedia, but also other open content and open source projects.

7. References


