

Where's the data from? Information source uses Wikipedia models

Bernice C. Crowell †

† Department of ALM, Uppsala University, Box 625, SE-751 26 Uppsala, Sweden

Paper type: Research article

Abstract

Introduction. Little is known about Wikipedia contributors' information behaviour and from where and how the information in the encyclopaedia originated. Even though a large number of texts in Wikipedia cite external sources according to the intentions of the verifiability policy, many articles lack references and in many others the references have been added afterwards.

Method. This article reports the results of a Web survey of information source use patterns, answered by 108 Wikipedia contributors in spring 2008.

Analysis. The qualitative questions were analysed using a close reading and grounded theory approach. The multiple-choice questions were analysed using descriptive statistics and bi-variate correlation analysis.

Results. The results indicate that there are several distinct groups of contributors using different information sources. The results also indicate a preference for sources available online. However, in spite of the popularity of online material a significant proportion of the original information is based on printed literature, personal expertise and other non-digital sources of information. The information source use of Wikipedia contributors is also illustrative of the complexity and life-world scope of human information behaviour.

Conclusions. Understanding the information source use of contributors helps us to understand how new Wikipedia articles emerge, how edits are motivated, where the information actually comes from and more generally, what kind of information may be expected to be found in Wikipedia.

Introduction

Blindness and sight impairments are common disabilities in all countries of the world. In 1996, the Wikipedia is a free and open collaborative online encyclopaedia based on a collaborative writing tool called Wiki (Leuf & Cunningham, 2001). Wikipedia was originally started as a redevelopment of an earlier online encyclopaedia called Nupedia. The Wikipedia project was started in 2001 and in 2003 the Wikimedia Foundation was formed as an independent institution to maintain and develop open content Wiki-based projects (Voss 2005). From the start, the Wikipedia project received considerable attention in the media (Remy 2002; Levack 2003) and has become one of the most popular Web sites in the world (Rainie & Tancer 2007).

The three fundamental policies guiding the development of Wikipedia are the "Neutral Point Of View" (Wikipedia 2008l), "No Original Research" (Wikipedia 2008m) and "Verifiability" (Wikipedia 2008l). The Neutral Point of View policy acknowledges the existence of multiple viewpoints and seeks to attain a fair (Wikipedia 2008l) representation of all contradictory viewpoints. The second, No Original Research, policy states that all information has to be attributable to an external published source to be considered reliable (Wikipedia 2008m) by the contributor. The third policy underlines the policy of verifiability by an external source (Wikipedia 2008l) and that the role of Wikipedia is to provide a comprehensive summary of existing knowledge rather than functioning as a source of new findings.

Although a large number of texts in Wikipedia cite external sources to meet the expectations of the verifiability policy (Wikipedia 2008l), articles may temporarily lack references if an editor is unable to find an appropriate one at the time of writing. If there are no ethical concerns, individual facts can be marked temporarily as requiring reference. There are also several projects for referencing previously unreferenced articles (Wikipedia 2009, 2008n). Even if the original contributors give references to their contributions at a later date, the original source of information and the cited reference need not to be the same. Therefore, it is not always altogether clear where the information presented in the articles actually originates. The Mediawiki system used in Wikipedia gives documentation for actual edits, but is obviously unable to document the sources of information used or decisions made by contributors if they are not explicitly given. Besides the occasional lack of references (Wikipedia 2009, 2008n), the pattern of providing references only after the first versions of an article have been written, reduces the transparency of the original sources of information and clouds the pattern of how the article texts emerge. In these cases the original source is not necessarily the cited one. The verifiability of individual facts does not necessarily suffer from this procedure if a reference can be found later. The viewpoint and emphasis of an article (i.e., what is generally important or interesting about the topic) and the contexts of individual facts (i.e., how one fact relates to others) might become obscured.

The use of dubious sources is one of the criticisms of Wikipedia (Wikipedia 2008a) although explicitly contrary to the official Wikipedia policy (Wikipedia 2008l). Several Wikipedia projects are working on inserting references to articles lacking citation mark-up and improving the quality of existing citations (Wikipedia 2008m,q). There is a reported general increase in the use of formal citation markup and agreement with scientific citation patterns (F. Å Nielsen 2007).

Currently there is little research available on information source use in Wikipedia. Nielsen notes that there is a slight tendency to cite articles in high-impact factor journals in Wikipedia in comparison to scientific citation patterns (Nielsen 2007) and Lih (2004b) underlines the importance of news sites as a source of information in Wikipedia. Understanding the information sources used by contributors could help to understand how new Wikipedia articles emerge, how edits are motivated, where the information actually comes from and more generally, what kind of information may be expected to be found in Wikipedia. The purpose of this article is to explicate the patterns of information source use of Wikipedia contributors and their implications for the contents of Wikipedia.

This paper reports results of a Web survey of information source use patterns, answered by 108 Wikipedia contributors in spring 2008. The results indicate that there are several distinct groups of contributors using different information sources. The results also indicate a preference for sources available online. However, in spite of the popularity of online material a significant proportion of the original information is based on printed literature, personal expertise and other non-digital sources of information. The findings shed light on information behaviour on the social Web and more precisely to the information source use of Wikipedia contributors. The second contribution of the study is in the field of information behaviour studies. The behaviour of Wikipedia contributors is an example of the diversity of strategies of finding and working with information and how different professional, hobby-related and personal interests and motivations coalesce into a single activity.

Information behaviour and information source use

Wilson has defined information behaviour as the totality of human behaviour in relation to the sources and the channels of information, including both active and passive information seeking, and information use (Wilson 2000: 49). Wilson sees information seeking behaviour as a constituent of information behaviour (Wilson 1999). Information source use, i.e., the choice and use of different information sources, can be seen similarly as a constituent of the broader phenomenon of seeking and using information. In fact, like the present study, the majority of the existing literature has focused on information seeking behaviour and information source use instead of a broader notion of information use (Wilson 1994; Taylor 1991). Case (2006) presents an overview of the recent research directions of information behaviour in an ARIST review.

Information behaviour depends on the context of information activity (e.g., Julien & Duggan 2000; McKechnie et al. 2002; Tenopir et al. 2003). Even though the context of activity, profession and disciplinary affiliation affects information behaviour, Heinström (2006) has pointed out that personality can be more significant explanatory factor of the differences in information behaviour between individuals than contextual factors such as the subject students are studying. Also Hyldegård's (2009) findings in a group-work context emphasise that personality traits have explanatory power on information behaviour.

In contrast to the majority of information behaviour research which is domain and context specific (Case 2006), Chatman (1991; 1996), Solomon (based on Dervin) (1997b; 1997a) and Wilson (2006), for instance, have emphasised that information behaviour is a complex life-world phenomenon. Approaches related to the general phenomenon of browsing (Case 2007: 89) such as accidental information encountering (Erdelez 1997), everyday life information behaviour (Savolainen 1995) and information foraging (O'Connor et al. 2003; Pirolli 2007) have emphasised that information is discovered in significantly more diverse and subtle ways than as purposeful fixes to explicit needs. Also, the recent studies on information behaviour in digital environments underline that personal preferences and expectations have a significant influence on how information is sought and used (e.g., Rowlands et al. 2008; Hargraves 2007). Regarding overarching proposals, Foster (2004), for example, has formulated a "non-linear" model of information-seeking behaviour and Spink and Cole (2006a) presented a tentative version of an integrated model of the information behaviour, which combines the three typically used approaches: everyday life information seeking, information foraging and the problem-solution perspective.

Earlier research on Wikipedia

The volume of Wikipedia related research has grown fast recently (cf. Ayers 2006). Wikipedia has been discussed frequently in the literature as an exemplary instance of the social Web phenomenon (e.g., Kolbitsch & Maurer 2006; Quiggin 2005), but there has also been an increasing number of specific Wikipedia related studies and research interest (Wikipedia 2008j; Wikimedia Foundation 2008b; Wikipedia 2008e; Wikimedia Foundation 2008d). Initially much of the discussion was focused on the risks of using Wikipedia (Denning et al. 2005). Thereafter the basic model of Wikipedia, its usefulness, biases, policies, information sources, exposure to vandals and political advocacies, privacy and quality, anonymity of the editors, diverse administrator and editor actions and copyright issues have been a subject of critique (Ayers 2006; Wikipedia 2008a). The use of Wikipedia as a reference is one further area of research (Ayers 2006), however, outside the direct scope of the present study. In spite of the evident doubts of impartiality, one of the most comprehensive sources of Wikipedia related criticism and debate is Wikipedia itself (see Wikipedia 2008a).

Accuracy and quality

The accuracy and quality of Wikipedia articles has been the most debated and consequently the most studied aspect of the collaborative encyclopaedia (Economist 2006a,b,c,d; Giles 2005; Korfiatis et al. 2006; Wikipedia 2008b). Generally Wikipedia has been recognised as reliable and typically more comprehensive than its traditional counterparts. Emigh and Herring conclude that Wikipedia maintains almost as high a standard of contributions as its traditionally edited print counterparts, although the editorial control in print encyclopaedias tend to make articles more formal and standardised (Emigh & Herring 2005). Brändle has concluded that such issues as topics of articles, anonymity of contributors and vandalism do not have a major influence on the overall quality of the German language Wikipedia (Brändle 2005). On an individual article level, however, significant omissions and errors may exist as the general disclaimer of Wikipedia clearly emphasises (Wikipedia 2008i).

In contrast to traditional encyclopaedias, a significant feature of Wikipedia is that the discussions on the quality of its content are accessible to the general public on the Wikipedia site and the discussion is directly related to the available Wikipedia data (Stvilia et al. 2008). Because of the transparency of the quality assurance process, a similar approach has been suggested to function as a model for revision of current peer-review system in sciences (E. W. Black 2008).

Lih (2004b), Korfiatis et al. (2006) and Dondio et al. (2006) have developed automated measures of estimating the quality of Wikipedia articles. Lih (2004b) suggests a rather simple method of picking the number of edits and number of individual users editing an article to establish a benchmark. Wilkinson and Huberman (2007) and Kittur and Kraut (2008) confirm the assumption that the articles with most editors tend to be of higher quality. Kittur and Kraut (2008), however, specify that the greater number of editors is beneficial only if their work is coordinated effectively either explicitly through communication between contributors or implicitly by the way they are editing pages. Korfiatis et al. (2006) have based their method on a network analysis approach. The Dondio et al. (2006) method is based on automatic calculation of eight different factors on the basis of the existing Wikipedia data on articles and contributors including their talk and history information. Brändle (2005) suggests more complex quality criteria based on content analysis of the articles.

Brändle (2005) found in his study that a wiki system works only if it is relevant for a critical mass of users and gathers enough attention from contributors. Anthony et al. (2007) suggest that the number of contributing editors, not only their quality, does count in the quality of articles. In contrast to Anthony et al. (2007), Stein and Hess (2007), in a study where they looked at featured articles (articles voted to be of exceptional quality) in the German language Wikipedia, concluded that the quality of Wikipedia articles depends more on who contributes than how many contribute. The excellence of users was measured by the number of excellent articles to which they had contributed (Stein & Hess 2007).

The anonymity of users has only a limited effect on article quality. Another interesting observation made in the report is that although the quality of contributions made by registered users is high, the highest quality contributions come from non-registered users (Anthony et al. 2007). Brändle (2005: 109-110) notes that anonymous users do not have a negative effect on article quality in the German language Wikipedia as long as the ratio of anonymous edits does not exceed 50 percent.

An often cited example of Wikipedia unreliability (Dondio et al. 2006; Wikipedia 2008a; Keen 2008: 40-41) is the Seigenthaler incident where a former administrative assistant to Robert Kennedy learned that in a Wikipedia article about himself he was mentioned as briefly having been suspected of involvement in the assassinations of John F. Kennedy and Robert Kennedy. The incorrect information was removed, but it was only discovered more than four months after its addition (Terdiman 2005). The Seigenthaler incident exemplifies the possibility of issues related to individual articles. Similarly, it illustrates a phenomenon observed by Lih (2004b) that the quality of Wikipedia articles tended to improve after an article had been cited in the press.

Another much debated showcase of Wikipedia reliability is the study of Giles published in *Nature*, which showed that experts were able to pinpoint a number of errors both in Wikipedia and in corresponding *Encyclopaedia Britannica* articles (Giles 2005). The study has been debated and questioned thereafter respectively by *Encyclopaedia Britannica Inc.* (2006), *Nature* (2006) and Wikipedia (2008h). In spite of the criticism, the study showed that, in general, Wikipedia may be considered to be fairly reliable and not substantially less reliable than its traditional counterparts. Also Chesney (2006) and Fallis (2008) report results that support the general credibility of Wikipedia articles.

In spite of the occasional doubts on article quality, the importance of Wikipedia as a publishing channel has started gaining more ground. The scholarly journal *RNA Biology* has been cited in the press that in future, authors will be required to submit a summary of their published article for Wikipedia. The Wikipedia summaries are peer-reviewed by the journal like the journal articles (Butler 2008). Another example of its perceived importance is a project conducted by the University of Washington Libraries with an aim of inserting links to the library collections from relevant Wikipedia entries in order to make the collections more findable (Lally & Dunford 2007).

Users and contributors

The cumulative number of people editing Wikipedia and the number of individual edits have grown steadily. According to statistics available on the Wikipedia page on editing frequency (Wikipedia 2008f), the peak of editor activity seems to have been at the end of 2007. The statistics are not entirely reliable, because only explicitly marked bots (automatic editor programmes) are excluded from the figures. The statistics should still be indicative of the general state of affairs, because generally speaking, bots are relatively easy to discover using measures similar to vandalism detection (Potthast et al. 2008).

Wikipedia can be edited both by using a registered user name or unregistered using the IP-address of the editor's computer as identification. The number of anonymous edits has increased simultaneously with the overall increase of edits although with a greater variance of growth (Wikipedia 2008f, Article_space_anons). Around a quarter of the edits in Wikipedia are made by contributors who have been active less than 100 days (Voss 2006a). Not all new editors are, however, new as such, because some contribute for a long period of time without registering and continue their participation later as registered users, which may obscure the statistics.

Bryant et al. (2005) identified several different user roles among active Wikipedia users. Most of the users write and edit articles, but there are plenty of other roles. Lurkers browse articles and make occasional anonymous edits. Administrators focus on answering questions by other contributors and cleaning up articles. Mediators try to discover extreme opinions in controversial topics and to formulate an article out of the discussion. One informant called herself a meta-user, someone who touches an article quite rarely and focuses more on other tasks within Wikipedia. A welcome committee goes to new users' talk pages and writes a message saying "Welcome". According to Bryant et al. (2005), new Wikipedia users tend to act as editors in topics they are familiar with. Later, after gathering more expertise, contributors tend to focus on administrative tasks and building the Wikipedia itself.

In spite of the optimistic suggestions (e.g., Tapscott & Williams 2006) that an Internet connection is enough to make people contributors in Wikipedia, Rask (2008) has shown that there is a stronger correlation between the level of human development and Wikipedia participation than between the latter and the level of technological development. Hargittai and Walejko (2008) have published similar findings on the factors affecting participation behaviour on the Web. The findings of Kittur et al. (2007) and the Wikipedia data on editing frequencies (Wikipedia 2008f) show that the number of contributions by individual editors follows approximately a power law (most edits are contributed by a very small group of individuals) (M. Newman 2005), although as Kittur et al. (2007) note, the influence of the elite might be weakening. Another interesting partially reported investigation by Schwartz (2006) suggests that outsiders mostly contribute the new text paragraphs. After a new paragraph is inserted the active Wikipedia insiders swarm in formatting and copy-editing the contribution (Schwartz 2006). This pattern is likely explain why the edit impact factor calculated by Priedhorsky et al. (2007) shows that versions edited by active Wikipedia editors are getting the highest number of views.

It has also been noted that not every Wikipedia user becomes a contributor. According to Pew Internet and American Life Project study, in 2007, 36% of adult Americans used Wikipedia (Rainie & Tancer 2007). Ortega et al. (2008) have shown further that approximately 90% of all surviving edits are written by 10% of contributors. Most Wikipedia users are not contributors and most of the contributors contribute very little. This general pattern seems to persist in different language editions although there are observable differences. Similarly, the pattern persists over time (Ortega et al. 2008). In another study, Ortega and Barahona (2007) also suggest that the number of active contributors and their patterns of activity change over time. In a small experiment, Schwartz (2006) counted the number of letters, which have survived to the most recent edition in a selection of articles in Wikipedia using a downloaded copy of the Wikipedia archives. In his preliminary findings, most of the surviving letters tended to have been written by contributors who made only a few edits to an article. Schwartz gives the Anaconda article as an example. The largest proportion of this article was written by a user who edited the article only twice in comparison to others who made thousands of edits.

Editing patterns

The patterns of editing Wikipedia articles have several distinct characteristics. According to Jones, the special structural demands placed by Wikipedia on its writers, might lead to unique revision patterns (J. Jones 2008). Viegas et al. (2004) developed a method called history flow to visualise these patterns. Besides studying the overall editing activity, Viegas et al. have noted that there are significant differences as to how much individual anonymous editors contribute to individual pages. The researchers could not, however, identify clear patterns or clusters of pages, which were more likely to be edited by anonymous users (Viégas et al. 2004). For instance, like Brändle (2005), Viegas et al. (2004) could not discern a clear correlation between the anonymity of users and vandalism.

Even though there is no clear statistical correlation, a small number of Wikipedia users referred to by Viegas et al. (2004) stated that they rely on authorship information when they browse Wikipedia trying to pinpoint anonymous and first time users who tend to represent a potential source of vandalism. While some stress that authorship is important in wiki projects, the neutral point of view (NPOV) policy and the collaborative form of Wikipedia articles is not in line with the so called thread mode editing (Cunningham & Cunningham, Inc. 2005). On talk related to other pages, however, the thread mode is the standard code of conduct (Viégas et al. 2004).

Although user activity has been one of the major concerns of investigation, various studies also present other observations and trends on the editing patterns. Lih (2004b) points out that edit peaks of different articles tend to coincide with specific events and news coverage on the topic. Viegas et al. (2004) found that the first text entered to an article tends to survive longer than later additions. Another observation made was that contributors tend to delete or add text instead of moving it. Besides content related trends, Pfeil et al. were able to distinguish clear cultural differences in editing Wikipedia pages in the German, French, Japanese and Dutch language versions of the encyclopaedia (Pfeil et al. 2006).

Summary of earlier research

The most central findings of earlier Wikipedia research within the scope of the present study relate to its diversity. Generally speaking, Wikipedia articles are of comparable quality to a printed encyclopaedia. The quality of individual articles, however, varies from low to high and depends considerably on individual contributors. The assets of Wikipedia are the transparency of its quality assurance and editing process. Also the studies of Wikipedia contributors and editing patterns strengthen the image that the role of individuals is significant. The number of the most active Wikipedia contributors is very low in comparison the total amount of Wikipedia users. The influence of the first and the most active contributors is generally very high. The amount of activity itself does not, however, fully reflect the influence or role of the individuals. Some contributors focus on writing new articles while some others have focused on editing minor details or, for instance, on dealing with administrative issues. It may be concluded that earlier studies have been able to discern a wealth of general patterns, but in many cases it seems that it is still the details that have the greatest explanatory power.

Empirical study

Aims and method

The purpose of this article is to explicate the patterns of information source use of Wikipedia contributors (what information sources are used and how) in order to discern patterns of activity and origins of Wikipedia article contents. This aids in understanding of how new Wikipedia articles emerge, how edits are motivated, where the information comes from and what kind of information may be expected to be found in Wikipedia.

The question was studied using a Web survey planned and executed considering the Association of Internet Researchers ethical guidelines (Ess & AoIR ethics working committee 2002) and earlier literature on Internet survey research (Deutskens et al. 2004; Kaplowitz et al. 2004; Yun & Trumbo 2000). The purpose of the survey was to gather data on the information source use patterns and actual sources used in Wikipedia articles. An invitation was sent to a convenience sample of Wikipedia contributors. There were two methods of recruiting the sample: (Group 1) An invitation to participate in the study was sent to wikiresearch, wikipedia-l and wikien-l mailing lists, to authors' personal blogs

and the survey was promoted using personal contacts; (Group 2) personal invitations were sent individually to eighty-three contributors who had made a non-minor edit to Wikipedia articles on April 29 2008 between 6 a.m. and 2 p.m. UTC and agreed in their user preferences that they may be contacted by e-mail using the internal mailing facility of Wikipedia. The facility allows registered users to send e-mail to other registered users without revealing their e-mail addresses to each other, thus protecting the anonymity of the recipient. The second method was used to get more data (N=63 after the first recruiting method) and to address an assumed overrepresentation of administrators (Schroer & Hertel 2007) and the consequent problems with the validity of the findings.

The latter approach of recruiting participants resulted a response rate of 52% (45 of 83) and as assumed, a more balanced representation of those contributors who do not participate in other Wikipedia activities than editing of the articles (see next section). The total number of informants who participated the survey was (N=) 108.

In the second sample, two of the contacted contributors expressed their displeasure at receiving the personal invitation and expressed their general concern about Wikipedia contributors being contacted in the described manner. The present study confirms the remarks of Voss (2006c) on the problems associated with the validity and acceptance of surveys in Wikipedia research. The first recruiting method produced a significantly biased sample (see next section). However, the latter approach of recruiting respondents assumed in the present study cannot be recommended either for future studies in Wikipedia context for community reasons (Ess & AoIR ethics working committee 2002) even though the approach itself is conventional in social science research. It is clear, however, that some effective means to reach at least relatively broad and representative samples of the Wikipedia contributor population are needed in future scholarly studies on both Wikipedia contributors and the contents of the encyclopaedia. The freely available data on committed edits and editorial discussions on Wikipedia and related fora are valuable in many respects, but it is only consequential evidence on many aspects of Wikipedia including contributor motivations and decisions, information source use, producer quality of the contents and the functioning of the community.

The qualitative survey approach was assumed in order to capture data not otherwise available. With the chosen approach, it was possible to reach a larger sample than by using more labour-intensive qualitative methods and at the same time, to get deeper insights (than with purely quantitative methods) into a research area that has not been studied before. For instance, interviewing an equal number of informants would not have been feasible for practical reasons and with little earlier information on the topic, planning a good quantitative survey would have been impossible. Similar approach has been used in earlier studies with similar aims than the present study (e.g., Patton 1987; Sammons & Speight 2008). Not all articles in Wikipedia have references and in many cases the existing references have been added later. In these cases the original sources of information are not necessarily the same as the cited ones. One further problem with referring to the existing references is that they reveal little of the motivation and complexity of information seeking processes leading to a citing.

The survey consisted of twenty-one questions (fourteen qualitative, open-ended questions and seven multiple choice questions) and was conducted using open source survey software LimeSurvey version 1.70+. After basic demographics, the respondents were asked about four of their latest contributions to Wikipedia (name of the article and a description, as precise as possible, where and how did you learn the information you contributed) excluding administrative edits, spelling corrections and similar technical changes. Respondents were also asked general questions about their information source use (what are the most useful sources), information seeking practices (whether the respondent seeks information explicitly for Wikipedia contributions, do they tend to ask someone for help), whether they tend to edit and write articles on familiar or unfamiliar, professional or hobby related topics, and their perception of quality and correctness (do contributions need to be absolutely correct, how did the respondents assess the quality of information in Wikipedia).

The demographic questions were developed using earlier Wikipedia user studies as a reference. The open-ended questions were designed to capture rich qualitative data on the information source use and information behaviour of the respondents.

In the survey, and consequently in the processing of the data, the focus was on open-ended questions and qualitative analysis. As a primary method of analysis, the descriptions of information source use from the survey were studied using close reading (DuBois 2003), a conventional method of critical analysis of texts and grounded theory approaches (Glaser & Strauss 1967). In the context of grounded research, the descriptions were analysed as texts in a similar manner to the interview answers. During the close reading of responses, several strategies of contributing and using information and five contributor types resulted in the constant comparison of how the respondents described their information source use. The evident limitations of qualitative data (Patton 1987) were acknowledged during the analysis. The multiple choice questions were analysed using basic descriptive statistics and bi-variate correlation analysis (Pearson and Spearman's rho) with 2-tailed significance in SPSS 15.0. The quantitative data is referred to whenever significant findings were made, otherwise the findings are based on the qualitative analysis.

Data

The representativeness of the sample is difficult to estimate as no comprehensive statistics of Wikipedia contributors exist (a survey of 106 contributors to the German language Wikipedia was carried out in 2005 (Schroer & Hertel 2007); the survey of 2,252 Hebrew language Wikipedia participants was directed at unregistered non-contributing users (Wikimedia Foundation 2008a); Hitwise carried out a survey of Wikipedia users and contributors (Rainie & Tancer 2007; Tancer 2007)). In general the present findings are in line with Schroer and Hertel (2007) and Hitwise studies (Rainie & Tancer 2007; Tancer 2007).

Of the initial sample (Group 1) of (N=) 65 contributors 48 (74%) were male and 13 (20%) female; 4 (6%) gave no response to this question. The median age group was 31-39. 58% (47) percent had at least a bachelor's degree. The most typical professions were student and software engineer and their related occupations. In Group 2 (N=43), 42 (98%) were male and only 1 (2%) female. Twenty-two (51%) had at least a bachelor's degree. Median age was 31-39, the same as in Group 1.

In the final sample (N=108) 90 (83%) were male and 14 (13%) female, 4 (4%) gave no answer, 69 (66%) had at least an undergraduate degree. The median age was similar to the initial sample (Group 1) although the average age was slightly higher. The most typical professions were student and various academic occupations. Otherwise, the variety of indicated professions was noticeably broader than in the initial sample. The sex distribution of the second sample was closer to the figures of Schroer and Hertel study (88% male, 10% female Schroer & Hertel 2007) than the Hitwise survey where 60% of the contributors and 50% of users were male (Tancer 2007; Rainie & Tancer 2007).

As Schroer and Hertel (2007) note, the mailing list recruiting method is likely to cause a significant overrepresentation of administrators (37% had administrative rights (Schroer & Hertel 2007) in comparison to 2% of users having administrative rights in the English language Wikipedia (Wikipedia 2008c) and 0.05% in the German language edition (Wikipedia 2008d) studied by Schroer and Hertel). This is likely to be true with the initial sample. The second sample may be expected to be more balanced especially because of the increase in the variation of professions and the significance of personal interest in the contributions. It is, however, to be noted that the typicality of contributions and contributors is somewhat controversial because the activity in terms of the number and length of contributions differs considerably among the contributors (Ortega et al. 2008).

The higher education level of the second sample may be coincidental or, for instance, caused by the fact that scholars are more likely to take surveys conducted by other scholars. Another limitation of all samples of Wikipedia contributors is that their reliability is impossible to estimate due to the lack of comprehensive demographic statistics of contributor population.

In summary, in spite of the known and unknown biases of the sample, the data provide indicative evidence of the information source use patterns of Wikipedia contributors.

Findings

Patterns of contributing

The analysis of the data shows that there is no one pattern of contributing, seeking information or participating in Wikipedia. One of the informants commented the survey making a similar remark on the basis of his/her own experience:

The wikipedia contributors, even the serious ones are a very heterogeneous group. For example, there are some who only do copy-editing of articles, who do not need any references, or even those who do what is called as background work such as fighting vandals, who deface authentic articles with misinformation. Such people when included in your sampling may give spurious results. (respondent R71)

Even though variation seems to be a rule, some general patterns are prevalent in the material. All but four informants seemed to focus on editing existing articles by adding details or at the most individual paragraphs. Only four respondents (R3, R21, R26, R27) indicated that the contributions they listed in the survey were new articles. Correcting details such as dates, adding names and places, rewriting unclearly expressed passages and adding references is common. Some contributors (two in the present material, R13, R33) focus on translating and moving content from different Wikipedia language versions to another, there are contributors who mostly do copy-editing (as stated by the cited informant) and those who concentrate on administrative background work. Yet there is also a group of contributors (two contributors in the present study) who have focused their entire attention on writing original articles from scratch on different topics either based on their expertise or personal interest. This is conceivable because articles are started only once and edited several times, 17.87 times on average (Wikipedia 2008c).

Fifty percent of the respondents who answered a question on the types of topics (personal, professional, familiar, unfamiliar) they edit, said that they tend to edit articles only on familiar topics. 15.4% indicated that they edit articles on areas on which they consider they have specific expertise. 34.6% specified that they tend to edit all kinds of articles and they seek information specifically for their contributions.

Over half of the respondents (54.0%) indicate that personal curiosity is important or very important when they seek information for Wikipedia. The motivation is, however, only seldom related to a practical need outside the context of Wikipedia. 18.0% answered that their contributions can be classed to some extent as by-products of their everyday life and work related problem solving. The majority (67.3%) indicated however that their topics of choice relate always or occasionally to their professional interests. 47% said that they mostly contribute on topics relating to their hobbies. For motivation there seems to be no perceivable patterns. They did not correlate significantly with each other or any other variables. There was a significant correlation between the familiarity of the topics edited (whether the contributor saw him/herself as an expert of the topic or started from the beginning) and that the information seeking for the contribution was also motivated by a hobby related (Spearman's rho ,528, Sig. 2-tailed 0.000) or other practical need (Spearman's rho ,508, Sig. 2-tailed 0.000).

The following groups of contributors (Table 1) could be discerned from the material by close reading and a grounded-theory-based analysis of the responses to the questions in which the informants were asked to describe the sources of information used for individual edits. The percentages indicated in Table 1 refer to the classification of the respondents on the basis of their answers. It needs to be emphasised that the figures are indicative of the current data only and are not suggested to be representative of the entire population of contributors. In the present data, there was no significant correlation between demographic variables and group membership although some generic tendencies could be observed.

Group	Description
Investigators	Contributions relate to personal interest or hobby related area (of expertise) based mostly on news sources, popular scientific or fact literature and/or visiting the local library. Some of them are following or browsing a distinct source (e.g., newspaper historical archive on the net) or a set of sources and contributing on interesting, possibly earlier unfamiliar, topics they find. They tend to have a distinct interest (a specific hobby like topic such as modern history or religion). They represent the hard core of contributors who start articles and make considerable contributions to existing ones. Investigators do in depth research on a topic of their interest and contribute their findings in form of new articles or sections in existing articles. They do not necessarily edit individual articles multiple times. Members of the group were mostly graduates, professionals working on topics other than those to which they are contributing. (38%)
Surfers	Contributions are based on easily findable sources available on the net. Surfers spend their time on using search engines and finding fitting material for articles. They search and browse the net for individual, often unrelated, topics pertaining to their personal life and sphere of experience (e.g., topic relating to their own nationality, a place visited). Personal curiosity seems not be as strong motivation for the <i>surfers</i> as it is to the <i>investigators</i> . Their personal interest on the topics they are editing is similar to the group of <i>investigators</i> , but their approach to information seeking is different. The seeking process is guided by ease of access, satisficing kind of searching and economy of effort rather than time consuming in-depth research work. Surfers are primarily secondary school educated, undergraduates and professionals. (32%)
Worldly-wise	These contributors tend to focus on topics relating to their own sphere of experience and knowledge. They do not tend to seek information explicitly for their Wikipedia contributions and tend to rely on serendipitous information seeking and information discovery. Background and the level of experience vary. (15%)
Scholars	Contributions on an academic or professional area of expertise. The archetypal contributor in this small, but quite distinct, group is a PhD student or a relatively young researcher who is contributing on the topics related to their research. PhD students generally need to familiarise themselves with a large amount of literature and topics related to the principal object of inquiry (e.g., Barrett 2005). Because not all of this information is used in the PhD thesis, it is possible to imagine that Wikipedia serves from an altruistic point-of-view the function of being an alternative publication channel for publishing insights from the secondary literature. (11%)
Editors	Some of the editors focus on administrative tasks, grammatical corrections, correction of inconsistencies between articles, and another group on translations from other language versions of Wikipedia. They do not generally seek information for their Wikipedia edits. The group was very small and rather heterogeneous in the present study, but they shared, broadly speaking, a professional background and a college level education. In the present material, the <i>editors</i> were teenagers. (4%)

Table 1: Groups of Wikipedia contributors according to a qualitative analysis of the research data.

Although the groups were relatively distinct, there was some individual variation and minor overlap, which the following comment exemplifies. In individual overlapping cases, the respondents were classified according to the primary pattern of using information sources indicated by their responses. One of the informants (R39) distinguished her professional and hobby contributions from each other.

My edits on the English Wikipedia are usually professional edits with information from research articles, while edits in my native language are "hobby" edits with information from news articles, books, etc. (R39)

The general diversity of information use of most of the informants shows similar patterns. Contributors seek familiar information in a different manner than they seek unfamiliar information. Similarly their knowledgeability of different topics varies in depth and detail. Especially in the group of worldly-wise, the respondents reported that their contributions were based both on an information they knew by heart and on information they knew about i.e., they knew where an answer could be found.

The contributing patterns were somewhat different in the initial sample (Group 1) than in the second one (Group 2). The results indicate that in the second group the contributions were oriented more typically towards the personal interests and profession of the contributors than to the general editorial issues. The initial group seemed to lean more towards the Wikipedia philosophy of collaborative writing and iterative improvement of articles instead of absolute correctness of all data in the first place, which seemed to be a more prevalent view in the second group. In the initial sample, the responses to the open-ended questions tended to be long and detailed. Similarly, the respondents often referred to administering procedures and official quality criteria, and rather often made technical changes and copy-edited the articles. All of these observations strengthen the impression that in the initial sample, there are substantially more advanced contributors than in the second.

Patterns of information source use

Different contributors use different kinds of sources. Almost all contributors mentioned online sources directly or indirectly. The most typical channels mentioned in the answers and comments to open-ended questions were Google, scholarly journals available online, databases, archives of major newspapers, online news services and sites related directly to the topic of the article.

General Web sources (49 mentions in all answers and comments to all questions).

Books (36 mentions, 7 explicitly from libraries).

News sources and online news archives (28 mentions).

Own personal sources and knowledge (21 mentions).

Journal articles (17).

Wikipedia (other articles and language editions) (11).

Own (degree) studies (8).

Own research work/results (professional experience) (5).

The sources already cited in the articles (2)

Because the survey was designed to gather qualitative information on the process of how and where the information for each contribution originated, the numbers of mentions of different types of information sources in the survey data are indicative. For instance, not all informants specified the source of books (36 mentions). Similarly, some informants indicated that 'Websites' and 'journal articles' as sources for their contributions. The figures do, however, shed light on the general patterns of how information sources are used by the contributors. The informants were asked to describe, as precisely as they could, where and how they learnt the information they contributed. Therefore the responses can be expected to reflect rather well the perceived principal sources of information for each contribution.

Websites, books and news seem to be the main information sources. One of the informants noted that,

online information is the handiest as it can be accessed simultaneously while writing (R21).

The contributors' personal knowledge and background motivates them to contribute, but also function as a major source of information. Although sources other than formal literary ones are in a minority, the results indicate that they can have a major impact on individual topics and articles. One respondent (R3) mentioned that his collection of CDs was an important source of information in one of his contributions. Informants also described how they recalled or knew something from their personal experience, sought a reference supporting their personal knowledge and made a contribution. Another noteworthy source of information mentioned by the informants was cross-Wikipedia translations. They can be an important source of information on article and topic level, even though they are a somewhat problematic category from an information source point of view as they do not represent an original, external source of information. The informants mentioned Google 42 times and, not surprisingly, it seems to function as the major tool for navigating the Web among the respondents.

Most of the respondents seemed to have a typical pattern of contributing and using information sources which they tended to follow rather consistently. One of the respondents (R21) used The Times News Archive as her major source of information, others referred to the Catholic Encyclopaedia (R28, R90), Archaeology journal (R69) and, for instance, to their chemistry classes (R12).

One third (34.1%) of the respondents indicated that they ask their friends about the topics they are editing at least occasionally. More than half (55.5%) do this only seldom or never. Fewer ask their colleagues. 61.9% indicated that they never ask and only 7.1% that they ask often. 53.5% never consult librarians, but 37.2% did at least occasionally. Different kinds of subject experts are consulted somewhat more often (at least occasionally 39.1%). The experts can be both subject experts and experienced Wikipedia contributors.

I write mainly about history but have no training as a historian, so I regularly seek the advice of trained or practising historians to ensure that I am not making obvious beginners' mistakes. While these experts are usually Wikipedia editors, I have also e-mailed academic experts with questions, and got answers back. (R26)

Correlation analysis revealed that respondents who asked for information were likely to ask almost anyone: friends, colleagues, family, subject experts and librarians. Only asking family members and asking librarians did not correlate significantly (Spearman's rho ,292, Sig (2-tailed) ,064).

One reason for not relying on social information is the Wikipedia Verifiability policy (Wikipedia 2008), which urges contributors to cite references.

I will use my public library, and will occasionally ask a friend to read an article if they are a subject matter expert. The people are not the reference source, however; if the information should be sourced, I will still seek printed or published information from a reliable source. (R29)

The required use of references is clearly a standard procedure for many contributors and in its vigour compares well to the scholarly practice of citing sources.

I use my reference books all the time, checking facts from Wikipedia. I add new ones, and add a reference to them, giving the book (down to the page number!) where I found the information. (R12)

The aspect of correctness is very important for contributors. 93.9% of the respondents indicated that it is important or very important. One informant commented;

I try to find several independent sources, e.g., two different books or three different Web-sites (R6).

One of the two respondents (R50), who disagreed on the importance of correctness, commented that the Wikipedia policy refers to verifiability (Wikipedia 2008) instead of correctness. There was a small correlation between information seeking for personal curiosity and the perceived importance of correctness of the contributions (Spearman's rho ,332, Sig. 2-tailed ,023).

Over half of the respondents (57.1%) indicated that they seek information specifically for their Wikipedia contributions. Only 10.2% indicated that they never seek information for Wikipedia contributions. Although the non-seekers indicated that they do not seek information for Wikipedia, their responses indicate that the information in their contributions comes from similar information sources, such as expert advice and literature, as the information contributed by the informants who identified themselves as explicit information seekers.

The patterns of information source use are strongly related to the patterns of participation in Wikipedia. Demographic variables did not explain information source use or membership in the five analytical groups of contributors (investigators, surfers, worldly-wise, scholars, editors) in the present study. The differences in contributing patterns between the two samples (Groups 1 and 2) can be explained by the populations. Wikipedia e-mail lists used for recruiting Group 1 could be estimated to be populated by the most active participants of the community (e.g., administrators, mediators and meta-users in Bryant et al. 2005). The method of approaching contributors directly with Group 2, on the other hand, can be expected to reach also those who participate only by editing topics of their interest without a specific interest in the Wikipedia project.

Discussion

The findings show that the information in Wikipedia comes from a variety of sources. The majority of information is retrieved from various documents available on the Web. Another large group of reference is the literature known by the contributors. The findings also show that the personal knowledge and social contacts of the contributors function as important sources of information even if they are not used as formal references.

Some aspects of the motivation of the contributors can be explained by their information seeking behaviour and their freeform comments. The groups of worldly-wise and scholars share the characteristic that the principal purpose of the contributors is seldom to find information primarily for Wikipedia. The information fitting for contributions is effectively a by-product of their work and leisure activity. In a way they represent the archetypal crowd, the anyone who can contribute their personal knowledge on various topics to Wikipedia (cf. Tapscott & Williams 2006). Investigators are the ones

who might be estimated to be most likely to see Wikipedia writing empathetically as a serious task (Wikimedia 2008). They enjoy being able to find and write information on topics nobody else has written about. In contrast, the surfers and editors are more likely to enjoy of collaboration and the possibility to contribute as broadly as possible. Wikipedia is a serious task also for them, but more on the social level than on the level of individual achievement. Whereas investigators receive credit for their original accomplishments, surfers and editors receive it for the community membership. Due to the small size of the group, the full variety of editor roles discerned by Bryant et al., (2005) could not be discerned in this study.

Of the information sources used, the online ones were highly popular, as might be expected. The most typical occupations of the informants, student and software engineer, are an explanatory factor. Wikipedia is an online resource and requires some level of computer literacy in the form of mastering the Wiki markup and editing functions. Therefore, it is likely that the contributors are also rather familiar with other Web services and effectively belong to the non-age specific Google generation discussed by Rowlands et al. (2008). Even those contributors who were referring to books and used their local library as a major source of information did their pre-study on the Web using mainstream search engines because of their ease of use and access to a wealth of resources, a pattern discussed earlier in the literature (e.g., Hargraves 2007; Pors 2008).

The group of surfers tended to work almost exclusively with search engines and the results ranked most relevant in the results. In other groups, the use of Web sources was not, however, confined to general sources. Many respondents indicated that they had used specialised repositories and digital archives and their search process incorporated a complex combination of sources and finding aids, resembling the information behaviour reported by Rowlands et al. (2008). A unique Wikipedia and open documentation kind of information use strategy is to translate articles from one language to another. The level of information literacy of many of the informants can be judged to be relatively advanced at least with the strategies they were relying on (Bruce 1997).

The use of social information sources was prevalent although not overwhelmingly so. This is in stark contrast to the general patterns of scholarly and everyday life information behaviour observed in many studies, and illustrates the particularity of contributing to Wikipedia from an information behaviour point of view. Many of the informants found their expert consultants within the Wikipedia community. This is a positive sign of the functioning and capability of the community. Investigators mostly consulted subject experts and information professionals. The worldly-wise were most likely to consult their friends and relatives although even in this group the tendency was rather low.

According to our findings, the information seeking behaviour of the contributors is very responsible, they scrutinise their sources with effort and are critical to the information they work with. Also in this respect the information literacy level of the respondents is high (Bruce 1997), although it is likely that the systematic aspects of their behaviour have become emphasised in the descriptions written after the actual information source use (Mather et al. 2000). Also, the effect of the social desirability bias might explain part of the observed systematicity, although as Fisher and Katz (2000) note, it simultaneously underlines the perceived importance of systematic search.

Many participants expressed specific interest in the results of the study and considered that the results are important. Therefore, most of the informants are likely to be conscious of the problems related to the use of sub-standard information sources. Although the purpose of this study was not to participate in the discussion on the quality of Wikipedia, the results confirm that quality information can be sought and found using multiple strategies and that the personal sphere of experience of the contributors functions both as a powerful motivational factor and a source of information. The local knowledge of people who live close to the places and phenomenon mentioned in Wikipedia, visits to remote places and personal information repositories in the form of books, lecture notes, photographs, comic books, audio discs and recorded TV-programmes form a broad and valuable information resource.

The groups resemble earlier findings in the literature. Schwartz (2006) observed investigator-like behaviour in his preliminary study of changes in a sample of Wikipedia articles. Also the group of editors or administrators discussed by Bryant et al. (2005) emerged clearly from the data in the form of the group of editors. Heinström presents three different groups of fast surfers, broad scanners and deep divers. In essence, the behaviour of the surfers identified in the present study, resembles how the fast surfers> behaved in Heinström's study. The investigators start with a broad scanning kind of behaviour and continue to deep diving. Scholars are clearly deep divers and worldly-wise fast surfers (Heinström 2006).

The information behaviour of Wikipedia contributors seemed also to be broadly dependent on their professional interests as has been suggested in the earlier literature (Julien & Duggan 2000; McKechnie et al. 2002; Tenopir et al. 2003). This tendency was strongest in the group of scholars and worldly-wise, who seek information, but rely mostly on familiar information sources. Similarly, the everyday-life interests of the informants were clearly present in many responses underlining the constituency of the practices of everyday-life information seeking (Savolainen 1995).

In a sense, the variety of strategies and information sources used resembles a foraging kind of information seeking, the ecology of hunting and gathering described by O'Connor et al. (2003). The strategy of investigators is reminiscent of bounty hunters and the way they develop their strategies and find their targets by a careful combination of intuition and information seeking (O'Connor et al. 2003, 45–94). It is a hunting and picking kind of an approach of purposeful information seeking for a specific need in a Taylorian manner of looking for answers (R. S. Taylor 1968). Scholars do a lot of coupling by creating links between what they have found in their work and what is lacking in Wikipedia. Surfers rely on indexing i.e., on trying to find information, which is reasonably likely to be relevant by relying on search engines. Their strategy likewise resembles the berry-picking strategy described by Bates (1989).

The strategies of investigators and surfers are related to the notion of reducing uncertainty (Belkin et al. 1982a,b; Kuhlthau 1991) on their topics of interest. Similarly they are coupled with the indirect aim of making sense of the personal life-world (Dervin 1998). The worldly-wise and the editors do not search for information, but their strategy on relying on readily available resources can be seen as a form of semi-purposeful grazing in an information space where relevant information is abundant (O'Connor et al. 2003: 127–129).

In general, it seems that the information behaviour of Wikipedia contributors incorporates a diversity of information seeking strategies and contexts discussed in the earlier literature from accidental information discovery (Erdelez 1997) to information seeking for everyday-life (Savolainen 1995) and professional interests (Julien & Duggan 2000; McKechnie et al. 2002; Tenopir et al. 2003). Most of the informants indicated that they seek information for the specific purpose of contributing to Wikipedia, but this purposeful information seeking was almost always combined to secondary motivations and interests springing from the personal and professional contexts of the contributors.

The information seeking patterns and, for instance, not seeking for information does not have to correlate with their general patterns of information behaviour and it probably says more of the kind of involvement in Wikipedia than how the contributors generally satisfy their information needs. The complexity, diversity and coexistence of motivations were characteristic of all informants. These findings provide strong support for the earlier observations on the significance of information behaviour as a life-world phenomenon and the coexistence and simultaneous effect of multiple information behaviour strategies and motivations.

Conclusions

The findings of this study show that the information in Wikipedia comes from a variety of sources online and offline. The majority of information is retrieved from the Web. The other large groups of reference are the literature known by the contributors and diverse news sources. The informants of the present

study indicated by the behaviour they described themselves that many of them are notably information literate at least with the strategies they are actively using. Further, the study provides illustrative and strong evidence on how the strategies and motivations of information seeking and use are intertwined and how information behaviour is a life-world wide phenomenon not limited to individual professional or hobby related contexts. The analysis of the data revealed five groups of contributors: investigators, surfers, worldly-wise, scholars and editors who all seek information using distinct strategies.

The qualitative approach assumed in the present study provides rich information on specific edits and the behaviour of individual contributors. It also makes possible to explain the process behind the explicit traces and references present in Wikipedia itself. Further, it provides necessary background information for designing effective quantitative surveys on the same topic. However, both qualitative and quantitative studies on references and referencing behaviour based on larger surveys and the material available on Wikipedia are needed to get a richer picture of the information sources used and cited. Further research on the general demographics of the contributors is also necessary to better understand the context of individual contributions and the general behaviour of those who participate in Wikipedia.

In spite of its limitations, the present study gives a strong indication that Wikipedia benefits from a broad variety of cited references and the work of capable information seekers who rely on various information seeking strategies. Although this does not directly imply anything about the quality of Wikipedia articles, the results give some indication that the community model of hunting and gathering information for Wikipedia has a clear potential at least as far as a significant share of the contributors subscribe to the overarching goals of the project. In this sense the present study supports earlier claims and findings that Wikipedia works. This gives some promise for the prospects of collaboration in the wider context of the participatory Web environment, although a specific emphasis needs to be placed on the particularity of the case of Wikipedia. Furthermore, this study adds to the earlier findings that the existence of different kinds of contributors and complementary strategies of finding relevant information from a diversity of sources contributes to the breadth and depth of the corpus of information available in Wikipedia.

References

- Anthony, D., Smith, S.W. & Williamson, T. (2007). The quality of open source production: zealots and good samaritans in the case of Wikipedia. Hanover, NH: Dartmouth College. (Dartmouth Computer Science Technical Report TR2007-606)
- Ayers: (2006). Researching Wikipedia - current approaches and new directions. Proceedings of the American Society for Information Science and Technology, 43(1), 252.
- Barrett, A. (2005). The information-seeking habits of graduate student researchers in the humanities. *The Journal of Academic Librarianship*, 31(4), 324–331.
- Bates, M.J. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online Review*, (13), 407–424. Retrieved 25 May, 2009 from <http://www.gseis.ucla.edu/faculty/bates/berrypicking.html> (Archived by WebCite® at <http://www.Webcitation.org/5roKSbO3x>)
- Belkin, N.J., Oddy, R.N. & Brooks, H.M. (1982a). Ask for information retrieval: part I.: background and theory. *Journal of Documentation*, 38(2), 61–71.
- Belkin, N.J., Oddy, R.N. & Brooks, H.M. (1982b). Ask for information retrieval: part I.: Results of a design study. *Journal of Documentation*, 38(3), 145–164.
- Black, E.W. (2008). Wikipedia and academic peer review - Wikipedia as a recognised medium for scholarly publication? *Online Information Review*, 32(1), 73–88.
- Brändle, A. (2005). Zu wenige Köche verderben den Brei: eine Inhaltsanalyse der Wikipedia aus Perspektive der journalistischen Qualität, des Netzeffekts und der ökonomie der Aufmerksamkeit. [Too many cooks spoil the broth: a content analysis of Wikipedia from the perspective of journalistic quality, the network effect and the economy of attention.] Unpublished licenciate dissertation, University of Zurich, Zurich, Switzerland.

- Britannica. (2006). *Fatally flawed: refuting the recent study on encyclopedic accuracy by the journal Nature*. Chicago, IL: Encyclopaedia Britannica, Inc. Retrieved 25 May, 2009 from http://corporate.britannica.com/britannica_nature_response.pdf (Archived by WebCite® at <http://www.Webcitation.org/5roKV4QaP>)
- Bruce, C. (1997). *The seven faces of information literacy*. Blackwood, South Australia: Auslib Press.
- Bryant, S.L., Forte, A. & Bruckman, A. (2005). *Becoming Wikipedian: transformation of participation in a collaborative online encyclopedia*. In *Proceedings of the 2005 International ACM SIGGROUP Conference on Supporting Group Work 2005*, Sanibel Island, Florida, USA (pp. 1-10). New York, NY: ACM Press.
- Butler, D. (2008). *Publish in Wikipedia or perish*. *Nature News*. Retrieved 25 May, 2009 from <http://www.nature.com/news/2008/081216/full/news.2008.1312.html> (Archived by WebCite® at <http://www.Webcitation.org/5roKX8fmp>)
- Case, D. (2006). *Information behavior*. *Annual Review of Information Science and Technology*, 40, 293–327.
- Case, D. (2007). *Looking for information: a survey of research on information seeking, needs, and behavior*. 2nd. ed. Burlington, MA: Academic Press.
- Chatman, E. (1991). *Life in a small world: applicability of gratification theory to information-seeking behavior*. *Journal of the American Society for Information Science*, 42(6), 438–449.
- Chatman, E. (1996). *The impoverished life-world of outsiders*. *Journal of the American Society for Information Science*, 47(3), 193–206.
- Chesney, T. (2006). *An empirical examination of Wikipedia's credibility*. *First Monday*, 11(11). Retrieved 25 May, 2009 from http://firstmonday.org/issues/issue11_11/chesney/index.html (Archived by WebCite® at <http://www.Webcitation.org/5roKZ8xZ7>)
- Cunningham & Cunningham, Inc. (2005). *Thread mode considered harmful*. C2.com. Retrieved 25 May, 2009 from <http://c2.com/cgi/wiki?ThreadModeConsideredHarmful>. (Archived by WebCite® at <http://www.Webcitation.org/5roKeyGNj>)
- Denning, P., Horning, J., Parnas, D. & Weinstein, L. (2005). *Wikipedia risks*. *Communications of the ACM*, 48(12), 152–152.
- Dervin, B. (1998). *Sense-making theory and practice: an overview of user interests in knowledge seeking and use*. *Journal of Knowledge Management*, 2(2), 36–46.
- Deutskens, E., de Ruyter, K., Wetzels, M. & Oosterveld, P. (2004). *Response rate and response quality of internet-based surveys: an experimental study*. *Marketing Letters*, 15(1), 21–36.
- Dondio, P., Barrett, S., Weber, S. & Seigneur, J. (2006). *Extracting trust from domain analysis: a case study on the Wikipedia project*. In Laurence Tianruo Yang, Hai Jin, Jianhua Ma, Theo Ungerer (Eds.), *Autonomic and Trusting Computing. Proceedings, Third International Conference, Wuhan, China, September 3-6, 2006*. (pp. 362-373). Berlin: Springer. (Lecture Notes in Computer Science, Vol. 4158).
- DuBois, A. (2003). *Close reading: an introduction*. In *Close reading: a reader*, (pp. 1–40). Durham, NC: Duke University Press.
- Economist (2006a). *Among the audience*. *Economist*, 379(8474), 3–6.
- Economist (2006b). *Battle of Britannica*. *Economist*, 378(8471), 65–67.
- Economist (2006c). *Open, but not as usual*. *Economist*, 378(8469), 73–76
- Economist (2006d). *The wiki principle*. *Economist*, 379(8474), 14–16.
- Emigh, W. & Herring, S.C. (2005). *Collaborative authoring on the Web: a genre analysis of online encyclopedias*. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS'05) - Track 4, vol. ;4, (p. 99a)*. Washington D.C.: IEEE Computer Society.
- Erdelez, S. (1997). *Information encountering: a conceptual framework for accidental information discovery*. In P. Vakkari, R. Savolainen & B. Dervin (Eds.) *ISIC '96: Proceedings of an international conference on Information seeking in context*, (pp. 125–146). London: Taylor Graham.
- Ess, C. & Association of Internet Researchers. *Ethics Working Committee*. (2002). *Ethical decision-making and internet research recommendations from the aoir ethics working committee*. Association of Internet Researchers. Approved by AoIR, November 27, 2002. Retrieved 25 May, 2009 from <http://www.aoir.org/reports/ethics.pdf> (Archived by WebCite® at <http://www.Webcitation.org/5roKzJ24c>)

- Fallis, D. (2008). Toward an epistemology of wikipedia. *Journal of the American Society for Information Science and Technology*, 59(10), 1662–1674.
- Fisher, R.J. & Katz, J.E. (2000). Social-desirability bias and the validity of self-reported values. *Psychology and Marketing*, 17(2), 105–120.
- Foster, A. (2004). A nonlinear model of information-seeking behavior. *Journal of the American Society for Information Science and Technology*, 55(3), 228–237.
- Giles, J. (2005). Internet encyclopaedias go head to head. *Nature*, 438(7070), 900–901.
- Glaser, B.G. & Strauss, A. L. (1967). *The discovery of grounded theory: strategies for qualitative research*. Hawthorne, NY: Aldine de Gruyter.
- Hargittai, E. & Walejko, G. (2008). The participation divide: content creation and sharing in the digital age. *Information, Communication and Society*, 11(2), 239–256.
- Hargraves, I. (2007). Controversies of information discovery. *Knowledge, Technology & Policy*, 20(2), 83–90.
- Heinström, J. (2006). Broad exploration or precise specificity: two basic information seeking patterns among students. *Journal of the American Society for Information Science and Technology*, 57(11), 1440–1450.
- Hyldegård, J. (2009). Personality traits and group-based information behaviour: an exploratory study. *Information research*, 14(2).
- Jones, J. (2008). Patterns of revision in online writing - a study of Wikipedia's featured articles. *Written Communications*, 25(2), 262–289.
- Julien, H. & Duggan, L. (2000). A longitudinal analysis of the information needs and uses literature. *Library and Information Science Research*, 22(3), 291–309.
- Kaplowitz, M.D., Hadlock, T.D. & Levine, R. (2004). A comparison of Web and mail survey response rates. *Public Opinion Quarterly*, 68(1), 94–101.
- Keen, A. (2008). *The cult of the amateur*. New York, NY: Doubleday.
- Kittur, A. & Kraut, R.E. (2008). Harnessing the wisdom of crowds in Wikipedia: quality through coordination. In *CSCW08: Proceedings of the ACM 2008 conference on Computer supported cooperative work*, (pp. 37–46). New York, NY: ACM Press
- Kittur, A., Pendleton, B.A., Suh, B. & Mytkowicz, T. (2007). Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. Paper presented at AltCHI at the 25th Annual ACM Conference on Human Factors in Computing Systems . Retrieved 12 August, 2010 from http://www.viktoria.se/altchi/submissions/submission_edchi_1.pdf (Archived by WebCite® at <http://www.webcitation.org/5rvgjaHxn>)
- Kolbitsch, J. & Maurer, H. (2006). The transformation of the Web: how emerging communities shape the information we consume. *Journal of Universal Computer Science*, 12(2), 187–213.
- Korfiatis, N.T., Poulos, M. & Bokos, G. (2006). Evaluating authoritative sources using social networks: an insight from Wikipedia. *Online Information Review*, 30(3), 252–262.
- Kuhlthau, C.C. (1991). Inside the search process: information seeking from the user's perspective. *Journal of the American Society for Information Science*, 42(5), 361–371.
- Lally, A.M. & Dunford, C.E. (2007). Using Wikipedia to extend digital collections. *D-Lib Magazine*, 13(5/6). Retrieved 25 May, 2009 from <http://www.dlib.org/dlib/may07/lally/05lally.html> (Archived by WebCite® at <http://www.Webcitation.org/5roL1s9JQ>)
- Leuf, B. & Cunningham, W. (2001). *The wiki way. Quick collaboration on the Web*. Reading, MA: Addison-Wesley.
- Levack, K. (2003). If two heads are better than one, try 7,000 with Wikipedia. *eContent*, 26(4), 12–13
- Lih, A. (2004). Wikipedia as participatory journalism: reliable sources? Metrics for evaluating collaborative media as a news resource. Paper presented at the 5th International Symposium on Online Journalism (April 16-17, 2004), University of Texas at Austin. Retrieved 12 August, 2010 from (Archived by WebCite® at <http://www.webcitation.org/5rvfbm5VX>)
- Mather, M., Shafir, E. & Johnson, M. (2000). Misremembrance of options past: source monitoring and choice. *Psychological Science*, 11(2), 132–138.
- McKechnie, L. M., Baker, L., Greenwood, M. & Julien, H. (2002). Research method trends in human information behaviour literature. *The New Review of Information Behaviour Research*, 3, 113–126.

- Nature (2006). Encyclopaedia britannica and nature: a response. *Nature*. Retrieved 25 May, 2009 from http://www.nature.com/press_releases/Britannica_response.pdf (Archived by WebCite® at <http://www.Webcitation.org/5roL3c8WL>)
- Newman, M. (2005). Power laws, pareto distributions and Zipf's law. *Contemporary Physics*, 46(5), 323–351.
- Nielsen, F. Å (2007). Scientific citations in Wikipedia. *First Monday*, 12(8). Retrieved 25 May, 2009 from http://www.firstmonday.org/issues/issue12_8/nielsen/ (Archived by WebCite® at <http://www.Webcitation.org/5roL5703L>)
- O'Connor, B., Copeland, H. & Kearns, J. L. (2003). *Hunting and gathering on the information savanna*. Lanham, MD: Scarecrow Press.
- Ortega, F. & Barahona, J.M.G. (2007). Quantitative analysis of the Wikipedia community of users. In *WikiSym 07: Proceedings of the 2007 international symposium on Wikis*, (pp. 75–86). New York, NY: ACM Press.
- Ortega, F., Gonzalez-Barahona, J. & Robles, G. (2008). On the inequality of contributions to Wikipedia. In *Proceedings of the Proceedings of the 41st Annual Hawaii International Conference on System Sciences*, (pp. 304–304). Washington, DC: IEEE Computer Society.
- Patton, M.Q. (1987). *How to use qualitative methods in evaluation*. Thousand Oaks, CA: Sage Publications.
- Pfeil, U., Zaphiris, P. & Ang, C.S. (2006). Cultural differences in collaborative authoring of wikipedia. *Journal of Computer-Mediated Communication*, 12(1), 88–113.
- Pirolli: (2007). *Information foraging theory: adaptive interaction with information*. Oxford: Oxford University Press.
- Pors, N.O. (2008). Traditional use patterns? An analysis of high school students' use of libraries and information resources. *New Library World*, 109(9/10), 431–443.
- Potthast, M., Stein, B. & Gerling, R. (2008). Automatic vandalism detection in Wikipedia. In Craig Macdonald, Iadh Ounis, Vassilis Plachouras, Ian Ruthven and Ryen W. White (Eds.). *Advances in information retrieval. 30th European Conference on IR Research, ECIR 2008, Glasgow, UK, March 30–April 3, 2008. Proceedings*, (pp. 663–668). Berlin, Heidelberg: Springer. (Lecture Notes in Computer Science, Vol. 4956)
- Priedhorsky, R., Chen, J., Lam, S.T.K., Panciera, K., Terveen, L., & Riedl, J. (2007). Creating, destroying, and restoring value in Wikipedia. In *GROUP07: Proceedings of the 2007 international ACM conference on Supporting group work*, (pp. 259–268). New York, NY: ACM Press
- Quiggin, J. (2005). *Blogs, wikis and creative innovation.*, Brisbane, University of Queensland, Schools of Economics and Political Science. (Australian Public Policy Program Working Paper P06#1)
- Rainie, L. & Tancer, B. (2007). Data memo. Washington DC: Pew Internet and American Life Project. Retrieved 10 August, 2010 from http://www.pewinternet.org/~media/Files/Reports/2007/PIP_Wikipedia07.pdf.pdf (Archived by WebCite® at <http://www.webcitation.org/5ruWkNWmd>)
- Rask, M. (2008). The reach and richness of Wikipedia: is wikinomics only for rich countries? *First Monday*, 13(6). Retrieved 25 May, 2009 from <http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2046/1970> (Archived by WebCite® at <http://www.webcitation.org/5roL73MNU>)
- Remy, M. (2002). Wikipedia: the free encyclopedia. *Online Information Review*, 26(6), 434–435.
- Rowlands, I., Nicholas, D., Williams, P., Huntington, P., Fieldhouse, M., Gunter, B., et al. (2008). The Google generation: the information behaviour of the researcher of the future. *Aslib Proceedings*, 60(4), 290–310.
- Sammons, C.C. & Speight, S.L. (2008). A qualitative investigation of graduate-student changes associated with multicultural counseling courses. *The Counseling Psychologist*, 36(6), 814–838.
- Savolainen, R. (1995). Everyday life information seeking: approaching information seeking in the context of Way of life Library and Information Science Research, 17(3), 259–94.
- Schroer, J. & Hertel, G. (2007). Voluntary engagement in an open Web-based encyclopedia: Wikipedians, and why they do it. *F/OSS Research Community*. Retrieved 25 May, 2009 from http://opensource.mit.edu/papers/Schroer_Hertel_Wikipedia_Motivation.pdf (Archived by WebCite® at <http://www.webcitation.org/5roL8ak4Q>)

- Schwartz, A. (2006). Who writes Wikipedia? Raw Thought (blog). Retrieved 25 May, 2009 from <http://www.aaronsw.com/Weblog/whowriteswikipedia> (Archived by WebCite® at <http://www.Webcitation.org/5roL9yxwN>)
- Solomon: (1997a). Discovering information behavior in sense making. I. Time and timing. *Journal of the American Society for Information Science*, 48(12), 1097–1108.
- Solomon: (1997b). Discovering information behavior in sense making. II. The social. *Journal of the American Society for Information Science*, 48(12), 1109 – 1126.
- Spink, A. & Cole, C. (2006). Human information behavior: integrating diverse approaches and information use. *Journal of the American Society for Information Science and Technology*, 57(1), 25–35.
- Stein, K. & Hess, C. (2007). Does it matter who contributes: a study on featured articles in the German Wikipedia. In *Proceedings of the 18th Conference on Hypertext and Hypermedia*, Manchester, UK, September 10 - 12, 2007 , (pp. 171–174). New York, NY: ACM.
- Stvilia, B., Twidale, M.B., Smith, L.C. & Gasser, L. (2008). Information quality work organization in Wikipedia. *Journal of the American Society for Information Science and Technology*, 59(6), 983–1001.
- Tancer, B. (2007, April 25). Who’s really participating in Web 2.0. Time. Retrieved 25 May, 2009 from <http://www.time.com/time/business/article/0,8599,1614751,00.html> (Archived by WebCite® at <http://www.Webcitation.org/5roLBUW1>)
- Tapscott, D. & Williams, A. (2006). *Wikinomics: how mass collaboration changes everything*. New York, NY: Portfolio.
- Taylor, R. (1991). Information use environments. In *Progress in Communication Sciences*, 10, 217–255.
- Taylor, R. S. (1968). Question-negotiation and information seeking in libraries. *College & Research Libraries*, 29(3), 178–194.
- Tenopir, C., Hitchcock, B. & Pillow, A. (2003). *Use and users of electronic library resources: an overview and analysis of recent research studies*. Washington, DC: Council on Library and Information Resources. (Report 120). Retrieved 8 August, 2010 from <http://www.clir.org/pubs/reports/pub120/pub120.pdf> (Archived by WebCite® at <http://www.Webcitation.org/5rpnR2Aa9>)
- Terdiman, D. (2005). Growing pains for Wikipedia. *CNET News.com*. Retrieved 25 May, 2009 from http://news.cnet.com/Growing-pains-for-Wikipedia/2100-1025_3-5981119.html (Archived by WebCite® at <http://www.Webcitation.org/5roLDDbq4>)
- Vižgas, F., Wattenberg, M. & Dave, K. (2004). Studying cooperation and conflict between authors with history flow visualizations. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, (pp. 575–582). New York, NY: ACM Press
- Voss, J. (2005). Measuring wikipedia. Paper presented at the International Conference of the International Society for Scientometrics and Informetrics: 10th, Stockholm (Sweden). Retrieved 8 August, 2010 from <http://eprints.rclis.org/3610/1/MeasuringWikipedia2005.pdf> (Archived by WebCite® at <http://www.Webcitation.org/5rpn5bwjD>)
- Voss, J. (2006a). Some statistics jimbo wish he had. *Wikimetrics (Blog)*. Retrieved 25 May, 2009 from <http://web.archive.org/web/20060613074526/http://wm.sieheauch.de/?p=44> (Archived by WebCite® at <http://www.Webcitation.org/5rwksJsOC>)
- Voss, J. (2006b). Workshop on wikipedia research. Paper presented at WikiSym 2006, Odense, Denmark. Retrieved 25 May, 2009 from <http://eprints.rclis.org/archive/00007102/01/FirstWorkshopOnWikipediaResearch.pdf>. (Archived by WebCite® at <http://www.Webcitation.org/5roLFeYsg>)
- Wikimedia (2008). *Meta:The Wikipedia Community*. Wikimedia Foundation. Retrieved 25 May, 2009 from http://meta.wikimedia.org/wiki/The_Wikipedia_Community (Archived by WebCite® at <http://www.Webcitation.org/5roLHgYTo>)
- Wikimedia Foundation (2008a). *Hebrew wikipedia satisfaction survey*. Wikimedia Foundation. Retrieved 25 May, 2009 from http://meta.wikimedia.org/wiki/Hebrew_Wikipedia_satisfaction_survey (Archived by WebCite® at <http://www.Webcitation.org/5roLIvxrB>)

- Wikimedia Foundation (2008b). Meta:research. Wikimedia Foundation. Retrieved 25 May, 2009 from <http://meta.wikimedia.org/wiki/Research> (Archived by WebCite® at <http://www.Webcitation.org/5roLKYEa9>)
- Wikimedia Foundation (2008c). Meta: research/social research collaborations. Wikimedia Foundation. Retrieved 25 May, 2009 from http://meta.wikimedia.org/wiki/Research/Social_Research_Collaborations (Archived by WebCite® at <http://www.Webcitation.org/5roLMMqOi>)
- Wikipedia (2008a). Criticism of Wikipedia. Wikimedia Foundation. Retrieved 25 May, 2009 from http://en.wikipedia.org/wiki/Criticism_of_Wikipedia (Archived by WebCite® at <http://www.Webcitation.org/5roLPbb3l>)
- Wikipedia (2008b). Reliability of Wikipedia. Wikimedia Foundation. Retrieved 25 May, 2009 from http://en.wikipedia.org/wiki/Reliability_of_Wikipedia (Archived by WebCite® at <http://www.Webcitation.org/5roLREgqg>)
- Wikipedia (2008c). Statistics. Wikimedia Foundation. Retrieved 25 May, 2009 from <http://en.wikipedia.org/wiki/Special:Statistics> (Archived by WebCite® at <http://www.Webcitation.org/5roLTgK0c>)
- Wikipedia (2008d). Statistik. Wikimedia Foundation. Retrieved 25 May, 2009 from <http://de.wikipedia.org/wiki/Spezial:Statistik> (Archived by WebCite® at <http://www.Webcitation.org/5roLVEcNC>)
- Wikipedia (2008e). Wikimedia research network. Wikimedia Foundation. Retrieved 25 May, 2009 from http://meta.wikimedia.org/wiki/Wikimedia_Research_Network (Archived by WebCite® at <http://www.Webcitation.org/5roLWYPay>)
- Wikipedia (2008f). Wikipedia in academic studies. Wikimedia Foundation. Retrieved 25 May, 2009 from http://en.wikipedia.org/wiki/Wikipedia:Wikipedia_in_academic_studies (Archived by WebCite® at <http://www.Webcitation.org/5roLYg4XR>)
- Wikipedia (2008g). Wikipedia: editing frequency. Wikimedia Foundation. Retrieved 25 May, 2009 from http://en.wikipedia.org/wiki/Wikipedia:Editing_frequency (Archived by WebCite® at <http://www.Webcitation.org/5roLa5uEn>)
- Wikipedia (2008h). Wikipedia:External peer review/Nature December 2005. Wikimedia Foundation. Retrieved 25 May, 2009 from http://en.wikipedia.org/wiki/Wikipedia:External_peer_review/Nature_December_2005 (Archived by WebCite® at <http://www.Webcitation.org/5roLbfJV6>)
- Wikipedia (2008i). Wikipedia:General disclaimer. Wikimedia Foundation. Retrieved 25 May, 2009 from http://en.wikipedia.org/wiki/Wikipedia:General_disclaimer (Archived by WebCite® at <http://www.Webcitation.org/5roLd0p9w>)
- Wikipedia (2008j). Wikipedia:Neutral point of view. Wikimedia Foundation. Retrieved 25 May, 2009 from http://en.wikipedia.org/wiki/Wikipedia:Neutral_Point_Of_View
- Wikipedia (2008k). Wikipedia:No original research. Wikimedia Foundation. Retrieved 25 May, 2009 from http://en.wikipedia.org/wiki/Wikipedia:Neutral_Point_Of_View (Archived by WebCite® at <http://www.Webcitation.org/5roLhRA50>)
- Wikipedia (2008l). Wikipedia: verifiability. Wikimedia Foundation. Retrieved 25 May, 2009 from <http://en.wikipedia.org/wiki/Wikipedia:Verifiability> (Archived by WebCite® at <http://www.Webcitation.org/5roLiNbon>)
- Wikipedia (2008m). Wikipedia:WikiProject critical source examination. Wikimedia Foundation. Retrieved 25 May, 2009 from http://en.wikipedia.org/wiki/Wikipedia:WikiProject_critical_source_examination (Archived by WebCite® at <http://www.Webcitation.org/5roLmrdVq>)
- Wikipedia (2008n). Wikipedia: WikiProject Fact and Reference Check. Wikimedia Foundation. Retrieved 25 May, 2009 from http://en.wikipedia.org/wiki/Wikipedia:WikiProject_Fact_and_Reference_Check (Archived by WebCite® at <http://www.Webcitation.org/5roLoawrI>)
- Wikipedia (2009). Wikipedia: unreferenced articles. Wikimedia Foundation. Retrieved 25 May, 2009 from http://en.wikipedia.org/wiki/Wikipedia:Editing_frequency (Archived by WebCite® at <http://www.Webcitation.org/5roLsVmlD>)

- Wilkinson, D.M. & Huberman, B.A. (2007). Assessing the value of cooperation in Wikipedia. *First Monday*, 12(4). Retrieved 25 May, 2009 from <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1763/1643> (Archived by WebCite® at <http://www.Webcitation.org/5roLvX6g1>)
- Wilson, T.D. (1994). Information needs and uses: fifty years of progress? In *Fifty years of information progress: a Journal of Documentation review*, (pp. 15–51). London: Aslib.
- Wilson, T.D. (1999). Models in information behaviour research. *Journal of Documentation*, 55(3), 249–270.
- Wilson, T.D. (2000). Human information behavior. *Informing Science*, 3(2), 49–55.
- Wilson, T. D. (2006). Information-seeking behaviour and the digital information world. *The Indexer*, 25(1), 28–31.
- Yun, G. W. & Trumbo, C.W. (2000). Comparative response to a survey executed by post, e-mail & Web form. *Journal of Computer Mediated Communication*, 6(1). Retrieved 25 May, 2009 from <http://jcmc.indiana.edu/vol6/issue1/yun.html> (Archived by WebCite® at <http://www.Webcitation.org/5roLyU15c>)