The People's Encyclopedia Under the Gaze of the Sages: A Systematic Review of Scholarly Research on Wikipedia

Okoli, Chitu; Mehdi, Mohamad; Mesgari, Mostafa; Nielsen, Finn Årup; Lanamäki, Arto

Link to article, DOI:
10.2139/ssrn.2021326

Publication date:
2012

Document Version
Publisher's PDF, also known as Version of record

The people's encyclopedia under the gaze of the sages: A systematic review of scholarly research on Wikipedia

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

Mohamad Mehdi
Computer Science, Concordia University, Montreal, Canada
mo_mehdi@encs.concordia.ca

Mostafa Mesgari
John Molson School of Business, Concordia University, Montreal, Canada
mmesgari@jmsb.concordia.ca

Finn Årup Nielsen
DTU Informatics, Technical University of Denmark, Kongens Lyngby, Denmark
fn@imm.dtu.dk

Arto Lanamäki
Information Science and Media Studies, University of Bergen, Bergen, Norway
arto.lanamaki@uib.no

Abstract

Wikipedia has become one of the ten most visited sites on the Web, and the world's leading source of Web reference information. Its rapid success has inspired hundreds of scholars from various disciplines to study its content, communication and community dynamics from various perspectives. This article presents a systematic review of scholarly research on Wikipedia. We describe our detailed, rigorous methodology for identifying over 450 scholarly studies of Wikipedia. We present the WikiLit website (http://wikilit.referata.com), where most of the papers reviewed here are described in detail. In the major section of this article, we then categorize and summarize the studies. An appendix features an extensive list of resources useful for Wikipedia researchers.

Keywords: Wikipedia, systematic literature review, encyclopedias, Web 2.0, social media, online collaboration, mass collaboration, information retrieval, information extraction, natural language processing, ontologies, open content, Creative Commons, motivations, online culture, Web references

Table of Contents

Introduction

With the dramatic increase in interest in Wikipedia during its ten-year history, it has become one of the ten-most visited sites on the Web, and the world's leading source of Web reference information. The encyclopedia is a prime example of Web 2.0 with its 23 million[1] articles based on the collaborative efforts of volunteers from around the globe.

Wikipedia has been, and remains, highly controversial among scholars. The broad swath of opinions is clearly reflected in the titles of some articles. On one extreme, there are the skeptical critics with a title like "Why you can't cite Wikipedia in my class" (Waters 2007), which accuses Wikipedia of "conflating facts with popular opinion." On the other extreme, there are the enthusiastic embracers, with a title like "A seismic shift in epistemology" (Dede 2008), which hails the rise of "Web 2.0 knowledge" as a "pure democracy" of knowledge in contrast to the "hierarchical meritocracy" of what is called "classical knowledge." In between, there are a wide range of opinions and perspectives; even the two articles mentioned from either end of the spectrum both recognize Wikipedia's strengths while cautioning against its weaknesses.

The popularity of the phenomenon attracts many researchers, and the populist tone of most Wikipedia articles makes them readily accessible to the "common researcher" in contrast to, for example, bioinformatics databases that typically require expert biomedical knowledge. The openness of the project and easy availability of data also make Wikipedia interesting to researchers. Typically, research on the Web requires crawling many sites; in contrast, each complete language versions of Wikipedia lies available for download as a single compressed XML file ready for use and analysis. Other Web 2.0 large-scale datasets, e.g., from Facebook, may simply not be available. Moreover, multiple language editions make it possible to explore cross-cultural issues. In addition, the availability of the entire revision history of all pages enables dynamic studies of content and contributors. In this aspect, Wikipedia is similar to free and open source software, whose publicly available code repositories also make similar research possible. However, rather than being restricted to the narrow interest of computer science, software engineering and related fields, Wikipedia spans literally all areas of human knowledge.

Much of the scholarly research can prove valuable in guiding Wikipedia contributors and managers on developing policies and best practices to improve the quality, performance, and overall value of Wikipedia. Moreover, such research is helpful to understand the implications of the burgeoning field of open content, which applies the same open-source development principles to the creation of non-software media such as books, music, video, and other information products. In order to consolidate and critically assess the current work on Wikipedia, as well as to offer a solid base for future targeted research, we have embarked on a systematic literature review on this rapidly-growing subject of research.

The Wikimedia Foundation, Wikipedia's non-profit sponsor, attempts to maintain an online catalogue of scholarly articles and researchers.[2] However, such a central resource can only track a fraction of the abundant body of work that has been conducted. A number of researchers have conducted literature reviews of various aspects of Wikipedia (and we review these reviews in the following section), but none has attempted a comprehensive review that examines all kinds of research conducted on Wikipedia. The vast diversity of research indicates that there is the crucial need for a comprehensive literature review of all kinds of scholarly work on Wikipedia to analyze particular trends in research and offer the basic groundwork for future work.

---

[1] http://stats.wikimedia.org/EN/TablesArticlesTotal.htm (September 2012)

[2] http://meta.wikimedia.org/wiki/Research

In response to this need, this article presents a comprehensive systematic review of scholarly research on Wikipedia. This review does not include studies about any wiki other than Wikipedia, nor about any Wikimedia Foundation project other than Wikipedia. However, we do cover many studies that primarily treat these and other related topics, as long as they include a significant component related explicitly to Wikipedia. We exhaustively cover all peer-reviewed journal articles (over 340) and doctoral theses (over 30) published before summer 2011, and additionally over 100 of the most highly cited conference papers. We only cover English-language publications, though they might have treated Wikipedia in any language. In the methodology section of this review, we explain our inclusion criteria in detail.

In the following section, we discuss previous literature reviews that have been conducted of scholarly research on Wikipedia, and we explain how this present review is different from and goes beyond previous efforts. We also acknowledge the limitations of this present review, indicating when these past have gone beyond the scope we are able to cover here. We then describe in detail the systematic review methodology we have followed based on Okoli and Schabram's (2010) guidelines to assure a comprehensive literature search and rigorous procedures for quality assurance. The subsequent section describes the WikiLit website (http://wikilit.referata.com) that we have established as a companion to this review. Next is the major part of this present review, a lengthy section that categorizes and describes the scholarly research conducted to date on Wikipedia. After concluding this review, we also include an appendix featuring an extensive list of resources useful for Wikipedia researchers.

Earlier Literature Reviews of Wikipedia Research

Ayers and Priedhorsky (2011) hosted a workshop at the 2011 WikiSym conference where they discussed the unique challenges that literature reviews of wikis and of Wikipedia entail. In particular, they note that the literature is extremely cross-disciplinary, and that the research is published in scholarly outlets that make it inaccessible to practitioners who could benefit from its practical applications. We try to address these two challenges both in this present review and in the WikiLit website[3], described later. Our review very carefully organizes studies by topic, including flagging when studies cover multiple topics. As we describe later, our topical categorization is directly derived from a structure designed by Wikimedia Foundation practitioners to facilitate diffusion of research results to the community. By summarizing the key findings in this review and by extracting essential details on the WikiLit website, practitioners and scholars alike can readily identify results that are most pertinent to them. In addition, we have identified the multiple knowledge domains of each study on the WikiLit website; thus, researchers and practitioners can readily identify the studies that apply to their knowledge domains of interest, regardless of the specific topics of the studies. Moreover, by publishing our review in open access outlets and by hosting the extracted data on an open wiki, the results of our analyses are readily accessible to the scholars and practitioners who might need them. Another challenge they note is that "research is available in a wide variety of both traditional and non-traditional venues" (2011, p.229). Although this is one of our most significant challenges in this review, we try to help other researchers by providing an extensive compilation of helpful resources for researchers in the appendix to this paper.

Ayers and Priedhorsky also note as a challenge that "literature about Wikipedia dominates the field" (2011, p.229), which makes literature reviews more difficult for researchers interested in other wikis or wikis in general. While this is not a disadvantage for our purposes here, it highlights the fact that there have been many attempts to review scholarly articles involving Wikipedia. We ourselves have executed such reviews in the past (Okoli 2009; Nielsen 2012), and this present review can be seen as a merger and extension of our previously distinct lines of work. We do note though, that Okoli (2009) summarized a few non-peer-reviewed or off-topic works that we do not discuss here, and Nielsen (2012) reviewed a

[3] Unfortunately, none of the authors of this paper was able to participate in Ayers and Priedhorsky's workshop. Although the

large number of wiki- and Wikimedia Foundation-related studies that are not directly related to Wikipedia. In this section, we describe previous reviews of scholarly research on Wikipedia by other researchers, and explain what this present review offers beyond what has been done in the past. In doing so, we also highlight the limitations of our present review and indicate when others' work is complementary to and goes beyond what we are able to cover here.

Early Review of Wikipedia as a Nascent Phenomenon (2006)

Ayers (2006) conducted the first standalone literature review of scholarly research on Wikipedia that we have identified. In her review of 18 studies, she focused on "social science and information science research studies done on and about Wikipedia," and expressly excluded studies that "mention Wikipedia as a case study" and those focused on technical infrastructure.

Ayers first examined the methodologies employed to study Wikipedia, grouping studies based on the aspect of Wikipedia they covered (e.g. article edit histories or user talk pages) and on if they adopted quantitative or qualitative analysis approaches. Then she grouped studies into two major topic categories, those that research Wikipedia content (corresponding to our Content category) and those that research the community (corresponding to our Participation category). Based on these categorizations, she discussed the articles in detail. She then concluded with suggestions for future research.

As an early literature review before this field of research exploded, Ayers' review is admirable for capturing the primary early trends and providing guidance to future researchers. Although the early date of the review resulted in covering only a small number of studies, her insightful summaries are worthwhile reading for researchers interested in general content and participation issues.

Wikipedia as a Textual Corpus (2009)

One of the most thorough literature reviews of Wikipedia conducted so far has been that executed by Medelyan et al. (2009). Nonetheless, their review does not attempt to be a comprehensive one: "It focuses on research that extracts and makes use of the concepts, relations, facts and descriptions found in Wikipedia, and organizes the work into four broad categories: applying Wikipedia to *natural language processing*; using it to facilitate *information retrieval* and *information extraction*; and as a resource for *ontology building"* (2009, p.716). This restricted focus corresponds exactly to our subset of Wikipedia studies that we label as "Corpus" topics; that is, research that rather than studying Wikipedia itself as a phenomenon, uses the textual products of Wikipedia as a textual corpus for conducting text-oriented research.

Medelyan et al.'s review begins with a detailed description of the technical characteristics of Wikipedia's textual structure (such as articles, categories, and intra-wiki links) that facilitate corpus-oriented research. This description is an invaluable introduction to Wikipedia for researchers, as they specifically focused on highlighting the features and characteristics that are most amenable to research analysis. After this introduction, the main body of Medelyan et al.'s review is an in-depth examination of specific studies grouped by the topic of their research questions. In fact, we borrow our subcategorization of the Corpus articles we present here from the structure of their review.

The primary difference between our present review and that of Medelyan et al. is that whereas they restricted their focus to corpus-focused Wikipedia articles, we applied no restriction whatsoever as to topic or domain of knowledge. In that sense, the articles they reviewed are a partial subset of those that we review. However, since Medelyan et al.'s review was specifically focused on Corpus studies, we did not bother to duplicate the description of any article that they described in detail; rather, we identify these articles and refer readers to Medelyan et al. for descriptions. On one hand, we did analyze all the Corpus articles to extract the research details such as specific topics, research methodologies, and so on; thus, they are fully included in our analysis in the WikiLit website. On the other hand, the only articles we describe in our Corpus section here are those that were not included by Medelyan et al. (mainly because

they were published after that review was conducted). In any case, we consider Medelyan et al.'s seminal review essential reading for researchers interested in using Wikipedia as a textual corpus. Our present review could be considered an extension or update of theirs with respect to that topic area, but it cannot replace it.

Attempted Frontend to a Mathematical Dissertation (2010)

Martin's (2010) review "was originally designed as a literature review for a doctoral dissertation focusing on Wikipedia." Apparently, this initial attempt was abortive[4], but it nonetheless yielded a valuable contribution to literature reviews of Wikipedia. Our review here covers most of the articles that he reviewed, except for a few conference papers.

Martin's review begins with a lengthy and very detailed description of the structural technical elements of Wikipedia from the perspective of three important database tables that store Wikipedia data: page (describing all kinds of Wikipedia pages, their namespaces and categories, and the links between pages); user (describing key user characteristics, as well as their watch lists, permissions, and log of their activities); and text (the actual content of the pages, including links to archives of previous versions). However, he did not cover images, which constitute another important table. Similar to Medelyan et al.'s (2009) extensive introduction to Wikipedia's structure, Martin's overview here is valuable to researchers whose quantitative work requires navigation of Wikipedia's database.

The rest of the review discussed a broad range of Wikipedia studies. Martin divided his coverage into six major categories: article quality issues, including vandalism and trends in quality improvement; trust or reliability of articles; semantic extraction; governance and society; economic implications; and epistemology. His review is valuable for researchers interested in these aspects of Wikipedia.

Bibliometric Analysis of Research Related to Wikipedia (2011)

The only other study that could be called a comprehensive literature review of Wikipedia studies is, in fact, actually a bibliometric analysis. Park (2011) investigated the extent to which scholars study Wikipedia and cite Wikipedia in their scholarly works. His number of identified articles is very different from ours because his methodology was quite different in a number of ways. To identify articles that treated Wikipedia, he searched for "Wikipedia" in the topic, title or reference of articles in the ISI Web of Science database and in the title, abstract, keyword and references of articles in the Scopus database; all such articles were included. However, our own search used in this article found that the vast majority of such articles do not actually treat Wikipedia; they only cite it or mention it in passing; we examined each individual article and excluded those that are not applicable. Park's study did not examine individual articles; he included all 1,746 identified publications in his bibliometric analysis. However, "the number should be taken with caution due to overlapping coverage of publications between [Web of Science] and Scopus," which seems to imply that he might not have removed duplicates. Another significant difference with our methodology is that Scopus yielded 921 conference papers; we only included around 100 of the most highly-cited conference papers. Moreover, Park's study included numerous non-peer-reviewed contributions, such as editorials and book reviews, which we excluded.

Park's study did not examine or discuss individual articles, but rather focused on reporting bibliometric measures: total numbers of studies, leading authors, their institutional affiliations, most frequent publication sources, main academic fields, and various statistics regarding the frequency that scholarly articles cite Wikipedia. We will not duplicate his bibliometric analyses in our study, and so refer readers to his article for these statistics. The only exception is that we did our own coding of academic fields per article, which we believe is more specific and hence more accurate than the per-journal coding of Web of Science and Scopus that Park used.

---

[4] His eventual dissertation (Martin 2012) mentions Wikipedia in passing only once.

Wikipedia as an Input-Process-Output System (2012)

Jullien (2012) conducted a very broad review that is one of the most extensive thus far conducted, with detailed insightful descriptions of over 250 scholarly works on Wikipedia. However, considering the enormous amount of literature, he screened out certain broad categories of studies: "the impact of the project on the environment ... such as how it is used to [accomplish] professional tasks (by the students, the researchers, the people in the industry), ... the analysis of the propositions to improve the tools (using it on mobile, creating a 3D Wikipedia), ... the use of Wikipedia as a database for information retrieval.... (2012, p.6). Moreover, his review was mutually exclusive in coverage with that of Medelyan et al. (2009), which uniquely focused on another large group of studies that Jullien expressly excludes from his scope, calling such topics "algorithm research, data-mining, computational intelligence, semantic, information retrieval" (2012, p.6). Our review, in contrast, includes all such studies in its scope. However, both Jullien and Medelyan et al. fully include conference proceedings in their literature scope, whereas our coverage of conference articles is limited to less than one hundred of the more highly cited articles. Jullien also discusses scholarly books on Wikipedia; our coverage of these is limited to brief summaries.

Jullien (2012) structured his review on a general input-process-output model. He grouped input-oriented studies as those who considered the Wikipedia environment and policies in place, and studies that investigated why people participate in Wikipedia. He grouped process-oriented studies (that is, those covering patterns of interaction) as those that investigated the activities and roles of Wikipedians; the structure and organization of the articles themselves; and the structure and governance of the Wikipedia community. From the output perspective, he discussed studies that investigated the quality and effectiveness of Wikipedia processes; users' experiences (both contributors' and readers'); and the quality of Wikipedia articles. Overall, despites its limitations in coverage, Jullien's review is one of the broadest thus far conducted. Although we include all the peer-reviewed journal articles that he does and many that he does not, he does cover many relevant conference articles that we could not include. Moreover, his organizing framework of Wikipedia research as an input-process-output model provides an alternate and valuable lens from which to consider the body of research.

Quantitative Studies of Wikipedia Participation (2012)

Yasseri and Kertész (2012) adopted an even narrower approach to reviewing Wikipedia studies than the others we have described: they restricted their focus to studies of participation in Wikipedia that adopted quantitative modes of analysis, which they call "computational social science". They begin with a general overview of Wikipedia for new researchers, briefly describing the Wikipedia articles, the Wikipedian community, and "accessories", by which they mean policies, talk pages, categories, and other ancillary pages. They also briefly describe various ways to access Wikipedia pages, in addition to the live website.

The main part of Yasseri and Kertész's working paper reviews various quantitative aspects of what we generally describe in our own review as "Participation" topics, related not so much to the Wikipedia encyclopedia articles as to the interactions of the Wikipedians who collaborate—and contest—in the community. Their core review is divided into two major sections: first, "Editorial habits" described a host of contribution patterns; and second, "Conflicts and edit wars", examined unpleasant editing scenarios in great detail. In fact, since this latter focus is the authors' primary research interest in Wikipedia, they provide the most comprehensive treatment of this body of studies that we have thus far encountered.

With their narrower focus, Yasseri and Kertész's review does not attempt to comprehensively span all Wikipedia studies. However, the significant advantage of their approach is that they are able to analyze their selected studies much more in-depth than most of the preceding reviews—and certainly more so than ours here—which provides an invaluable resource for researchers interested in that scope and approach to Wikipedia research. This is similar to the restricted attention of Medelyan et al. (2009) to corpus topics. We believe that because of the sheer bulk of the studies—and their continuing growth in

number—future literature reviews of Wikipedia studies would like adopt a similar approach of deeper review of a narrower subset of studies.

Wikimedia Research Newsletter (Since 2011)

For many years, the Wikimedia Foundation has maintained the wiki-research-l mailing list, which has been and remains the primary communication channel among Wikipedia and wiki researchers. Researchers would often announce newly published papers to each other on this list. In addition, since 2005 the Wikimedia Foundation has maintained a weekly newsletter called *The Signpost* (originally *The Wikipedia Signpost*) that reports news related to Wikipedia and other Wikimedia Foundation projects[5]. This has also been a regular forum for announcing published research on Wikipedia.

In July 2011, the Wikimedia Foundation launched the monthly Wikimedia Research Newsletter (WRN)[6], dedicated to announcing and summarizing scholarly research on Wikipedia and other Foundation projects. While it does not provide exhaustive searches of relevant research, inclusion of studies is based on the alertness of the Wikimedia research community. Anyone is invited to provide summaries of research, or they can alert the Wikimedia Research Committee to write the summaries.

The WRN provides clear summaries of research on any Wikimedia Foundation project, so its inclusion is broader than our scope here. However, it is a living and ongoing project, and has published regularly since its inception.

Because of the enormous scope of our review, and because of our limited resources for carrying out, we have set the cut-off date for the inclusion of studies in our review to July 2011, when the WRN was launched. This enables us to focus on extracting, summarizing and detailing the studies before that date. Beyond July 2011, we refer readers to the WRN for more recent scholarly work on Wikipedia.

Future Literature Reviews (After 2012)

The primary limitation of any literature review is that once it is published, its information remains static, whereas new relevant research continues to be published. Although it is impossible for us to maintain an ongoing dynamic review of Wikipedia research, we have attempted the next-best thing with a particular feature of our WikiLit site: Google Scholar forward citations. We describe this feature in detail in our description below of the WikiLit site, including how it can support future literature reviews of Wikipedia research.

Systematic Review Methodology

To assure a rigorous review, we closely followed Okoli and Schabram's (2010) detailed guidelines for conducting a systematic review. In fact, although those guidelines apply generally to information systems research, they were developed specifically with this review project in mind, and so they are most appropriate.

Research Questions

For this systematic review of research on Wikipedia, our questions are fairly broad, as they intend to cover the breadth of research that has been conducted on this vast field. We have the following specific research questions:

---

[5] https://en.wikipedia.org/wiki/Wikipedia:Wikipedia_Signpost/About

[6] https://meta.wikimedia.org/wiki/Research:Newsletter

1. *What high-quality research has been conducted with Wikipedia as a major topic or data source? Who (which authors) have published, when (year), and where (journal, conference)?* We intend to seek out and catalogue the existing high-quality research (as defined later in this protocol) that has been conducted on Wikipedia, for the benefit of helping potential consumers of this research become aware of its existence. We shall also analyze research publication trends during the first decade of Wikipedia.
2. *What findings have been drawn from existing research?* What are the conclusions and contributions to knowledge that researchers have drawn from their research?
3. *What are the details of research designs and approaches that have been adopted to study Wikipedia?* As a guide for future studies, we want to identify how researchers have approached Wikipedia detail. This involves many details, such as research designs, data collection approaches, and so on.

Through our review, we have collected a comprehensive list of available articles, as well as their locations in literature databases. We identify and organize the lines of thought that have been pursued thus far, to understand the history and direction of the field, and to offer a guide for potential future research and practical application.

Protocol and Training

To establish a plan for the execution of this review, we prepared a detailed review protocol, following Kitchenham et al.'s recommendations (Brereton et al. 2007; Kitchenham & Charters 2007). We obtained peer review of this protocol, published it at as a working paper (Okoli & Schabram 2009a), and we presented it in a poster session (Okoli & Schabram 2009b) and at an intensive research workshop (Lanamäki et al. 2011). This protocol provided a detailed manual for training our review team to assure the rigor of our work. All reviewers who searched for articles and extracted details read Okoli and Schabram's (2010) methodology manual. In subsequent descriptions of our methodology, we describe the measures we took to assure the consistency and accuracy of the execution of this review.

Practical Screen

To assure an unbiased selection process, we specified in advance in the protocol the criteria for inclusion and exclusion of articles from the final study. This "practical screen" (Okoli & Schabram 2010) is carried out at the initial stage to weed out articles, not based on their quality, but rather on two practical criteria: whether the study's content is applicable to the research questions; and whether it meets other explicit practical constraints. Based on Fink's (2005) criteria for the practical screen, we restricted our included studies to content limited by our scope; to research found through our English-language databases; and to peer-reviewed journal articles and doctoral theses.

First, we addressed the first research question: "What high-quality research has been conducted with Wikipedia as a major topic or data source?" When available as a search option, we limited our search to the articles' title or abstract, since we included only articles that treated Wikipedia as a significant subject, rather than those that merely refer to it. When such a refined search was not an available option for a given database, we searched the full text and examined the article for appropriateness.

Second, because of the practical limitation of the research milieu of our research team, and because of the vast number of studies under consideration, we had to restrict our search to studies in English-language databases. It is unfortunate that this study had to exclude the significant work being conducted in other languages, such as German or French; however, we are unable to do adequate justice to the literature in other languages. Nonetheless, we hope that the explicit reporting of this review would serve as a model for its replication in other languages.

Finally, we choose to include work only peer-reviewed journal articles and doctoral theses. Doctoral theses are reviewed by qualified academics, and we are aware of some significant work that has been

done by students; it would be in the interest of identifying quality research to exclude these arbitrarily. However, for the sake of restricting the practical scope of our study, we eliminated non-peer-reviewed journal articles from our systematic search.

We initially tried to include peer-reviewed conference articles, but after locating over 1,500 such publications, we were forced to make the difficult decision to exclude these from our systematic search. We are simply unable to cover such a vast body of research. Nonetheless, no review of scholarly research can be comprehensive that ignores the very significant work that has been published only in conferences. Thus, although we did not include these in our systematic search, we carefully selected almost 100 key conference publications in our supplementary search, whose methodology we describe below. Nonetheless, we recognize our inability to systematically include conference papers is an important limitation of our review methodology.

In our desire to be as exhaustive as possible, we did not apply any other practical screening criterion beyond those listed above.

Searching for the Literature

Because of the relative recency of the Wikipedia phenomenon (Wikipedia was launched in January 2001), we conducted only electronic searches, since virtually all related publications are electronically indexed. To assure exhaustiveness in our search, we did not assume the appropriateness of any particular subject domain. Consequently, we searched through all 484 English-language databases of scholarly literature available at Concordia University, Montreal (as of 2009). These databases span almost all areas of inquiry: business/commerce, the fine arts, humanities, science and engineering, and the social sciences. Note that we did not search Google Scholar, since its results were undiscriminating and hence useless for our searches.

In all databases, we searched only for three words: "Wikipedia," "Wikipedian," and "Wikipedians." (A "Wikipedian" is a person who contributes to Wikipedia.) Other than these three keywords, no synonym is appropriate, as our study is uniquely on Wikipedia. We are not studying wikis other than Wikipedia; we are not studying MediaWiki (Wikipedia's wiki platform); we are not studying other Web 2.0 phenomena such as blogs; we are not studying any Wikimedia Foundation project other than Wikipedia. For our purposes, these three simple keywords are sufficiently inclusive (to capture all relevant studies) and exclusive (to not capture unrelated studies). Depending on the specificity of the database, when possible we restricted our search to titles, abstracts and keywords or subject entries. Searching the full text yielded thousands of irrelevant entries (such as citations to Wikipedia articles or only passing mentions). The irrelevancy of such searches bore out for those databases that only permitted full-text searching; we did search the full text in such cases. This more focused keyword search distinguishes our review from Park's (2011) analysis; he included all articles that had "Wikipedia" anywhere, without verifying the appropriateness of such articles.

In a preliminary search following this methodology, we were able to eliminate the majority of 484 databases as not containing a single relevant article; we were left with a total of 74 applicable databases that contain one or more articles on Wikipedia.

We selected the articles for inclusion by following this procedure: three reviewers from our team worked through around 30 or so of the articles together, to ensure that everyone understood the working principles; these articles were randomly selected from fairly diverse databases, to give a taste of the variety of what might be encountered. Next, two reading reviewers each randomly received half the articles; the randomization was achieved by sorting the articles alphabetically by title. The alphabetical sorting and distribution was helpful in record keeping and tracking which reviewer was assigned which article, yet assured random assignment of articles by topic or subject matter.

Each reviewer scanned their articles, and decided whether to include or exclude them. Reasons for exclusion of specific articles were recorded. Then each reviewer verified the articles that the other excluded. Brereton et al. (2007) found that, in a large systematic review, having one reviewer score and another verify the decision is more or less as effective in decision quality, yet more efficient in time, than having both reviewers score all articles in detail. Any article that one of these reviewers felt should be included was retained; thus, at this stage, we favored the retention of articles in the study.

By December 12, 2010, we collected 6,107 articles from all the databases, before we removed duplicates across databases. After removing duplicates, we had 2,678. In addition, AISEL added 2 valid journal articles (1 original, 1 duplicate) and 10 valid conference articles (all original); 2 invalid conference articles in the AISEL search were not included. Thus, after duplicates, we had 2,689 articles. This number was before removing articles that failed the practical screen.

To validate the exhaustiveness of our databases, we verified our results by consulting with subject "experts" (2005). For our topic, we had two sources of "experts" who are competent to validate our search. First, the Wikimedia Foundation offers two bibliographies of research on Wikimedia projects[7,8] (including non-Wikipedia studies). These lists were compiled by various researchers who self-reported their work and that of others to the Foundation wiki pages.

We compared these lists of studies with the list that we retrieved from our searches. By exhaustively examining each item posted on these pages, we identified only 13 peer-reviewed journal articles and 4 doctoral theses that were not already located by our prior searches. The reasons that we missed them were that 3 were either forthcoming or indexed past our cut-off date of November 2010; 3 North American articles and 2 European articles were not indexed in any database at all (except perhaps Google Scholar); 4 European articles were published in journals that normally publish non-English articles; 1 article was marginally relevant to Wikipedia, and did not mention it in the title or abstract; 3 of the 4 theses were published outside North America; and 1 North American thesis (published in 2010) was indexed after November 2010.

In addition to these, we added 1 article we personally knew about that was relevant, yet did not mention Wikipedia in the title or abstract; we also added 1 forthcoming article we learnt about from the Wikipedia Signpost weekly newsletter.[9] This gave a total of 603 peer-reviewed journal articles, 29 doctoral theses, and 50 conference articles for a total of 682 items.

We merged the articles from these pages with our own search results and posted them back onto the page listing academic studies of Wikipedia[10].

The second source for experts is the Wikimedia researchers' mailing list (wiki-research-l) hosted by the Wikimedia Foundation. The subscribers to this list are active Wikipedia researchers; we presented our compiled list up to this point to them, and asked them to identify any research they are aware of that meets our criteria which we might have missed. We thus identified a further five to ten articles through this source. Yet another source came through the list of nominees for the inaugural Wikimedia France Research Award in 2012[11]. From this source, we included eight additional conference papers nominated by miscellaneous individuals as "the most influential research paper on Wikimedia projects and free knowledge projects in general"; we also added one nominated journal article that was otherwise published after our July 2011 cut-off.

---

[7] http://en.wikipedia.org/wiki/Wikipedia:Academic_studies_of_Wikipedia

[8] http://meta.wikimedia.org/wiki/Research

[9] http://en.wikipedia.org/wiki/Wikipedia:Signpost

[10] http://en.wikipedia.org/wiki/Wikipedia:Academic_studies_of_Wikipedia

[11] http://meta.wikimedia.org/wiki/Research:Wikimedia_France_Research_Award

Supplementary Literature Searches

The advantage of a systematic literature search is that the methodology identifies a higher number of appropriate publications than any other approach (Petticrew & H. Roberts 2006). However, if the practical screen is too strictly adhered to, a purely systematic search will invariably miss some relevant studies. The "systematicness" of the search should be seen as a tool for including more sources than would be otherwise; it should never be used as a legalistic guide to excluding appropriate sources that are not identified by the systematic methodology. Thus, it is appropriate to supplement any systematic search with non-systematic approaches that can hopefully identify some additional appropriate sources.

Selective Search of Relevant Papers

One of the authors of this paper independently located many references to Wikipedia, scholarly and otherwise, that he judged relevant to Wikipedia researchers (Nielsen 2012). He located references mainly from the following sources:

- His Danish popular science summary of Wikipedia research from April 2008, containing 87 peer-reviewed and non-peer-reviewed references (Nielsen 2008b);
- A "Wikipedia" e-mail alert service from the E-Print Network of the United States Department of Energy's Office of Scientific and Technical Information (http://www.osti.gov/eprints/);
- The wiki-research-l mailing list;
- Other press sources, in particular Google News.

Backward Citations

Normally, in a systematic literature review, it is advisable to search the bibliographies of all located articles to identify further relevant articles; these are called backward citations. Unfortunately, because of the sheer quantity of studies in this review, we were unable to do this. However, we did examine the reference lists of all the literature reviews we had identified, and added any missing peer-reviewed journal articles papers that we located. In addition, to look for more reviews, we checked the list of literature reviews at WikiPapers[12]. However, we had independently identified all the reviews listed there.

Conference Papers

After acquiring over 1,500 conference articles, we were uncertain whether or not to include conference papers at all in the review—it is impractical to review so many articles in one study. We considered excluding conference articles completely, but after feedback from the wiki-research-l mailing list (a list of researchers of Wikipedia and other Wikimedia Foundation projects), we realized that some of the best and highest-cited Wikipedia research is actually published in conferences. Moreover, in some subfields of computer science such as Human-Computer Interaction, conference articles are terminal publication targets of the best quality work, rather than journals.

We thus used two sources to identify highly-cited conference articles. First, we made a citation-per-year ranking of the conference articles about Wikipedia, using citation data from the ACM Digital Library. After this we decided to include the top 50 highest-cited conference articles in our review, as well as a few other core publications suggested by wiki-research-l members. Second, we identified the 100 top-cited publications from a Google Scholar search on "Wikipedia." We included all 15 conference articles that were on that list that we did not already have, and that were not covered by Medelyan et al. (2009).

Based on these supplementary searches, we eventually selected over 60 peer-reviewed conference articles for full extraction and summarization in our review. In addition, we include many other conference

[12] http://wikipapers.referata.com/wiki/List_of_literature_reviews

articles in our summaries below, though we did not extract their full research details. We consider this a fair compromise, since we unfortunately were limited in personnel and resources to include more studies in this already extensive review.

Data Extraction

To answer the other research questions, we examined the papers more carefully than was done in our original scan for inclusion. Because of the sheer quantity of the studies, it was not practical to read each paper thoroughly. However, with a focused data extraction protocol, the necessary answers to the research questions were systematically obtained. The readers read the title, abstract, introduction and conclusion of all their assigned articles, then also whatever was necessary to extract the data needed to extract pertinent details. We list and explain the details of what we extracted in the section that describes the WikiLit website, where we eventually posted all the extracted research data.

To assure consistency in data extraction, three reviewers worked through around 30 of the articles together. Once everyone was satisfied that the data extraction procedures were understood, the two reading reviewers was randomly assigned half of the articles (alphabetically sorted) and extracted data from them independently. Each week, these two reviewers met with another reviewer to verify random extractions and to discuss problematic issues together and resolve discrepancies.

Synthesis

There are three major aspects of our synthesis of the studies we have located. First, we present the summarized findings from the studies we have identified in a comprehensive, organized and accessible format. For this purpose, we have used the hierarchical topics of the scholarly research as our organizing principle; we present these synthesized findings in the section on "Findings from Scholarly Research on Wikipedia." For these research summaries, we include not only the systematically extracted data that we present on the WikiLit website, but also the appropriate supplementary research that we have located. Thus, our summaries in this review article cover more studies than we parsed and included in the WikiLit website.

The next two steps of our synthesis are still in progress, and will be published in future versions of this review. Second, we will quantitatively analyze the words found in the abstracts of keywords from systematically extracted articles. The goal of this synthesis is to identify major patterns and associations in Wikipedia research. Third, we will carefully analyze and compare the detailed research data from our systematically extracted articles. We will outline and graph trends, directions and associations of the research. We expect that insights from this analysis will help direct researchers in new Wikipedia research and in research on related matters.

Having explained our methodology in detail, we now proceed to describe in more detail the WikiLit website, on which we present all the details of our systematically extracted articles.

WikiLit: A Semantic MediaWiki of Wikipedia Research

We have posted and described all the studies that we identified in our systematic review on a website we call WikiLit (http://wikilit.referata.com). This site serves as a project website, and has supported the collaborative work process of our geographically distributed research team. The website has also provided transparency and openness to the research process. Following Constantinides et al.'s (2012) suggestions for "the production of various subtexts in other places around the core research article" (2012, p.16), the WikiLit website has served as a platform for anyone to become informed about the project, to contribute with missing articles, to correct information, and to provide feedback.

Many of the articles that we identified through the supplementary literature search procedure we described earlier are also included on WikiLit. In addition, since it is an open wiki, many authors have added their own studies to this website. However, as many of the studies added by external authors do not contain complete research details, we do not summarize all of these in this article. In other words, the WikiLit website documents partial or full details of many studies that are not included in this paper that presents research summaries.

The WikiLit website runs on MediaWiki, the same wiki software that runs Wikipedia. However, we use the Semantic MediaWiki (SMW) extension (which Wikipedia does not use), which extends the wiki with database-like features that enables many advanced analysis features.

Unlike permanent sites like AcaWiki and Brede Wiki (which compile summaries of scholarly studies of all subjects) and WikiPapers (which compiles bibliographic and other details of scholarly research specifically on wikis), WikiLit is only a temporary project site to support this literature review of studies that focused on Wikipedia. We intend to host our extracted literature review data there during the process of the review, and then when completed, eventually export all the data to long-term sites like AcaWiki and WikiPapers and anyone else that wants it.

Extracted Research Data from the Publications

For each article that we identified for extraction, we extracted very detailed information pertinent to researchers; we described the extraction methodology earlier. In almost every case, individual articles might fit multiple options for any extracted research data field (for example, an article might cover multiple topics, and might use more than one research design). In addition to the following extraction details listed here, we also compiled summaries of each article. These summaries are not included on the WikiLit website, but are organized and presented in this review in the section on "Findings from Scholarly Research on Wikipedia."

► Humanities	► Interdisciplinary	► Social sciences
- History (6)	- Anthropology (1)	- Business administration (4)
- Leisure studies (1)	- Computer science (174)	- Communications (18)
- Linguistics (8)	- Health (21)	- Economics (10)
- Literature (1)	- Industrial ecology (1)	- Education (48)
- Music (1)	- Information science (63)	- Geography (7)
- Philosophy and ethics (19)	► Logic and mathematics	- Information systems (120)
- Rhetoric (9)	- Logic (0)	- Journalism (3)
- Theology (1)	- Mathematics (2)	- Knowledge management (14)
- Visual arts (0)	► Natural sciences	- Law (9)
	- Biology (3)	- Library science (35)
	- Chemistry (3)	- Marketing (1)
	- Forestry (1)	- Political science (8)
	- Geography (7)	- Psychology (11)
	- Physics (3)	- Sociology (29)

Number of studies in each domain are in parentheses

Table 1. Knowledge Domains Covered by Wikipedia Research in WikiLit

- **Topics** that are covered by various research articles on Wikipedia. The topics are hierarchically arranged, and provide the primary organizing principle for our description of the studies, as detailed in the major section "Findings from Scholarly Research on Wikipedia." Many articles covered multiple topics.

- **Domains** of human knowledge covered by the publication, such as natural sciences or humanities. It is represented by a category tree (Table 1)[13] with the following major groups: Humanities, Interdisciplinary, Logic and mathematics, Natural sciences and Social sciences. Our top-level domains are adapted from Gordana Dodig-Crnkovic (2002). However, we subsume her "Culture" domain into "Humanities." We defined the lower level categories ourselves based on the subject matter of the articles. Some sub-domains belong to multiple top-level domains; we classify these provisionally as Interdisciplinary.
- **Research questions** that the authors of the article have explicitly posed. Very often, this field consists of direct quotations from the article.
- **Theory type** is based on Gregor's (2006) theory types found in information systems research:
  - *Analysis*: Answers only the "what?" question. This category does not include causal relationships.
  - *Design and Action*: Describes the creation of an artifact or process, or presents a practical solution.
  - *Explanation*: Presents a causal relationship with an explanation of why and how it operates.
  - *Prediction*: Projects expected future results
- **Research questions** that the authors of the article have explicitly posed. Very often, this field consists of direct quotations from the article.
- **Wikipedia coverage** refers to the sense in which the study is "about Wikipedia" in order to qualify for inclusion in this review. What is the nature of coverage of Wikipedia in the study?
  - *Main topic*: Wikipedia is explicitly and unambiguously the primary focus of the study.
  - *Case*: Wikipedia is one substantive case treated in this study, but it is not the only one.
  - *Sample data*: The study's purpose is not about Wikipedia per se, but a significant portion of the study's data is taken from Wikipedia in order to achieve the study's objectives.
  - *Other*: The study is about Wikipedia in some other sense not captured by the other three categories. Note that this option is not for studies that only marginally refer to Wikipedia (e.g. gives a definition or description for a term taken from Wikipedia)—such studies are explicitly excluded from this review.
- **Theories:** Various theoretical bases, frameworks and perspectives that the study draws upon or builds. These data were mainly copied from the article, and thus are not structured.

Research design refers to the general methodological approach for conducting a research study. In this review, we use the following classifications based on various sources (Experiment-Resources.com 2008; MBA Knowledge Base 2011; Ratcliff 2004; Myers 2008; Järvinen 2008)

- to describe a research design:
  - Non-empirical studies
    - Mathematical modeling (without empirical data)
    - Theoretical/Conceptual
  - Quantitative empirical
    - Mathematical modeling
    - SEM (Structured equation modeling) or PLS (Partial Least Squares)
    - Econometrics and time series
    - Experimental and Quasi-experimental Research
    - Other Statistical Analysis (not otherwise specified)
    - Meta-analysis
  - Qualitative empirical
    - Action research
    - Case study
    - Ethnography

[13] http://wikilit.referata.com/wiki/Category:Domains

        - Grounded theory
        - Historical analysis
        - Qualitative literature review (including Systematic Review)
        - Semiotics
        - Discourse analysis
        - Hermeneutics
        - Narrative and metaphor
        - Typology/taxonomy
        - Content analysis
        - Phenomenology
- **Collected datatype** is the nature of the collected data. The possible values are:
    - *Archival records:* Any kind of data that has been recorded or stored prior to the commencement of the study. In particular, this includes secondary datasets.
    - *Computer usage logs:* Logs that track computer or Internet usage, such as webpage browsing logs and search logs.
    - *Direct observation:* Direct live observation of peoples' behaviors. In the context of Wikipedia research, this does not include reading of community activities as documented on various pages on the website; these would fall under "Wikipedia pages."
    - *Interviews:* Synchronous interviews of live persons, whether face-to-face, online, by telephone, or by other means.
    - *Survey:* Questionnaire sent out for respondents, whether online, by mail, or by telephone.
    - *Documents:* Other documents not included in the other categories.
    - *Websites:* Any website data other than Wikipedia, though Wikipedia may be one among others.
    - *Experiment:* An experiment set up under controlled conditions to investigate a particular phenomenon while minimizing potentially confounding factors.
    - *Literature review:* Search and aggregation of articles in a scholarly literature review.
    - *Wikipedia pages:* Any kind of page from any namespace of Wikipedia.
- **Collected data time dimension** is the time dimension of the collected data. The possible values are:
    - *Cross-sectional*: Data is collected at one fixed point in time, that is, data is a snapshot of the current state of Wikipedia.
    - *Longitudinal:* Data is collected at two or more distinct points over time
    - *Both:* Both cross-sectional and longitudinal data is collected
    - *N/A:* The time dimension of collected data is not meaningful in the context of the study under consideration
- **Unit of analysis of the study:** A single study can treat more than one unit of analysis at a time. The possibilities are:
    - *Article:* Treats each individual encyclopedia article in Wikipedia. This could include focusing on the title of each article, or on the talk page of each article; however, in either case, the article is still the unit of analysis.
    - *Article view:* Treats each page view of an article as a unit.
    - *Category:* Treats each category defined in Wikipedia as a unit.
    - *Edit:* Treats each individual edit made to Wikipedia as a unit.
    - *Language:* Treats each language version of Wikipedia (e.g. en, de, fr, cn) as a unit
    - *Scholarly article:* Treats each scholarly article published about Wikipedia as a unit. This unit of analysis is unique to literature reviews.
    - *Subject:* Treats each subject or topic (e.g. sports, musicians, etc.) as a unit. This does not include studies that examined only articles within one subject (e.g. psychology articles)—in such a case, the article might be the unit of analysis.
    - *User:* Treats each Wikipedian as a unit.
    - *Website:* Treats each website as a unit. This is usually the case for studies that compare Wikipedia with other websites.

- **Wikipedia data extraction** refers to the general means by which Wikipedia data was obtained for the purpose of the study. The options are:
  - *Live Wikipedia:* Data was extracted from accessing the live Wikipedia website. This includes data extracted from history pages on the live Wikipedia, as long as a local version of Wikipedia was not reproduced to obtain the data.
  - *Clone:* A local version of Wikipedia was installed and analyzed, usually based on historical data dumps.
  - *Secondary dataset:* A preprocessed dataset of Wikipedia was used to obtain the data for analysis. That is, the researchers depended on someone else's reprocessing of a Wikipedia clone.
- **Wikipedia page type** refers to the type of Wikipedia page that is analyzed in the study. We list the following page types:
  - Article
  - Article:talk
  - User
  - User:talk
  - Policy
  - Discussion and Q&A
  - Log
  - Collaboration and coordination
  - Conflict resolution
  - Information categorization and navigation
  - Quality management
  - Other
- **Wikipedia language** refers to which language version of Wikipedia was used for the study. The following notes are pertinent:
  - "English" means that the English Wikipedia (en) was explicitly mentioned.
  - "Not specified" means that there is no mention of which language version was used. However, since all the studies included are in English, we believe it is safe to assume that "Not specified" means English in the context of this review.
  - In general, we list each individual language version used in the study up till the first five. When more than five studies are included, we use the label "Multiple."
  - "All languages" is used only when all the existing languages are compared in a study (though perhaps a few minor ones were excluded for some given reason).
- **Conclusions** that the authors of the article have drawn from their study. Very often, this field consists of direct quotations from the article.
- **Comments** that we compilers have made or copied from the article.

Navigating the WikiLit Site

The main page gives brief instructions on how to browse the WikiLit site. The primary recommended means of navigation are as follows:

Searching for keywords: The fastest way to locate any specific item is to enter search words (for example, keywords, an article title, or a researcher's name) in the search box on the top of any page.

Browse for articles by topic: If interested in a general topic of interest, the topical directory (http://wikilit.referata.com/wiki/Category:Topics) provides a hierarchical, expandable list that eventually lists every article on the site according to its relevant topics. Articles are listed multiple times when they treat multiple topics. This is the same topic hierarchy used in the Findings section of this article. However, this article includes summaries of many articles that are not included on the WikiLit site, especially for conference papers and non-peer-reviewed work. Moreover, the WikiLit site includes some

articles that do not appear in this review, when added by people other than the authors of this review. These articles include conference articles and articles published after our cutoff date.

Browse for articles by domain: If interested in a domain of knowledge, the directory of domains (http://wikilit.referata.com/wiki/Category:Domains) likewise provides a hierarchical, expandable list that lists every article on the site according to domain. Articles are listed multiple times when they involve multiple domains.

Using the WikiLit Site for Literature Reviews and Forward Citations

As we mentioned earlier, traditional literature reviews are limited by the fact that they become outdated once they are published, since any research conducted after the publication date cannot be included (except by an update of a previous review, which, once published, itself henceforth suffers the same weakness). To partially alleviate this weakness, we have included the Google Scholar link for forward citations for each of the articles on WikiLit, as following forward citations is an established practice for capturing the latest publications within a research area (Webster & Watson 2002, p.xvi).

A backward citation is what is normally known as a reference within a published article; it refers to an article that the present article cites. A forward citation, in contrast, is a future publication that cites an article. For example, if article B (published 2012) cites article A (published 2009), then A is a backward citation of B. If later article C (published 2014) cites A and B, then C is a forward citation of A and B.

Databases of scholarly literature often store the references in an article. Storing these linkages enables backward and forward citations to be easily traced and navigated. Google Scholar not only features the world's largest database of scholarly publications (thus recording the most citations), but it is also freely accessible on the Internet. Thus, we chose its database for linking citations between articles on WikiLit. We label this item "Google Scholar citations" in the infobox that lists publication details on the page for any article. The item displays the Google Scholar ID for the article, and links to the Google Scholar page that lists all recorded forward citations of the article.

The immediate value of this item is that researchers can readily locate any article that cites a given article. For the purpose we have mentioned—conducting future literature reviews—this feature can be used to not only search for forward citations for a single article, but to carefully expand a Wikipedia topic area to find new articles related to the topic.

We will illustrate this usage with a topic area that we consider interesting, but thus far little covered in Wikipedia. We have identified only five articles that objectively evaluate to what degree Wikipedia articles are up-to-date; these are listed in the Currency topic category of our review. A researcher interested in this area of research could take the following steps to explore further:

1. Obtain the articles listed on WikiLit by navigating to the Currency topic (http://wikilit.referata.com/wiki/Category:Currency). That page lists all identified articles according to our criteria for that topic. (In addition, the Currency category of this review includes two additional studies that we identified that do not meet the WikiLit website criteria that we described in our inclusion criteria earlier; one is an article in Danish (Bekker-Nielsen 2011) and the other is an unpublished conference video (Wedemeyer et al. 2008).)
2. Go to each relevant article on WikiLit, and obtain the papers. For example, one of the listed papers is "Philosophy democratized?" (Elvebakk 2008)[14]. Often, as in this case, the URL for the article is available under the "List" item of the Publication infobox. Sometimes, the Document Object Identifier (DOI) is available, whose link leads to the publisher's official page for the article. If none of these leads to an available copy of the article (e.g. the publisher version is

[14] http://wikilit.referata.com/wiki/Philosophy_democratized%3F_A_comparison_between_Wikipedia_and_two_other_Web-based_philosophy_resources

protected by a paywall), some of the search options, especially the Google Scholar search, might lead to a freely downloadable version of the article. If none of these Web options leads to the article, then the researcher should contact a librarian for help.

3. After reading or examining the papers, the researcher can find related research by scanning the bibliography of the actual papers (electronic or print) for backward citations.
4. For forward citations, the WikiLit article page has an option labeled "Google Scholar citations" which records the article's Google Scholar ID and links directly to the article's citation page. By scanning the list of subsequent articles that cite the given articles, the researcher can see if there are any possible future articles that also objectively measure the currency of Wikipedia. (As of the time of publication of this review, there are none, but in coming years there most likely will be some.)

The same procedure described can also be used for searching for other articles that treat any topic of Wikipedia. It can also be used to search for articles that cover Wikipedia in any domain of knowledge, such as Music (http://wikilit.referata.com/wiki/Category:Music).

Findings from Scholarly Research on Wikipedia

Table 2 displays the topic categories of studies in our sample, with the number of studies in each category. Note that the numbers of articles in the leaf (terminal) categories do not add up to any meaningful total, since most articles cover more than one topic category and are thus counted multiple times.

We have grouped the articles into six general categories, which are not of our own origination. It is very important that the results of this literature review be effectively disseminated, not only to Wikipedia researchers, but even more so to the practitioners who develop and administer Wikipedia on a day-to-day basis. To facilitate this practitioner dissemination, we have drawn our four major categories from the Wikimedia-pedia (http://strategy.wikimedia.org/wiki/Wikimedia-pedia), a collection of strategy documents maintained by the Wikimedia Foundation whose goal is to compile knowledge that is currently known concerning the Foundation's projects and to also field research questions for desired knowledge. The Wikimedia-pedia is grouped under four major categories, which we have paralleled in our categorization, though we have renamed them to more appropriately reflect the contents from a research topic perspective: First, they have Reach, which we call Readership, since this category concerns reaching readers around the world with Wikimedia content. Second is Quality, which we call Content, referring to studies concerned with actual content of the encyclopedia articles; this includes the quality of articles but also issues such as the overall size of Wikipedia. Third is Participation, whose name we retain unchanged, referring to studies concerning to Wikipedians as community members, their contribution to articles and other kinds of collaboration. Fourth, they have Operations, which we call Infrastructure, referring to studies about the organizational, legal, and technological infrastructure underlying Wikipedia.

In addition to these four categories from the Wikimedia Foundation, we have added the Corpus category, concerning the use of Wikipedia as a textual corpus for scholarly research; this is a category quite unique to Wikipedia researchers; our sub-classifications here are mainly based on Medelyan et al.'s (2009) literature review that focused exclusively on this category of Wikipedia research. Finally, we classify some research articles as "General," when they look at Wikipedia as a whole in a way that is not effectively captured strictly by one of the other categories.

Based on these six major topic categories, the two data coders among us initially assigned each article they coded a sub-category. All of us co-authors then got together to verify these categorizations, and then redistributed the articles for verification and recategorization. After a couple of rounds of verification together, we finalized our categorizations of each article. As can be expected, many articles were assigned

multiple topics in our classification, since they covered multiple aspects of Wikipedia. In the following subsections, we discuss and synthesize findings from each of the topic categories in our study.

- **Content**
  - Other content topics (8)
  - Quality
    - Antecedents of quality (17)
    - Comprehensiveness (22)
    - Currency (5)
    - Featured articles (20)
    - Readability and style (10)
    - Reliability (31)
  - Size of Wikipedia (12)
- **Corpus**
  - Information retrieval
    - Cross-language information retrieval (5)
    - Data mining (7)
    - Geographic information retrieval (3)
    - Information extraction (15)
    - Multimedia information retrieval (4)
    - Other information retrieval topics (10)
    - Query processing (6)
    - Ranking and clustering systems (14)
    - Text classification (10)
    - Textual information retrieval (5)
  - Natural language processing
    - Computational linguistics (7)
    - Other natural language processing topics (7)
    - Semantic relatedness (17)
  - Ontology building (21)
  - Other corpus topics (10)
- **General**
  - Encyclopedias (10)
  - Epistemology (20)
  - Ethics (6)
  - Literature review (7)
  - Miscellaneous topics (4)
  - Research platform (10)
  - Wikipedia as a system (6)
- **Infrastructure**
  - Legal infrastructure (6)
  - Technical infrastructure (21)
- **Participation**
  - Antecedents of participation
    - Contributor motivation (34)
    - Cultural and linguistic effects on participation (10)
    - Other antecedents of participation (8)
    - Societal antecedents of participation (11)
  - Collaborative culture
    - Community building (13)
    - Contributor engagement (11)
    - Culture and values of Wikipedia (12)
    - Deliberative collaboration (21)
    - Other collaboration topics (26)
    - Policies and governance (33)
    - Quality improvement processes (16)
    - Scholarly contribution (3)
    - Social order (15)
    - Student contribution (17)
    - Vandalism (15)
  - Participation outcomes
    - Contributor perceptions of credibility (4)
    - Other participation outcomes (12)
    - Participation trends (16)
  - Software for participation
    - Collaboration software (9)
    - Reputation systems (6)
- **Readership**
  - Commercial applications (10)
  - Knowledge source
    - Health information source (14)
    - Judiciary use (1)
    - Knowledge source for scholars and librarians (14)
    - News source (3)
    - Other knowledge source topics (2)
  - Ranking and popularity (11)
  - Reader perceptions of credibility (21)
  - Software for readership
    - Computational estimation of reliability (5)
    - Reading support (3)
  - Student readership
    - Cross-domain student readership (16)
    - Domain-specific student readership (13)
    - Student information literacy (12)

Number of studies in each domain are in parentheses

Table 2. Categorization of Topics of Wikipedia Research in WikiLit

The topic categories in this article closely mirror the topic hierarchy on the WikiLit website (http://wikilit.referata.com/wiki/Category:Topics); as much as possible, we use exactly the same topic headers. Thus, interested readers can easily find more details on the relevant articles on the website. However, to facilitate our presentation, we sometimes subdivide our summary descriptions in this article beyond the subcategories in the website.

Because of the sheer vastness of the research on Wikipedia, we recognize that it might be somewhat difficult to sequentially read the descriptions of all the articles in this review. Thus, readers should feel

free to skip around in this section, reading the subheadings and only reading the article descriptions of those topics that suit a reader's particular interests.

General: About Wikipedia in General

In our sample of 477 studies, 59 General articles (12%) covered Wikipedia-related issues very broadly, covering a wide swath of aspects of Wikipedia in the study that cannot be confined to any of our other major categories. The studies we describe here treated Wikipedia as an encyclopedia; studied epistemological issues related to Wikipedia; discussed Wikipedia as a platform for scholarly research; considered Wikipedia as a system; and covered other miscellaneous topics not suitably categorized elsewhere in this review.

Encyclopedia

A number of articles treated Wikipedia from a functional perspective straightforwardly as an encyclopedia; these articles mainly examined to what extent Wikipedia has succeeded in building a 21st century encyclopedia that incorporates "the sum of all human knowledge" (Wales 2004). In 2011, Benjamin Mako Hill presented his research about Wikipedia and other crowd-sourced online encyclopedia, Interpedia, The Distributed Encyclopedia, Everything 2, h2g2, The Info Network, Nupedia and GNUpedia, trying to answer what distinguished the successful Wikipedia from the failed or less successful projects. Hill noted that Wikipedia offered low transaction costs in participation and initial focus on substantive content rather than technology (Garber 2011).

Reagle (2008) analyzed the history and ethnography of Wikipedia to understand its "good faith" culture. His arguments aimed to support Wikipedia being "the closest realization yet of a long held aspiration for a universal encyclopedia" (2008, p.3). First, he framed Wikipedia historically as the newest producer of universal encyclopedia. He discussed similarities and differences between Wikipedia and its antecedents and highlighted reasons for Wikipedia success. Second, he presented projects such as Gutenberg, Interpedia and Nupedia. The major difference between these projects and Wikipedia is the incorporation of the new wiki technology. Even though wikis helped realizing the vision of a universal Wikipedia, the issue of experts versus amateurs goes on. Third, he discussed "encyclopaedic impulse" by presenting Daniel Pink's model of the three periods of encyclopedic production. Then, he presented five characteristics of open content communities: open products, integrity, transparency, non-discrimination and non-interference, emphasizing Wikipedia's collaborative good faith culture and its leadership development. Reagle concluded that apart from technology, Wikipedia's success is due to its collaborative culture. In general, this doctoral dissertation is very positive towards Wikipedia, notwithstanding the chapter on carefully presented points of criticism. Reagle's dissertation was later published as a popular book, *Good Faith Collaboration* (Reagle 2010b).

In her dissertation, Kennedy (2009) compared the creative process of Ephraim Chambers' 1728 *Cyclopædia* with that of Wikipedia, focusing on the notion and process of authorship in the two encyclopedias. She found that the two were remarkably similar in being collaboratively created works whose goal was not creation of original content so much as the reliable compilation of existing knowledge. Although constrained by the limitations of 18th century communication technology, Chambers did make a concerted effort to incorporate contributions of others, including of readers of the encyclopedia, into his work. Moreover, he eschewed considering himself the primary author, viewing his compilation as a collaborative work. Kennedy considers the development process of both encyclopedias most accurately described as a work of "curation," that is, a dynamic activity of categorizing and displaying various bits of knowledge as they were collected, rather than aiming to produce a final, static product. She considered that the hypertext mode of both encyclopedias (the 1728 Cyclopedia achieved this with multiple internal cross-references) made each reader's reading experience unique, thus making each reader an author of their own version of the encyclopedias' readings. Finally, she explored the ambiguity of authorship found in the creation of Wikipedia content by bots, which have dual authorship

in the human bot creators and in the software bots themselves. With their limited creative input, she considers both bot creators and bots themselves to be "compilers" rather than "authors" in the more traditional sense.

Some other contributions have situated Wikipedia as part of the evolving encyclopedic tradition. Stakić (2009) introduced wiki technology and Wikipedia in particular to Serbian librarians. A notable part of his article is his brief tracing of the history of encyclopedic compilation, which he frames as cumulating in wiki technology and the wiki way of mass collaboration, as manifested in Wikipedia, as the state of the art of the accumulation of encyclopedic knowledge. Kohn (2010) evaluates past editions of the *Encyclopaedia Judaica*, and considers future models for its republication. He suggested that "the model offered by Wikipedia could work well for the *Encyclopaedia Judaica*, allowing it to retain the core of the expert knowledge, and at the same time channel the energy of volunteer editors" (2010, p.249). Haider and Sundin (2010) argued that Wikipedia, in its striving to maintain a neutral point of view, has come to symbolize contemporary views on knowledge. It enhances the status of lay people by functioning as a space for our cultural memory. Wikipedia has successfully combined the enlightenment ideals of encyclopedias with contemporary ideals of knowledge construction that "breaks with the tradition of controlled expertise." Kolbitsch and Maurer (2004) proposed adding community building futures to collaborative knowledge development systems like Wikipedia, as well as features to make the encyclopedia more flexible and dynamic like omnipresent annotations and active documents. Through interviews with Swedish Wikipedia administrators, Mattus (2008) explored the nature of the collaboratively shaped encyclopedia. She concluded that "Wikipedia is not ready-made as are traditional encyclopaedias; it is a product collaboratively constructed in present time" (2008, p.197) and thus it should "be interpreted and considered on its own terms" (2008, p.183).

In contrast to these generally positive evaluations, Wikipedia has also been considered a definite step backwards in encyclopedic evolution. Larry Sanger, the co-founder of Wikipedia who later became one of its most vocal critics, discussed the role of experts in the development of Wikipedia content (2009). He set up a straw-man proposition—one that no Wikipedian seriously claims—that he called the "Wikipedia Potential Thesis" where Wikipedia would become so excellent that there would be no need for experts to have any role in human society to vet what is accepted as knowledge. He then proceeded to demolish his straw man, arguing that Wikipedia cannot continue to improve in excellence without according experts a special arbitratory role. In contrast, he presented his newer open encyclopedia project, Citizendium, as a model of developing a high-quality open encyclopedia by according recognized subject experts a privileged role.

Notwithstanding Sanger's criticism of disprivileging experts, the scholars who have assessed Wikipedia according to its primary claim to be an encyclopedia have generally considered it a positive development in this direction. Even then, they all considered it a radical departure from the previous norms and approaches in this literary genre.

Epistemology

Beyond its specific role as an encyclopedia, Wikipedia has also been evaluated more generally as a source of knowledge. Ironically, although many scholars have questioned Wikipedia's reliability when considering traditional criteria, almost all the studies that have scrutinized Wikipedia from an epistemological perspective have strongly validated its epistemic qualities as a valuable source of knowledge (Schiltz et al. 2007). Epistemology, the theory of knowledge, is "the branch of philosophy concerned with the nature and scope (limitations) of knowledge" (Wikipedia contributors 2012). A significant sub-stream of research 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 that has been called one of "seismic" proportions (Dede 2008).

Some scholars pointed out that Wikipedia is so different from anything before it, that it merits new approaches for its fair and meaningful assessment. Magnus (2009) argued that Wikipedia defies traditional approaches to assessing credibility, and thus its accurate and fair assessment "requires new epistemic methods and strategies" (2009, p.74). Fallis (2008) 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. Moreover, when compared to its more likely alternate sources on the Web (such as via search engine queries), Wikipedia is strikingly superior as a source of knowledge. 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. Writing from the perspective of the epistemology of testimony, Tollefsen (2009) argued that "Wikipedia involves an odd mix of individual testimony and group testimony where, at times, the group testifying is Wikipedia itself" (2009, p.22). She argued that as Wikipedia matures, people will increasingly consider its "testimony" as given in good faith, yet will continue to need to critique it, as with any other kind of testimony.

Dede (2008), Eijkman (2008), and Matychak (2008) 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" (Dede 2008). 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. Eijkman (2008, p.93) "argues that the continuing dominance and therefore likely application of conventional old paradigm foundational learning theory will work against ... the powerful affordances Web 2.0 social media provides for learning focused on social interaction and collaborative knowledge construction." He argued that educators need to broaden their perspective on the epistemological basis of student learning, in order to appreciate the need and value of their being accultured into the emerging knowledge landscape that Web 2.0 phenomena such as Wikipedia present.

In general, the research on the Wikipedia infrastructure indicates that the open source model, combined with enabling collaborative technologies like the wiki, are fundamental to enabling the survival and growth of such a large encyclopedia. However, Stettler (2008) argued that "the institutional and social functions/tasks of the so-called 'knowledge architects or designers' who are behind Wikipedia should be looked at more profoundly" to ensure that all the stakeholders, including readers and community participants, obtain value from the Wikipedia system in order to keep the enterprise viable. Rodríguez (2007) similarly argued that the Wikipedia epistemology of communally generated knowledge parallels the collaborative approach of U.S. Latino/a liberation theology; both forms of knowledge stand against the traditional foundationalist expert-oriented perspective of knowledge. He argued that these forms of truth-seeking can provide more accurate understandings of truth than the classical scientific rationalist view.

Knowledge as presented in Wikipedia is not necessarily a finished product, but can be seen as a dynamically shaped epistemology. In Santos' (2009) dissertation on the relationship between Levinas' ethics-oriented rhetoric and Web 2.0, he argued: "Wikipedia's 'end' is not necessarily as concerned with producing a finished product as its mission statement might suggest. ... [Wikipedia's rules] seek as their principal goal to support an other's response ability and to invite them to speak. Wikipedia's primary obligation is not to create objective Truth, but rather to foster, support, and maintain 'neutral' relationships." (2009, p.196) Hartelius (2010) hypothesized Wikipedia "as a model of dialogic expertise." Based on the Bakhtinian theory, she argued that Wikipedia confronts the "monologic" expertise "by facilitating an ongoing chain of interdependent and multivocal 'utterances'" (2010, p.506).

Despite these positive evaluations, Wikipedia is not universally applauded on epistemological grounds. Wray(2009) argued that Wikipedia is unreliable because there is no good epistemological reason for

readers to trust what they read—there is no incentive for contributors to be reliable because of their relative anonymity. We note, though, that this philosophical assessment was not based on any formal empirical assessment of the content of Wikipedia articles—which, of course, is the same for the more positive articles we have summarized in this topic category.

A noteworthy study is Mendoza's (2009) proposal for the creation of WikiID, a dynamic Body of Knowledge for interior design practice. While this article refers to Wikipedia mainly as a reference model for another wiki—and we generally exclude such articles in this review—hers is interesting for its highlighting of the dynamic nature of what is culturally accredited as knowledge, and her comments pointing out that savant-compiled knowledge is often even more biased based on scholarly traditions than Wikipedia, whose neutral point of view and active invitation of alternative points of view provides a mechanism for uprooting systematic bias in what is accepted as "knowledge."

In another study that considered Wikipedia as a knowledge base, Pentzold (2009) interpreted Wikipedia "as a global memory place […] where memorable elements are negotiated" (2009, p.255). He argued that Wikipedia "provides an ideal example of the discursive organization of remembrance and the different observable steps of memory work as they evolve online" (2009, p.267).

Two studies used Wikipedia as a model for reforming the academic peer review system. Fitzpatrick (2009) argued that the collaborative mode of communication can offer insights for the process to develop further into peer-to-peer review. This would allow "not just the results of our research and vetting processes, but the very processes themselves to become an open, accessible part of the published record" (2009, p.127). Black (2008) discusses how academic peer review could be reformed by emulating Wikipedia's collaborative creation model. He highlighted many shortcomings of the traditional system, 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. However, this would have serious implications for traditional academic practices such as tenure and grant awards. He proposed that this could liberate knowledge creation from the sacred cathedral of cloistered academia to a broader collaboration by experts of various capacities. However, we feel that his comments are questionable, since the original research that is subject to academic peer review is expressly forbidden on Wikipedia—this is not merely a matter of subject difference: in fact, Wikipedia's Neutral Point of View principle cannot operate with original research, which by definition cannot be verified by supporting sources because it is original. However, the MediaWiki software could certainly serve as an "original-research" peer review platform in academia.

Other articles with bearing on epistemology are discussed elsewhere in this review (Eijkman 2010; Veltman 2005; Gunnels 2007; Geiger & Ribes 2010; Garud et al. 2008; Cimini & Burr 2012).

Ethics

Some studies have investigated various ethical issues surrounding Wikipedia, including questions of identity and representation, information privacy and transparency.

Probably the most widely researched ethical issue on Wikipedia is the so-called "Essjay controversy," where a very active Wikipedia contributor and administrator misled the community by lying about his real-life credentials. Brown (2009) elaborates on the controversy particularly with respect to the concept of "ethos." He distinguished between situated ethos and invented ethos arguing that Wikipedia contributors are asked to rely on the invented ethos created by the trail of citations to other texts, not on the situated ethos based on the real life expertise. He concluded that "Wikipedia gestures toward an emerging rhetoric that offers us ways to rethink the intersections of ethos, identity, intellectual property, and textual origins" (2009, p.W255). Santana and Wood (2009) criticize Wikipedia as socially irresponsible because the anonymity of its contributors (screen names of registered users are still anonymous without further identifying information) makes Wikipedia content not so transparent as to its sources. They mainly argued from principles of ethical theory where opacity through anonymity is often

used by powerful actors (such as world leading websites) to mislead their clients. They illustrate their intrepidation with the Essjay controversy.

For O'Neil (2011), the Essjay controversy is a case where Wikipedia did not apply a critique of expertise uniformly. In contrast, he saw Wikipedia as "the most radical form of anti-credentialism." O'Neil further discusses the unique power structures in Wikipedia, which renounces traditional structures of aristocracy and entitlement divorced from actual demonstrated performance arguing for a "hacker authority" linked to the non-transferable characteristics of the individual which may have extraordinary skills or personality. He saw Wikipedia "barnstars" as a public token of appreciation from one participant to another with respect to this hacker authority. O'Neil also argued for an "online collectivist authority" with two central components: roles and rules. He believes that administrators in their roles "are determining what deserves to be included in the encyclopedia," in contrast with Monaci's (2009) view where administrators have no particular social privileges, but are merely users with increased responsibility for technical tasks. O'Neil regards Wikipedia governance as self-similar where rules are written with the same software as the articles themselves in a "writing about writing" process. In commons-based peer production, he furthermore saw a unity between consumers and producers (readers and writers in Wikipedia) and between experts and amateurs.

In his article on ethics and trust in open source communities, de Laat (2010) argued that Wikipedia did not have an a priori ethic in the beginning, but it "had to create an appropriate ethic along the way. In the interim, the assumption simply had to be that potential contributors were trustworthy; they were granted 'substantial trust'" (2010, p.327). Furthermore, he argued that Wikipedia is "likely to converge in the future towards a mid-level of discretion. In such a design the anonymous user is no longer invested with unquestioning trust." (2010, p.327)

One ethical consideration that receives surprisingly little attention is privacy in Wikipedia. In their discussion on privacy in Web 2.0 in general, Pekárek and Pötzsch (2009) noted the minimal access control in Wikipedia where third parties regardless of status level have access to user pages and may write on user pages. They argued that the issues they "found arise mainly due to collapsing contexts, i.e. users' personal data used in contexts other than the original and intended one. The finding that social software lacks fine-grained and user-determined access control options aggravates this source of privacy issues."

Literature Review

Although this is a dedicated topic in the WikiLit website, we described the literature reviews on Wikipedia in an earlier dedicated section of this article, and so we refer readers there.

#REDIRECT [[Earlier Literature Reviews of Wikipedia Research]]

Research Platform

A few studies have aimed to present Wikipedia to researchers as a platform for their scholarly research, introducing and describing its characteristics that make it most valuable and amenable to the discovery of original knowledge, focusing on its potential and implications for their respective fields of research. This category is distinct from articles that investigated Wikipedia as a Knowledge Source by Scholars and Librarians, which we describe separately.

Kane and Fichman (2009) discuss various ways in which wikis can be incorporated in the teaching and research of information systems professors. In particular, they discuss experiences and recommendations about how to go about conducting research on Wikipedia. In a specific examination of Wikipedia's value for history students and scholars, Rosenzweig (2006) proposed Wikipedia as a model for the collaborative writing of history, a normally highly individual craft. In addition to these, many other articles that studied and presented Wikipedia as a general research platform are discussed elsewhere in this review, particularly in the section on Wikipedia as a corpus (Bizer et al. 2009; Voss 2005; Suchanek et al. 2007;

Denoyer & Gallinari 2006; Schenkel et al. 2007; Suchanek et al. 2007). In addition, Ahn et al. (2005) is covered by Medelyan et al. (2009).

Wikipedia as a System

Some studies have taken a systems approach to analyzing Wikipedia, applying systems theory to consider it a composite of interacting sub-systems with properties at various changing states. Pamkowska (2008) presented Wikipedia as an autopoietic system, that is, one that self-creates, self-adjusts and self-develops; she presented it as a model for managers to administrate similarly autopoietic organizations. Müller-Birn et al. (2010) provide an approach to organizing data in online production systems, with "two practical implications: first, available data from online production systems can be obtained and evaluated more easily. Second, results are comparable because the generic vocabulary serves as a shared understanding of online production systems."

Although they take a software engineering approach on the design of systems, Garud et al.'s (2008) study on designing systems with a view to incomplete specifications has significant epistemological implications. They proposed that the design of a system has two distinct components: design of the process of creation and design of the outcome. They argued that when designing moving targets such as open source software like Linux and open content projects like Wikipedia, designers ought not insist on planning out a completed system, which would result in rapid obsolescence. They should rather aim at a moving target, a system which at any point in realization of its development is incomplete. The epistemological implication is that knowledge should not only be seen as the way to ascertain an absolute truth, but rather a process of discovery which is never completed. Wikipedia is quite amenable to this view of the nature of knowledge.

Miscellaneous Topics

Finally, a few studies that generally covered Wikipedia defy classification into any other category that we have specified in this review, mainly because they focused on topics other than Wikipedia, yet discussed Wikipedia in non-trivial ways.

Morse (2008) 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. He encouraged their use for knowledge management, though he cautioned that hierarchical organizational structures could limit their potential.

In his short essay, Lawler (2006) listed several crises and conflicts that Wikipedia has gone through. He stated that Wikipedia deals with exactly the same kind of problems that occur between humans in offline social contexts.

Purdy's (2006) dissertation on digital archives uses Wikipedia as a major case (the journal archive JSTOR and the plagiarism service Turnitin are the others) of a non-traditional archive. Wikipedia is both valuable and problematic in saving every version of each article: this provides a rich dynamic archival record, but it makes references to archival copies challenging since the current version changes constantly. He also comments unfavorably about the lack of expert vetting of article content.

Leinonen et al. (2009) explored the potential of wikis for education by examining Wikiversity, a Wikimedia project aimed for building a free learning community. This study introduced three learning metaphors: acquisition, participation, and knowledge creation. It also differentiated between Wikiversity and Wikipedia's focus, structure and policies. For instance, the Wikipedia community focuses on the content. However, Wikiversity's focus should be on its community members and their educational development. Moreover, policies such as NPOV work well for the purpose of Wikipedia but not necessarily for Wikiversity.

Content: The Content of Wikipedia

The Content category, with 91 articles (19%), includes studies related to Wikipedia content, its growth, its depth, breadth, and reliability, mainly focusing on the encyclopedia articles and the structure in which they are presented. This topic group corresponds to the Quality topic in Wikimedia-pedia. The two major issues related to Wikipedia content are its quality and its size, though some studies treated other topics related to Wikipedia content.

Quality

Quality of Wikipedia articles has been one of the main concerns of academic and user communities about Wikipedia, mainly due to the non-expert and openly participatory nature of Wikipedia development. Lewandowski and Spree (2011) found that Wikipedia results shown on search engines are quite dependent on the quality of articles; thus quality of articles has important direct consequences for its readership and existence on the web. Thus, Wikipedia has attracted significant attention from researchers to investigate this important aspect. Such studies typically select a sample of Wikipedia articles and "manually" read and judge the quality, sometimes in comparison with other encyclopedias or other resources.

We distinguish the studies we describe in this section from those on Contributor Perceptions of Credibility and Reader Perceptions of Credibility. These latter two sections describe subjective assessments of Wikipedia's trustworthiness from the perspective of contributors and readers, respectively. In contrast, in this section we describe formal studies that have used some sort of external expert source to try to evaluate the quality of Wikipedia using objective standards.

Actually, "quality" is a very complex concept, and so different studies have investigated various aspects of this idea. More precisely, we have found that studies of Wikipedia's "quality" have investigated one or more of the following aspects: Reliability or accuracy (that is, absence of factual errors); comprehensiveness or breadth of coverage of subject matter, whether within an individual article or across multiple articles; currency or up-to-dateness of the article contents; and readability and quality of writing style. In addition to these precise quality topics, some articles have studied antecedents to these various aspects of quality. A final important group of articles investigated various aspects of featured articles, those articles vetted by the Wikipedia community as being of high quality.

Antecedents of Quality

There has been great interest in understanding the factors that lead to the quality of Wikipedia articles. In this review, we discuss those antecedents particularly related to patterns of collaboration in the section on Quality Improvement Processes. Almost all antecedents to Wikipedia quality are related to other topics (especially to Quality Improvement Process); thus, many other related articles are discussed in other sections in this review (Geiger & Ribes 2010; Santana & Wood 2009; Jones 2008; Ransbotham & Kane 2011; Carillo & Okoli 2011; Klemp & Forcehimes 2010; Ehmann et al. 2008; Rahman 2006; Rahman 2007; Lih 2004; Aniket Kittur & Kraut 2008; Stvilia et al. 2008; Arazy et al. 2011; Duguid 2006; Adamic et al. 2010; Anthony et al. 2009).

Comprehensiveness

Wikipedia is purportedly aimed at incorporating all human knowledge within an encyclopedia (Wales 2004), so comprehensiveness has been always a major point of inquiry about Wikipedia and an important aspect of its quality to see how much of the knowledge from different fields of human knowledge is represented in Wikipedia. Researchers have investigated comprehensiveness of Wikipedia in variety of fields from art, philosophy, and science to medicine, history, and psychology. This stream of research captures the fields that are underrepresented (or overrepresented) in Wikipedia, and sometimes come up with complementary or conflicting results. Of course, Wikipedia coverage of topics has grown over time. Thus, we arrange the articles in this section mainly chronologically, since the results of earlier studies might very well not be representative of Wikipedia's current condition.

Altmann (2005, p.755) found that "Medical Informatics is not represented sufficiently since a number of important topics is missing."

Looking at outbound scientific citations in 2007, Nielsen (2007) found that astronomy and astrophysics articles in the English Wikipedia were significantly more frequently cited compared to Journal Citation Reports. The Journal of Biological Chemistry was undercited, but that changed after automated mass-insertion of genetic information (Nielsen 2008a). One peculiarity with the sample occurred with Australia botany journals. A Wikipedia project had produced a number of well-sourced articles on Banksia, some attaining featured article status.

On a cross-section of 446 articles randomly picked from *Encyclopædia Britannica*, Wikipedia articles lacked entries for 15, e.g., "Bushman's carnival," "Samarkand rug" and "Catherine East" (Wedemeyer et al. 2008). All 192 random geographical articles picked from *Britannica* had corresponding articles in Wikipedia. Of 800 core scientific topics selected from biochemistry and cell biology text books, 799 could be found in Wikipedia. Wedemeyer et al. concluded that science is better covered than general topics and that Wikipedia covers nearly all encyclopedic topics.

By sampling 3,000 articles from the 2006 English Wikipedia and categorizing them against the Library of Congress categories, Halavais and Lackaff (2008) found categories such as social sciences, philosophy, medicine and law underrepresented in Wikipedia compared to statistics from Books in Print. The two latter categories, however, had on average comparably large article sizes. They identified science, music, naval studies and geography as overrepresented, with music probably benefiting from fan contributions and other categories from the mass-insertion of material from public data sources such as the United States Census. When compared to three specialized encyclopedias in linguistics, poetry and physics, they found many expected articles to be missing. Halavais and Lackaff also noted some peculiarities in Wikipedia, such as extensive list of arms in the military category, comics fans to some extent driving the creation of articles in the fine art category, and voluminous commentary on the Harry Potter series in the literature category.

Schweitzer (2008) examined the coverage of psychology-related topics on Wikipedia. These were not only well covered, but the articles also displayed on top of the major search engines. Students were found to use Wikipedia for personal and school-related activities, but generally not as academic references.

For twentieth century philosophers, Elvebakk (2008) compared Wikipedia against two online peer-review resources, The Stanford Encyclopaedia of Philosophy and the Internet Encyclopedia of Philosophy, with respect to coverage of gender, nationality and discipline. She concluded that Wikipedia in 2008 represented philosophy topics essentially the same way as more traditional resources. Wikipedia had far more articles about the philosophers than the two other resources and only some minor differences in fractions, such as a smaller fraction of German and French philosophers. Similarly, Bragues (2009) tested "the quality of Wikipedia, [by] sampling ... articles relating to seven top Western philosophers" (2009, p.117). However, he found out that on "average, the online encyclopedia captured 51% of the expert consensus surrounding the seven philosophers examined" (2009, p.151). All of the analyzed philosophers' pages had a strong biography section, "arguably too strong" (2009, p.152). "This could reflect the fact that contributors to Wikipedia's philosophy pages have less experience and confidence grappling with philosophical analysis. It may be that, compared to academic philosophers, Wikipedians on average find it less pleasurable to engage philosophic arguments [*sic*] and prefer to focus on the characters and histories of famous personages." (2009, p.152) Bragues concluded that he "was unable to uncover any outright errors," and that the "sins of Wikipedia are more of omission than commission" (2009, p.152).

Kittur et al. (2009) developed an algorithm that would assign a topic distribution over the top-level categories to each Wikipedia article. After evaluating the algorithm on a human labeled dataset, they examined the English Wikipedia and found that "Culture and the arts" and "People and self" to be the most represented categories. Between the 2006 and 2008, they found that "Natural and physical sciences"

and "Culture and the arts" categories grew the most. By combining the algorithm with a method for determining degree of conflict of each article (Aniket Kittur, Suh, et al. 2007), they could determine that "Religion" and "Philosophy" stood out as the most contentious topics.

Royal and Kapila (2009) compared the number of words in sets of Wikipedia articles with the year associated with the articles. They found that articles associated with recent years tended to be longer, that is, recency could somewhat predict coverage. The length of year articles between 1900 and 2008 and the year as a predictor variable had a Spearman correlation coefficient on 0.79. The results were not homogeneous, as the length associated with articles for Time's person of the year had a correlation of zero with the year. Academy award winning films and "artist with #1 song" had correlations on 0.47 and 0.30, respectively. They also examined other sets of articles in Wikipedia and the correlation with column inches in Micropædia of the *Encyclopædia Britannica*, country population and company revenue. The correlations were 0.26, 0.55 and 0.49, respectively. In their comparison with 100 articles from Micropædia, they found that 14 of them had no Wikipedia entry, e.g., "Russian Association of Proletariat," "League for the Independence of Vietnam" and "urethane."

Rush and Tracy (2010) argued for measuring Wikipedia presence of an academic field as a proxy for the public impact of the field, as presence and accessibility is the necessary condition for having impact. They thus concluded that communication research does not have the impact it is supposed to have and offered suggestions to improve this situation.

Kim et al. (2010) examined the usefulness of Wikipedia content in covering the Pathology Informatics educational curriculum, and found that it covers 90% of the curriculum with high-quality, comprehensive and current articles beneficial for both beginning and advanced learners.

Atanassova (2011) looked into how bioengineering topics are covered in Wikipedia. The study identified many article categories covering the field topics, as well as variety of Wikipedia projects and portals related to bioengineering topics.

In addition to these, other articles that studied Wikipedia's comprehensiveness are discussed elsewhere in this review (Radtke & Munsell 2010; Rosenzweig 2006; Michael R Laurent & Vickers 2009; Korosec et al. 2010; Jancarik & Jancarikova 2010; Rector 2008; Clauson et al. 2008; A. Leithner et al. 2010; P. T. Johnson et al. 2008; Stvilia et al. 2007). In addition, Milne et al. (2006) is covered by Medelyan et al. (2009).

Currency

Currency refers to the degree to which Wikipedia articles reflect up-to-date information about their topics. Its live, continuous online publishing model has generally proven a major strength in comparison to other encyclopedias, both online and offline.

In a comparison between Wikipedia and Medscape, Clauson et al. (2008) found four factual errors in Medscape among 80 articles examined. Two of these occurred due to lack of timely updates. They found no factual errors in Wikipedia. In a study on twentieth century philosophers, Wikipedia had far more articles on philosophers born after the Second World War than two other online encyclopedias, The Stanford Encyclopedia of Philosophy and The Internet Encyclopedia of Philosophy (Elvebakk 2008).

Although Wikipedia is often current, its propensity to import large bodies of work in the public domain (which is often many decades old) sometimes compromises its currency. The Danish Wikipedia has a large number of bibliographies copied more or less unedited from two old reference works with expired copyrights: *Dansk biografisk Leksikon* and *Salmonsens Konversationsleksikon*. The age of the works affects the language and viewpoint of the Wikipedia articles (Bekker-Nielsen 2011). Such risks might also occur in the English Wikipedia, where many articles feature imports from the 1911 edition of *Encyclopædia Britannica*. However, Wedemeyer et al. (2008) found that Wikipedia was much more up-to-date than the present-day *Britannica*.

In addition to these, other studies related to Wikipedia's currency are described elsewhere in this review (Michael R Laurent & Vickers 2009; Stvilia et al. 2007).

Featured Articles

Featured articles are those which have been carefully examined and approved by the community as being of all-round high quality. In the English Wikipedia, such articles are honored by being "featured" as the article of the day on the English Wikipedia main page. Most other language Wikipedias have a parallel concept of their highest quality class of articles, though they bear different names.

Some studies have accepted the community's evaluation of these articles as high quality, and then used them to validate their proposed quality assessment method for Wikipedia articles. Dondio and Barrett (2007) developed a method for computationally differentiating featured articles from others by predicting trustworthiness of articles. Blumenstock (2008) offered word count as a predictor of article quality. He argued that this simple metric is "considerably more accurate than the complex methods proposed in related work, and performs well independent of classification algorithm and parameters." (2008, p.1096).

Stvilia et al. (2007) presented a context-independent framework for information quality (IQ) assessment. They tested their framework with two large datasets, of which the other was English Wikipedia. In the Wikipedia case, they used featured articles as benchmarks for article information quality. Featured articles "were valuable resources for designing and validating individual IQ metrics as well as for testing the entire IQ measurement model" (2007, p.1732). They evaluated their model using 236 featured articles and 828 random articles. They conclude that the "IQ measurement model was shown to be successful in discriminating high-quality articles" (2007, p.1732).

However, the Wikipedia community's vetting process cannot be considered a sufficient assurance of quality. Lindsey (2010) examined the quality control procedures for Wikipedia featured articles, and found that 12 out of 22 investigated English Wikipedia featured articles actually do not follow Wikipedia's own criteria; thus, he concluded that the quality control procedures are ineffective.

Other studies have examined factors that led to such articles to be considered as high quality articles. Poderi (2009) investigated the correlation between revision patterns and quality of featured articles of Wikipedia, and found that such correlation is meaningful and having a main author would increase the consistency and quality of Wikipedia articles. Moreover, Jones (2008) investigated the relationship between patterns of editing and article quality on Wikipedia by comparing featured and non-featured articles. He found that revision history of quality articles are populated equally by both content and surface related edits; however, the lower quality articles were populated by more content edits and fewer surface edits. In another similar study, Stvilia and Gasser (2008) employed activity theory to analyze the pattern of change in information quality of featured articles. They found out that the article structure and article revision patterns are influential on article quality, and that the presence of principal author enhanced the quality of an article.

In addition to these, other articles that studied featured articles are discussed elsewhere in this review (Viégas, Wattenberg & McKeon 2007; Stvilia et al. 2009; Goldspink 2009; Goldspink 2010; Sara Monaci 2009; Ransbotham & Kane 2011; Carillo & Okoli 2011; H. Zeng et al. 2006; Wilkinson & Huberman 2007a; Wilkinson & Huberman 2007b).

Readability and Style

A number of studies have examined the writing style and readability of Wikipedia articles. The findings generally show that Wikipedia articles are at least as easy to read as, if not more so, than their online and offline counterparts; however, the writing style is found to be less consistent, especially concerning the international topics. Medelyan et al. (2009) discussed Emigh and Herring (2005); we discuss other articles here.

Some studies examined Wikipedia's readability in its own right, without making external comparisons. Positively, these studies found many well-written articles. Negatively, the same studies found many poorly-written articles, and found that the overall quality is rather inconsistent. Dalby (2007) commented generally on the language versions of Wikipedia, focusing mainly on the English Wikipedia. He noted that the quality of English is very inconsistent, and that many non-native English speakers contribute, leading to poor quality writing in some articles, especially those on international topics. West and Williamson (2009) inquired about the credibility and quality of Wikipedia articles, and found that Wikipedia articles are objective, clearly presented, reasonably accurate, and complete. However there are some poorly written articles containing unsubstantiated information and providing shallow coverage of the topic. Clark et al. (2009) found that not only topics can distinguish Wikipedia articles, but also their structural form, i.e., genre. The genre may evolve as editors extends and change the articles.

Some studies compared Wikipedia's readability with that of other comparable online resources. These studies varied in their results: some found Wikipedia articles generally equally readable, some less so, and others found Wikipedia generally more readable. Elia (2009) compared the readability and maturity of Wikipedia articles with those of the Britannica Online encyclopedia and found no significant difference, at least from a quantitative point of view. Korosec et al. (2010) compared student use of the German Wikipedia and the chemistry encyclopedia Rompp Online in the area of chemical thermodynamics. They found that while students use both, Rompp Online is victim of its exactness and academic writing style. Wikipedia is more comprehensive and easily readable, which are more important to students. They concluded that while both resources are good for initiating research, students should learn how to use both peer-reviewed and non-peer-reviewed material in their learning. Comparing Wikipedia's cancer information from August 2009 with the US National Cancer Institute's Physician Data Query (PDQ), Rajagopalan et al. (Rajagopalan et al. 2010; Rajagopalan et al. 2011) found that Wikipedia had lower readability as measured by the Flesch–Kincaid readability test.

Den Besten and Dalle (2008) studied the Simple English Wikipedia, a distinct Wikipedia (that is, with completely separate articles from the regular English Wikipedia) that limits its vocabulary sense and grammatical structure to facilitate reading by children and by learners of English. They investigated the editorial process that implemented the rules instated to keep articles "simple."

In addition to these, other articles that studied Wikipedia's readability and style are discussed elsewhere in this review (Purdy 2009; Stvilia et al. 2007).

Reliability

Reliability of Wikipedia papers has been always one of the main concerns of Wikipedia users; thus this is one of the most widely investigated aspects of Wikipedia research. This is variously called reliability, accuracy, and freedom from errors. We distinguish this from measures of trustworthiness, which we generally call "credibility"; such topics are treated in other sections of this review (Contributor Perceptions of Credibility and Reader Perceptions of Credibility). We also distinguish this section from Computational Estimates of Reliability, which uses computational methods to estimate reliability. Here we treat studies where subject experts empirically evaluated the reliability of Wikipedia articles.

We group this body of work into three subsections. Many studies examined the reliability of articles either on their own, or in comparison with other reference sources. A second group of studies examined the quality of citations from Wikipedia articles to external sources. A third group examined trends in reliability of articles over time.

With continuous revision, the reliability of Wikipedia articles has generally improved over time (at least, for existing articles; new articles start from ground zero in terms of quality). Thus, we arrange the articles in this section mainly chronologically, since the results of earlier studies might not accurately represent Wikipedia's current condition.

Reliability Assessment of Wikipedia

Some of the most popular Wikipedia studies—that is, those that have received the most press attention—are those that face of Wikipedia with authoritative sources of information to compare their respective reliabilities. Although very many such comparisons have been conducted, here we describe only the scholarly ones. Results have been mixed, with some studies evaluating Wikipedia quite favorably, and others not as much. We group these studies accordingly.

Positive or Equivalent Evaluations: Some empirical studies have found Wikipedia at least equal in reliability to well-established reputable sources. Some have found Wikipedia even superior.

The most famous scholarly assessment of Wikipedia is a comparison of selected science articles in Wikipedia and *Encyclopædia Britannica* conducted by *Nature*, the leading science journal (Giles 2005). Giles found Wikipedia's accuracy comparable to those of *Britannica*. The articles were masked as to their provenance and evaluated by scientists who had published in *Nature*. The scientists found that among 42 articles, Wikipedia contained more factual errors, omissions and misleading statements (162, with an average of 4 per article). However, *Britannica* was not far behind (123 errors, average 3 per article). Both encyclopedias contained "serious errors, such as misinterpretations of important concepts" (2005, p.900). Although *Britannica* fared better on this examination, its finding was shocking and was widely considered a major blow against *Britannica* and a boon to Wikipedia, considering that Wikipedia was only four year old at the time, compared to *Britannica* at 232 years old in 2005. The study was not itself peer-reviewed, and was vociferously contested by *Encyclopædia Britannica*; however, *Nature* defended its analysis.

Chesney (2006) identified Wikipedia articles as highly credible, though experts perceive the credibility of Wikipedia articles different from what non-experts perceive. However, this study was not a comparison with other sources, and was not blinded.

Compared to some other resources, Rosenzweig (2006) found Wikipedia to be accurate in reporting names, dates, and events in U.S. history; in 25 biographies only four clear-cut factual errors, mostly small and inconsequential, were found: "Wikipedia, then, beats Encarta but not American National Biography Online in coverage and roughly matches Encarta in accuracy" (2006, p.129).

Devgan et al. (2007) surveyed medical doctors concerning 39 common surgical procedures. They could find 35 corresponding Wikipedia articles, with all of them judged to be without overt errors. The researchers could recommend 30 of the articles for patients (22 without reservation), but also found that 13 articles omitted risks associated with the surgical procedure (2007).

In a small comparison study on medical information with just three topics, blinded experts found some factual errors in Wikipedia, at around the same level as medical online resources UpToDate and eMedicine (2008). AccessMedicine was found to have no factual errors in the three articles examined.

Shachaf (2009) examined the quality of answers on Wikipedia Reference Desk in comparison with those on library reference services, and found that they are actually quite comparable.

Rajagopalan et al. (Rajagopalan et al. 2010; Rajagopalan et al. 2011) examined Wikipedia August 2009 cancer information and US National Cancer Institute's Physician Data Query (PDQ). They found that Wikipedia had similar accuracy and depth compared against the professionally-edited resource.

Negative or Inferior Evaluations: Some empirical studies have found Wikipedia of inferior quality to well-established reputable sources. Some have concluded that Wikipedia is of such poor quality that it is inadvisable to use it; these more strongly negative recommendations tend to accompany unfavorable evaluations in healthcare topics.

Mercer (2007) reviewed some key mental health topics in Wikipedia and found them generally lacking in quality, mainly because of what he perceived to be the influence of contributors lacking genuine professional expertise on the subjects. However, he recognized Wikipedia's importance and potential and

recommended a number of measures that could hopefully improve the quality of articles. Unfortunately, most of these involved contributors revealing their real-world identities, which conflicts with Wikipedia's strong policy of permitting anonymous participation.

Rector (2008) examined comprehensiveness and accuracy of Wikipedia articles in comparison to three other reference resources. She found Wikipedia a wealth of information and a proper model for peer-production of reference material. However, it does not fare as favorably as do others in terms of accuracy and comprehensiveness of articles.

Cautionary notes have been made for the open-wiki model in cases where potentially hazardous procedures are described (2006). In particular, medical procedures and pharmaceutical compounds may call for complete and accurate description. Clauson et al. (2008) compared Wikipedia and Medscape Drug Reference (MDR), a free online "traditionally edited" database, for medical drug information. They found that Wikipedia could answer fewer drug information questions, e.g., about dosage, contraindications and administration. In the evaluated sample, Wikipedia had no factual errors but had a higher rate of omissions compared to MDR. However, they found a marked improvement in the entries of Wikipedia over a just 90 days period. The study went on to mainstream media with headlines such as "Wikipedia often omits important drug information" and even "Why Wikipedia Is Wrong When It Comes To Prescription Medicine." However, as noted by some Wikipedians[15], the study neglected the fact that one of the Wikipedia manuals of style explicitly requests: "Do not include dose and titration information except when they are notable or necessary for the discussion in the article." Thus in one of the eight examined question categories in Clauson et al.'s study, the Wikipedia omissions were quite possibly intentional.

Luyt and Tan (2010) investigated the credibility of Wikipedia articles about America's history and found that most of the article contents were either not verifiable, or the resources were not valid ones like academic publications. They concluded that Wikipedia is not appropriate for serious reference work, and that readers should be taught to evaluate its content.

Leithner et al. (2010) investigated the scope, completeness, and accuracy of information for osteosarcoma on English Wikipedia in April 2009, compared with patient and professional sites of the US National Cancer Institute (NCI). Three independent observers scored the answers to twenty questions. Although they judged Wikipedia's information as generally good, it scored lower compared to the two NCI versions (though this was statistically significant only for the professional version). Thus, they suggested adding external links to these websites on Wikipedia articles.

Lavsa et al. (2011) compared the drug information for twenty of the most frequently prescribed drugs in the United States with the drug package information and certain authoritative databases. They found that the Wikipedia articles were all incomplete in providing full drug information, often missed important details, and were often inaccurate. They recommended against its use by pharmacology students for drug information.

Citing Other Sources

Some studies examined an important proxy measurement of an encyclopedia's reliability: the quality of supporting citations to other sources. Wedemeyer et al. (2008) found that most well-developed articles had sufficient references comparable to a scientific review article, but some articles, even two featured ones, had insufficient referencing. Haigh (2010) examined the health related Wikipedia articles to evaluate the quality of their source and supporting information. She found that Wikipedia health resources are clearly identifiable reputable ones that make it an appropriate resource for use of nursing students. Stankus and Spiegel (2010) compared Wikipedia with a peer-reviewed online encyclopedia,

[15] http://en.wikipedia.org/wiki/Wikipedia_talk:WikiProject_Medicine#Drug_Information_in_Wikipedia

Scholarpedia, to see how they reference books. They reported that although Wikipedia references books 40% less frequently, the books and authors referenced are as legitimate as those of Scholarpedia. However, Wikipedia more frequently references more publicly accessible material and undergraduate books.

Quality-Related Trends

Some studies observed the evolution of Wikipedia reliability measures over time. Luyt et al. (2008) investigated how errors are spread out through the life of Wikipedia articles. They found that a significant number of erroneous edits occur in the earlier article edits, with 20% on the first day. Nielsen (2008a) studied scientific citations in Wikipedia through time. He found an increasing use of structured citation markup, especially after mass insertion of gene and protein information and citations by a bot.

In addition to the articles described in this section, other articles that studied Wikipedia's reliability are discussed elsewhere in this review (Page 2010; K. West & J. Williamson 2009; Willinsky 2008; Korosec et al. 2010; Ehmann et al. 2008; Elvebakk 2008; J. Y. Kim et al. 2010; Elia 2009; Fallis 2008; Lewandowski & Spree 2011; Magnus 2008; Bragues 2009; Lindsey 2010; Radtke & Munsell 2010; Stvilia et al. 2007).

Size of Wikipedia

Wikipedia is one of the largest encyclopedias in the world, with articles on almost any topic. Many studies have measured its size and growth trends, with some attempting to explain these trends and investigating the factors enhancing or hindering Wikipedia's growth.

Voss (2005) presented measures of various Wikipedia elements such as articles, authors, edits and links, providing a quantitative analysis of Wikipedia's growth and complexity. He showed that Wikipedia grows exponentially with varying growth rates among different languages. He also highlighted using Wikipedia as a thesaurus since each article covers a single concept with links to related articles.

Spinellis and Louridas (2008) concluded that Wikipedia's rapid growth is self-sustaining as there is a constant ratio of complete and incomplete articles. Moreover, most new articles are spurred by the creation of links to non-existing articles added by other users. Since their study involved the development of articles over time, they needed to download the database, in order to longitudinally observe changes.

Greenstein (2007) commented on Wikipedia's "long tail": because there are no space restrictions in this encyclopedia, many articles treated relatively obscure subjects, sometimes at great length.

Rahman (2006; 2007) examined the reasons behind Wikipedia success in terms of size and quality. He explained Wikipedia's size based on "equilibrium contributions depending on the differences in types" (2006, pp.91–92). Moreover, Wikipedia's reliability is maintained by its definition as public good, besides free-riding and free-editing. He argued that Wikipedia's uniqueness "as a public good, combined with free-riding and free-editing help to maintain the [large size and] reliability of Wikipedia" (Rahman 2007, p.96) relative to other open source systems.

Zlatic et al. (2006) analyzed several language versions of Wikipedia as complex networks. They concluded that "it is very likely that the growth process of Wikipedias is universal" (2006, p.9), based on results from eight characteristics of article networks. Discrepancies were greatest in Polish Wikipedia, originating from the use of calendar pages, a characteristic uncommon in other language versions.

Other articles that also discussed Wikipedia's size are discussed in other sections in this review (Lih 2004; Niederer & Dijck 2010; Lam & Riedl 2009; Rask 2008; van Dijk 2009; Shim & J. Yang 2009).

Other Content Topics

There are many other studies that covered topics related to Wikipedia content, but that do not fit well in our other categories.

Some studies have suggested or developed robots to automatically extend and improve the Wikipedia content. Adar et al. (2009) described Ziggurat, a system for detecting inefficiencies in Wikipedia infoboxes in different language versions of an article. Articles from different languages are aligned using cross-language links. Then, the infoboxes are aligned using a supervised learning classifier, logistic regression. Matches are detected and infoboxes gaps are filled using its match from another language version.

Researchers may be quite attentive to what is written on Wikipedia about themselves: In a 2011 Nature poll that recruited 840 researchers through email and social web sites, 19% responded that they check Wikipedia once or more per week for citations to themselves or their work (Reich 2011; Nature 2011). In comparison, 30% checked citation-counting sites (such as ISI Web of Science) and 38% checked search engines with the same frequency. The poll reported that 9% of the researchers had edited or inserted a reference to their work on Wikipedia within the last 12 months, and around 3% has edited their own Wikipedia biography. In her analysis of online philosophy resources, Elvebakk (2008) speculated that philosophers have added Wikipedia entries for themselves as a form of self-promotion.

Luyt (2011) examined Wikipedia's articles on the histories of Singapore and the Philippines to compare their reports with the dominant historiographic records of these two nations. He found that the record of Singapore, like the dominant historiography, is fairly straightforward in reflecting the well-accepted history. In contrast, as the history of the Philippines has multiple alternative scholarly interpretations, its Wikipedia record reflects conflicting perspectives and resorts extensively to quoting historiographical sources to record different views.

Kimmerle et al. (2010) visualized the evolution of Wikipedia articles and that of the authors contributing to them. They observed that the pattern of evolution of both articles and authors are the same, affirming co-evolution of the Wikipedia social system and the individuals' cognitive systems.

In addition to these, other articles that studied other aspects of Wikipedia's content are discussed elsewhere in this review (Wagner 2005; Huvila 2010; Krötzsch et al. 2007).

Corpus: Use of Wikipedia as a Textual Corpus

The Corpus category, with 131 articles (27%), discusses researchers using Wikipedia as a textual corpus for various text analysis studies. There is no corresponding topic group in Wikimedia-pedia. What is distinctive about this major category is that the goals or outcomes of these studies are usually not focused on Wikipedia itself; they usually use Wikipedia content (both direct article text and metadata) as a textual data source for some other scientific analysis. We divide this topic category into four subcategories, three of which are obtained from a literature review which "focuses on research that extracts and makes use of [the information] found in Wikipedia" (Medelyan et al. 2009); we discuss their review in the section on literature reviews of Wikipedia research. In these three sub-categories, we identified research that used Wikipedia's content, including articles, hyperlinks, and statistical data for developing new methods, frameworks, techniques and systems, within three major areas: information retrieval (IR), natural language processing (NLP), and ontology building (OB). Our fourth subcategory comprises corpus topics that do not fit neatly into the other three.

Information Retrieval

The enormous collection of articles available on Wikipedia has encouraged many IR researchers to use corpora (the plural of corpus) extracted from Wikipedia. IR is a broad area of study that aims to build systematic approaches to solve various challenges related to providing information search and access.

Among the large topics under the IR umbrella are textual or multimedia retrieval, information extraction, text classification, query processing, and data mining, as well as others. The majority of the articles we found developed new methods or algorithms to enhance the performance of IR systems in terms of the relevance of the information retrieved and the query execution time.

Cross-language Information Retrieval

Studies on cross-language IR used Wikipedia to improve the task of retrieving information in a language different from that of the user query. A descriptive outcome of studies in this category is WikiWord, a system that extracts lexical and semantic information from Wikipedia to build a multilingual thesaurus (Kinzler 2008; Kinzler 2009). Wikipedia inter-language links provided a rich tool to improve cross-language IR. Erdmann et al. (2009) used these links to extract bilingual terminology. They implemented a Support Vector Machine (SVM) classifier to train a manually labelled data of term pairs and to test the performance of other extracted terms. Lin et al. (2009) and Lin et al. (2010) described a Japanese-Chinese cross language IR system which is composed of four components; segmentation, translation, disambiguation, and retrieval and re-ranking. The translation component consists of a Japanese-Chinese bilingual dictionary and Wikipedia inter-language links to translate query terms. To enhance Korean-Chinese IR (KCIR) tasks, Wang et al. (2008) suggested a hybrid named entities translation from Korean to Chinese. The proposed system uses Wikipedia inter-language links as a translation tool to expand the bilingual dictionary and learns translation patterns directly from Google search results. The results of the experiments showed an improved KCIR performance in comparison with another method that only uses an offline dictionary. Potthast et al. (2010) surveyed and compared the models for cross-language plagiarism detection dealing with analysis of similarities between texts from different languages. They found and reported the better performing algorithms.

Another study that examined cross-language question answering using Wikipedia (S. Ferrandez et al. 2009) is covered by Medelyan et al. (2009).

Data mining

Data mining, also referred to as data or knowledge discovery, is basically the process of extracting patterns from a large dataset. This is also another main task in IR that motivated many researchers from the IR and machine learning communities to use Wikipedia as a data source to develop new mining systems. Different approaches were proposed in the articles of this section to mining information from large knowledge sources including Wikipedia. Extracted information provides additional knowledge by discovering new patterns from the available data.

In his dissertation, Zhang (2009) proposed a new graph-based text mining system. A collection of text is first represented using a graph. This representation makes use of an ontology map or Wikipedia categories. Then, the structure of the graph with its nodes and edges was analyzed to uncover patterns to be used to enhance text clustering. For instance, Zhang studied the effect of different types of linkage on text clustering. Afterwards, the use of Wikipedia ontology was analyzed and compared to other methods when applied in text clustering systems. The graph-based methods presented herein were tested with two applications in the biomedical literature context: text clustering and summarization.

Denoyer and Gallinari (2009) used a corpus of 100,000 Wikipedia XML documents along with their internal structure and the link information between documents in the XML Mining Track at INEX 2008. The focus of this track was on two tasks applied to IR: classification and clustering of XML documents. Their article reported the results of experiments done by different participants. Pöllä and Honkela (2010) applied English Wikipedia in their study on “the combination of symbol frequency analysis and negative selection algorithm for anomaly detection of discrete sequences” (2010, p.1256), concluding that “the baseline result of the Wikipedia edit detection experiment is promising” (2010, p.1265).

Medelyan et al. (2009) also described other data mining studies using Wikipedia (Bhole et al. 2007; Milne et al. 2006).

Geographic Information Retrieval

Geographic Information Retrieval (GIR) extends the IR task by associating a geographic location feature to the treated documents. Some Wikipedia articles have markup with geographical coordinates that can be extracted and used with rendered maps such as in Google Earth and Danish Findvej.dk. The studies included in this section proposed different methods to solve the GIR task using the geospatial information available from Wikipedia.

Quack et al. (2008) described an approach for mining images available on the web using unsupervised learning. The proposed system starts with a pool of geo-tagged images from Flickr and a grid of geospatial tiles to build "a database of mined objects and events, many of them labeled with an automatically created and verified link to Wikipedia." (2008, p.55). Overell and Ruger (2008) used Wikipedia corpus to generate co-occurrence models for place names disambiguation. These models proved to enhance the performance of GIR systems in terms of their mean average precision. Using Wikipedia corpus, Stokes et al. (2008) investigated the success of NLP approaches to GIR tasks. They found that a careful choice of weighting schemes in the IR engine can minimize the negative impact of severe errors like toponym detection errors, toponym resolution errors, and query overloading.

Information Extraction

In this sub-category of IR, studies used Wikipedia to extract structured information. Documents used for information extraction include text, HTML and XML pages.

Named Entity Recognition

One of the basic tasks of information extraction (IE), named entity recognition (NER), deals with identifying named entities such as personal names, names of organizations or genes from freeform text. NER often relies on a machine learning algorithm and an annotated dictionary (gazetteer). Several researchers have used Wikipedia for NER (Kazama & Torisawa 2007; Balasuriya et al. 2009). Bunescu (2007) also aimed to derive new IE techniques with higher performance than existing ones using NER, named entity disambiguation and relation extraction. For NER, he considered the correlations between candidates named entities. These correlations were captured using Relational Markov Networks. Named entity disambiguation was achieved by detecting matches between proper names and named entities compiled from Wikipedia. A ranking function was used to compute the similarity value between proper names and named entities. Extracting relations between pairs of entities was solved using two types of supervised learning; single and multiple instance learning. Mika et al. (2008) used a NER tool to semantically annotate Wikipedia corpus linking articles texts to its infoboxes. The resulting annotations were then linked to DBpedia to enrich its metadata. This mapping between the semantic annotations and DBpedia was employed to generate additional sentences used to improve the initial annotation task.

Keyword Extraction

Wikipedia was used by Grineva et al. (2009) as a knowledge base to derive semantic information for a new competitive key terms extraction method. A document is first represented by a graph of semantic relationships among its terms. The dense part of the graph depicts the document's main topics while sparse part represent the less important terms. Afterwards, the graph is partitioned using graph community detection techniques. A criterion function is then used to select groups with important terms. Wikipedia is utilized to extract information necessary to compute the terms weights and their semantic relatedness. The main advantages of this approach include the elimination of a training phase and the effectiveness with noisy and multi-theme documents. Mihalcea and Csomai (2007) and Csomai and Mihalcea (2008)

proposed Wikify, a system for keyword extraction and word-sense disambiguation using Wikipedia. "Specifically, given an input document, the Wikify system identifies the important concepts in the text and then links them to the corresponding Wikipedia pages" (2008, p.34). The tests employed demonstrated Wikify's improvement in the time taken to answer questions.

Devereux et al. (2009) showed "the usefulness of three types of knowledge in guiding the [feature] extraction process: encyclopedic, syntactic and semantic" (2009, p.137). They also proposed a new feature extraction method using class-based information. The Wikipedia corpus was used to evaluate the accuracy of the proposed method.

Information Extraction for Query Systems

Relevancy and execution time are two main quality attributes of query retrieval systems. Researchers have long tried to optimize such systems to improve their quality, especially when dealing with large amount of data. Hence, the large number of Wikipedia articles presents a challenging dataset for such tasks. For instance, WIQA 2006 is a task which can be described as the process of accessing Wikipedia content to answer specific queries. The choice of Wikipedia documents for this task was justified by Wikipedia being "one of the largest reference works ever, making it a natural target for question answering systems" (Jijkoun & Rijke 2007).

Chu (2008) proposed new approaches for handling sparse relational datasets, specifically data extracted from unstructured documents. Chu addressed the RDBMS issues in handling sparse data, beginning by the construction of a workbench to extract and query structured from unstructured data. Various tools were then provided to query and process data. The new way of processing data stems from the "pay as you go" concept which helps processing the data incrementally. Experiments to examine structured queries over Wikipedia were designed to test the performance of the workbench. Results showed that users were able to establish sophisticated queries. Chu also argued that his approach significantly eased the transition from extracting attributed from documents to querying these attributes.

Kasneci et al. (2008) proposed the NAGA query engine for the YAGO ontology described in Suchanek et al. (2008). YAGO facts are based on Wikipedia infoboxes and category names. YAGO-NAGA is proposed to extract information for building large scale knowledge bases. This is an ongoing work of maintaining and extending YAGO and providing a toolkit to extract information from it. The usefulness of YAGO has been demonstrated by its usage by various knowledge management projects such as DBpedia, SUMO, and UMBEL.

Open Information Extraction

Weld et al. (2008) explored the challenges and benefits of open IE in the context of Kylin. Kylin is an IE system that "uses self-supervised learning to train relationally-targeted extractors from Wikipedia infoboxes" (2008, p.67). Kylin's goal is to help scaling to the Web the task of converting unstructured text to relational form. The study highlighted the importance of combining the relation approach used in Kylin with the structural approach to potentially improve the precision and recall of open IE systems.

In addition to these articles, Medelyan et al. (2009) discuss other studies on information retrieval in detail (Milne & Witten 2008; Wu & Weld 2007; Wu et al. 2008; Auer et al. 2007; Cucerzan 2007).

Multimedia Information Retrieval

Multimedia databases, including images and videos available online, have exponentially increased, raising the need for new techniques to search these large collections. Wikipedia with both its textual and multimedia contents was a target application for the studies included in this section.

Ah-Pine et al. (2009) investigated multimedia information access. They proposed two novel approaches for hybrid text-image information processing that can be readily applied to the more general multimodal

scenarios. They extended the principle of trans-media feedback into a metric view. The new similarity measures of cross-content enables us to find expressive images for a text, to annotate an image, cluster or retrieve multi-modal objects. Rahurkar et al. (2010) proposed a two-component application for image interpretation. The first component is responsible of keyword disambiguation using the titles of Wikipedia articles. The second one consists of an "image-to-semantic-concept" mapping which is achieved by extracting semantic knowledge from Wikipedia. An image sorting system was developed based on the previous approach and on an image sorting algorithm.

A large collection of videos is publicly available on the web. Classifying these videos assists the video search and retrieval tasks. Therefore, Perea-Ortega et al. (2010) used the articles and Google searches to add more informational sources to assist the classification task of video data. VideoCLEF 2008 dataset and several supervised machine learning algorithms were used in various experiments to prove the enhancement of the video classification results using the web content.

Another useful tag for images is the locations in which they were taken. Kalantidis et al. (2010) proposed a new application, VIRal, for finding the location where a photo is taken using its visual content and Wikipedia geo-referenced articles. Using VIRal geo-tagged images are first clustered into groups containing the same scene but from different views. Then, a two dimensional scene map is constructed on which an indexing algorithm is directly applied. The underlying clustering and mining solutions were challenged by a one million urban image dataset and proved efficient.

Query Processing

The performance of IR systems is heavily affected by user queries. In this section, we summarize studies that aimed to expand queries dynamically by mining Wikipedia.

Milne et al. (2007) presented and discussed a new search interface called Koru. To understand the subject of both queries and documents, Koru derives a thesaurus for each document collection from Wikipedia. Wikipedia's articles are then used to model the building blocks of the thesaurus. The wiki and its hyperlinks were used to determine the connections between the thesaurus blocks.

Elsas et al. (2008) explored the blog feed retrieval task from two viewpoints; retrieval models and query expansion algorithms. The models developed in this study emphasized the importance of modeling the topical relationship between the feed and its entries. Moreover, a Wikipedia link-based query expansion method for feed retrieval proved to outperform other methods with no query expansion.

Theobald et al. (2008) described TopX, a system that intends to merge two points of view for processing top-k query for semi-structured data; database systems and IR. TopX's components are categorized as data-entry time or query-processing time. At the former, the documents' contents are indexed and the concepts and semantic relations are identified. At the latter, queries are decomposed and query keywords are mapped into available concepts. The Wikipedia corpus was used as a test bed and results showed that TopX performs better than existing systems in terms of effectiveness and efficiency.

Machine learning techniques have also found their ways to query classification and segmentation. Hu et al. (2009) used Wikipedia articles and categories to solve the challenges of the query intent classification problem. Compared to other machine learning approaches, this method decreases the human effort to train a query intent classifier and improves the classification accuracy. Wikipedia knowledge was also used by Tan and Peng (2008) to augment their proposed unsupervised learning approach to query segmentation. Wikipedia and the Expectation-Maximization (EM) algorithm used to optimize the proposed approach showed 46% improvement in comparison with other segmentation methods.

Hwang et al. (2010) used the full English Wikipedia dataset exported in October 2007 to test and support the performance claims of a dynamic authority-based searching system. They proposed an approximation of the ObjectRank algorithm which materializes small subsets of the entire data graph. This helps reduce the query execution time as the algorithm needs to run only on one of the generated sub-graphs.

Ranking and clustering systems

Being a large collection of data, Wikipedia has been used to enrich texts from various sources to improve the performance and accuracy of clustering and ranking processes.

Lizorkin et al. (2010) described a technique to estimate the accuracy when iteratively computing SimRank, a "simple and intuitive measure of ... similarity between objects ... [used in] information retrieval" (2010, p.422). They illustrated this by computing "SimRank scores on a subset of English Wikipedia corpus, consisting of the complete set of articles and category links" (2010, p.422).

To help narrow down the retrieved results of search engines, Gollapudi and Sharma (2009) proposed a set of axioms for result diversification which can be viewed as a re-ranking process for the search results. The Disambiguation pages in Wikipedia were used as an evaluation dataset used to test the methods presented. The titles of these pages were used to draw ambiguous queries which are keyed in a search engine. The results of each search are then used to test the diversification algorithm.

Banerjee et al. (2007) improved the classification task of short texts such as blog feeds by extending each feed with additional features extracted from the titles of related Wikipedia articles. The results of the experiments proved that this approach "can substantially improve clustering accuracy" (2007, p.788).

In a procedure that may be called automated link discovery, tools suggest intrawiki links from a word in a Wikipedia article to an appropriate wiki page. Adafre and de Rijke (2005) proposed a method to discover missing links in Wikipedia which "could be used as an online authoring aid by revealing a ranked list of important candidate links, and the associated Wikipedia links" (2005, p.96).

The Wikipedia knowledge base was used by Carmel et al. (2009) to enhance cluster labelling. Their approach begins by extracting a number of representative terms from the texts of documents. Then, these terms are used to query Wikipedia and get the pages relevant to the corresponding cluster of documents. The meta-data of the returned pages such as the titles and the categories are then used to label the cluster. For subjects that are well covered by Wikipedia, this method proved to assign very good labels for clusters of documents.

Zaragoza et al. (2007) examined the problem of ranking entities of different heterogeneous sets of types. They employed "a statistical entity recognition algorithm to identify many entities (and their corresponding types) on a copy of the English Wikipedia." There are two noteworthy observations: first, the notion of inverted entity frequency is important to discount general types in entity containment graphs. Second, the rank of the documents in the computation of correlations enhances the performance of web methods. Pehcevski et al. (2010) implemented a new approach for entity ranking systems using the categories and link structure of Wikipedia. This approach was also extended by including a topic classification based on extracted features from an INEX topic definition. The experiments conducted using the 2006 Wikipedia XML Corpus illustrated the advantages of using the categories and semi-structured data of Wikipedia to increase the effectiveness of entity ranking systems.

Bai et al. (2010) presented a "Supervised Semantic Indexing (SSI) which defines a class of nonlinear (quadratic) models that are discriminatively trained to directly map from the word content in a query-document or document-document pair to a ranking score" (2010, p.1). The proposed methods were tested with Wikipedia documents taking advantage of Wikipedia's links structure.

One of the important aspects of query retrieval systems is the organization of the search results into a hierarchy of labeled clusters. A set of the list of ambiguous Wikipedia entries was used to help "analyzing the subtopic (rather than the topic) relevance of Web search results" (Carpineto et al. 2009).

Formulas with algorithms have been put forward that quantitatively characterize the concepts and they have been applied to a diverse set of networks, e.g., the network of movie actors, power grid, neural network and the World Wide Web. Wikipedia researchers have also examined the quantitative characteristics for the networks inherent in Wikipedia.

Networks can be represented in matrices, thus matrices can also be constructed from content and metadata in Wikipedia articles. Mathematical operations can be performed on the matrices to examine aspects of Wikipedia or to test computational algorithms on large-scale data. Buntine (2005) built a matrix from the within-wiki links between 500,000 pages of the English 2005 Wikipedia and used a discrete version of the hubs and authority algorithm to find topics in Wikipedia. For example, one topic would display the Wikipedia articles "Scientific classification" and "Animal" as the top authorities and "Arterial hypertension" and "List of biology topics" as the top hubs.

By augmenting the normalized adjacency matrix with an extra term the so-called Google matrix can be formed. The first eigenvector associated with the Google matrix determines the PageRank of an article. The adjacency matrix may be transposed, normalized and augmented. Its first eigenvector may be found to yield what Zhirov et al. (2010) called the CheiRank. They used CheiRank and PageRank to "analyze the properties of two-dimensional ranking of all Wikipedia English articles and show that it gives their reliable classification with rich and nontrivial features" (2010, p.523).

Instead of working from the links, the words of a Wikipedia articles may also be used as features in the construction of a matrix, so that the resulting matrix is a document-term matrix. A decomposition of such a matrix is often termed latent semantic analysis, particularly if singular value decomposition is the decomposition method. For assessing the performance of newly developed algorithms Řehůřek (2010) constructed a document-term matrix from the entire English Wikipedia with the resulting size of 100,000 times 3,199,665, corresponding to a truncated vocabulary on 100,000 words and almost 3.2 million Wikipedia articles.

Nielsen (2008a) used non-negative matrix factorization in a hierarchical mode to cluster Wikipedia articles and scientific journals based on the scientific citations in Wikipedia. His algorithm identified scientific areas such as "cancer" and "immunology", each associated with a set of Wikipedia articles and a set of scientific journals.

In addition to the above articles, another study that treated ranking and clustering systems is summarized elsewhere in this review (J. Hu et al. 2008).

Text Classification

Text classification is a common problem in IR systems in which a classifier is trained to associate documents to appropriate classes. Studies in this section examined various methods to solve this problem benefiting from the large collection of documents available from Wikipedia.

Several studies used Wikipedia knowledge base to enhance the text classification task. In his dissertation, Gabrilovich (2006) used linguistic information from Wikipedia and the Open Directory Project to improve text categorization performance. He used feature generation technique to empower the training instances with "new, more informative and discriminating features" (2006, p.7). Wang and Domeniconi (2008) used Wikipedia to improve document classification by defining concept-based kernels. The representation of documents is augmented by extracted knowledge from Wikipedia in a semantic kernel form. The proposed approach works in both supervised and unsupervised learning settings. In other words, it works even if class labels of documents are not available. Testing this approach with four different datasets such as Reuters-21578 and OHSUMED showed better accuracy results than the bag of words (BOW) techniques. Meyer et al. (2008) compared the use of Wikipedia as a corpus for IR of learning resources with the use of traditional corpora. They found that the use of Wikipedia successfully determined general topics, specific topics and subtopics of learning resources. Wang et al. (2009) extended the BOW method with a thesaurus derived from Wikipedia to improve the text classification task. The summary of this study is available in the Semantic Relatedness category.

Wikipedia content is also organized using categories and templates which can be further exploited for text classification. Overell et al. (2009) utilized Wikipedia's categories and templates as two structural

patterns to extend the WordNet lexicon and develop a system, ClassTag, for classifying tags. The first component of the system classifies Wikipedia articles based on the aforementioned structural patterns and lexicon. Tags are then mapped into the resultant categories in the second component. Two measures, recall and precision, were separately optimized to test the efficiency of ClassTag. Results showed improved performance when compared to WordNet.

The remaining studies we discuss in this category used various methods in different classification applications. For instance, Murugeshan et al. (2010) presented a profile based method for Wikipedia XML document classification, using negative category information. Farhoodi et al. (2009) presented an automatic web page classification method, which they tested in Persian Wikipedia. They demonstrated "the usefulness of using content-based and context-based web page features in a linear weighted combination" (2009, p.264). Ray et al. (2010) discussed automatic question classification, a module of a question answering system. They proposed a "solution for answer validation where answers returned by open-domain Question Answering Systems can be validated using online resources such as Wikipedia and Google" (2010, p.1935). Xiang et al. (2010) proposed new algorithms for text analysis and retrieval to address the gap between different knowledge areas and transfer the knowledge from one domain to another one. Wikipedia was used as a supporting data source to assist the classification task using semi-supervised learning.

In addition to the above articles, another study that dealt with text classification is summarized elsewhere in this review (Adar et al. 2009), and Medelyan et al. (2009) described another (Gabrilovich & Markovitch 2006).

Textual information retrieval

Studies in this section aimed to improve the major textual retrieval tasks. These include query processing, computing relevance feedback and employing disambiguation techniques using Wikipedia. Liu (2006) modeled a new IR system by designing and incorporating a word sense disambiguation algorithm and expanding queries using Wikipedia and WordNet dictionaries. The experimental results showed an increase in the performance of the proposed systems in terms of recall, precision, mean and geometric mean average precisions. Bast et al. (2007) presented ESTER, an efficient search engine that works based on a combination of full text and ontology search. "For the Wikipedia collection combined with the YAGO ontology, ESTER can process a variety of complex queries in a fraction of a second, with an index size of only about 4 GB" (2007, p.678). In a similar line of research, Vechtomova (2010) proposed new models for "retrieving blog posts containing opinions about an entity expressed in the query" (2010, p.71) by building a number of faceted queries (disjunctions of a list of short queries) using Wikipedia. She argued the importance of using Wikipedia "for finding concepts related to the opinion targets" (2010, p.87). Clark et al. (2009) aimed to "to analyze how texts are used in different contexts with the final goal of retrieving structured texts" (2009, p.1). To achieve this goal, they examined Wikipedia articles considering various aspects. They analyzed the development and evolution of genre in Wikipedia. They also discussed whether Wikipedia articles are composed of purpose and form. "This research has the potential to show how human categorization behavior can be emulated computationally by a machine that actually 'understands' the meaning of a text for automatic retrieval" (2009, p.15).

Other information retrieval topics

In addition to the IR topics described above, there are some articles concerning retrieving data or information from Wikipedia that do not fall under any of the labeled IR topics.

Pak and Chung (2010) proposed "a new strategy for matching contextual ads [using] Wikipedia articles as reference points to establish matching between ads and pages" (2010, p.273). Zhou et al. (2008) attempted to solve one of the main challenges in peer-to-peer file sharing systems: supporting content-based search. They proposed "a novel adaptive indexing approach …, which can identify importance of

terms without keeping global knowledge." They validated their approach "on both benchmark and Wikipedia datasets" (2008, p.381)..

To accomplish a folksonomy visualization, Lee et al. (2008) proposed a statistical model based on the frequency of each tag in Wikipedia articles to derive subsumption relations between tags. The neighboring tags were used to disambiguate the sense of a tag, since one word can be associated with multiple articles in Wikipedia. This method was tested with the del.icio.us tags and "managed to display the subsumption relationships between tags in an intuitive way" (2008, p.1094).

Csomai (2008) proposed a new approach for automated keyword extraction and its application to the back of the book indexing. The goal of this study is to solve the keyword extraction problem with less resources and higher performance. After examining various supervised and unsupervised keyphrase extraction techniques, Csomai found that keyphrase extraction can definitely be used to automate the back of the book indexing task. The indexing process should be modularized where each module handles different stages of the process. Such modules include candidate extraction, phrase ranking and phrase filtering. In addition, combination of different candidate extraction methods lead to better results than the state of the art tf-idf method. This dissertation also considered new features based on statistical measures and linguistic features based on semantic analysis.

In his doctoral dissertation on event modelling in time using cascades of Poisson processes, Simma (2010) built "a model of the revision history of Wikipedia, identifying how the community propagates edits from a page to its neighbors and demonstrating the scalability ... to very large datasets" (2010, p.1).

Krizhanovsky and Smirnov (2009) proposed a method for indexing wiki texts. They implemented this method in Russian, English, and German Wikipedias.

Demartini et al. (2010) proposed a formal model for describing and ranking entities to solve the problem of entity retrieval (ER). Wikipedia page links and categories were employed for query-category assignments. Combined with other natural language processing (NLP) techniques, the performed tests showed an improvement of the ER task.

Friedlin and McDonald (2010) investigated the medical knowledge of Wikipedia and used it to improve a laboratory and clinical observation database (LOINC). They found the medical knowledge of Wikipedia very extensive and useful as a scientific medical informatics resource, and the software they developed could satisfactorily add descriptions from Wikipedia articles to LOINC part names.

Natural Language Processing

The ambiguous nature of natural language raises the need for computational linguistic analysis for the processing of languages in a range of applications. Natural language processing (NLP) can be applied to the translation of a text into another language, paraphrasing a text, and answering questions about the content of a text. Being a multilingual online encyclopedia, Wikipedia offers NLP researchers a semantically rich dataset.

Computational Linguistics

Computational Linguistics is a subfield of NLP that aims to derive functions to investigate and evaluate various facts about human language. A wide variety of studies have investigated this topic.

Gurevych and Wolf (2010) identified various sources in a Wikipedia article to provide lexical semantic information (LSI) including the first paragraph, the article redirects, infoboxes, article links, disambiguation pages, history pages, and categories. The first paragraph, for example, includes a short definition of what the article is about. Article "redirects cover plural forms, spelling variations, and abbreviations." Infoboxes present a summary of the common features shared by a set of articles. All articles about cities, for example, enclose a table summarizing the same set of attributes about different cities around the world. Such attributes comprise the government type, city area, elevation, city

population, and time zone. “Thus, articles including infoboxes of the same class are of interest for an automatic extraction of ontological relations.” Another important LSI is the article links that use highlighted text to connect articles to each other. The format of the article links appends additional information about the relatedness of two concepts by dividing the article link between a link target and a link label. If the link target and link label belong to the same concept, the article link format is “[<link target>].” Otherwise, the article link format is defined as “[<link target> | <link label>].” Terms having different meanings in different contexts are also captured through disambiguation tags. This additional information is beneficial in various NLP tasks including machine translation and matching of semantically related words. Paraphrasing is also another NLP task that can make use of another Wikipedia LSI, the history pages that allow the comparison of different revisions of the same article.

Several other studies have investigated different aspects of computation linguistics. Turdakov and Kuznetso (2010) studied the literature of word sense disambiguation (WSD). This study discussed several WSD problems and described available algorithms and methods used to solve them. They examined the method used in the Texterra system and compared it to other methods in the literature. As Texterra used it, they highlighted Wikipedia as a suitable corpus for such methods because of its structured document network and different types of pages such as disambiguation and redirection pages. Ganter and Strube (2009) described a system for detecting linguistic hedges using Wikipedia weasel tags. This system is based on words preceding weasels and added syntactic patterns. “The experiments show that the syntactic patterns work better when using a broader notion of hedging tested on manual annotations” (2009, p.176). Mihalcea (2007, p.196) described “a method for generating sense-tagged data using Wikipedia as a source of sense annotations.” The results of the experiments designed in this study showed “that the Wikipedia-based sense annotations are reliable and can be used to construct accurate sense classifiers.” Furbach et al. (2010) described LogAnswer, “a German language question answering system which combines computational linguistics and automated reasoning to deduce answers from a knowledge base derived from Wikipedia” (2010, p.51). A semantic network representation of a snapshot of the German Wikipedia and 12,000 logical rules were used as the knowledge base of the system. Zesch et al. (2008) developed two java-based APIs to extract information from Wikipedia and Wiktionary (JWPL and JWKTL). These APIs aim to mine the lexical information of the aforementioned knowledge bases. They also provide useful tools to support NLP studies.

Semantic Relatedness

Computing semantic relatedness among a set of documents or terms is a challenging task which assigns a similarity value based on the semantic content of these documents. The studies grouped in this section used Wikipedia knowledge base to compute the semantic relatedness of words and documents.

Semantic relations have been shown to enhance the performance of clustering algorithms. Hu et al. (2008) built a concept thesaurus on the semantic relations extracted from Wikipedia to be used in a new text clustering method. Compared to traditional text clustering methods based on “bag of words,” this method showed an enhanced clustering performance when applied with Reuters and OHSUMED datasets. In a similar study, Wang et al. (2009) developed an automatic thesaurus of concepts from Wikipedia to enrich the “bag of words” representation of texts. This thesaurus aimed to capture the semantic relations between the words of a text to improve the text classification results. Several experiments were conducted using three different datasets; Reuters, 20NG, and OHSUMED. The classification performance of the proposed approach was measured using precision-recall metrics. The results showed the effectiveness of the added thesaurus.

Turdakov and Velikhov (2008) proposed a similarity measure based on Dice’s measure to compute the semantic relatedness between Wikipedia articles. Two articles are considered to be related if their Dice measure is high. This measure is computed as the ratio of the number of links the two articles have in common to the total number of links of both articles.

Pantel et al. (2009) proposed using distributional and entity set expansion to improve the computation task of the semantic term similarities. "The pairwise similarity between 500 million terms was computed in 50 hours using 200 quad-core nodes" (2009, p.946). Holloway et al. (2007) investigated the semantic map of Wikipedia and found that although the category structure of Wikipedia is constructed by varied people and bots with different motives, it is actually well developed and maintained. Several groups have extracted information from the templates of Wikipedia and other built databases. For instance, the YAGO system, proposed by Suchanek et al. (2007), extracted data from Wikipedia and combined it with WordNet. Grineva et al. (2009) used Wikipedia to compute semantic information for a new competitive key terms extraction method. This study is summarized in the Information Extraction section.

Gabrilovich and Markovitch (2009) proposed a new method for representing natural language semantics, Explicit Semantic Analysis (ESA), that represents the meaning of any text in terms of concepts based on Wikipedia articles. They argued that the main advantage of their contribution is in handling synonymy and polysemy. ESA was tested in the text categorization context. When compared with previous methods, ESA enhanced the assessment of semantic relatedness of words and texts.

Li et al. (2010) proposed a new approach for keyphrase extraction using topic relevance and term association. They represented a document as a weighted graph. The graph's vertices correspond to selected terms from the document and the weights denote the semantic relatedness among these terms. The use of Wikipedia in this method was selecting keyphrase candidates and computing their semantic relatedness. Different algorithms were employed then to relate documents by their topics and compute the term association. Experimental results showed that the keyphrase extraction approach proposed in this paper outperforms other approaches.

Zesch and Gurevych (2009) and Zesch et al. (2008) investigated the literature aiming at developing measures for computing semantic relatedness of word pairs. The identified measures were categorized into four types: path based, information content based, gloss based, and vector based measures. They compared these measures in relation with the datasets used (such as WorldNet or Wikipedia), the measure type, and the language (English or German). They concluded that the "'wisdom of the crowds' based resources are not superior to 'wisdom of linguists' based resources" (2009, p.25). A higher precision but lower recall can be obtained by using the first paragraph of a Wikipedia article rather than the whole article. In addition, their study presented two freely available systems: DEXTRACT, which assists the construction of "corpus-driven semantic relatedness datasets" (2009, p.25), and JWPL, a Java-based Wikipedia API "for building natural language processing (NLP) applications" (2009, p.25).

Other studies that examined various aspects of computing semantic relatedness using Wikipedia, (Gabrilovich & Markovitch 2007; Ponzetto & Strube 2007a; Ponzetto & Strube 2007b; Yun Li et al. 2008) are available in Medelyan et al. (2009).

Other Natural Language Processing Topics

Other studies explored different NLP applications using Wikipedia. Coursey (2009) described a new machine learning algorithm, WikiRank, developed to assign values to each entry of an encyclopedic knowledge source. Based on links that associate the entries of an encyclopedia, these assigned values can be used in various NLP applications: automatic topic identification, text-based paraphrases recognition, and ontology terms recognition.

Dorji et al. (2010) presented a new method for Field Association (FA) terms, that is, "words or phrases that serve to identify document fields ... in document classification" (2010, p.141). Their method will "extract, select and rank FA Terms from domain-specific corpora using part-of-speech (POS) pattern rules, corpora comparison and modified tf-idf weighting" (2010, p.141). The method was evaluated using "306MB of domain-specific corpora obtained from Wikipedia dump" (2010, p.157).

Mehler et al. (2010) presented "an approach to automatic language classification based on complex network theory" (2010, p.716). They tested their variant of the Sapir-Whorf Hypothesis using a corpus of

160 Wikipedia-based social ontologies. This corpus allowed them "access to structural analyses of linguistic networks ... as a new resource of language classification" (2010, p.737).

Stone et al. (2010) used Wikipedia corpus to compare different models for paragraph similarity analysis, and also to automatically generate similar smaller corpora. When comparing single paragraphs, the results favored the use of simple models such as word overlap over more complex ones such as Topic Model and LSA.

Ferschke et al. (2011) presented the Wikipedia Revision Toolkit, an open source toolkit which is usually used with the Java Wikipedia Library (JWPL). The main features of this toolkit include the reconstruction of past states of Wikipedia and the access to all article revisions. Moreover, this toolkit provides a vital knowledge source based on Wikipedia's edit history to enhance the natural language processing algorithms.

In addition to those discussed here, Medelyan et al. (2009) described other NLP studies whose details we extracted on the WikiLit website, but that we do not summarize here (Strube & Ponzetto 2006; Ruiz-Casado et al. 2007; Cucerzan 2007).

Ontology Building

Ontology, in the information science context, can be simply defined as the description of a set of concepts within a domain and of the relationships between those concepts. Ontology building (OB) has attracted the interest of a large number of researchers in the last decade especially with the exponential increase of available data online. Researchers have recognized Wikipedia as a major data source to support their work in developing new web ontologies. The foremost reason for building ontologies is analyzing and enabling the reuse of domain knowledge. The articles grouped in this subcategory presented different approaches to OB using Wikipedia.

Hepp et al. (2007) presented a "quantitative analysis of Wikipedia entries and their properties," a proof "that the URIs of Wikipedia entries are reliable identifiers for ontology concepts" and a presentation "of how the entries available in Wikipedia can be used as ontology elements." Wikipedia, as a collaborative work of an enormous number of volunteers, has helped the move away from traditional approaches to ontology construction in which the source of knowledge emerged only from experts. In the Web 2.0 and Wikipedia age, researchers shifted towards collaborative bodies of knowledge as a source for building ontologies. Moreover, Wikipedia was also used to evaluate the ontology available in its category structure to support browsing in Wikipedia. In this case, Wikipedia can benefit from its own structure to improve its content. The goal of the category structure in Wikipedia is to provide a navigation mechanism that allows its users to see all pages which relate to a particular page of their interest.

Kim et al. (2007) proposed an approach to merging and matching ontologies. They used the Wikipedia philosophy ontology as one of the input ontologies, along with "oriental philosophy ontologies, western philosophy ontologies, [and the] Yahoo western philosophy dictionary." To overcome the non-scalability limitation of existing ontology learning methods, Wong et al. (2007) proposed a new clustering algorithm, called Tree-Traversing-Ant (TTA). They used the TTA algorithm along with two measures for term similarity and dissimilarity: normalized Google distance and number of Wikipedia which is based on the cross-linking of Wikipedia articles. Their empirical tests showed 48% ontological improvement. Guo et al. (2009) proposed an ontology learning technique which relies on socially emergent bodies of knowledge like Wikipedia to build ontologies rather than the traditional expert knowledge. The resulting ontologies were comparable to the traditional ones.

Two major projects are noteworthy for OB: YAGO and DBpedia. Suchanek et al. (2008) described the YAGO ontology. YAGO is based on the concepts derived from Wikipedia infoboxes and the taxonomies available from WordNet. The evaluation of YAGO showed a 95% precision according to the type

checking techniques they employed. YAGO has been exploited in various applications: semantic search, entity organization, information extraction and ontology construction.

Bizer et al. (2009) provided a detailed overview of the DBpedia project, which has produced one of the largest knowledge bases extracted from Wikipedia. It contains the descriptions of "more than 3.64 million things out of which 1.83 million are classified in a consistent Ontology, including 416,000 persons, 526,000 places, 106,000 music albums, 60,000 films, 17,500 video games, 169,000 organisations, 183,000 species and 5,400 diseases"[16]. DBpedia can handle complex queries against Wikipedia via SPARQL query builders and interfaces. It also links other available online datasets to Wikipedia information. Among others, the British Broadcasting Corporation uses DBpedia for linking documents across their web site (Kobilarov et al. 2009).

Cantador et al. (2011) used Wikipedia categories as a semantic knowledge base for the purpose of transforming social tags into ontology concepts in the task of automatic tag categorization. Furthermore, Wikipedia entries and URIs were used for "adding machine readable annotation to existing Web content." Each entry in the English version of Wikipedia is considered a unique identifier for the concept described in the corresponding entry, and so can be exploited as an ontology component. The approach they proposed was evaluated on a dataset collected from Flickr. The results showed the improvement achieved using content and context based tags instead of subjective and organizational ones. Moreover, Hu (2010) used Wikipedia to enrich ontologies with Wikimantics which can be described as vectors extracted from Wikipedia articles. Hu referred to theses vectors as "Wikipedia-enhanced concept descriptors" (2010, p.470). Wikimantics were shown to be useful to several applications including ontology matching, with the limitation of being strongly tied to one repository, Wikipedia.

According to Muchnik et al. (2007), content is not the only factor that dictates the hierarchy of concept. Context is also an important element in such hierarchies. Directed networks of terms were employed to handle context. They proposed five different statistical methods designed to construct a hierarchy in networks of related terms. These methods were applied to Wikipedia and an excellent fit was shown for the comparison of "the hierarchy obtained from the article network to the complementary acyclic category layer of the Wikipedia."

McCrae and Collier (2008) presented a method for automatically generating regular expression patterns and developing a thesaurus. A classifier was used to classify terms as synonymous or non-synonymous. This classifier was trained using the BioCaster ontology in the biomedical domain. The proposed method was compared with Wikipedia and WordNet and experiments showed promising performance. In the field of nucleic acid research, Gardner et al. (2010) presented the Rfam database which "aims to catalogue non-coding RNAs [ribonucleic acids] through the use of sequence alignments and statistical profile models" (2010, p.D141). They discussed the pros and cons of using Wikipedia for community-driven annotation. In conclusion, they "highly recommend other curation efforts turning to Wikipedia for their annotation," while also warning that "you will lose the tight control of the data allowed by in-house curation" because "Wikipedia is built by consensus" (2010, p.D142).

Capocci et al., investigated "the nature and structure of the relation between imposed classifications and real clustering in a particular case of a scale-free network" (2008, p.1), Wikipedia. While they found a statistical similarity between the two methods, there are also differences. They attributed this to "the nature and presence of power laws ... and cannot be used as a benchmark to evaluate the suitability of a clustering method" (2008, p.1).

Yu et al. (2009) presented ROMEO, a "requirement-oriented methodology for evaluating ontologies." ROMEO imposed five ontology requirements on Wikipedia. Since there is no strict hierarchy imposed on the Wikipedia category structure, the first requirement was to have an "adequate level of category

[16] http://blog.dbpedia.org/2011/09/11/dbpedia-37-released-including-15-localized-editions/

intersection." To ensure that the category structure is useful and efficient in browsing articles, the second requirement provides a guideline on how to group categories adequately. The third stresses that cycles should be avoided in the category structure as they can lead to users being lost in a cycle of navigation. The fourth ontology requirement is for "ensuring the set of categories available is complete." This will solve problems such as the lack of an appropriate set of categories for an article. The last requirement calls for "ensuring the set of categories associated is correct," that is, no articles are incorrectly placed in multiple categories.

Banchuen (2008) developed a geographical analogue engine that aimed to compute the similarity within textual information and combine the results with those of the inadequate numeric. Wikipedia articles were used to create an ontology using the Web Ontology Language (WOL) that computer algorithms can manipulate. Banchuen explored techniques from various fields including artificial intelligence, linguistics, cognitive science, and knowledge engineering. The experimental results highlighted several observations related to different semantic measures, such as the statistical description, template description, complete stop-words list, and complete vocabulary.

Sigurdsson and Halling (2007) used Wikipedia topics related to music for the MuZeeker search engine, grouping search results according to Wikipedia categories.

Syed (2010) proposed a knowledge base derived automatically from Wikipedia and other similar information sources that organizes world knowledge in a standard machine readable format. This would allow computer applications to better access and exploit knowledge in different forms.

In addition to those discussed here, another article elsewhere in this review discusses other aspects of ontology building in Wikipedia (Krötzsch et al. 2007).

Other Corpus Topics

In addition to the corpus topics described above, there are some articles concerning using data or information from Wikipedia as a textual corpus that do not fall under any of the previous corpus topics.

Weiss et al. (2010) presented a commutative replicated data type (CRDT) algorithm with an "undo anywhere, anytime" feature. This algorithm operates on highly dynamic content in a peer-to-peer network. "It ensures the CCI (Causality, Convergence, and Intention) consistency model and tolerates an unknown number of peer, a high degree of churn and network failures" (2010, p.1172). Letia et al. (2010) addressed the design of CRDT to solve the consistency problem in large-scale systems. The CRDT aimed to "make concurrent updates commute." The proposed CRDT in this article was called treedoc. Wikipedia revision pages were stored as treedocs with each revision being the result of one of two operations: insert or delete. CRDT performed better than traditional approaches.

Curino et al. (2008) presented a new system, PRISM, to solve the time-consuming and error-prone problems of the schema evolution task. "Continuous validation against challenging real-life evolution histories, such as the one of Wikipedia, proved invaluable in molding PRISM into a system that builds on the theoretical foundations laid by recent research and provides a practical solution to the difficult problems of schema evolution." (2008, p.772). Curino et al. (2008) aimed to provide a deep analysis of the evolution of databases in Web information systems. For instance, they studied the evolution of Wikipedia database and schema. They concluded by highlighting the need for automation tools for documenting database and schema evolution especially in the case of open and more dynamic Web information systems.

Capocci et al. (2006) conducted a social network analysis of Wikipedia topics and hyperlinks. They "measure a scale-invariant distribution of the in and out degree and ... [were] able to reproduce these features by means of a simple statistical model" (2006, p.1). Silva et al. (2010) used Wikipedia content to construct a network of mathematical theorems. They employed the diversity entropy method to identify the centrality of each theorem. According to their modeling, oldest theorems have higher values of

diversity entropy which give them more importance than frontier theorems, the ones recently added to the network.

Denoyer and Gallinari (2006) described a corpus they compiled of articles from eight language Wikipedias converted to XML. The corpus consists of article pages and categorization of these articles arranged in various useful configurations, and has proven extremely popular for IR and ontological research.

Schenkel et al. (2007) exploited Wikipedia structure to build YAWN, a Wikipedia XML corpus with semantic tags. The main sources exploited for semantics include categories and lists of similar pages.

In addition to the studies described here, Ahn et al. (2005) is covered by Medelyan et al. (2009).

In summary, Wikipedia holds many characteristics that make it a target data source for a range of applications in the information retrieval, natural language processing, and ontology building areas. These characteristics include the large amount of textual and semantic information available in Wikipedia, its category structure, its semi-structured data represented in the XML documents and the graph structure that can be constructed by its entries (nodes) and the article links (edges). Medelyan et al. (2009) provide an excellent description of these particular characteristics of Wikipedia for interested researchers to understand and exploit.

Infrastructure: The Legal and Technical Support for Wikipedia

27 articles (6%) treated the infrastructure underlying Wikipedia—here we include studies concerning two rather distinct fundamental factors that enable Wikipedia to exist in its current form: legal and technological infrastructure. Both aspects of this topic group are included in the corresponding Operations topic in Wikimedia-pedia, which is why we group them together in our topic categorization.

Legal Infrastructure

Wikipedia's slogan is "The Free Encyclopedia," where "free" has the same meaning as in "free software"—it is an encyclopedia whose content is licensed for free or open-source distribution with copyleft, meaning that anyone is legally authorized to reuse its content (including for commercial purposes), as long as they make their derivatives available under the same sharing license terms. It was originally licensed under the GNU Free Documentation License (GFDL), and then in 2009 added dual licensing with the Creative Commons Attribution-ShareAlike license (CC-BY-SA), with similar legal features but with compatibility with other content that uses this latter license. In fact, since the two licenses are legally incompatible, the Free Software Foundation (FSF), curator of the GFDL, created a special clause that permitted Wikipedia to achieve this license extension. Re-users of Wikipedia content are permitted to choose either or both of the two licenses for compliance[17]. Wikipedia's legal infrastructure is a critical aspect of what has enabled the growth and survival of Wikipedia, and a number of studies investigated this aspect.

Polukarova (2007) described the copyright laws that Wikipedia depended on (prior to their adoption of CC-BY-SA). His article gives an overview of Wikipedia's legal infrastructure, and its analysis of Wikipedia's operation in the context of copyright law is still helpful and elucidating. Maracke (2010) then discussed the overall legal framework of the Creative Commons licenses, "a way to protect creative works while encouraging certain uses of them, tailored to each creator's individual preference" (2010, p.13).

Wielsch (2010), discussing the legal governance of Wikipedia, mentioned the roles taken by the Wikimedia Foundation, the FSF and Wikipedia contributors during the license transition from GFDL to

[17] http://wikimediafoundation.org/wiki/Terms_of_use#7._Licensing_of_Content

CC-BY-SA and the role of FSF as license steward. He argued: "Traditional copyright law ... is organized around the idea of a single creative entity," and as such "is not well equipped to accommodate the needs of these forms of collaboration." He argued that the "production of global knowledge commons is in need of a transnational law for networks."

Nov and Kuk (2008) argued that in this legal framework, users with intrinsic motivations are less likely to withdraw than those with extrinsic motivations. Wikipedia, however, is not merely an altruistic venture; it provides assets of value to various stakeholders, which has been characterized as a "value chain" with "triggers, that is a mix of end users and producers/creators (Cedergren, 2003). The driving forces are not only a part of the value chain, but also parts of personal motivations and benefits to the society."

In addition to those discussed here, another article elsewhere in this review discussed various aspects of Wikipedia's legal infrastructure (Famiglietti 2011).

Technical Infrastructure

Wikipedia is built upon the custom-designed MediaWiki wiki server. We consider here studies that examined the technological platform upon which Wikipedia operates, many of which proposed extensions. However, in accordance with the scope of this review, we cover here only articles that treated MediaWiki specifically in the context of Wikipedia. In addition, there are some studies about software extensions that are specifically focused on helping improve collaboration or reading; we discuss these in the sections on Software for Participation and Software for Reading, respectively.

Accessibility

A few studies have examined aspects of Wikipedia that make it accessible to disabled readers. Lopes and Carriço examined 100 Wikipedia articles and 265 non-Wikipedia Web articles cited by Wikipedia (R. Lopes & Carriço 2008). They looked for the articles' levels of accessibility, that is, to what extent they fulfilled the Web Content Accessibility Guidelines of the World Wide Web Consortium for people with disabilities. They found that Wikipedia articles are usually more accessible on average than the Web articles that the articles cite. This finding is not surprising, as Wikipedia's HTML content is automatically constructed from its wiki markup, and the software can be easily programmed to ensure accessibility (for example, automatically setting the "alt" field of the HTML <img> tag). However, Lopes and Carriço argued that citing poorly accessible pages could lower the credibility of Wikipedia, and so they suggested measures to increase the use of accessible sources.

Buzzi and Leporini (2009) went beyond considering readability to also consider editability—a defining characteristic of Wikipedia. They demonstrated that the Wikipedia interface, though readable, was not easily editable via screen reader, thus limiting the ability of blind Wikipedians to contribute content. They proposed specific modifications in conformity with World Wide Web Consortium accessibility standards that would greatly enhance blind Wikipedians' ability to participate.

Automatic Creation of Content

Some articles have investigated mechanisms for automatically creating encyclopedic content. Within Wikipedia, the Protein Box Bot has added several thousand articles on genes (Huss et al. 2008), with automated construction of an "infobox," free-text summary and relevant publications aggregated from Entrez Gene and a gene atlas database (Su et al. 2004). The WikiOpener MediaWiki extension creates a framework for querying external databases such that their material can be merged into the wiki (Brohée et al. 2010). Kinsella et al. (2008) proposed that web developers use semantic development technologies to resolve integration issues with social networks and open content projects in order to ease the reuse of content on other websites.

Other than these attempts, most studies concerning the automatic creation of content were restricted to the automatic creation of links, which is not quite as complex an endeavor. Kaptein et al. (2010) built a

system that would predict the links in the "External Links" section found in around 45% of Wikipedia articles. By using the ClueWeb category B data consisting of 50 million English Web pages, anchor text index, document priors (length, anchor text length and URL class priors), document smoothing and Krovetz stemmer, they could reach 0.68 in performance measured with the so-called Mean Reciprocal Rank (MRR). By furthermore using the social bookmarking site Delicious, they could improve MRR to 0.71.

Graphical Extensions

Several studies have examined technical measures to aid the visual aspects of Wikipedia's presentation.

Suh et al. (2007) proposed a user conflict model to uncover patterns of conflicts in Wikipedia articles. Based on this model, they developed a visualization tool, Revert graph, to depicts the revert relationships between groups of users. This tool was capable of discovering conflict patterns such as vandalism and mediation. However, "sources of disagreements, types of conflicts, and motivation for editing" were not addressed.

The Java program WikiStory constructs interactive timelines based on Wikipedia material. The Web application HistoryViz displays events related to a queried person on a timeline. Apart from this visualization, the system also features a Java applet graph visualization of Wikipedia pages (Sipos et al. 2009). The system relies on algorithms for categorization of Wikipedia articles into persons, places or organizations (Bhole et al. 2007).

The "Copernicus" system makes an attempt at creating a 3D wiki (Jankowski & Kruk 2008; Jankowski 2008a; Jankowski 2008b). A two-layer interface presents the Wikipedia article in a transparent foreground, while the background presents a 3D model related to the Wikipedia article. The user can trigger predefined camera movements and adjust the transparency.

Perona (2010) proposed a visual interface for Wikipedia that enables visual knowledge contribution, organization, and queries. It allows users to click on different part of an image to get the related information.

Other Aspects of Technical Infrastructure

Krötzsch et al. (2007) presented Semantic MediaWiki as "an extension to be integrated in Wikipedia, that allows the typing of links between articles and the specification of typed data inside the articles in an easy-to-use manner" (2007, p.584). They argued that "Semantic Wikipedia can become a platform for technology transfer that is beneficial both to researchers and a large number of users worldwide" (2007, p.585). Despite its potential, as of 2012 it is still uncertain whether Semantic MediaWiki will be incorporated into Wikipedia[18]

In another article that examined aspects of Wikipedia's technical infrastructure, Urdaneta et al. (2009) traced traffic on the English Wikipedia to identify issues that could help optimize the international distributed network load. They identified issues of concern, and then proposed optimal configurations for Wikipedia and other decentralized wikis.

In addition to those discussed here, many other articles elsewhere in this review discussed various aspects of Wikipedia's technical infrastructure (Kolbitsch & Maurer 2004; Hahn 2009; Hahn 2010; Langlois 2008; Priedhorsky 2010; Cosley 2006; Cosley et al. 2006; Cosley et al. 2007; Yuan et al. 2009; Cross 2006; Geiger & Ribes 2010; Muller-Seitz & Reger 2009; Curino, Moon & Zaniolo 2008; Roth 2007). As we have mentioned, the topic categories Software for Collaboration and Software for Reading also describe related studies.

---

[18] http://en.wikipedia.org/wiki/Semantic_MediaWiki#Potential_use_on_Wikipedia

Participation: About Contributors and their Activities

200 articles (42%) studied issues related to participation or collaboration in the Wikipedia community, including studies on contributors that create or edit Wikipedia articles and studies about other collaborators that actively participate in the online community life, such as voting for featured articles or resolving disputes among contributors. This topic group corresponds to the topic in Wikimedia-pedia that bears the same name. Participation is the most popular Wikipedia research category: more than 40% of the studies we identified and whose details we have extracted are about participation. We have divided this topic area into five major subcategories: factors that precede and lead to participation in Wikipedia; a wide variety of topics related to Wikipedia's culture of collaboration; outcomes arising from participating in Wikipedia; and software tools targeted to helping Wikipedia contributors.

Antecedents of Participation

Many studies have focused on the characteristics of Wikipedians, either from an individual or from a societal perspective. We examine here the contributors' characteristics that they bring into the Wikipedia community, and that affect their interactions with Wikipedia and with other Wikipedians. Specifically, we group these studies into four subcategories: Contributor Motivation, Cultural and Linguistic Effects on Participation, Societal Antecedents of Participation, and Other Antecedents of Participation.

Contribution Motivation

Contributor motivation is one of the most popular research topics within Wikipedia-related research. It is intuitively interesting why people dedicate their time and effort into a project which doesn't provide any monetary compensation in return. The articles in this section are focused on why people contribute voluntarily to Wikipedia, with reasons ranging from fun and socializing to more ideological reasons.

Brown (2008) discussed online cultural production from the hacker ethic perspective. He considered the phenomenon in relation to labour and leisure, arguing that online cultural production lies between these two notions. Shao (2009) presented an analytical framework that describes three distinct yet closely interrelated reasons that people participate in user-generated media: "they consume contents for fulfilling their information, entertainment, and mood management needs; they participate through interacting with the content as well as with other users for enhancing social connections and virtual communities; and they produce their own contents for self-expression and self-actualization" (2009, p.7). Similarly, Cho et al. proposed "a research model that specifies theoretical intersections among key social, motivational, and belief factors pertaining to knowledge-sharing behavior" in computer-mediated environments (2010, p.1209). Based on interviews of 22 Wikipedians, Forte and Bruckman (2005) noted that the Wikipedia's incentive system resembled that of the scientific community with its cycle of credit. Although Wikipedia articles are not signed by authors, their editing histories are available and Wikipedia contributors "recognize one another and often claim ownership of articles".

In Schroer and Hertel's study, Wikipedians' "motives derived from social sciences (perceived benefits, identification with Wikipedia, etc.) as well as perceived task characteristics (autonomy, skill variety, etc.) were assessed as potential predictors of contributors' satisfaction and self-reported engagement" (2009, p.96). Munk (2009) argued that Wikipedians are motivated by three types of motivations: feeling of self-efficacy, experience of self-esteem, and egalitarian ideology. Prasarnphanich and Wagner study concluded that "while Wikipedians have both individualistic and collaborative (altruistic) motives, collaborative motives dominate" (2009b, p.33). Baytiyeh and Pfaffman found that "Wikipedia administrators are largely driven by motivations to learn and create" and that altruism is "one of the most important factors" (2010, p.128). In a similar line of research, Zhu (2008) questioned the traditional understanding that indirect network effects are the primary force driving Wikipedia progression. He found that quality and altruism are more important, especially in the early stages of the platform development.

Other studies analyzed the relationships between the users' contributions and their motivations. In other words, how do different motivations influence the contributions to Wikipedia? Yang and Lai (2010b) surveyed Wikipedians to investigate how their self-concept-based motivations affected their contribution to Wikipedia. They found that internal motivations congruent with wikipedians' personal standards were the strongest motivating factor to contribute. In addition, people were more likely to contribute when they perceived that Wikipedia's information was of higher quality, that the system itself was of higher quality, and when they had a positive attitude towards Wikipedia. In a similar tone, Antin's (2010) study examined how information about the ways online systems operate affects participation. He found that the more people know about the system, the more they are willing to participate. Additionally, experimental results showed that people with high relative competence feedback contributed more to collective good than others with less relative competence feedback. Moreover, Ha and Kim (2009) studied the motivation structure for online mass collaboration; they theorized that there are different dominant motives for different types of collaboration. For instance, Wikipedia participation as is a kind of active cooperation needs more hedonic and social-psychological rewards than monetary rewards. George (2007) argued that Wikipedia has tackled non-contribution, a problem often associated with commons, by providing lock-in practices to keep the core group motivated and gaining more status within the community.

Motivations of Wikipedia contribution have been compared to various other settings. Comparisons have been drawn to open source software development and corporate wikis, among others. Oreg and Nov (2008) compared the motivation for contributing in open content and open software development. They found that while the software contributors are more motivated by reputation-gaining and self-development, content contributors are more encouraged by altruistic motives.

Prasarnphanich and Wagner (2009a) argued that wiki technology mobilizes participants with a wide range of interests and motivations. More specifically, Arazy et al. (2009) addressed the usefulness of wikis in corporate settings where the participant population is smaller and less diverse than in other wikis like Wikipedia. They analyzed the reasons of success of wikis at IBM, and the motivations for participation in these wikis. Results revealed similar motivations to Wikipedia contributors. They also highlighted many advantages of using wikis in corporate environments such as global collaboration, employee empowerment, and low barriers to adopt wiki technology.

Müller-Seitz and Reger (2009) studied how open source software principles apply in two related non-software projects, Wikipedia and OScar. After analyzing the two projects, they found that "many parallels to the OSS arena can be drawn in both cases," though "several factors limit the applicability of OSS principles to non-software-related arenas" (2009, p.372). They subsequently conducted a qualitative study that examined Wikipedia in the context of two research questions (Muller-Seitz & Reger 2010). First, what are the participation motivations of Wikipedians in connection to open innovation? Second, what can possibly decrease their contributions? A content analysis of multiple Wikipedia articles as well as 22 interviews with Wikipedians and OSS developers revealed that "OSS-related motivational mechanisms partially apply to Wikipedia participants" (2010, p.457).

Nov (2007) conducted a qualitative study on Wikipedia including the characteristics of its contributors and their motivations for participation. This study uncovered tensions around negative stereotypes of contributors as geeks, nerds, or hackers. Nov argued that the view of online collaboration should shift towards highlighting intrinsic motivations such as passion and interest to collaborate. He 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. In addition, Nov (2009) studied the motivations for information sharing in relation with different types of information. Wikipedia, Flickr and other open source projects were analyzed to highlight why, what, and where information sharing occurs. In conclusion, he recommended helping contributors maintain their efforts by recognizing and highlighting intrinsic motivations for participation, which he argued helps decrease the tendency of contributors "to withdraw efforts as a result of future external appropriation" (2009, p.9).

Yang and Lai (2010a) randomly sampled 219 English Wikipedia users. Using structural equation modeling to describe knowledge sharing behavior in terms of intrinsic motivation, extrinsic motivation, external self-concept and internal self-concept, they found that internal self-concept-based motivation was the most important factor for the knowledge sharing behavior. This factor was associated with questions such as "I consider myself a self-motivated person" and "I like to share knowledge which gives me a sense of personal achievement." In contrast, intrinsic motivation was found to rarely motivate, identified with questions such as "I enjoy sharing my knowledge with others" and "Sharing my knowledge with others gives me pleasure". Nov and Kuk (2008) found that users with intrinsic motivations are less likely to withdraw than those with extrinsic motivations.

Do Wikipedians have a special personality type? Amichai-Hamburger et al. (2008) surveyed 139 Wikipedians and non-Wikipedians with a personality questionnaire. Wikipedians scored lower on "Agreeableness" and higher on "Openness." Scores on "Extroversion" and "Conscientiousness" personality dimensions depended on the gender of the subject. Based on previous research, they hypothesized that Wikipedians would score lower on extroversion, but their results found only female Wikipedians to score lower. They suggested that Wikipedians scored lower on agreeableness because contribution to Wikipedia is an apparent pro-social behavior, linked to egocentric motives such as "personal expression, raising self-confidence, and group identification" (2008, p.680). Wikipedia participants "locate their real me on the Internet more frequently as compared to non-Wikipedia members" (2008, p.679).

In addition to the studies summarized here, other studies elsewhere in this review have also examined some aspects of contributor motivation (X. (Michael) Zhang & Feng Zhu 2011; Okoli & Oh 2007; Cosley 2006; Anthony et al. 2009; Shachaf & N. Hara 2010; Preece & Shneiderman 2009; Otto & Simon 2008; Auray et al. 2007; Ciffolilli 2003; Roth 2007; Miquel Ribé & H. Rodríguez 2011).

Cultural and Linguistic Effects on Participation

A number of studies investigated the effects of participants' culture or language on their participation in Wikipedia. This topic focuses on contributors' prior culture, as distinct from the shared culture that contributors build within Wikipedia, which we cover in Culture and Values of Wikipedia. The great majority of studies have been conducted on a single language version of Wikipedia, mostly on the English language. However, some studies have taken multiple languages under scope. Two main multilingual approaches were evident: studies that involve multiple Wikipedia language versions together, and studies that involve multiple languages or dialects (but maybe only one Wikipedia language version).

Wikipedia provides a large multi-lingual corpus that has been examined in different contexts such as users' behavior, and content quality. Pfeil et al. (2006) examined the correlation between Hofsted's cultural dimensions and the contributors' behavior on Wikipedia with different languages. They found a significant correlation, indicating the influence of people's cultural background on their contribution behavior on Wikipedia, and the Internet in general. Stvilia et al. (2009) found that "different Wikipedia communities may have different understandings of and models for quality" (2009, p.232). They also demonstrated "the feasibility of using some article edit-based metrics for automated quality measurement across different Wikipedia contexts" (2009, p.232). Hara et al. (2010) examined the normative and behavioral differences between various Wikipedia language versions. They found that different language versions of Wikipedia demonstrate different patterns of cultural behavior, such as differences in community well-being postings. Miquel Ribé and Rodríguez (2011) developed an "autoreferentiality" measure, by which they meant the degree to which articles in a particular Wikipedia language are primarily targeted to the unique interests of members of that language community, with less regard to issues of external interest. They contended that this is an important missing factor in the Wikipedia contribution literature (which they claimed is overwhelmingly biased towards motivations for participation in the English Wikipedia only). They analyzed twenty Wikipedia language editions and

found that autoreferentiality varies widely, with Icelandic, Japanese and Swahili the most internally focused of their sample, and Catalan, Dutch and Chinese the least so.

Liao (2009) examined the case of the Chinese Wikipedia and observed how four different regional variations of Chinese language were able to be united as one Chinese Wikipedia. He found that it "has shown some of the potentials of remixing citizenship or media citizenship that are not only enabled by the Internet but also unmatched by other state and market players. Chinese Wikipedia's attempt to create an 'unbounded citizenship' based on shared yet different Chinese language and knowledge through cross-boundary discussion is arguably unprecedented. In conclusion, participatory user-generated culture has the potential to reconnect participants across the existing polity boundaries within a linguistic space." (2009, p.56)

Some studies examined why the language editions of Wikipedia differ in size. Van Dijk (2009) cited the linguistic community population, literacy, Internet availability, freedom of speech, and established tradition of encyclopedias as the main factors influencing growth of Wikipedia language editions. Rask (2008) analyzed eleven Wikipedia language editions with respect to creation date, number of speakers of the language, Human Development Index, Internet users, Wikipedia contributors and edits per article; he found several correlations between these variables. For example, the Internet penetration and level of human development were correlated with the number of contributors. Other variables that may affect the size of different language editions are culture of volunteering, willingness to translate (from other language Wikipedia), and problems with non-Latin characters. Concerning the relatively small sizes of the Korean and Chinese Wikipedias, Shim and Yang (2009) suggested the competition faced by Wikipedia from other knowledge-sharing Web-services: Korean question/answering site Jisik iN and Chinese online encyclopedia Baidu Baike.

Jancarik and Jancarikova (2010) examined the appropriateness of Wikipedia material for preparing teachers of mathematics and biology in Czech. They demonstrated that the English Wikipedia properly covered the topics with highly detailed articles, but the Czech Wikipedia, whose scientific topics mostly consisted of English translations, included less detail and covered fewer titles, which made it insufficient for use in e-learning.

In addition to these, other studies of cultural and linguistic effects on Wikipedia participation are discussed elsewhere in this review (Gehl 2010; Baxter 2009).

Societal Antecedents of Participation

Many studies examined the motivations for contributing to Wikipedia at the societal level, as distinct from the individual level, which we cover in the Contributor Motivation topic. Zhang and Zhu (2011) found that the more a participant values social benefits, the less probably they would return after they had been forced to quit temporarily, for example in case of a government-initiated Internet block. In addition, many other articles from other sections of this review (especially in Contributor Motivations) discussed various aspects of Wikipedia's societal antecedents of participation (Nov 2007; Reagle 2008; Stvilia et al. 2008; H. Cho et al. 2010; Prasarnphanich & Wagner 2009a; Prasarnphanich & Wagner 2009b; W. Zhang & Kramarae 2008; Schroer & Hertel 2009).

Other Antecedents of Participation

A number of studies covered other factors that lead to participation that are not included in our other categories.

A significant stream of research led by Daniel Cosley examined how to use technology to help users contribute to Wikipedia. Cosley (2006) examined the challenge of motivating contributions to online communities. He used social science and public good theories to understand what motivate people to contribute to group works. These theories suggest that people contribute more when they trust their contributions will be valued by the community. Therefore, Cosley built a review system to encourage

good content and restrain bad content. Another proposition from theory is that people are motivated to contribute when the cost of contributing is lower. To cut down contribution costs, Cosley used task routing algorithms to assign people tasks they are more likely to perform. The review system and the task routing algorithms proposed in this dissertation were tested with a movie database and Wikipedia. Results showed the effectiveness of these algorithms in improving people's motivation to contribute.

Cosley et al. (2006) proposed two research questions related to contributions to online communities: How does assigning specific tasks to different contributors affect the quantity of contributions? How does a pre-publishing review process affect the quality of contributions? A field experiment showed that task assignments can increase the quantity of contributions. The quality of contributions is almost the same regardless of pre- or post-publishing review. However, reviewing process before accepting a contribution tends to slow the growth of contributions. Building on this research, they created SuggestBot, an intelligent agent that recommends tasks for the volunteer contributors of Wikipedia (Cosley et al. 2007). They found that SuggestBot's recommendations resulted in four times as many edits by volunteers as were made without its suggestions. In a subsequent study, they further found that "although the SuggestBot innovation saw limited distribution, adopters made significantly more contributions to Wikipedia after adoption than nonadopter counterparts in the comparison group" (Yuan et al. 2009, p.32).

Hardy (2010) examined the social production and implications of volunteered geographic information. Wikipedia articles that included geo-tags (metadata with geographic information) were used to build a dataset to answer the following questions: What are the similarities or differences between contributions of geographic versus non-geographic content? How do the spatial distributions of articles and participants influence contributions? The main results of this dissertation indicated that distance influences anonymous contributions. Registered user contributions were less influenced by proximity.

In addition to these, another article that studied other antecedents of Wikipedia participation is discussed elsewhere in this review (Buzzi & Leporini 2009).

Collaborative Culture

Collaboration "involves two or more contributors discussing, cooperating, and working together to create something or share information" (Preece & Shneiderman 2009, p.20). Typically, collaboration takes place outside of the encyclopedia articles in the "back narratives" such as talk pages and discussion threads. This topic covers a wide breadth of articles that deal with phenomena that, although unseen by most readers, drives the Wikipedia community.

Community Building

This section comprises articles that focused on the means, perceptions, and impacts of building a sense of community and thus helping participants increase their contributions. These studies aimed to highlight the reasons that people have the greater common good before their individual interests.

Lin (2006) explored the formation, development and sustainability of an Opensource Opencourseware Prototype System (OOPS), whose purpose is to translate the MIT OpenCourseware project into Chinese. Lin used the concepts of narrative authority and knowledge community to analyze the motivation for participation and the expression of individual narrative authority involved in building communities through volunteer work. Lin thus discovered what she called "experience asymmetry," which "exists when people have diverse experiences resulting in different and, at times, competing understandings" (2006, p.ix). Lin also examined other open source projects such as Wikipedia to uncover the relation between the online and offline lives of contributors. After analyzing OOPS, Wikipedia and other open source projects, Lin recommended various ways to sustain online communities through encouraging commitment of participants and distributing leadership.

Otto and Simon (2008) examined the effects of the changes in social characteristics on the online community network evolution. They found that removing commitment building policies does not lower performance, and that structural control is needed for sustaining the credibility and content value of Wikipedia.

Zhang and Kramarae (2008) discussed "the potentials of new collaboration technologies in supporting feminist collaboration on language and gender studies built on invitational collaboration" (2008, p.9). They emphasized the role of new technologies such as wikis and blogs in promoting invitational collaboration among feminist scholars. Related to Wikipedia, they highlighted the Feministing Wikipedia campaign with its proposal to "create Femipedia, a Wikipedia on knowledges and issues of women in different languages and from different disciplines" (2008, p.15).

McGrady (2009) argued that "us-ness," the commitment into a shared effort towards a greater good, is what makes Wikipedia work. He studied an aspect of how people act to the contrary, to tear down the sense of community. In Wikipedia, a "sock puppet" refers to an alternate account a user might create to hide their identity, when the user acts as if the sock puppet is a different person, hiding behind the anonymity of the Internet—such behavior is strictly forbidden. By their multiplicity, sock puppet accounts can create a sense of consensus, significantly bias poll results, and work around Wikipedia's three-revert-rule for fighting other editors—instances of "gaming the system." Through text analysis, looking for repeated spelling errors and idiosyncrasies in sentence construction, socket puppets may be identified.

Konieczny (2009) reviewed Wikipedia's infrastructure, participation, policies and governance to answer the following questions; is Wikipedia a community? Is it a social movement? The answers revealed that Wikipedia is more than an encyclopedia. With its developed policies, philosophies, and values, Wikipedia can definitely be considered to be a community. Also, "[by] educating its editors, fostering a collective identity tying Wikipedia with the [Free and Open Source Software Movement Industry] (FOSSMI) and diffusing those values to a wider public, Wikipedia seems to be, at least, an Internet-era 'community of print', or a part of the social movement community (SMC) surrounding the FOSSMI" (2009, p.219).

Antin and Cheshire (2010) contested the characterization of Wikipedia readers as passive "free-riders" with three arguments against this perspective. First, they argued that many readers do not contribute to Wikipedia not because they want to take advantage of other people's labour, but because they do not understand how to contribute. To support this argument, they demonstrated by a student survey that their sample's amount of Wikipedia contributions was strongly correlated to their knowledge of how to contribute. Second, they argued that "reading itself constitutes a form of contribution" (2010, p.127), since more active contributors are largely motivated to contribute because of the size of the reading audience. Third, they argued that "reading Wikipedia is a form of legitimate peripheral participation through which individuals gain entrée and can move towards more active participation" (2010, p.127).

Pentzold (2011) employed grounded theory to analyze online conversations between Wikipedia editors. His analysis aimed to uncover the way these editors "construct their self-understanding and self-description as 'community'" (2011, p.2). He analyzed the Wikipedia-l mailing list to answer 23 questions related to the Wikipedia community, concluding that users "understand their collective as an ethos-action community tying community membership not to admission procedures but to the personal acceptance of a set of moral obligations and rules of conduct" (2011, p.13).

In addition to these, other articles related to community building are described elsewhere in this review (Baytiyeh & Pfaffman 2010; Beer 2008; Kolbitsch & Maurer 2004; Lam et al. 2011).

Contributor Engagement

Contributor engagement refers to social and technical mechanisms to help participants who are already active to remain engaged in the community.

Some Wikipedians mentioned in interviews that they started editing Wikipedia because they discovered errors or omissions in an article that they knew something about (Bryant et al. 2005). However, "as their participation becomes more central and frequent, Wikipedians adopt new goals, new roles, and use different tools although they are doing so in the same 'place.' ... They move from a local focus on individual articles to a concern for the quality of the Wikipedia content as a whole and the health of the community." (2005, p.9)

Kriplean et al. (2008) studied the role of barnstars, a means for Wikipedians to show appreciation to each other. They found out that barnstars have a role in "a wide span of non-authoring work, including social support and the acknowledgment of articulation work" (2008, p.9). Ferriter (2009) studied how Wikipedia's processes and its structure help sport fans communicate. She found that users educate fellow fans in relevant social and sport meanings by creating article narratives. She believed that participation in Wikipedia allows users to reach consensus on facts and events.

Lam et al. (2011) found that Wikipedia has a male-dominated gender gap, as was claimed by The New York Times in January 2011: "there is a substantial male-skewed gender imbalance in English Wikipedia editors that does not appear to be closing at any appreciable rate" (2011, p.9). They also found indications of "a culture that may be resistant to female participation" (2011, p.9).

In addition to these, other articles related to contributor engagement are described elsewhere in this review (Cosley 2006; Hickerson & Thompson 2009; Schroer & Hertel 2009).

Culture and Values of Wikipedia

Many studies focused on the culture and values of Wikipedians in their participation in Wikipedia. This is distinct from the category Social Antecedents of Participation, whose studies examined the external culture of contributors that affects their participation in Wikipedia. Two approaches have generally been employed: comparing the culture of Wikipedia to something else, or studying a particular aspect of Wikipedia's culture. An important aspect of Wikipedia's culture is its individual values and shared ethos. Although written policies often reflect values, this topic category focuses more on unwritten values, whereas the Governance and Policies category indicates those that are enshrined in the Wikipedia policy documents. Some of the most thorough treatments of these topics were in doctoral dissertations, which we describe here.

Famiglietti's (2011) dissertation examined how Wikipedia came to take its present form as a centralized information utility co-owned by multitudes of distributed users. He analyzed the theory, put forth by some such as Yochai Benkler, that Wikipedia evolved with the ideal of the "cyborg individual." This perspective was adopted by hackers afraid to lose the benefits of their computing resource, and so created the free software community that assures that every contributor has the rights and resources to fork the product and maintain their computing autonomy. Famiglietti argued that this does not accurately describe Wikipedia, since in its present form forking is pointless and futile. Rather, Wikipedia has evolved into an information utility—a non-duplicable public resource—that is symbiotically linked with the search engine, another information utility. He examined the case of the "Gaza War" article, where the NPOV policy permitted the chronicling of a controversial subject, partly by excluding would-be contributors who were unwilling to abide by the NPOV policy; this policy permits the existence of Wikipedia as an information utility.

Forte (2009) dedicated a chapter of her dissertation to analyze governance in Wikipedia. She based her study on interviews with twenty individuals who played various roles within Wikipedia. The interviews focused on answering "why the participants contributed to Wikipedia, how they had gotten started, how they perceived their role, and how their perception of Wikipedia and their participation in it had changed over the course of their engagement with the site" (2009, p.56). A categorization of Wikipedians and their roles evolved from answers to the previous questions. Wikipedians range from novices who edit only the little they know to experts whose goal is building up Wikipedia. The perception of community, rules, and

division of labor also vary between novices and experts: early contributors to Wikipedia often don't understand the community concept of Wikipedia, consciousness of which grows with each contribution.

Gehl's (2010) dissertation argued that Web 2.0 sites offer users a surface image of unlimited creative potential where they are free to create and contribute content as they like. However, these sites maintain a not-so-apparent underlying structure that attempts to control and direct users' contributions to the interests of the site owners, be they commercial or other. Whereas most sites try to keep this underlying hidden structure hidden from users, Wikipedia is notable in its transparency that gives users full access and a controlling voice in the structural infrastructure that shapes the direction of the online encyclopedia. Although early on (in 2002) Wikipedia did toy with the idea of featuring profit-garnering advertisements, the fork of the Spanish Wikipedia and other widespread community protest killed that consideration, and steered Wikipedia in a not-for-profit direction, for the primary interest of its users (both readers and contributors) rather than commercial enterprises such as Jim Wales's Bomis, Inc.

Friesen and Hopkins (2008) presented an ethnographic study of Wikiversity by investigating its cultural aspects. This investigation was directed by an eleven-week course designed and delivered via Wikiversity. The results of this analysis were interpreted by comparing them to Wikipedia's cultural success. The conclusions drawn addressed the questions of "open culture, education, and accreditation".

Auray et al. (2007) found that in the French Wikipedia, despite the "inequality of contributions and of authority between many passerby contributors and a handful of core members, … there remains a democratic atmosphere, in a sense of a social mobility: to join the 'core team' is very simple" (2007, p.197).

In addition to these, other articles related to the culture and values of Wikipedia are described elsewhere in this review (McGrady 2009; Pentzold 2011; Reagle 2008; Hoffman & Mehra 2009; Goldspink 2010; Lam et al. 2011).

Deliberative Collaboration

Deliberation—discussion and consideration of all sides of an issue—is an essential aspect of how Wikipedia works. Researchers have often studied this aspect of the community.

Some studies observed how the features of wiki technology permit Wikipedia to be a platform where knowledge and understanding can be developed through rational dialogue. Cress and Kimmerle (2008) developed a theoretical model based on Luhmann's systems theory with Piaget's cognitive theory, to describe "how learning and collaborative knowledge building take place" (2008, p.105). This model demonstrates "the interplay of the social system wiki and individuals' cognitive systems" (2008, p.119) in the context of Wikipedia knowledge building. Klemp and Forcehimes (2010) appreciated Wikipedia's potential to offer new opportunities for deliberative democracy. They argued that "the collaborative editing process found within Wikipedia ought to be viewed as a promising supplement to traditional deliberation," promoting "the virtues of inclusion and accuracy at large scales." Benkler and Nissenbaum (2006) analyzed the ethical aspects of peer production with the goal of promoting virtuous behavior. Wikipedia, being a large platform of common based collaboration, provides a good case for examining peer production. Wikipedia was found to "enforce[s] the behavior it requires primarily through appeal to the common enterprise in which the participants are engaged, coupled with a thoroughly transparent platform that faithfully records and renders all individual interventions in the common project and facilitates discourse among participants about how their contributions do, or do not, contribute to this common enterprise" (2006, p.398).

Along these lines, Cimini (2010) investigated the impact of online dialogues on the meaning of Down's syndrome and the extent to which these dialogues can change the way that "disability" is theorized. He concluded with a message of hope by highlighting the potential for change to the best representation of Down's syndrome. Cimini and Burr (2012) examined online deliberation through Habermas' universal pragmatics and Bakhtine's dialogism theories. Wikipedia, a sample of online social interaction, was

studied as a case for online deliberation. They found that Wikipedia "indexes and accentuates struggles over authority, power, and control, as well as the attempts to overcome or profit from these" (2012, p.159).

Also drawing from Habermas, Hansen et al. (2009) presented Wikipedia as a platform that approximates Habermas' ideal of rational discourse as it surmounts the effects of authority and control. Although it is falls short on some points, the rational discourse of Wikipedia nonetheless demonstrates several features of Habermas' theory. They attributed its achievements to Wikipedia's social norms applied with the emancipatory characteristics of wiki technology.

Pentzold and Seidenglanz (2006) used Foucault's discourse theory to portray the content construction practices in Wikipedia. They proposed ways for analyzing Wikipedia's structure, community, and editing processes. They used a sample article, "Conspiracy theory", to analyze its discursive material by examining "its rules of production and limitation. [The result of this analysis] showed that the discourse unfolds according to the most of the regularities listed by Foucault" (2006, p.67).

Chon (2012) presented the operation of various aspects of the effect of a romantic collective author within the collaborative authorship practices. Contributions to Wikipedia were analyzed as a form of collaborative digital authorship. "This analysis reveals that these social practices of collaborative expression give rise to both the collective genius and collective authority aspects of the romantic author" (2012, p.832). This study's conclusion helps understand the construction of knowledge which can consequently lead "to more inclusive and reliable forms of knowledge" (2012, p.848).

Some studies noted different aspects of how Wikipedia's deliberative culture is used to develop its policies and governance structures. Black et al. (2011) examined how the "No Personal Attacks" policy has evolved as an example of a small group collaborating to develop policies that guide their work interactions. They found that the group process accorded with theory in small group research and deliberative discussion, which could serve as a model for other online communities and collaborative groups. In employing content analysis and social network mapping, they contributed methodological tools that other researchers could use to investigate collaboration in Wikipedia and similar wikis. Hilbert (2009) demonstrated how some Wikipedia governance rules can be used for design and development of e-democracy applications. He argued that Wikipedia and related applications "have the potential to fulfill the promise of breaking with the longstanding democratic trade-off between group size (direct mass voting on predefined issues) and depth of argument (deliberation and discourse in a small group)" (2009, p.87).

Although deliberative collaboration is normally considered positive, one study demonstrated how its agenda-oriented application can be used to systematically undermine Wikipedia's Neutral Point of View principle. In their study of political NGOs in Wikipedia, Oboler et al. (2010) showed "a systematic use of criticism elimination" and categorized four types of editors. They showed "that some types use criticism elimination to dominate and manipulate articles to advocate political and ideological agendas" (2010, p.284).

Some other articles related to deliberative knowledge collaboration are described elsewhere in this review (Hickerson & Thompson 2009; Tollefsen 2009; Auray et al. 2007; Lorenzen 2006; Leskovec et al. 2010a). In addition, another relevant article (Viégas, Wattenberg, Kriss, et al. 2007) that we do not describe here—but is included on the WikiLit website—is described by Medelyan et al. (2009).

Policies and Governance

Policies are the explicit normative standards which guide activities in Wikipedia. Governance refers to development and application of policies. In this section, we discuss studies that dealt with these interrelated aspects of Wikipedia.

Policies

Wikipedia has five core policies, known as the Five Pillars of Wikipedia[19]:

1. "Wikipedia is an online encyclopedia" (that is, it is not a newspaper, scientific research journal, textbook, or other genre of writing);
2. "Wikipedia is written from a neutral point of view" (which excludes writing in an opinionated matter, but explicitly welcomes objective reports of documented opinions);
3. "Wikipedia is free content that anyone can edit, use, modify, and distribute" (this is expressed in the GFDL and CC-BY-SA licenses; this principle also permits anonymous contributions);
4. "Editors should interact with each other in a respectful and civil manner" (this sets the tone for discourse and provides a basis for expelling users who manifest undesirable behaviors); and
5. "Wikipedia does not have firm rules" (other than these five principles, all other policies are constantly in flux and generally based on dynamic consensus).

In addition, there is a plethora of other policies derived from the five pillars, sometimes complementing, and even ridiculing them.

Many researchers have been interested in explaining how a consensus is reached in Wikipedia. Kriplean et al. (2007) employed a grounded theory approach to analyze the role Wikipedia policies play in the consensus process. A dump of Wikipedia active talk pages was analyzed from the articulation work perspective. Three aspects of articulation work were revealed as the result of this study: process, content, and mutual support. Moreover, the results showed that policy plays a major role in easing collaboration conflicts and challenges. The conclusions stressed the need for developing more effective systems to support mass collaboration by accommodating the following social relations: community, consensus, coercion, and control.

Butler et al. (2008) further described the nature of policies and guidelines in Wikipedia. Observing Wikipedia policies and guidelines and their involvement among the community expressed the power of wikis in providing rich organizational structures. "As a result [wikis] are capable of truly supporting a much broader range of structures and activities than many of the other more structured, collaborative platforms" (2008, p.1108).

Various studies examined the evolution of specific Wikipedia policies, such as verifiability and No Personal Attacks. Konieczny (2009) analyzed the edits and editors of the Wikipedia verifiability policy page to investigate the presence of oligarchy in Wikipedia. He found that oligarchy holds in Wikipedia if a small number of the most active editors who also hold high positions (such as administrators) win disputes on the page: "Wikipedia's editors are constantly tweaking the site's policies, so far successfully coping with the site's growing popularity, retaining their idealistic goals, and preventing a rise of any noticeable oligarchy" (2009, p.189).

Hoffman and Mehra (2009) argued that the "dispute resolution process is an important force in promoting the public good [that Wikipedia] produces, i.e., a large number of relatively accurate public encyclopedia articles" (2009, p.151). They statistically demonstrated how Wikipedia "functions not so much to resolve disputes and make peace between conflicting users, but to weed out problematic users while weeding potentially productive users back in to participate" (2009, p.151).

In his study of Wikipedians' discussive editing styles, Goldspink (2010) found that even though mutual encouragement is strongly encouraged by Wikipedia etiquette, this rarely happens in communicative interactions. Instead, he considered the style of communication to be most often non-collegial. He concluded that despite "a large number of rules, etiquettes and guidelines, explicit invocation of rules

[19] http://en.wikipedia.org/wiki/Wikipedia:Five_pillars

and/or use of wider social norms appeared to play a small role in regulating editor behavior" (2010, p.652).

Hemphill (2008) studied Wikipedia from the distinctive perspective of economic regulation theory. He argued that although Wikipedia competes with fee-based encyclopedias, the argument for regulation in its case is nonetheless narrow.

Reagle (2010a) argued that "Wikipedia, with its hundreds of norms, might be representative of a new type of large and verbose online community where such an undertaking is necessary to properly appreciate the scope of the community and its culture. Also, such an undertaking might reveal new questions for researchers. For example, the ambiguities and conflicts in the notion of neutrality, the recurrent motif of conflict and drama as being addictive and intoxicating, and the role of humor and sarcasm all merit further investigation."

Governance

Wikipedia governance was studied by various researchers providing different perspectives.

Roth (2007) explored the dynamics of wiki communities and the influence of its members and the wiki content on its growth. In analyzing the various factors that distinguish Wikipedia from other wikis and allow it to remain viable, she found population and content dynamics to be the most descriptive and most predictive factors of the viability of wiki communities. Farrell and Schwartzberg (2008) similarly examined rules in online communities and their influence on the decisions made by the participants. Based on case studies of Wikipedia and Daily Kos, they concluded that avoiding "the tyranny of the minority from consistently overwhelming the majority" is the major problem of such websites. They divided online communities into majority and minority groups. Each group is then affected by different kinds of rules. Moreover, they differentiated between online communities seeking to generate knowledge and those seeking to generate political actions. Tolerance to the diversity of points of views of participants should be higher in the former than the latter.

Okoli and Oh (2007), working on the basis that peer recognition within the community is one motivation for continued contribution, investigated the aspects of contributors' social capital that enhanced their status in the community, as measured by their administratorship (admin) status. They calculated the social capital of Wikipedians who participated in article creation, within the social networks of their co-collaborators on various articles. By deriving measures of network closure and structural holes, they found that close-knit networks among other Wikipedians are the primary factor in getting someone promoted to admin status. Whereas diversity in Wikipedia experience had some effect, these effects were mixed. In another study on election to admin status, Leskovec et al. (2010a) investigated various aspects of the deliberative procedure by which Wikipedians are elected to admin status in the community. They found that the most influential factors on a user voting for a new admin had more to do with the users' personal interactions with the candidate and with the difference in "merit" between the voter and the candidate. Specifically, they found that voters are more likely to support a candidate's promotion when they have exchanged more direct conversations with the candidate, when the candidate has contributed more edits to Wikipedia then the voter has, and when the candidate has received more barnstars than the voter has. In addition, they found that the first few votes in an election have a disproportionate effect on the final outcome of the election.

Forte and Bruckman (2008) found that Wikipedia governance tends to decentralize: "As the community grows, it has become necessary for governance mechanisms to shift outward into the community. This decentralization was not entirely accidental; self-organization was dependent in part on the design of the technology and embedded in the philosophy of the community's founder and early participants." Forte et al. (2009) further demonstrated "how governance on the site is becoming increasingly decentralized as the community grows and how this is predicted by theories of commons-based governance developed in offline contexts" (2009, p.49). More specifically, they showed how governance "relies heavily on

community-generated social norms, which are articulated in artifacts of governance called 'policy'" (2009, p.70).

Goldspink (2009) presented a qualitative and quantitative analysis of the governance mechanisms in the context of Wikipedia. In analyzing discussion pages related to both controversial and featured articles, he observed a correlation between the article group and the communication style. He found that utterances are ignored in general, positive utterances are validated more than negative ones, and there is no need for frequent invocation of rules in Wikipedia to achieve regulation. However, he noted that further analysis was needed to obtain more conclusive results.

To better understand the motives for participating in online social activities, Preece and Shneiderman (2009) synthesized and analyzed a broad literature on technology-mediated social participation, including studies on Wikipedia. They proposed a framework which characterizes the "evolution from reader, to contributor, to collaborator, and finally, to leader," and provided examples on how to apply this framework for research.

Malone et al. (2010) described four elements of the "collective intelligence genome," elements they believe are essential for successful crowdsourcing projects. In Wikipedia, they assessed that the creation, retention and deletion of article content is conducted mostly by the crowd (though administrators make final deletion decisions). Moreover, all activity is motivated by "love" and "glory" (not by "money"), and a variety of collaborative decision models are employed, including collaboration, voting and hierarchical authority.

Konieczny (2010) characterized Wikipedia governance as largely "adhocratic," implying that strategy emerges from action rather than being predetermined.

Kostakis (2010) reflected on the problem of peer governance by selecting a Wikipedia popular conflict between inclusionists and deletionists. He analyzed Wikipedia internal mailing lists, email interviews with (ex-) Wikipedians, external websites concerning Wikipedia and other domain experts. Other than several reflections on peer governance, the main conclusion was that Wikipedia should go back to its inclusionist roots. He also recommended the implementation of functional and scientific conflict resolution techniques.

Magrassi (2010) argued that free and open source software (FOSS) requires more top-down coordination than Wikipedia because the software needs to be coherent while Wikipedia can have low coherence: a Wikipedia article may still be good even if other articles are of low quality or non-existent. In major FOSS systems one may see a hierarchical structure with a "benevolent dictator" on the top, followed by "co-developers" and ordinary developers. For the "development" of an article on Wikipedia, this is not necessary.

Other articles that discuss various aspects of Wikipedia policies and governance are discussed in other sections in this review (Cedergren 2003; Geiger & Ribes 2010; Gehl 2010; Famiglietti 2011; Hilbert 2009; L. W. Black et al. 2011; den Besten & Dalle 2008; Muller-Seitz & Reger 2010; Pekárek & Pötzsch 2009; Reagle 2008; Chon 2012). An additional study, (Aniket Kittur, Suh, et al. 2007), that examined Wikipedia governance is also discussed by Medelyan et al. (2009).

Quality Improvement Processes

Many have investigated how such an unorganized crowd can create content that is sometimes of higher quality than expert-created content. Some related efforts focused on approval of articles through peer review. There are already mechanisms for peer review on Wikipedia itself (e.g., the featured article process), but combinations with external systems have also been suggested (Murray 2007).

Group Characteristics

Many articles have used group theory to understand the article improvement process, considering the set of contributors to an article to be a virtual work group.

Duguid (2006) discussed two laws believed to be the reason for quality of the products of software peer-production. First, the more people participate, the higher the resulting quality. Secondly, when more people participate, good elements remain and poor ones go away. However, he then argued that these two laws are not sufficient for knowledge peer-production systems like Wikipedia.

Viégas et al. (2007) provided a qualitative analysis of the process through which a Wikipedia article becomes a "Featured article." They argued how various aspects of the wiki technology, instead of creating chaos and anarchy, create formalized rules and well-defined processes. Benkler and Ostrom's frameworks were used to analyze these rules and processes. They concluded that "the vast number of policies in Wikipedia and the existence of robust, formal processes such as [featured articles] have been devised and modified over time according to a set of collective-choice rules makes Wikipedia a fascinating example of self-governing institutions."

In their study of coordination processes on discussion pages, Kittur and Kraut (2008) demonstrated how "Wikipedia is both an existence proof and a model for how complex cognitive tasks with high coordination requirements can be effectively achieved through distributed methods" (2008, p.45). They asserted that coordination of the group of editors is an essential factor affecting article quality. This coordination may be explicit, by discussing how to write different parts of the article, or implicit, with a few editors structuring the entire article beforehand. They argued that having higher number of contributors would increase the quality of the article only if coordination exists between them.

Stvilia et al. (2008) accessed Wikipedia's extensive discussion pages to observe the nature and directions of the conversations between the contributors, and how these collaborations operate to assure information quality in Wikipedia. They found Wikipedia particularly 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.

Arazy et al. (2011) demonstrated that more cognitively diverse groups of contributors produce higher quality articles. They found that when contributors to an article are more focused on the content itself, the article is of higher quality than when the contributors are more focused on administrative activities. Finally, they also found that task conflict negatively affects article quality.

Ransbotham and Kane (2011) investigated the effect of contributor retention and turnover on the eventual promotion and demotion of articles to featured status. They found that it is optimal to have a mix of both new and long-term contributors to attain and maintain featured status. This is notable as it contradicts widely-held views that long-term retention of participants is an unmitigated good in online collaboration. They also noted that in Wikipedia, "knowledge creation and knowledge retention are actually distinct phases" (2011, p.613) of the article lifecycle.

Carillo and Okoli (2011) investigated the group process mechanisms that contribute to the quality of articles in Wikipedia. Applying the Input-Process-Output approach and the Time, Interaction, and Performance Theory, they found evidence for the positive effects of group size and shared experience on both group process variables and group effectiveness; of group heterogeneity on group production; organizational support and member activeness on group well-being; member activeness on member support; and of organizational support and member activeness on group effectiveness.

Individual Aspects

A couple of studies examined more individual aspects of the quality development process. Adamic et al. (2010) found a significant correlation between focus and quality of individual contribution across a range of traditional and modern knowledge sharing media including Wikipedia: narrower individual

contribution domains resulted in higher and more consistent quality. In addition, Anthony et al. (2009) argued that "registered participants, motivated by reputation and commitment to the Wikipedia community, make many contributions with high reliability."

Other Factors

Wagner (2005) drew from knowledge management theories to analyze Wikipedia as a knowledge creation system. 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.

In examining the extent of Wikipedia's citation of the open-content Stanford Encyclopedia of Philosophy, Willinsky (2007) found that whereas Wikipedia widely references external scholarly resources, most of these references are not open access (that is, not freely accessible to the Internet public). He argued that whereas the scholarly citations do increase Wikipedia's source quality, the citation of non-open-access sources (particularly when open access alternatives are available) limits Wikipedia's educational value. He subsequently studied how Wikipedia editors have drawn on the open-access and peer-reviewed Stanford Encyclopedia of Philosophy to enhance the reliability and quality of articles (Willinsky 2008). He demonstrated that Wikipedia has drawn on 80% of the entries in this scholarly encyclopedia. Moreover, most of the scholarly material in Wikipedia leads to academic journal and databases.

Lichtenstein and Parker (2009) addressed concerns about the quality of Wikipedia content by proposing a model of collective intelligence whereby the Wikipedia community is encouraged to formulate policies that encourage the intervention of privileged subject experts to verify the content and settle content disputes.

Some other articles related to quality improvement processes are described elsewhere in this review (den Besten & Dalle 2008; Geiger & Ribes 2010).

Scholarly Contribution

Whereas some scholars are quite cynical towards Wikipedia, others strongly encourage their colleagues to incorporate Wikipedia into their scholarly practice and to develop it into a first-class resource. Bateman and Logan (2010) argued that "scientists who receive public or charitable funding should seize the opportunity to make sure that Wikipedia articles are understandable, scientifically accurate, well sourced and up-to-date". Logan et al. (2010) guided scientists on contributing to Wikipedia. For example, they suggested that scientists register an account for privacy, security and reputation building as well as to gain access to the "watchlist" feature that helps them keep track of when pages in which they are interested might be edited. They suggested that scientists "avoid shameless self-promotion" by not writing their bibliography page on Wikipedia (they should let others do that for them).

Most articles encouraging scholars to contribute to Wikipedia have focused on specific subject areas. In discussing available Web resources that enable creative collaboration on industrial ecology information, Davis et al. (2010) noted that Wikipedia could function as a central hub of scholarly information, similarly to how a group of molecular biologists took ownership of the RNA-related articles by forming the RNA WikiProject to develop such articles and maintain them at a high state of scientific quality (Daub et al. 2008). They also illustrated how DBPedia (which we describe in the Ontology Building topic) can be used to query Wikipedia with industrial ecology research questions.

In a similar application, Huss et al. (2010) reported on building Portal:Gene Wiki, an organized grouping of gene-related information on the English Wikipedia. Rather than developing a separate wiki for the annotation and function of human genes, it is presented as a subset of Wikipedia and thus takes advantage of Wikipedia's technical, human, and knowledge resources.

19 members of the WikiProject Medicine, mainly consisting of medical doctors, coauthored an article that called on medical professionals, especially doctors, to contribute to Wikipedia (Heilman et al. 2011).

They argued that Wikipedia is in fact an extremely important public health information source (we describe this aspect of the study in the section on "Health Information Source"), and so as a service of high-impact public education, qualified professionals should help improve its quality.

Bond (2011) called on ornithologists to appropriate Wikipedia "as a teaching and outreach tool. … [For example,] professors can replace essays and reports assigned to students with the creation or improvement of a taxonomic Wikipedia entry." He recommended the WikiProject Birds as a starting point.

Social Order

Many studies have examined the social order of Wikipedia, examining the roles, identity and power in the community. Roles refer to both explicitly defined status levels (e.g. administrator, bureaucrat, bot) and implicit emergent roles such as "core group," elite, developer, vandal, "sock puppets," and benevolent dictator. Identity refers to questions on anonymity and pseudonomity, and how these aspects affect authority of information. Power refers to discussion about social hierarchy and division of labour.

Several papers take a quantitative approach to examine the social order, for example, by downloading the Wikipedia dumps and analyzing the contribution of editors with respect to some user role. Kittur et al. (2007) studied the distribution of Wikipedia content with respect to two groups of editors; "common" and "elite." The results of analyzing a history dump of Wikipedia generated in 2006 lead to the following conclusions. 50% of the edits were contributed by elite users in 2002. However, elite users accounted for only 20% of the edits by mid-2006. This is due to the increase of participation of common users (with fewer than 100 edits). Furthermore, the elite group was found to contribute more to the content change of articles than the common group.

Using a similar methodology, Ortega and Gonzalez-Barahona (2007) performed a quantitative study on the contributions through time made by the different user groups, also showing results for sysops (admins). They reported having "found that the analysis of sysops is not a good method for estimating different levels of contributions, since it is dependent on the policy for electing them (which changes over time and for each language). Moreover, we have found new activity patterns classifying authors by their contributions during specific periods of time, instead of using their total number of contributions over the whole life of Wikipedia." The dependency on the sysop policy is demonstrated with the Swedish Wikipedia where sysops are re-elected every year. Ortega and Gonzalez-Barahona argued that these sysops need to maintain a constant level of effort to get re-elected and that is why the contribution of the sysops is large on the Swedish Wikipedia.

In their dynamic social network analysis Iba et al. (2010) identified two main categories of editors by studying participation trends. "Coolfarmers" are "the prolific authors starting and building new articles of high quality," while "egoboosters" "use Wikipedia mostly to showcase themselves." Among the coolfarmers they found two subtypes: zealots that engage in one-to-one fights with other editors and mediators which have a more diversified dialogue. For egoboosters they noted three types of networks: snake, wheel and star, where, e.g., the star network appears for a user controlling an article by relentlessly editing it after other editors make changes. With the "metrics for identifying the most valuable contributors to Wikipedia," Iba et al. argued that it "has direct practical applicability beyond finding the egobooster, by e.g. proposing alternate ranking systems for the quality of articles based on the quality of contributors." They displayed the editing patterns with a network visualization.

Other papers discuss the social order in more qualitative terms. In a study on Italian Wikipedia, Monaci (2009) noted that the "community seems to avoid any kind of individual authorship and [tends] to hold in higher regard technical roles as Administrators, Burocrati and Checkusers who are devoted to daily management activities. Those roles, especially in the words of the directly involved people, are considered the most important for the encyclopaedia's development and maintenance." (2009, p.156) She argued that administrators and other formally recognized roles get no privileges in the editing and reviewing of articles; she merely considered these roles as performing technical tasks. She concluded that

quality "doesn't depend on a progressive definition of roles and competences as observed in other [commons-based peer production experiences]."

After discussing motivational aspects of Wikipedia contributions, Müller-Seitz and Reger (2010) regarded signs of bureaucratization, where differentiating between users (e.g., visitors and administrators) "thwarts anarchic ideals." They noted the "comparatively independent" decisions the programmers make in installing technology-related features that control permissions on Wikipedia, and also noted that Jimmy Wales has been referred to as a "benevolent dictator." Based on interviews with Wikipedians, they claimed that he has "ultimate control over crucial decisions that embrace Wikipedia." Müller-Seitz and Reger also saw increasing institutionalisation within the Wikimedia Foundation, Arbitration Committee and Association of Members' Advocates.

A number of papers primarily discuss other aspects of Wikipedia, but touched on the concept of social order. George (2007) argued from the notion of "core community" explaining that open source communities have a clear hierarchy with a core group of developers. He saw the same pattern in Wikipedia, where a "core group watches closely over the project."

Other articles that also discussed aspects of Wikipedia social order are discussed in other sections in this review (Geiger & Ribes 2010; Ciffolilli 2003; Santana & Wood 2009; O'Neil 2011; Lam et al. 2011; Leskovec et al. 2010a).

Student Contribution

Whereas many teachers and professors have frowned on their students' use of Wikipedia in assignments, others have embraced it as an essential learning platform for the 21st century that gives students the opportunity to learn actively by contributing to the creation of articles based on objective knowledge. By August 2006, over twenty different universities were listed on the School and university project Wikipedia page (Konieczny 2007); in 2012, there were 31 current projects listed[20]. A wide variety of Wikipedia assignments have been described: creating a new article, translating, copy editing, reviewing or adding references to an existing one (Konieczny 2007; Witzleb 2009) as well as monitoring how their own contribution was changed by other Wikipedia editors (Chandler & Gregory 2010). These assignments been reported for various domains: history (Chandler & Gregory 2010; Pollard 2008; Nix 2010), ecology (Callis et al. 2009), law (Witzleb 2009) and chemistry (Moy et al. 2010). In this section, we discuss studies where students have been organized to contribute to Wikipedia. This is distinct from the Student Readership set of topics, which only treated students reading Wikipedia without contributing to it.

Wannemacher (2011) reviewed the experiences educators have had with the many projects on student assignments in Wikipedia contribution that has been listed on Wikipedia. On the English Wikipedia he found 132 university projects and a number of other projects on other language versions of Wikipedia. He found projects as early as 2002. Most projects were in the humanities (68), followed by social sciences (17) and engineering (12), medical (12) and natural sciences (12). Wannemacher noted the aims of the projects: increasing students' state of knowledge, student awareness of contesting in knowledge production, learning collaborative writing, increased motivation, knowledge introductions, conducting research, editing and bibliographic processing. Some projects attempted to improve articles to the statuses of "Good Article" or "Featured Article" status. One project failed, with a large portion of the articles being deleted. Wannemacher also reported that issues that teachers faced, such as making students understand the implications of the ShareAlike licences, choosing Wikipedia texts to work on and making warm-up assignment, e.g., to wiki syntax experimentation. He concluded: "Diverse factors such as the authentic learning environment, the didactically activating method of collaborative work and the text

[20] http://en.wikipedia.org/wiki/Wikipedia:School_and_university_projects

production for a very large audience create strong motivational impulses for students to carry out Wikipedia assignments."

Konieczny (2007) illustrated how wikis and Wikipedia can be used in teaching and learning, e.g., with the benefits of making the students contribute to society and getting visible results. He noted that Wikipedia assignments can include creating a new article, translating, copy editing, categorizing, or adding references to an existing article. He also noted the distinction between document style editing and thread style editing. Compared to editing on local university wikis, he argued that Wikipedia contribution has advantages as Wikipedia creates a "global newsgroup"; better connects theory that are taught with "real life" (e.g., by hyperlinking to real-life phenomenon and using free images); allows the students to use tools of the Wikipedia community (e.g., categories and templates); and makes the students realize that their efforts benefit others.

Kupiainen et al. (2007) discussed social media's role in higher education. They argued that "Wikipedia and similar digital tools provide both challenges to and possibilities for building learning sites in higher education and other forms of education and socialisation that recognise various forms of information and knowledge creation" (2007, p.128). They argued that "in higher education it is possible to save and renew higher learning's critical and revolutionary function by applying various digital information and communication technologies and using them wisely to create abilities or literacies" (2007, p.128).

Noveck (2007) supported the use of wikis and Wikipedia in teaching law and suggested different ways of facilitating such use. She argued that wikis and Wikipedia can be an important part of contemporary education for students in law and other domains.

Pollard (2008) described how using Wikipedia in history courses could provide students with "twenty-first-century learning skills such as digital-age literacy, inventive thinking, effective communication, and high productivity." She listed the different task in creating a new Wikipedia article: creating an account on Wikipedia (so the teacher can assess the contribution); reviewing Wikipedia rules; researching the topic; determining conflict and consensus; and including references and links internally on Wikipedia and to reliable external web sites, as well as citing snippets from primary sources.

For a course in comparative law, Witzleb (2009) gave assignments on Wikipedia article writing and on article reviewing. He noted possible student improvements in computer literacy, Internet resource critique, and collaborative work preparation. He also noted that lack of a textbook with good coverage of the course content is a reason to let students expand Wikipedia. Contributing to Wikipedia can also be a helpful practice for students to learn how to adapt their writing to their audience.

Crovitz and Smoot (2009) proposed that, rather than banning students from using Wikipedia, teachers should teach them about its downsides and dangers. Students should be able to benefit from the opportunity to write for real audiences, establish credibility, and discuss the nature of truth, accuracy, and neutrality.

Chandler-Olcott (2009) argued that teachers should encourage students to write in digital collaborative environments such as Wikipedia.

Forte and Bruckman (2009) studied how high school students adopted MediaWiki as a learning and writing platform. They unveiled the central role of Wikipedia as an information source for students. The conventions of Wikipedia writing also strongly affect students' actions in adopting other wiki-based platforms.

In a graduate seminar on plant-animal interactions, participants assessed the quality and content of ecology content on Wikipedia (Callis et al. 2009). They found Wikipedia generally limited in depth and breadth and that it had too few citations. Then they proceeded to edit Wikipedia in their domain and found the process "straightforward and efficient, particularly once we learned the protocol for proposing and implementing changes."

Moy et al. (2010) described how they improved Wikipedia chemistry information as a university course assignment. They had encouraging results, and offered guidelines on how Wikipedia can be further applied for educational purposes: "Students appeared to assess the material they added to the chosen entry more critically compared to when they were simply studying for the class, perhaps because of the visible nature of Wikipedia."

Chandler and Gregory (2010) shared their experiences using Wikipedia in a college classroom teaching history with Wikipedia article writing assignments. They detailed the small exercises before the major contribution: creating an account, sandbox editing, making a small change to an existing article and adding a reference. The major contribution was first submitted to the teacher as a paper. This would aid plagiarizing detection. Chandler and Gregory also mentioned problems they experienced. For example, some students were banned because of copyright violation. Student reactions to this activity were quite varied, including worry, anxiety, irritation, pride and indignation ("how dare someone make changes to our article?!"). They concluded that "students came to appreciate what Wikipedia is and what it is not. Students expressed that they think Wikipedia is acceptable for a quick reference, and that the references for the individual articles can be quite helpful, but they were quick to point out that Wikipedia is not the be all and end all of research."

Nix (2010) also shared her experiences in using Wikipedia in history teaching. In the process, she found that other Wikipedia editors added notices, for example, about missing citations to the contributed articles. She also found the vast majority of articles contributed by the students were deleted within a week. She suggested that students should engage in editing Wikipedia, but the learning process should not stop there. "The strength of this exercise comes from having students observe, discuss, and write about what happens to their articles after they publish them on the site."

Purdy (2010) proposed that the distinctions between student research (i.e., reading up on a topic) and writing are blurred by Web 2.0 services like Wikipedia, JSTOR, ARTstor and del.icio.us. He suggested that JSTOR and ARTstor could adopt Wikipedia's functionality by associating discussion pages with archived texts and images.

Radtke and Munsell (2010) described the quality and extensiveness of forestry articles in Wikipedia. Since these articles were originally limited, they assigned students to create and improve forestry articles, and were pleased to find that even after the student assignment, numerous Wikipedia contributors actively continued to develop the articles.

For smaller Wikipedia language versions, students' work may not catch the attention of a sufficient number of other Wikipedians. In a study on the Danish Wikipedia, historian Bekker-Nielsen (2011) monitored 17 Wikipedia articles in 380 days after his students had created the initial articles, e.g., for Demetrios Poliorketes and Carausius. The highest number of edits an article received in the 380 days was 26. Of the total number of edits only 6% was content-related. Bekker-Nielsen also noticed one edit by a Wikipedian that deteriorated the article quality, as well as inaccurate references in one article created by a student but not being removed by any Wikipedian before the study was made public.

In addition to these, another article that studied student contribution is discussed elsewhere in this review (Antin & Cheshire 2010).

Vandalism

In the study of Wikipedia vandals and vandalism, some articles are focused more on the behavior of vandals as malicious participants, and others are more focused on vandalism as undesirable content. In the latter case, such research might be considered a subtopic of content. Usually, however, such content does not remain a permanent part of the Wikipedia articles, and so we classify both research concerning the vandals and research concerning the content of their activities under the same topic.

In one of the first journal articles about Wikipedia, Ciffolilli (2003) presented some of the collective production principles Wikipedia is based on. In particular, he addressed "the problem of graffiti attacks—the submission of undesirable pieces of information." He argued that "Wiki technology reduces the transaction cost of erasing graffiti and therefore prevents attackers from posting unwanted contributions." He saw authority on Wikipedia as gained through reputation.

For obvious vandalism, such as when large parts of an article are deleted, Viégas et al. (2004) found that it typically only takes a couple of minutes before an article gets reconstructed. Subtle vandalism, however, may remain for much longer: Magnus (2008) made an experiment of adding purposely faulty information to Wikipedia article anonymously and from various IP addresses across separate philosophy articles, e.g., the edit "Kant's poetry was much admired, and handwritten manuscripts circulated among his friends and associates" in the "Immanuel Kant" article. (Researchers should note, though, that such vandalism "experiments" are against Wikipedia policy[21].) After 48 hours, 18 of 36 occurrences of this kind of vandalism remained. Magnus believed that a monolithic argument about whether Wikipedia is "reliable or not" is not valid; rather, "interacting with Wikipedia involves assessing where it is likely to be reliable and where not."

In an ethnographic study, Lorenzen (2006) observed the "Wikipedia:Vandalism in progress" page in two time periods in October and November 2005 to see how vandals were dealt with. Although he observed hundreds of reports of vandalism, not all were in fact vandalism. He reported 16 false reports and 39 user bans in the time period. He also discussed the issue of subtle vandalism escaping detection. He concluded that "Wikipedia does have a good system in place that can protect the integrity of articles in many instances."

Buriol et al. (2006) tracked the number of reverts through time from 2002 to the beginning of 2006. The number of reverts, including fast reverts rose almost monotonically from below 1% to over 6%, and "that may signal an increasing amount of vandalism per page." The trend was interrupted at the introduction of the three-revert rule established in November 2004 which almost halved the so-called double-reverts. (The three-revert rule forbids more than three reverts of content to a page in any 24-hour period.)

The "recent change patrol" watches over the recent changes in Wikipedia on an entirely voluntary basis and edits or deletes vandalism. Wikipedians have constructed many tools for monitoring and semi-automatically editing and reverting vandalism, e.g., Cluebot. These tools are mostly rule-based applying simple heuristics, but vandalism detection may also be viewed as a machine learning problem where the task is to classify an edit as a vandalism or not (Potthast et al. 2008; Smets et al. 2008; A. G. West et al. 2010). One approach used bag-of-words and a naive Bayes classifier as well as probabilistic sequence modeling, though could not improve upon ClueBot results (Smets et al. 2008). Potthast et al.'s (2008) approach detects vandalism in Wikipedia based on logistic regression. The classification task is accomplished based on various features extracted to quantify the characteristics of vandalism in Wikipedia articles. These features included term frequency, character distribution, edit anonymity, edits per user, and size ratio. This approach achieved 83% precision with 77% recall.

In her doctoral dissertation, Sara Javanmardi (2011) mined Wikipedia's history pages in order to uncover the patterns of the users' contributions. She developed an automatic detection mechanism of Wikipedia vandalism by modeling its users' reputations. These reputations along with textual features were found to detect low quality contributions and vandalism with higher accuracy when compared to previous approaches.

[21] See http://en.wikipedia.org/wiki/Wikipedia:Research#Advice_for_researchers and http://en.wikipedia.org/wiki/Wikipedia:Do_not_disrupt_Wikipedia_to_illustrate_a_point

With their method to predict the trustworthiness of Wikipedia articles based on revision history and Bayesian modeling, Zeng et al. (2006) showed an example with a marked drop in trustworthiness when a user performs vandalism with mass deletion.

Javanmardi, Lopes and Baldi (2010) described models for user reputation on wikis and applied their system on the revision history of the English Wikipedia. They distinguished between admins, good users, vandals and blocked users and their model for classifying between vandals and admins has a performance of around 97.5% area under the ROC curve. The reputation is based on how much of a user's insertion is maintained. Two of their models take the temporal aspect of deletions into account such that content that is quickly removed negatively affects the user's reputation, that is, the user is more likely to be a vandal. Their third model also takes into account the reputation of the user that deletes the content.

Priedhorsky et al. (2007) speculated that "the widespread use of anti-vandalism bots" has been a major reason why "the exponential increase in the probability of encountering damage was stopped" (2007, p.268). However, they stated that it "is likely that vandals will continue working to defeat the bots, leading to an arms race" (2007, p.268). In his doctoral dissertation, Priedhorsky (2010) presented Cyclopath, a geowiki for cyclists. In its design, he drew numerous principles from Wikipedia, and did some particular analysis of Wikipedia in the process. Most notable is his analysis of extent of damage and vandalism in Wikipedia, following up on his earlier paper (Priedhorsky et al. 2007). In a dataset based on revisions up to 2006, he found that nonsense (53%) and offensive (28%) contributions were the most frequent; however, misinformation (20%) and offensive contributions were potentially the most harmful in their effect on the integrity of the wiki. This work also identified ratio between the number of damaged revisions and total revision, 5%, and considered the metric "damaged article view" that "measures the number of times an article was viewed while damaged."

Regarding automatized vandal fighting, Geiger and Ribes (2010) argued that "often-unofficial technologies have fundamentally transformed the nature of editing and administration in Wikipedia" (2010, p.117) and that in "large part to a vast array of interoperable tools, bots, and standards, the process of vandal fighting is becoming increasingly automated." Furthermore, they claimed that bot-performed vandal fighting not only speeds up existing processes, but also transforms them. They argued that "technological tools like bots and assisted editing programs are significant social actors in Wikipedia, making possible a form of distributed cognition regarding epistemological standards—independent of what those standards happen to be" (2010, p.124). Based on earlier research, they reported that 16.33% of all edits in 2009 was made by bots (fully-automated software agents). Geiger and Ribes also described how a vandal gets reverted and blocked using a combination of assisted editing tools and bots, showing how multiple Wikipedian vandal fighters use Huggle and Twinkle (anti-vandalism software) and how the tools interacted together with ClueBot over a fifteen minute period.

A few studies have investigated what motivates vandals. George (2007) noted that the "Wikipedia community has been remarkably successful" in combatting vandalism, "developing sophisticated methods and programs to detect and corrected vandalized content." He offered suggestions for motivations of vandals: "sheer joy of subversion" and "propagating false or misleading information as part of an agenda." Shachaf and Hara (2010) studied Wikipedia trolls' behaviors and motivations, and compare them with hackers. From interviews with administrators on the Hebrew Wikipedia, they identified eleven trolls and observed their activity on various Wikipedia pages. They found that boredom, attention seeking, and revenge motivate trolls, who are entertained and find pleasure from causing damage. Their behaviors are repetitive, intentional, and harmful actions undertaken in isolation and under hidden virtual identities.

In addition to these, another article that studied vandalism is discussed elsewhere in this review (Suh et al. 2007).

Other Collaboration Topics

This category comprises topics that do not neatly fall in the other collaborative culture topic areas.

Collaboration has also been discussed in the context of content quality of Wikipedia. Ehmann, Large and Beheshti (2008) found that "add link" and "add information" were the most common instances of collaboration, and talk pages played an integral role in facilitating the collaboration process. They argued that areas which have attracted less attention are of lower quality. Tumlin et al. (2007) investigated the collectivism and collaborative knowledge development in Wikipedia, and discussed how the negative aspect of such collaborative knowledge generation threatens the information quality. They emphasized the vital role of librarians for evaluating the accuracy and integrity of information sources. Moreover, Wilkinson and Huberman (2007a) proposed that "Wikipedia article quality continues to increase, on average, as the number of collaborators and the number of edits increases which explains that topics of high interest or relevance are naturally brought to the forefront of visibility and quality."

In their methodological paper, Meishar-Tal and Tal-Elhasid (2008) argued that measuring collaboration is different in Wikipedia than in educational wikis. They suggested that in order to produce more accurate measurements of group collaboration and to more accurately compare and rank users by the intensity of collaboration, it is better to measure the intensity of member-to-member collaboration in addition to the number of editors per page.

Leskovec et al. (2010b) analyzed the signs of relationships (links) in online social networks such as Epinions, Slashdot and Wikipedia. A positive sign indicates a friendship relation while a negative sign indicates an opposition relation. They found "that the signs of links in the underlying social networks can be predicted with high accuracy, using models that generalize across this diverse range of sites" (2010b, p.641).

Purdy (2009) conducted an in-depth scholarly study of Wikipedia's writing composition characteristics. He argued that Wikipedia represents an important form of writing today—online collaboration. His analysis observed that although Wikipedia represents a new form of dynamic, unstable knowledge, it nonetheless manifests traditional writing composition elements of revision, collaboration and authority.

Madison, Frischmann, and Strandburg (2010) argued that "understanding the origins and operation of beneficial constructed commons requires detailed assessments that recognize that they operate simultaneously at several levels, each nested in a level above, and that each level entails a variety of possible attributes" (2010, p.37).

Mateos-Garcia and Steinmueller (2008) employed "the concepts of Epistemic Community, Legitimate Peripheral Participation and Distributed Authority to elaborate a model for the analysis of social and organizational dynamics in Free/Libre/Open Source projects."

In her doctoral dissertation, Langlois (2008) examined how Amazon.com and Wikipedia use an interaction of their web technologies and cultural practices to shape meaning from user contributions. Specifically, she analyzed how the technical capabilities of the wiki software affects the shaping of meaning in Wikipedia and other MediaWiki-powered sites, producing meanings through interactions between the host organization, the technology, and the users.

Discussing the Breton Wikipedia version, Baxter (2009) argued that Wikipedia can be relevant in developing and evolving minority and lesser-used languages.

In addition to these, other articles that studied other collaboration topics are discussed elsewhere in this review (Müller-Birn et al. 2010; Preece & Shneiderman 2009; Gehl 2010; M. Lin 2006; Poderi 2009; Muller-Seitz & Reger 2009; Ha & Y.-H. Kim 2009; Pfeil et al. 2006; Kaplan & Haenlein 2010; Mattus 2008; de Laat 2010; Yasseri & Kertész 2012; Kimmons 2011). In addition, Wilkinson and Huberman (2007b) is covered by Medelyan et al. (2009).

Participation Outcomes

A number of articles treated some particular outcomes of participating in Wikipedia, some intentional and some unintended, other than the obvious outcome of producing a Web-based encyclopedia. Some of these can be grouped as contributor perceptions of credibility and as participation trends, though others covered a wide variety of other outcomes.

Contributor Perceptions of Credibility

Some studies examined the credibility of Wikipedia from its contributors' perspective. We categorize articles here that examined Wikipedians' insider perceptions of the credibility of the articles they work on. This is distinct from Wikipedia readers' perceptions of the encyclopedia's credibility (which we classify as Reader Perceptions of Credibility) and also distinct from any attempt at objective, unbiased evaluation of reliability (which we classify as Reliability).

Francke and Sundin (2010) studied the credibility assessment of Wikipedia editors. They concluded that the "situations and purposes for which the editors use Wikipedia are similar to other user groups, but they draw on their knowledge as members of the network of practice of Wikipedians to make credibility assessments, including knowledge of certain editors and of the MediaWiki architecture." In Sundin's later paper (2011), he studied the verifiability policy. He concluded that active editors "can be seen as akin to janitors of knowledge, as they are the ones who, through their hands-on activities, keep Wikipedia stable" (2011, p.840).

An important Wikipedia principle is "no original research." As Wikipedia is not a place where novel research is published, article credibility is dependent on citations to published external sources. By studying Wikipedia reference sources, Huvila (2010) showed that "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." He argued that this finding helps 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."

In addition to these, another article that studied contributor perceptions of credibility is discussed elsewhere in this review (Brown 2009).

Participation Trends

Quite a few studies have examined the trends of Wikipedian participation over time.

Buriol et al. (2006) used the hyperlink structure between Wikipedia articles to build a Wikigraph and study the evolution of web graphs over time. The purpose of this study was twofold. First, it aimed to uncover any trend in the evolution of Wikipedia articles. Second, it aimed to highlight the temporal evolution of the topological properties of Wikigraphs. The results showed that the number of articles and of Wikipedia editors was growing exponentially. In addition, the number of articles and hyperlinks in each article was still growing linearly. Other results, such as a constant average number of edits per user, showed the level of maturity of the evolution of Wikipedia.

Almeida et al. (2007) studied user behavior using statistical modeling in relation to Wikipedia's evolution. They found: First, the evolution of Wikipedia follows a "self-similar process" rather than the Poisson process that governs the evolution of most web pages. Second, Wikipedia is growing exponentially due to the continuously increasing number of its users. Finally, they found that the number of changes to Wikipedia articles follow a power law distribution.

Priedhorsky et al. (2007) studied the measure of value of a single edit, using the number of page views of the edited version to generate a metric that they called persistent word view (PWV). Based on these computations, they observed that frequent editors are increasingly more represented in the versions of

articles that readers actually see. They also observed that erroneous versions of articles, before they are corrected, are increasingly visible to readers.

Brandes and Lerner (2008) developed software for visualizing the revisions made to an article. Their visualization readily identifies who the main contributors are, who edits whose revisions, and general patterns of revisions made. They presented it as a useful tool to quickly visualize the pattern of revisions on an article, which could suggest point-of-view conflicts or vandalism (though their tool could not readily distinguish between these different phenomena).

Ortega's (2009) doctoral thesis quantitatively analyzed the top-ten language editions of Wikipedia. He found that as of 2007, the number of contributors had tapered off as had the number of monthly contributions. However, he found that there was increasing activity in talk pages, which was related to increasing quality of the articles. He also found that active contributors were active for 200 to 400 days, after which they reduced their activity. Combined with the finding that there was a decreasing rate of new editors, he cautioned that the Wikipedia community needs to actively try to increase the number of active contributors, or else the quality of Wikipedia might suffer from the trend. In a related study, Ortega et al. (2008) examined contribution inequality in the ten biggest language versions. Using Lorenz curves and Gini coefficients, they found "large differences in the number of contributions by different authors ..., and a trend to stable patterns of inequality in the long run" (2008, p.304).

Lam and Riedl (2009) investigated the growth of Wikipedia over the years. Although they observed that the number of new articles developed are decreasing, they argued that the main challenge for Wikipedia would be how to attract new editors and retain the existing ones, which has gotten harder due to increasing editing conflicts.

Another study on multiple Wikipedia language versions by Hara et al. (2010) examined the norm and behavioral differences between various Wikipedia instances. They found that different language versions of Wikipedia demonstrate different patterns of cultural behavior, such as difference in community well-being postings.

Kimmons (2011) analyzed the revision histories of all 3.4 million articles in English Wikipedia. His analysis consisted of eight different measures: rigor, diversity, diversity index, revision chaining, collaborative rigor, revision lengths, contributions made by registered users, and contribution index. Among his findings was a strong participation inequality between users: "51 percent ... made only two or fewer revisions to Wikipedia" with "78 percent of all revisions made by registered contributors being made by the top one percent (n=31,914) of contributors". He concluded that "the typical article in Wikipedia reflects the efforts of a relatively small group of users (median of 12) making a relatively small number of edits (median of 21)".

In addition to these, other articles that studied participation trends are discussed elsewhere in this review (Lam et al. 2011; Suh et al. 2007). Moreover, a literature review that we described earlier reviews studies on Wikipedia participation trends in considerable depth (Yasseri & Kertész 2012).

Other Participation Outcomes

This category comprises articles that treated miscellaneous outcomes of participating in Wikipedia.

Several studies have viewed Wikipedia as a role model for other settings. Holley (2010) considered Wikipedia a role model for libraries to develop crowd-sourced services. In her short essay, Miller (2005) asked, "What does it mean to author a piece of writing?" (2005, p.37). She claimed that the line between reading and writing is deliberatively blurred in Wikipedia. She argued that this is the contemporary view on authorship in general, where "we no longer say we 'are' authors. Instead we periodically author, read, and share information" (2005, p.40). In a similar vein, Tkacz (2007) argued that whereas the Internet is sometimes used as a tool of oppressive surveillance, Wikipedia provides a positive kind of visibility that

makes the political process of creating knowledge open to public view. This distributes power among both the creators of knowledge and its consumers.

Thom-Santelli (2010) discovered that Wikipedia contributions involve territoriality—the expression of ownership towards an object. She concluded that some expert participants "express territoriality regarding their expertise through higher levels of participation," and they more likely "vote down novice-generated tags in a defensive manner" (2010, p.4). She suggested that territoriality could be used as a design resource to generate more contributions.

In an article on popular music culture, Beer (2008) claimed that Wikipedia and other Web 2.0 platforms allow new participatory and collaborative cultural forms. One of his examples of these forms is "flickering friendships"—connections between "'like-minded' people who have never actually met" (2008, p.224).

Tseng and Huang (2011) examined various aspects of Wikipedia such as its content and technical and social values. They explored these aspects from knowledge sharing and job performance perspectives. The results of a statistical analysis "indicated that Wikipedia is significantly associated with the degree of attainment of job performance and knowledge sharing" (2011, p.6122). Consequently, "enterprises can encourage knowledge sharing among its employees and enhance their job performance" (2011, p.6122).

In addition to these, other articles that studied other participation outcomes are discussed elsewhere in this review (Wagner 2005; Page 2010; W. Zhang & Kramarae 2008; Klemp & Forcehimes 2010; A. Rubin & E. Rubin 2010; Pentzold 2009).

Software for Participation

A number of studies investigated software specifically developed to support Wikipedia's participation-related activities. This is beyond general software and extensions that are built on the MediaWiki platform, which we discuss in the section on Technical Infrastructure. For those focused on helping participation, we consider studies in two general categories: collaboration software and reputation systems.

Collaboration Software

Some studies have been conducted on software that is developed to support collaboration in Wikipedia. Mostly these are various kinds of bots (software agents) that automate, support, and transform Wikipedia editing processes. Additionally, visualization software can be used in collaboration support.

Niederer and Van Dijck (2010) argued that analyses of Wikipedia collaboration should include considerations of the role of bots as non-human agents. "Bots are systematically deployed to detect and revert vandalism, monitor certain articles and, if necessary, ban users, but they also play a substantial role in the creation and maintenance of entries" (2010, p.1383). They pointed out that "human editors would never be able to keep up the online encyclopaedia if they were not assisted by a large number of software robots" (2010, p.1377). Through an analysis of various Wikipedia language versions, they found a somewhat inverse correlation between the size of a Wikipedia language version and the percentage of bot edits. For example, the German Wikipedia "has only 9 per cent bot activity" (2010, p.1378), whereas "Wikipedias of small and endangered languages show a high dependency on bots and a relatively small percentage of human edits. Oriya, for instance, depends 89 per cent on automated software programmes" (2010, p.1381). They concluded that Wikipedia should be seen as "a sociotechnical system" (2010, p.1384), and its "nature and quality should be evaluated in terms of collaborative qualities … of its human *and* non-human actors" (2010, p.1383).

A page will almost always have multiple authors. The revision history records each author's contributions, but the format of the revision history makes it nontrivial to determine who contributed what and the most, since text may be reformulated, moved, deleted and reintroduced. To get an overview of the

edits, the convenient program history flow takes the revision history as input and visualizes the entire history of an article with colorings determined by author (Viégas et al. 2004). Another related tool, WikiDashboard (Suh et al. 2008) generates a visualization of the edit activity of each Wikipedia page. It embeds the generated plot in a proxy copy of the Wikipedia article showing the amount of edits of each author through time for the given article.

Another noteworthy tool is SuggestBot, which recommends tasks for Wikipedians to take on. However, because of its theoretical background, we describe its related studies in the section on Other Antecedents of Participation (Cosley et al. 2006; Cosley et al. 2007; Yuan et al. 2009).

In addition to these, other articles that studied collaboration software are discussed elsewhere in this review (Priedhorsky et al. 2007; Geiger & Ribes 2010).

Reputation Systems

One of the ongoing challenges with Wikipedia is the assurance that articles are kept to a high standard of reliability. Because of the enormous number of articles, there have been several attempts to produce automatic metrics to compute the predicted reliability of an article; these are covered in the Computational Estimation of Reliability category. Studies covered there focused on algorithms targeted to Wikipedia readers. In contrast, here we cover software extensions targeted to Wikipedia contributors to aid their editing activities. The strategy for reputation systems is to use the reputation of the contributors to an article as a proxy for an article's reliability, with the assumption that when contributors with proven track records of working on high quality articles contribute to unconfirmed articles, these other articles will likely be of relatively high quality.

Arazy et al. (2010) have argued that reputation systems and other tools can change wiki editors' behavior. Research has provided several kinds of approaches to how reputation systems can be developed and implemented.

Javanmardi, Lopes and Baldi (2010) described designs for reputation systems that can be used in Wikipedia. They offered three reputation models, which all differ in complexity, accuracy and robustness. Korsgaard and Jensen (2009) presented a Wikipedia reputation system design for user-contributed article ratings. Their solution does not require changes to Wikipedia, because the software is installed on the user's own computer (2009, p.81).

Adler and de Alfaro (2007) have demonstrated the applicability of a user reputation system. After implementing the system in Italian and French Wikipedias, they found out that it has "good predictive value: changes performed by low-reputation authors have a significantly larger than average probability of having poor quality, and of being undone" (2007, p.1). In a subsequent study, Adler et al. (2008) implemented a trust assignment algorithm to provide a content-based reputation system for Wikipedia. The trust values of an article are associated to each word of each revision of that article. The additions of new words, the deletion and alteration of existing words, and the reputation of the authors are factored into the trust computation. These values proved to be good indicators of articles stability.

Readership: About Readers of Wikipedia

94 articles (20%) studied issues related to Wikipedia readers (as distinct from contributors), how they perceive and use Wikipedia, and the purposes of their use. This topic group corresponds to the Reach topic in Wikimedia-pedia. The major categories here cover studies about commercial applications of Wikipedia content; the use of Wikipedia as a general source of knowledge on the Internet; Wikipedia's ranking and popularity compared to other knowledge sources; the extent to which Wikipedia readers consider it credible; software tools targeted to helping Wikipedia readers; and various topics related to students as readers of Wikipedia.

Commercial Applications

Although Wikipedia is an open content project funded by a not-for profit foundation, its Creative Commons Attribution-ShareAlike license explicitly permits commercial reuse of its content. Several studies investigated phenomena that try to leverage or take advantage of this permission.

Two studies investigated aspects of how companies might profit directly from Wikipedia. Langlois and Elmer (2009) investigated how Wikipedia content is being used anywhere across the Internet. They found that it is mostly used for generating commercial content or for increasing traffic through search engines links. Plaza (2011) investigated how Wikipedia entries can get traffic to a tourism website in comparison with other traffic sources like Google. She found that Wikipedia entries are quite effective in getting people to visit and navigate through the sample website she studied.

Kaplan and Haenlein (2010) introduced businesses to using social media. They noted that Wikipedia is very restrictive in permitting commercial participation in its community, yet urged businesses to pay attention to it because "although not everything written on Wikipedia may actually be true, it is believed to be true by more and more Internet users" (2010, p.62). However, they warned that trying to gloss corporate image by getting third parties to edit Wikipedia articles is probably futile at best and at worst, could likely backfire.

Hickerson and Thompson (2009) considered the potential of Wikipedia as a tool for public relations, investigating how "wiki sites uphold dialogic principles and encourage dialogue" (2009, p.9). The fact that the site is open, free, and does not serve financial interests of any single party at the expense of others, contributes to participants' feeling of partial ownership. This, in turn, "may encourage repeat visits to the site and an increased investment in the organisation" (2009, p.9).

In related work, Rubin and Rubin (2010) hypothesized that the degree of Web activity about a company correlates with the extent to which investors are generally informed about their companies. To test this, they investigated the frequency of edits of Dow Jones Industrial firms' entries on Wikipedia in relation to analysts' forecasts and recommendations, and confirmed that Wikipedia edit frequencies are indeed correlated with the accuracy of corporate analysts' forecasts.

In addition to these, other articles related to commercial applications of Wikipedia are described elsewhere in this review (Cedergren 2003; Gehl 2010; Rahman 2006; Rahman 2007; Forte et al. 2009).

Knowledge Source

In this section, we discuss the use of Wikipedia as a source of various kinds of knowledge, including health information, the judiciary, as a resource for scholars and librarians, current news, and for other information purposes.

Health Information Source

Many articles discussed the use of Wikipedia as a source of health information for the general public as well as for health professionals. This is distinct from its use by medical students, which we cover in Domain-Specific Student Readership. An important article in this topic category was coauthored by 19 members of the WikiProject Medicine, mainly consisting of medical doctors (Heilman et al. 2011). They reviewed the literature in this area and concluded "that the medical information on Wikipedia is found in articles on many topics that contain few factual errors, although the depth of individual articles and the ease of understanding need to be improved substantially".

Hughes et al. (2009) examined how Web 2.0 tools like Wikipedia were being used in clinical contexts. They found that although medical practitioners were aware of credibility deficiencies of Wikipedia, they employed different strategies to cope with the risk while meeting their background information needs. Younger (2010) studied the potential of wikis and Wikipedia as an information source for nurses. She argued that although it is not likely for Wikipedia to replace the traditional printed valid information

sources, it would be a promising starting point for nurses in searching for evidence-based patient related information.

A few studies have investigated how Wikipedia compares with other online sources of healthcare information. Since this involves not only popularity, but also the responsible provision of information that could affect people's health, accuracy was of primary concern in these comparisons. In a study on the efficiency of Web resources for identifying medical information for clinical questions, Wikipedia failed to give the desired answer in around one third of the cases, whereas Web search engines, especially Google, were much more effective. However, Wikipedia was more efficient than medical sites such as UpToDate and eMedicine in terms of failed searches and number of links visited, and it proved to be the most frequent "end site" that provided the ultimate answer from a Google search (2008). Mühlhauser and Oser (2008) found the German Wikipedia comparable in quality to the websites of two major German statutory health insurance providers for content and presentation of patient information. However, in their assessment based on the standards of evidence-based medicine, none of the three sources proved satisfactory. Yermilov et al. (2008) compared the quality of Internet sources of surgery information. Wikipedia's average information quality was less than that of professional societies, government and hospital sites, but it was higher than the average quality of universities and manufacturer or pharmaceutical sites.

Using search engine optimization techniques, Laurent and Vickers (2009) investigated the Google ranking of the English Wikipedia for health topics. Queries based on 1,726 keywords from an index of the American MedlinePlus, 966 keywords from a NHS Direct Online index and 1,173 keywords from the United States National Organization of Rare Diseases, they compared Wikipedia to .gov domains, MedlinePlus, Medscape, NHS Direct Online and a number of other domains. They found the English Wikipedia as the Web site with the most top rankings. Using data from stats.grok.se for June and January 2008, they also examined health-related topics with probable seasonal effects, such as frostbite, hypothermia, hyperthermia and sunburn. They found a clear effect in the page views. They also analyzed the page view statistics of three articles describing melamine, salmonella and ricin. These examples were associated with official health alerts in 2008, and page view statistics showed a marked increase correlating with the timing of announcements.

In addition to these, other studies related to Wikipedia as a health information source are described elsewhere in this review (Hickerson & Thompson 2009; Clauson et al. 2008; A. Leithner et al. 2010; Cimini 2010). In summary, the studies on this topic have generally found Wikipedia useful for general health information, but unsurprisingly, do not consider it as a reliable source for healthcare decisions. Wikipedia, for its part, explicitly disclaims giving medical advice[22], and beyond attempting to provide generally useful information, has no goal of being a primary source for medical decisions.

Judiciary Use

Three legal studies examined the judiciary use of Wikipedia and discussed the controversy of using Wikipedia as an authority (Breinholt 2008; Stoddard 2009; Peoples 2009). Breinholt (2008) classified the different uses of Wikipedia into four categories:

1. Wikipedia as a dictionary. For example, Wikipedia was used to answer what "candy striper" means.
2. Wikipedia as a source of evidence. In the most perilous uses of Wikipedia, judges rely on Wikipedia for evidence (for example, to determine whether or not the United States Interstate 20 passes through California).
3. Wikipedia as a rhetorical tool. This involved innocuous uses, such as for literary allusions.

[22] http://en.wikipedia.org/wiki/Wikipedia:Medical_disclaimer

4. Judiciary commentary about Wikipedia. One case involved a judge cautioning against citing Wikipedia in an appellant brief.

Peoples (2009) examined the quality of the Wikipedia articles cited by American judicial opinions. He found that the "majority of citations to Wikipedia entries in cases were not significant to the case but were merely collateral references." Peoples proposed a number of best practices for citing Wikipedia. He presented some cases and scenarios where Wikipedia should not be cited and others where citing Wikipedia could be deemed appropriate.

We suggest that Wikipedia articles can be appropriately used for definitions, when such definitions have been considered by many readers and multiple editors have edited the article over time—these could be considered consensus definitions. Indeed, in some cases Wikipedia may be one of only a few references available, as in the "candy striper" case. Wikipedia, of course, explicitly disclaims giving legal advice[23], and is not suitable for such purposes.

Knowledge Source by Scholars and Librarians

Somewhat surprisingly, the largest number of articles related to Wikipedia as a knowledge source studied its use for research purposes by scholars and librarians. While it is true that Wikipedia's most vociferous critics hail from these ranks, a very large number of academicians in fact have quite positive, if nuanced, perceptions of Wikipedia's value.

Source for Scholarly Research

A Wikimedia Foundation survey has found researchers to be generally quite positive towards Wikipedia: Over 90% of 1743 self-selected respondents were "very favorable" or "somewhat favorable" (2009). Among Public Library of Science (PLoS) authors, the result was 96%. Other results showed that 68% answered, "Yes, on a large scale," to the question, "Would you be in favor of efforts to invite scientist to add or improve Wikipedia articles?". Such results are very positive for Wikipedia, but may be biased due to the self-selection of respondents and because the publisher web site with initial reference to the survey was open access.

A few studies have surveyed scholars to understand their usage and attitudes towards Wikipedia. Dooley (2010) surveyed 105 university faculty members and found that 54.4% considered Wikipedia to be moderately or very credible, 26.6% considered it having some credibility, and 20% considered that it had "no credibility." 45 of 105 respondents said they used Wikipedia moderately to frequently in their teaching or research, 40 only occasionally, and 20 said they never used Wikipedia for teaching or research. Despite the controversy with students, many professors and other researchers do in fact cite Wikipedia. Concerning citations of Wikipedia, Dooley examined 250 research reports published in 2009 and early 2010 from the Academic OneFile electronic database that contained "Wikipedia" in their text. She found that 27 of the papers featured Wikipedia as the main topic and 62 had only brief mentions of Wikipedia. 249 of these papers cited Wikipedia as a source.

Chen (2010) found that although academics extensively use online information resources and databases for teaching and research purposes, they are often concerned about credibility of Wikipedia content. Those who use Wikipedia are more likely to also be Wikipedia contributors. Eijkman (2010) observed how academics cautiously use Wikipedia along with other sources of knowledge. He found that although they are aware that it disrupts their traditional power as knowledge providers, academics are not generally as antagonist towards Wikipedia as is commonly assumed.

Page (2010) compared the current state of Wikipedia's documentation of biological species with E. O. Wilson's vision of an "encyclopedia of life." He contended that in its dominance of search result

[23] http://en.wikipedia.org/wiki/Wikipedia:Legal_disclaimer

rankings, contributor size, and potential linkage to other data, Wikipedia is currently the closest achievement of this vision. Curiously, he made no mention whatsoever of WikiSpecies, the Wikimedia Foundation's project whose goal is more closely tailored to that vision.

In his article on Israel-Lebanon conflict on Wikipedia and Wikinews, Hardy (2007) argued that while "Wikipedia is not a threat to the peer reviewed publications of academia, it is definitely a competitor, not only by dint of the number of people who consult it, but the quality of some of the articles in their own right" (2007, p.22). We believe, however, that this assessment indicates a misunderstanding of the role of Wikipedia—it is an encyclopedia that only documents accepted knowledge, and explicitly excludes the publication of original research from its mission. Thus, we believe there is no competition between these complementary roles.

Other scholars were more cautious in their appraisal of Wikipedia, and pointed out some challenges that its use presents scholars. In her paper about conducting art history research, Chen (2009) recommended caution in using Wikipedia: "the content of the Wikipedia article can be used as tips for possible approaches to this object, but not as a source for the actual paper" (2009, p.123). Knapp (2008) discussed the challenges that arise from citing amorphous web sources such as Wikipedia, where the source materials are constantly changing. She explored possible resolutions, such as a "virtual bookshelf" which includes electronic attachments of source materials along with a published work.

In addition to these studies, those described elsewhere in this review (especially in the section on Scholarly Contribution) also discuss aspects of scholars using Wikipedia as a knowledge source (Davis et al. 2010; Huss et al. 2010; Bateman & Logan 2010; Bond 2011; Kubiszewski et al. 2011).

Source for Librarians

Several studies discussed issues particularly pertinent to librarians. We normally describe such articles in other categories if they are more specific about the focus of the article, but here we discuss those that are essentially focused on librarianship. (Note that in our WikiLit website, we do identify all library-related research in the "Library science" domain—we identified over 30 such studies.) These studies unanimously called on librarians to consider Wikipedia a positive phenomenon, and to take advantage of it.

Some studies considered the use of Wikipedia inevitable, and so called on librarians to thus embrace it and even use it as an opportunity to teach information literacy. Choolhun (2009) documents that Wikipedia is increasingly being used as the first source for legal information inquiries by lawyers and law students. She thus calls for increased engagement with Web 2.0 use by legal librarians. Gunnels and Sisson (2009) cautioned against avoiding Wikipedia and other Web 2.0 tools for research, but urged instead teaching students to be critical about the information found on these sources and how to validate them through reliable sources.

A few studies went further to regard Wikipedia as a unique opportunity to promote libraries and librarianship, and to spearhead their relevance to the forefront of the information age. Belden (2008) presented a case of how a university library was able to gain "dramatic increases in Web usage and reference requests by harnessing the power of social networks such as Wikipedia and MySpace" (2008, p.99). She explained that sites such as Wikipedia "provide the tools to allow dynamic, interactive means of sharing information and helping connect the dots," and that the skills needed in these activities are "the very abilities that librarians and scholars hope to inculcate in our educational endeavors" (2008, p.110). Luyt et al. (2010) found that many librarians view Wikipedia as an opportunity for the profession rather than a threat. They saw this encouraging result as an opportunity to connect with non-Western users and content, and also as an opportunity for librarians to forge a leading role in the emerging information society. Jacobs (2009) commented on Hahn's (2009) study of student use of Wikipedia on iPods (described in the section on Cross-Domain Student Readership). She examined how new information

technologies like Wikipedia influence librarians and academic libraries, suggesting that librarians engage in and promote the new technologies to the library world.

News Source

Wikipedia has been investigated in its role as a source of current news. In one of the earliest academic studies of Wikipedia, Lih (2004) gave an introductory sketch of the then three-year-old endeavor. He described Wikipedia as "the largest form of participatory journalism to date" (2004, p.1), and demonstrated that the average size of Wikipedia articles and the number of article edits had been steadily increasing during those years. He analyzed Wikipedia articles cited in the news in a thirteen-month period to compare their quality before and after citation in the press. Although his quality measures were rudimentary computations based on the numbers of edits and of contributors, his analysis did show that increases in his quality measure were correlated with citation in the press (though he did not demonstrate that this increase was not due to the timeliness of the events rather than due to the press citation).

Wikipedia was used as one of many sources of documentation on details of the delayed response in 2005 to Hurricane Katrina in the United States (Chua et al. 2007). Thelwall and Stuart (2007) found that "Web 2.0 resources such as Wikinews, the Wikipedia, and the Flickr picture sharing site" (2007, p.523) are important secondary sources of information provision and sharing in the event of natural disasters and similar crises. Nonetheless, traditional mass media remain the predominant sources of information.

Some studies have examined how newspapers frame Wikipedia and use it as a source. Shaw (2008) reported that *Philadelphia Inquirer* instructs journalists never to use Wikipedia "to verify facts or to augment information in a story," and that one reporter complained: "there is no way for me to verify the information without fact-checking, in which case it isn't really saving me any time." Some other news organizations, such as *Los Angeles Times*, do occasionally permit citation of Wikipedia as a source. Messner and South (2010) reported that although newspapers have not referenced Wikipedia very much in the past, their reliance on this source has recently been increasing, and they tend to present it as a generally accurate source.

Other Knowledge Source Topics

In addition to the topics described above, there are some articles concerning readers' use of Wikipedia as a source for knowledge that do not fall under any of the other knowledge source topics.

Koolen et al. (2009) found that high-frequency Web search queries often directly relate to Wikipedia pages. Within a large sample of web queries, 38% exactly matched the title of a Wikipedia page. The content and context of the matched Wikipedia page could then be used to expand the query. Wikipedia pages can also form an intermediary between a user query and a collection of books being searched.

Other articles that touch on other knowledge source topics also relate to other topics, and so are described elsewhere in this review (Langlois & Elmer 2009; A. Rubin & E. Rubin 2010).

Ranking and Popularity

This category includes studies that compared the use of Wikipedia with other knowledge sources for getting information, as well as studies that investigated the popularity of topics within Wikipedia. These studies have consistently confirmed that Wikipedia is a premier source of knowledge on the Internet.

Some studies compared Wikipedia's ranking with that of other important websites. Höchstötter and Lewandoski (2009) compared search results of four major search engines: Google, Yahoo, Live.com and Ask. They found that Wikipedia is the most frequently represented website in all search engines. However, there are some differences in how different search engines rank Wikipedia pages: "Yahoo and MSN place the most Wikipedia results on their results pages. Google boost Wikipedia result mostly on first position but shows less Wikipedia links in total [*sic*]" (2009, p.1810). Lewandowski and Spree

(2011) noted that Wikipedia results shown on search engines are quite dependent on the quality of articles.

A few studies examined what is popular on Wikipedia. Ratkiewicz et al. (2010) provided a quantitative analysis of the dynamics of online popularity of Wikipedia content. They found that the dynamics of popularity are characterized by "bursts, displaying characteristic features of critical systems such as fat-tailed distributions of magnitude and inter-event time" (2010, p.1). Spoerri (2007b) examined which were the most popular articles and topics on Wikipedia. He found that over half of the most visited pages are related to entertainment and sexuality, that popularity of Wikipedia pages is related to search behavior on the Web, and that search engines—especially Google—fuel Wikipedia's growth, and thus shape what is popular on Wikipedia. He also examined the 100 most visited Wikipedia articles for five consecutive months, finding that 40% of these—mostly related to sexuality and entertainment—were highly visited in all five months, and 25% were highly visited only in a single month (2007a). Waller (2011) investigated the search queries that directed Australians to Wikipedia pages, and found that they search more for lighter topics such as entertainment rather than for more serious information.

Bar-Ilan (2006) studied a case of "Google-bombing," where the top results to the search keyword "Jew" yield the Wikipedia article and an anti-Semitic website. She observed that the Google ranking is primarily due, not to pages that actually discuss these two websites, but rather to links from discussions and blogs purposely inserted to influence search engine rankings.

Some articles compare Wikipedia's ranking with other sources of health information; we discuss these in the topic "Health Information Source" (P. T. Johnson et al. 2008; Mühlhauser & Oser 2008; Michael R Laurent & Vickers 2009). In addition another article that dealt with issues related to Wikipedia's ranking and popularity is described elsewhere in this review (Langlois & Elmer 2009)..

Reader Perceptions of Credibility

Some studies examined the credibility of Wikipedia from its readers' perspective. We categorize articles here that examined readers' perceptions of credibility without attempting some kind of objective evaluation of reliability, such as by subject experts (we classify those in the Reliability section); this is also distinct from Wikipedians' inside-view perceptions of credibility of the articles they create (which we classify as Contributor Perceptions of Credibility).

Some studies examined characteristics of articles' presentation that affect readers' perceptions of their credibility. Veltman (2005) argued that access to the entirety of knowledge is becoming feasible with the open source movement. She highlighted the central importance of quality along with open access in terms of quantity. She proposed techniques for presenting knowledge on the Internet that facilitate readers' rapid assessment of its credibility. Kubiszewski et al. (2011) performed an experiment on "whether certain webpage characteristics affect academics' and students' perception of the credibility of information presented in an online article" (2011, p.659). They concluded that "compared to Encyclopedia Britannica, article information appearing in both Encyclopedia of Earth and Wikipedia is perceived as significantly less credible" (2011, p.664). They also found that the appearance of a biased sponsor lowered credibility.

Chen (2009) examined how information technology professionals used Wikipedia information for work-related purposes. He found that they treat Wikipedia as a ready reference for general information, but did not consider it sufficiently developed for professional use. They considered that Wikipedia needs to improve its contribution and editorial process in order to raise its quality.

A few studies have noted and suggested various means for rapidly estimating the quality of an article, usually through the observations of reliable proxies. Blumenstock (2008) found that the word count of an article performs surprisingly well as a predictor for article quality, at least when distinguishing between featured and random articles, with an error rate of around 96% on a corpus of 1,554 featured and 9,513 randomly selected articles. He suggested setting a cut-off at 2,000 words between the two sets. Cross

(2006) offered a text-colorizing software as "a visual cue that enables [users] to see what assertions in an article have ... survived the scrutiny of a large number of people, and what assertions are relatively fresh, and may not be as reliable." The number of editors having an article on their watch list could possibly also make a good indicator of the article quality. However, this number has not been available to researchers.

Although there are very many articles that discussed readers' perceptions of Wikipedia's credibility, most such articles normally treated other subjects more substantially; thus, we describe them elsewhere in this review (S. Lim & Kwon 2010; S. Lim 2009; Sundin & Francke 2009; Messner & South 2010; Page 2010; Calkins & Kelley 2009; Kaplan & Haenlein 2010; H. Chen 2010; Sanger 2009; Magnus 2009; Luyt et al. 2010; Luyt, Zainal, et al. 2008; Eijkman 2010; Dooley 2010; Head & Eisenberg 2010; H. Zeng et al. 2006).

Software for Readership

A number of studies investigated software specifically developed to help Wikipedia readers. Some of these are targeted to alerting Wikipedia readers to the trustworthiness of articles, and others are designed to enhancing external content by automatically identifying relevant Wikipedia content.

Computational Estimation of Reliability

A number of studies developed computational methods for estimating the reliability of articles, mainly to help readers assess whether articles were trustworthy. We discuss these articles here since they are directed to readers. This is distinct from the Reliability topic, in which human experts assess the quality of Wikipedia articles.

Zeng et al. (2006) developed a method to predict trustworthiness of Wikipedia articles based on the revision history of the articles, validated using featured articles. They concluded that Wikipedia is generally trustworthy, and that visualizations of article trustworthiness can enable users to access the more trustworthy versions of the articles and to avoid vandalism and malicious content. They also designed and implemented a trust management layer for collaborative information repositories in general, and Wikipedia in particular (McGuinness et al. 2006).

Korfiatis et al. (2006) investigated the development of quality articles in Wikipedia by using social network analysis to determine the authoritativeness of articles. They developed an approach to calculating social network measures such as centrality. They used a Web crawler (before these software agents were banned on Wikipedia because of the excess server load they cause). They argued that as Wikipedia keeps growing, it will be more challenging to keep the content reliable.

Hu et al. (2007) proposed three models for assessing quality of Wikipedia articles based on the interaction data between articles and their contributors. Adding article length to their model could also improve model performance.

Dondio and Barrett (2007) developed a method to predict trustworthiness of Wikipedia articles using computational trust techniques and domain-specific analysis; then they validated their method by differentiating featured articles from others using the method.

The Wikiganda, formerly available from www.wikiwatcher.com, used automated text analysis to detect biased edits (2009). The opinion mining and sentiment analysis technique uses a lexicon of over 20,000 words from General Inquirer and Wiebe wordlists so that each revision can get a Propaganda Score labeled as negative, positive or "vague" propaganda. In conjunction with the WikiTrust system and evaluated against 200 manually labeled revisions, the system showed a precision/recall performance of 52%/63%.

In addition to these, Adler et al. (2008) discussed reader-oriented reliability algorithms. However, since this is integrated with their contributor-targeted system, we describe that study in the section on Reputation Systems.

Reading Support

Encountering knowledge gaps while reading is an issue that people face daily. This problem motivated several studies to develop reading support tools using Wikipedia to fill these gaps. In other words, Wikipedia articles were extracted to fill the missing information resulting from knowledge gaps.

Jordan and Watters (2009) designed a prototype to bring up the single most relevant Wikipedia article when a user selects part of a text. The most successful model could accurately find the best article in 70% of the cases and help readers to fill the gap in their personal knowledge using Wikipedia articles. As an application for the previous study, Jordan (2009) proposed a system to help people reading academic abstracts to be able to highlight part of them, and then a pop-up would appear with a single Wikipedia article explaining the highlighted part. The system tries to suggest the most related article based on understanding the context of the abstract and article categories.

With the increase of blogs and social networks comes the need for a support system to fill the content holes. Nadamoto et al. (2010) suggested a new method to search for these content holes, defined as "the user's unawareness of information." Wikipedia articles were used to extract and present the holes in community-type content. Their proposed method differs from otherwise related information retrieval tasks in that it searches for different information instead of similar one.

Student Readership

Many studies investigated how students use Wikipedia, both as a general source of information and in student projects where they were assigned work that explicitly involved reading Wikipedia articles. As with all research fields where survey and experimental research is common, students are a common subject for Wikipedia research. On one hand, they are a demographic of young and intelligent people, and so are highly relevant for forward-looking topics of inquiry, such as the use of Web 2.0 resources like Wikipedia. On the other hand, they provide a convenience sample, since professors can easily persuade their students to participate in studies, either purely voluntarily or for course credit, if some learning component is incorporated into the exercise.

These studies treated students in secondary school, undergraduate and post-graduate education. Articles we classify here mainly involve student information literacy in critically reading and using information from Wikipedia articles. In contrast, we classify articles concerning projects where students are assigned to contribute to and develop Wikipedia articles as Student Contribution, which we discuss separately. Since all Student Contribution activities necessarily require students to read Wikipedia, we discuss here only those articles that have no substantial component involving the students contributing content to Wikipedia. We categorize the following kinds of student readership articles: those concerning citing Wikipedia; those that treated students and Wikipedia articles across various domains of knowledge or field of study; those that are restricted to a specific domain; and those that deal with general matters of student literacy.

Cross-Domain Student Readership

Many articles treated student readership in general, regardless of their domain of knowledge or field of study. Some of these generally investigated how students use Wikipedia.

One of the perennially controversial questions about Wikipedia is whether or not it should be allowed as a citation, especially for student work. In early 2007, a department at Middlebury College decided to hold students responsible for using Wikipedia as a source after a batch of students had used erroneous information on Wikipedia about certain topics in the history of Japan (Waters 2007). Media reports

implied that the department of Neil Waters, the teacher of the class involved, "was at war with Wikipedia itself." However, Waters himself actually told students "that Wikipedia is a fine place to search for a paper topic or begin the research process." The department adopted the following policy:

> Whereas Wikipedia is extraordinarily convenient and, for some general purposes, extremely useful, it nonetheless suffers inevitably from inaccuracies deriving in large measure from its unique manner of compilation. … Students are responsible for the accuracy of information they provide, and they cannot point to Wikipedia or any similar source that may appear in the future to escape the consequences of errors. (Read 2007)

This policy is actually in line with the opinion of the Wikimedia Foundation. However, Jimmy Wales, founder of Wikipedia, later said that he saw no problem in younger students using Wikipedia as a reference, and that it should be used as a stepping stone to other sources (Coleman 2007).

Tann and Sanderson (2009) examined the web-based information-seeking practices of university students. They found that many queries that have been previously considered informational by past research have since taken on a more navigational nature. Moreover, IMDb and Wikipedia have both accumulated a sufficient level of information to address the users' information needs. Head and Eisenberg (2010) investigated how college students use Wikipedia for course-related research. They discovered that college students do use Wikipedia, but are aware of its limitations in credibility and depth. It is more often used in the initial stages of research to obtain background information, and then is complemented with scholarly resources. Students mainly appreciate Wikipedia for its coverage, currency, comprehensibility, and convenience. Luyt et al. (2008) interviewed young people concerning their perception and use of Wikipedia. They found that Wikipedia played only a minor role in the lives of their interviewees, but these young people were quite aware of its drawbacks when they did use it. The researchers thus concluded that common concerns about Wikipedia's negative effects on young people are exaggerated. However, only 15 subjects were interviewed, apparently all in Singapore.

Maehre (2009) explored various pedagogical principles to encourage instructors to allow the use of Wikipedia in their students' projects. He argued in favor of producing and engaging in information creation. Moreover, Maehre promoted the focus on the content of a resource rather than the credibility of authors. According to Maehre, the world outside of universities and colleges classrooms is an interactive world. Thus, educators should work towards having information creators rather than information readers or finders.

In two unique studies, Hahn (Hahn 2009; Hahn 2010) observed that undergraduates running the iPod and iPod touch Wikipedia app mainly search for recreational and short factual information. However, they were all satisfied with the experience and found it useful for preparing a research paper.

Other articles related to cross-domain student readership are described elsewhere in this review (Eijkman 2010; Rand 2010; Jennings 2008; Gunnels 2007; Harouni 2009; Patch 2010; Sundin & Francke 2009; Kubiszewski et al. 2011).

Domain-Specific Student Readership

A number of articles focused on student use of Wikipedia articles within a specific knowledge domain. Many of these considered usage by medical students, who are being trained to make life-and-death decisions based on their evaluation of information. Wedemeyer et al. (2008) asked students to evaluate Wikipedia biochemistry articles. One third responded that they never use Wikipedia. Among the remaining two thirds, 12% used Wikipedia as their primary source and 31% used their textbook and Wikipedia equally. The remaining 57% used Wikipedia only as a supplement. The majority of the students preferred Wikipedia to the textbook. Judd and Kennedy (Judd & G. Kennedy 2009; Judd & G. Kennedy 2010) studied Australian biomedical students' on-campus use of internet web sites. Wikipedia's use increased "from only 2% of sessions in 2005 to 16% in 2008 and 2009" (2010, p.1568). They concluded, among other issues, that students "are increasingly reliant on generalist information retrieval

tools, particularly Google and Wikipedia, to support their learning activities" (2010, p.1570). Fiore (2011) found that 47% of 186 medical students who recently completed psychiatric clinical clerkship used Wikipedia as one of the primary sources for preparing for psychiatry exams. Question books (88%) and the peer-reviewed website Up-to-Date (59%) were more frequently used, but textbooks (10%) less used. Among the students using Wikipedia, 84% also used question books.

Some researchers examined how and why journalism and mass communication university students use Wikipedia. Lim (2009) affirmed many other research findings that although students commonly use Wikipedia for finding background information with acceptable qualities, they are fully aware of its quality issues and do not use it blindly. However, they do not verify the information on Wikipedia, but rather use its sources and links to get further information. Lim and Kwon (2010) compared student usage of Wikipedia by gender. They found that while male students used Wikipedia more frequently and had a positive attitude towards it, "female students displayed more cautious or conservative attitudes, emotions, and behaviors."

Other articles related to domain-specific student readership are described elsewhere in this review (Korosec et al. 2010; Jancarik & Jancarikova 2010; J. Aycock & A. Aycock 2008; Schweitzer 2008; Lavsa et al. 2011; Haigh 2010).

Student Information Literacy

A number of studies considered student information literacy. These studies unanimously called on teachers and professors to embrace rather than ban Wikipedia, urging them to seize the opportunity to educate students in information literacy skills needed for the 21st century.

Some studies discussed information literacy in general. Rand (2010) argued that students can learn critical thinking skills using Wikipedia. Jennings (2008) argued the necessity of information literacy skills for 21st century students to become lifelong learners in using all information resources. He highlighted the importance of Wikipedia, as it facilitates teaching and learning such skills. Gunnels (2007) proposed using Wikipedia as a starting point towards a new way of teaching information literacy skills which enhances the quality of both the users and creators of information resources. Judd and Kennedy (2009) looked into how medical students use Wikipedia and other online resources to acquire their needed information. They concluded that higher emphasis on information literacy skills training is required to make sure students are able to locate and use the best available information.

Some studies described teaching experiences engaging Wikipedia for information literacy. Harouni (2009) reported using Wikipedia in a literacy class to teach students critical reading skills. After the lessons, students could clearly articulate their reference choices, and they were able to discern and use more comprehensive and unbiased Wikipedia articles. Patch (2010) argued that as students are already using Wikipedia, writing teachers should follow suit and incorporate Wikipedia into their teaching. She described some of her experiences with students and Wikipedia, and argued that "many students are 'underprepared' to consume and use online texts responsibly" (2010, p.282). She concluded that by employing Wikipedia, students can "have an easier time making the leap to higher-level inquiry and responsible scholarship" (2010, p.282). In sharing their respective experiences in teaching about Wikipedia in computer science and anthropology courses, Aycock and Aycock (2008) proposed using Wikipedia to teach students not only about the use and interpretation of information resources, but also about management of rapidly changing collaborative information resources.

Two studies featured in-depth investigation of how students handle Wikipedia information. Calkins and Kelley (2009) examined history students' perception of Wikipedia credibility and collaborative work. Although students are aware of factual errors in Wikipedia, they nonetheless believe that it is getting better as more people contribute and correct errors; they are mostly in favor of accuracy and collaboration on Wikipedia. Sundin and Francke (2009) carefully investigated how secondary school students negotiate the credibility of information in their learning process. Although these students used Wikipedia

information, they were uncertain about its credibility because they employed traditional methods for credibility assessment based on authorship and origin, neither of which is clear in Wikipedia. Sundin and Fracke suggested that an update to these methods is needed.

Other articles related to student information literacy are described elsewhere in this review (Choolhun 2009; Gunnels & Sisson 2009; Chandler-Olcott 2009).

Conclusion

Wikipedia is a collaborative ecosystem, in which user participation results in content of considerable quality and quantity. Whenever there is content that someone finds interesting, there will be people to read that content. Whenever there are enough readers, some of them will take the next step to becoming participants in content production. The more there are participants, the more content will be produced, which will reach even more readers. In this way participation, content, and readership form an ongoing cycle. All this is made possible by an infrastructure of software, hardware, and human capabilities. The infrastructure enables creation, mediation, and archiving of information. The better the infrastructure, the better the cycle of participation, content and readership can revolve.

Beyond these uses, Wikipedia's abundance of open and free content has enabled the site to be used for various kinds of research purposes, such as information retrieval, that take advantage of its huge datasets.

Wikipedia's phenomenal success has attracted the interest of scholars who desire to understand the inner workings of this exemplary open content application. Much of this research can prove valuable in guiding Wikipedia governance and development, on improving policies and best practices to improve the quality, performance, and overall value of Wikipedia.

We believe that scholarly research is a critical contributor to thoroughly understanding the workings of Wikipedia, an important and widely used global resource. This study helps bring this understanding to the people who actively construct Wikipedia, enabling them to leverage this valuable knowledge base. This systematic review helps make sense of the varied research that has been done to date, and set directions for future research on the fascinating and important phenomenon that is Wikipedia.

Acknowledgments

We want to note that all five co-authors were intensely involved in this project, and each one of us spent hundreds of hours on its execution. We would thus ask that all five co-authors be listed whenever this paper is cited; if that is impractical, then please refer to the paper as "the WikiLit review" in lieu of using an "et al." formulation.

We thank Kira Schabram for her invaluable assistance in developing the systematic literature review methodology used (Okoli & Schabram 2010), and in conducting the pilot study (Okoli & Schabram 2009a). We thank Bilal Abdul Kader for his assistance in the pilot study (Okoli et al. 2009). We thank Richard Wong for his assistance in collecting author data.

We thank Emilio J. Rodríguez-Posada (emijrp) for his model WikiPapers site, upon which much of WikiLit was based. We thank the innumerable researchers on the wiki-research-l and authors of included studies for their many comments and revisions, and for their wikified peer-review. In particular, we thank John Vandenberg, Ivan Lanin, Oren Bochman, Oren Bochman, Ofer Arazy, Jens Lehmann, Diomidis Spinellis, Pierre Lindenbaum, Stefano Mizzaro, Andrew Krizhanovsky, Igor Nikolic, Roderic D. M. Page, Carlos Castillo, Howard T. Welser, Cullen J. Chandler, Ziko van Dijk, Shane Greenstein, Alexander Halavais, Piotr Konieczny, Angela Rand, Heiko Haller, Chen Mei Hui, Dmitry Lizorkin, Andrew G. West, Robert Neal Baxter, Carolyn Watters, as well as many anonymous contributors to the

WikiLit website. We also thank Daniel Kinzler, Torsten Zesch, and Felipe Ortega for pointing to references and tools.

Earlier versions of this study have been previously published in a conference (Okoli 2009) and as a working paper (Nielsen 2012). The protocol for this study was presented in a conference (Okoli & Schabram 2009b), published as a working paper (Okoli & Schabram 2009a), and discussed in a workshop (Lanamäki et al. 2011).

This study was funded by the Social Sciences and Humanities Research Council of Canada; the Lundbeck Foundation Center for Integrated Molecular Brain Imaging (CIMBI); the Concordia University Aid to Scholarly Activity fund; and the Danish Council for Strategic Research through the Responsible Business in the Blogosphere project.

References

Adafre, S.F. & Rijke, M. de, 2005. Discovering missing links in Wikipedia. In *Proceedings of the 3rd international workshop on Link discovery*. New York, NY, USA: ACM, pp. 90 – 97. Available at: http://0-portal.acm.org.mercury.concordia.ca/citation.cfm?id=1134271.1134284&coll=DL&dl=GUIDE&CFID=112025803&CFTOKEN=32336862&preflayout=flat [Accessed November 21, 2010].

Adamic, L.A. et al., 2010. Individual focus and knowledge contribution. *First Monday*, 15(3), p.14 pp.

Adar, E., Skinner, M. & Weld, D.S., 2009. Information arbitrage across multi-lingual Wikipedia. In *2nd ACM International Conference on Web Search and Data Mining, WSDM'09, February 9, 2009 - February 12, 2009*. Proceedings of the 2nd ACM International Conference on Web Search and Data Mining, WSDM'09. Barcelona, Spain: Association for Computing Machinery, pp. 94–103. Available at: http://dx.doi.org/10.1145/1498759.1498813 [Accessed November 6, 2010].

Adler, B.T. et al., 2008. Measuring Author Contribution to the Wikipedia. In *Proceedings of WikiSym 2008*. ACM. Available at: http://www.soe.ucsc.edu/~luca/papers/08/wikisym08-users.pdf.

Adler, B.T. & Alfaro, L. de, 2007. A content-driven reputation system for the wikipedia. In *Proceedings of the 16th international conference on World Wide Web*. New York, NY, USA: ACM, pp. 261 – 270. Available at: http://0-portal.acm.org.mercury.concordia.ca/citation.cfm?id=1242572.1242608&coll=DL&dl=GUIDE&CFID=112015986&CFTOKEN=43703661&preflayout=flat [Accessed November 21, 2010].

Ah-Pine, J. et al., 2009. Crossing textual and visual content in different application scenarios. *Multimedia Tools and Applications*, 42(1), pp.31–56.

Almeida, R.B., Mozafari, B. & Cho, J., 2007. On the evolution of wikipedia. In *International Conference on Weblogs and Social Media*. Available at: http://www.icwsm.org/papers/2--Almeida-Mozafari-Cho.pdf [Accessed October 15, 2012].

Altmann, U., 2005. Representation of Medical Informatics in the Wikipedia and its Perspectives. *Studies in Health Technology and Informatics*, 116, pp.755–760.

Amichai-Hamburger, Y. et al., 2008. Personality Characteristics of Wikipedia Members. *CyberPsychology & Behavior*, 11(6), pp.679–681.

Anthony, D., Smith, S.W. & Williamson, T., 2009. Reputation and Reliability in Collective Goods. *Rationality and Society*, 21(3), pp.283–306.

Antin, J., 2010. *Social operational information, competence, and participation in online collective action*. United States -- California: University of California, Berkeley. Available at: http://proquest.umi.com/pqdweb?did=2125353481&Fmt=7&clientId=10306&RQT=309&VName=PQD.

Antin, J. & Cheshire, C., 2010. Readers are not free-riders: reading as a form of participation on wikipedia. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*. CSCW '10. New York, NY, USA: ACM, pp. 127–130. Available at: http://doi.acm.org/10.1145/1718918.1718942 [Accessed October 22, 2012].

Arazy, O. et al., 2011. Information Quality in Wikipedia: The Effects of Group Composition and Task Conflict. *Journal of Management Information Systems*, Forthcoming.

Arazy, O. et al., 2010. Recognizing Contributions in Wikis: Authorship Categories, Algorithms, and Visualizations. *Journal of the American Society for Information Science and Technology*, 61(6), pp.1166–1179.

Arazy, O. et al., 2009. Wiki Deployment in Corporate Settings. *IEEE Technology & Society Magazine*, 28(2), pp.57–64.

Auer, S. et al., 2007. DBpedia: A nucleus for a Web of open data. In *6th International Semantic Web Conference, ISWC 2007 and 2nd Asian Semantic Web Conference, ASWC 2007, November 11, 2007 - November 15, 2007*. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Busan, Korea, Republic of: Springer Verlag, pp. 722–735. Available at: http://dx.doi.org/10.1007/978-3-540-76298-0_52 [Accessed November 6, 2010].

Auray, N., Poudat, C. & Pons, P., 2007. Democratizing scientific vulgarization. The balance between cooperation and conflict in french Wikipedia. *Observatorio (OBS*)*, 1(3). Available at: http://obs.obercom.pt/index.php/obs/article/view/152 [Accessed January 18, 2011].

Aycock, J. & Aycock, A., 2008. Why I Love/Hate Wikipedia: Reflections upon (Not Quite) Subjugated Knowledges. *Journal of the Scholarship of Teaching and Learning*, 8.

Ayers, P., 2006. Researching wikipedia - current approaches and new directions. *Proceedings of the American Society for Information Science and Technology*, 43(1), pp.1–14.

Ayers, P. & Priedhorsky, R., 2011. WikiLit: collecting the wiki and Wikipedia literature. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration*. WikiSym '11. New York, NY, USA: ACM, pp. 229–230. Available at: http://doi.acm.org/10.1145/2038558.2038612 [Accessed October 23, 2012].

Bai, B. et al., 2010. Learning to rank with (a lot of) word features. *Information Retrieval*, 13(3), pp.291–314.

Balasuriya, D. et al., 2009. Named entity recognition in Wikipedia. In *Proceedings of the 2009 Workshop on the People's Web Meets NLP*. Association for Computational Linguistics, pp. 10–18.

Banchuen, T., 2008. *The geographical analog engine: Hybrid numeric and semantic similarity measures for U.S. cities*. United States -- Pennsylvania: The Pennsylvania State University. Available at: http://proquest.umi.com/pqdweb?did=1637577661&Fmt=7&clientId=10306&RQT=309&VName=PQD.

Banerjee, S., Ramanathan, K. & Gupta, A., 2007. Clustering short texts using wikipedia. In *30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR'07, July 23, 2007 - July 27, 2007*. Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR'07. Amsterdam, Netherlands: Association for Computing Machinery, pp. 787–788. Available at: http://dx.doi.org/10.1145/1277741.1277909 [Accessed November 5, 2010].

Bar-Ilan, J., 2006. Web links and search engine ranking: the case of Google and the query "Jew." *Journal of the American Society for Information Science and Technology*, 57(12), pp.1581–1589.

Bast, H. et al., 2007. ESTER: Efficient search on text, entities, and relations. In *30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR'07, July 23, 2007 - July 27, 2007*. Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR'07. Amsterdam, Netherlands: Association for Computing Machinery, pp. 671–678. Available at: http://dx.doi.org/10.1145/1277741.1277856 [Accessed November 6, 2010].

Bateman, A. & Logan, D.W., 2010. Time to underpin Wikipedia wisdom. *Nature*, 468, p.765.

Baxter, R.N., 2009. New technologies and terminological pressure in lesser-used languages: The Breton Wikipedia, from terminology consumer to potential terminology provider. *Language Problems and Language Planning*, 33(1), pp.60–80.

Baytiyeh, H. & Pfaffman, J., 2010. Volunteers in Wikipedia: Why the Community Matters. *Educational Technology & Society*, 13(2), pp.128–40.

Beer, D., 2008. Making Friends with Jarvis Cocker: Music Culture in the Context of Web 2.0. *CULTURAL SOCIOLOGY*, 2(2), pp.222–241.

Bekker-Nielsen, T., 2011. Historie på Wikipedia. *Noter*, 188, pp.48–52.

Belden, D., 2008. Harnessing Social Networks to Connect with Audiences : If You Build It, Will They Come 2.0? *Internet Reference Services Quarterly*, 13(1), pp.99–111.

Benkler, Y. & Nissenbaum, H., 2006. Commons-based Peer Production and Virtue. *The Journal fo Political Philosophy*, 14(4), pp.394–419.

den Besten, M. & Dalle, J.-M., 2008. Keep it simple: A companion for simple wikipedia? *Industry and Innovation*, 15(2), pp.169–178.

Bhole, A. et al., 2007. Extracting named entities and relating them over time based on Wikipedia. *Informatica (Ljubljana)*, 31(4), pp.463–468.

Bizer, C. et al., 2009. DBpedia - A crystallization point for the Web of Data. *Journal of Web Semantics*, 7(3), pp.154–165.

Black, E.W., 2008. Wikipedia and academic peer review : Wikipedia as a recognised medium for scholarly publication? *Online Information Review*, 32(1), pp.73–88.

Black, L.W. et al., 2011. Self-Governance Through Group Discussion in Wikipedia Measuring Deliberation in Online Groups. *Small Group Research*, 42(5), pp.595–634.

Blumenstock, J.E., 2008. Size Matters: Word Count as a Measure of Quality on Wikipedia. In *Proceedings of the 17th International World Wide Web Conference (WWW2008). April 21-25, 2008. Beijing, China*.

Bond, A.L., 2011. *Why ornithologists should embrace and contribute to Wikipedia*,

Bragues, G., 2009. Wiki-Philosophizing in a Marketplace of Ideas: Evaluating Wikipedia's Entries on Seven Great Minds. *MediaTropes eJournal*, 2(1), pp.117–158.

Brandes, U. & Lerner, J., 2008. Visual Analysis of Controversy in User-Generated Encyclopedias. *Information Visualization*, 7(1), pp.34–48.

Breinholt, J., 2008. *The Wikipediazation of the American Judiciary*, Available at: http://nefafoundation.org/miscellaneous/justiceforall/nefawikipedia0108.pdf.

Brereton, P. et al., 2007. Lessons from applying the systematic literature review process within the software engineering domain. *Journal of Systems and Software*, 80(4), pp.571–583.

Brohée, S., Barriot, R. & Moreau, Y., 2010. Biological knowledge bases using Wikis: combining the flexibility of Wikis with the structure of databases. *Bioinformatics*.

Brown, J.J., 2009. Essjay's "Ethos": Rethinking Textual Origins and Intellectual Property. *College Composition and Communication*, 61.

Brown, J.J., 2008. From Friday to Sunday: the hacker ethic and shifting notions of labour, leisure and intellectual property. *Leisure Studies*, 27(4), pp.395–409.

Bryant, S.L., Forte, A. & Bruckman, A., 2005. Becoming Wikipedian: Transformation of participation in a collaborative online encyclopedia. In *2005 International ACM SIGGROUP Conference on Supporting Group Work, GROUP'05, November 6, 2005 - November 9, 2005*. Proceedings of the International ACM SIGGROUP Conference on Supporting Group Work. Sanibel Island, FL, United states: Association for Computing Machinery, pp. 1–10. Available at: http://dx.doi.org/10.1145/1099203.1099205 [Accessed November 5, 2010].

Bunescu, R., 2007. *Learning for information extraction: From named entity recognition and disambiguation to relation extraction*. United States -- Texas: The University of Texas at Austin. Available at: http://proquest.umi.com/pqdweb?did=1375541921&Fmt=7&clientId=10306&RQT=309&VName=PQD.

Buntine, W., 2005. Static Ranking of Web Pages, and Related Ideas. In M. Beigbeder & W. G. Yee, eds. *Open Source Web Information Retrieval*. Available at: http://www.emse.fr/OSWIR05/2005-oswir-p23-buntine.pdf.

Buriol, Luciana S. et al., 2006. Temporal Analysis of the Wikigraph. In *Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence*. pp. 45–51. Available at: http://0-portal.acm.org.mercury.concordia.ca/citation.cfm?id=1248823.1249047&coll=DL&dl=GUIDE&CFID=112020990&CFTOKEN=50968312&preflayout=flat [Accessed November 21, 2010].

Butler, B., Joyce, E. & Pike, J., 2008. Don't look now, but we've created a bureaucracy: The nature and roles of policies and rules in Wikipedia. In *26th Annual CHI Conference on Human Factors in Computing Systems, CHI 2008, April 5, 2008 - April 10, 2008*. Conference on Human Factors in Computing Systems - Proceedings. Florence, Italy: Association for Computing Machinery, pp. 1101–1110. Available at: http://dx.doi.org/10.1145/1357054.1357227 [Accessed November 6, 2010].

Buzzi, M. & Leporini, B., 2009. Editing Wikipedia content by screen reader: easier interaction with the Accessible Rich Internet Applications suite. *Disability and Rehabilitation. Assistive Technology*, 4(4), pp.264–275.

Caddick, S., 2006. Wiki and other ways to share learning online. *Nature*, 442, p.744.

Calkins, S. & Kelley, M.R., 2009. Who Writes the Past? Student Perceptions of Wikipedia Knowledge and Credibility in a World History Classroom. *Journal on Excellence in College Teaching*, 20.

Callis, K.L. et al., 2009. Improving Wikipedia: educational opportunity and professional responsibility. *Trends in Ecology & Evolution*, 24(4), pp.177–179.

Cantador, I., Konstas, I. & Jose, J.M., 2011. Categorising social tags to improve folksonomy-based recommendations. *Web Semantics: Science, Services and Agents on the World Wide Web*, 9(1), pp.1–15.

Capocci, A. et al., 2006. Preferential attachment in the growth of social networks: The internet encyclopedia Wikipedia. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 74(3), p.036116.

Capocci, A., Rao, F. & Caldarelli, G., 2008. Taxonomy and clustering in collaborative systems: the case of the on-line encyclopedia wikipedia. *Europhysics Letters*, 81(2), pp.28006–1.

Carillo, K. & Okoli, C., 2011. Generating quality open content: A functional group perspective based on the time, interaction, and performance theory. *Information & Management*, 48(6), pp.208–219.

Carmel, D., Roitman, H. & Zwerdling, N., 2009. Enhancing cluster labeling using wikipedia. In *32nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2009, July 19, 2009 - July 23, 2009*. Proceedings - 32nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2009. Boston, MA, United states: Association for Computing Machinery, pp. 139–146. Available at: http://dx.doi.org/10.1145/1571941.1571967 [Accessed November 5, 2010].

Carpineto, C. et al., 2009. Mobile information retrieval with search results clustering: Prototypes and evaluations. *Journal of the American Society for Information Science and Technology*, 60(5), pp.877–895.

Cedergren, M., 2003. Open content and value creation. *First Monday*, 8(8).

Chandler, C.J. & Gregory, A.S., 2010. Sleeping with the Enemy: Wikipedia in the College Classroom. *History Teacher*, 43(2), pp.247–257.

Chandler-Olcott, K., 2009. Digital Literacies. A Tale of Two Tasks: Editing in the Era of Digital Literacies. *Journal of Adolescent & Adult Literacy*, 53.

Chandy, R., 2009. Wikiganda: Identifying Propaganda Through Text Analysis. *Caltech Undergraduate Research Journal*, pp.6–11.

Chen, C.-J., 2009. Art history: a guide to basic research resources. *Collection building*, 28(3), pp.122–125.

Chen, H., 2010. The perspectives of higher education faculty on Wikipedia. *Electronic Library*, 28(3), pp.361–73.

Chen, H.-L., 2009. The use and sharing of information from Wikipedia by high-tech professionals for work purposes. *Electronic library*, 27(6), pp.893–905.

Chesney, T., 2006. An empirical examination of Wikipedia's credibility. *First Monday*, 11(11). Available at: http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1413/1331.

Cho, H., Chen, M. & Chung, S., 2010. Testing an Integrative Theoretical Model of Knowledge-Sharing Behavior in the Context of Wikipedia. *Journal of the American Society for Information Science and Technology*, 61(6), pp.1198–1212.

Chon, M., 2012. The Romantic Collective Author. *Vanderbilt Journal of Entertainment and Technology Law*, 14. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2076959 [Accessed October 11, 2012].

Choolhun, N., 2009. Google: to use, or not to use. What is the question? *Legal Information Management*, 9(03), p.168.

Chu, E., 2008. *Sparse relational data sets: Issues and an application*. United States -- Wisconsin: The University of Wisconsin - Madison. Available at: http://proquest.umi.com/pqdweb?did=1599589551&Fmt=7&clientId=10306&RQT=309&VName=PQD.

Chua, A.Y.K., Kaynak, S. & Foo, S.S.B., 2007. An analysis of the delayed response to hurricane katrina through the lens of knowledge management. *Journal of the American Society for Information Science and Technology*, 58(3), pp.391–403.

Ciffolilli, A., 2003. Phantom authority, self-selective recruitment and retention of members in virtual communities: The case of Wikipedia. *First Monday*, 8(12). Available at: http://www.firstmonday.org/Issues/issue8_12/ciffolilli/.

Cimini, N., 2010. Struggles online over the meaning of "Down"s syndrome': A "dialogic" interpretation. *Health (London, England: 1997)*, 14(4), pp.398–414.

Cimini, N. & Burr, J., 2012. An Aesthetic for Deliberating Online: Thinking Through "Universal Pragmatics" and "Dialogism" with Reference to Wikipedia. *The Information Society*, 28(3), pp.151–160.

Clark, M., Ruthven, I. & Holt, P.O., 2009. The evolution of genre in Wikipedia. *Journal for Language Technology and Computational Linguistics*, 24(1), pp.1–22.

Clauson, K.A. et al., 2008. Scope, completeness, and accuracy of drug information in Wikipedia. *The Annals of Pharmacotherapy*, 42(12), pp.1814–1821.

Coleman, A., 2007. Students `should use Wikipedia'. *BBC NEWS*. Available at: http://news.bbc.co.uk/2/hi/technology/7130325.stm.

Constantinides, P., Chiasson, M.W. & Introna, L.D., 2012. The ends of information systems research: a pragmatic framework. *MIS Quarterly*, 36(1), pp.1–20.

Cosley, D.R., 2006. *Helping hands: Design for member-maintained online communities*. United States -- Minnesota: University of Minnesota. Available at: http://proquest.umi.com/pqdweb?did=1212797001&Fmt=7&clientId=10306&RQT=309&VName=PQD.

Cosley, D.R. et al., 2007. SuggestBot: using intelligent task routing to help people find work in wikipedia. In *Proceedings of the 12th international conference on Intelligent user interfaces*. pp. 32 – 41. Available at: http://0-portal.acm.org.mercury.concordia.ca/citation.cfm?id=1216295.1216309&coll=DL&dl=GUIDE&CFID=112020990&CFTOKEN=50968312&preflayout=flat [Accessed November 21, 2010].

Cosley, D.R. et al., 2006. Using intelligent task routing and contribution review to help communities build artifacts of lasting value. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. pp. 1037 – 1046. Available at: http://0-portal.acm.org.mercury.concordia.ca/citation.cfm?id=1124772.1124928&coll=DL&dl=GUIDE&CFID=112031225&CFTOKEN=18535462&preflayout=flat [Accessed November 21, 2010].

Coursey, K., 2009. *The value of everything: Ranking and association with encyclopedic knowledge*. United States -- Texas: University of North Texas. Available at: http://proquest.umi.com/pqdweb?did=2005595041&Fmt=7&clientId=10306&RQT=309&VName=PQD.

Cress, U. & Kimmerle, J., 2008. A Systemic and Cognitive View on Collaborative Knowledge Building with Wikis. *International Journal of Computer-Supported Collaborative Learning*, 3.

Cross, T., 2006. Puppy smoothies: improving the reliability of open, collaborative wikis. *First Monday*, 11(9). Available at: http://www.firstmonday.org\-/issues\-/issue11_9\-/cross\-/index.html.

Crovitz, D. & Smoot, W.S., 2009. Wikipedia: Friend, Not Foe. *English Journal*, 98.

Csomai, A., 2008. *Keywords in the mist: Automated keyword extraction for very large documents and back of the book indexing*. United States -- Texas: University of North Texas. Available at: http://proquest.umi.com/pqdweb?did=1597616811&Fmt=7&clientId=10306&RQT=309&VName=PQD.

Csomai, A. & Mihalcea, R., 2008. Linking documents to encyclopedic knowledge. *IEEE Intelligent Systems*, 23(5), pp.34–41.

Cucerzan, S., 2007. Large-scale named entity disambiguation based on Wikipedia data. In *Proc. 2007 Joint Conference on EMNLP and CNLL*. Joint Conference on EMNLP and CNLL. Prague, Czech Republic, pp. 708–716. Available at: http://acl.ldc.upenn.edu/D/D07/D07-1074.pdf.

Curino, C.A., Moon, H.J., Tanca, L., et al., 2008. Schema evolution in wikipedia - Toward a web Information system benchmark. In *ICEIS 2008 - 10th International Conference on Enterprise Information Systems, June 12, 2008 - June 16, 2008*. ICEIS 2008 - Proceedings of the 10th International Conference on Enterprise Information Systems. International Conference on Enterprise Information Systems. Barcelona, Spain: Inst. for Syst. and Technol. of Inf. Control and Commun., pp. 323–332. Available at: http://www.cs.ucla.edu/~zaniolo/papers/ICEIS2008.pdf.

Curino, C.A., Moon, H.J. & Zaniolo, C., 2008. Graceful database schema evolution: the PRISM workbench. In *Proceedings of the VLDB Endowment VLDB Endowment Hompage*. pp. Volume 1 Issue 1, August 2008. Available at: http://0-portal.acm.org.mercury.concordia.ca/citation.cfm?id=1453856.1453939&coll=DL&dl=GUIDE&CFID=112025803&CFTOKEN=32336862&preflayout=flat [Accessed November 21, 2010].

Dalby, A., 2007. Wikipedia(s) on the language map of the world. *English Today*, 23(02), p.3.

Darren Hardy, 2010. *Volunteered geographic information in Wikipedia*. Ph.D. University of California, Santa Barbara. Available at: http://www2.bren.ucsb.edu/~dhardy/HardyThesis2010.pdf.

Daub, J. et al., 2008. The RNA WikiProject: Community annotation of RNA families. *RNA*, 14(12), pp.2462–2464.

David Ahn et al., 2005. Using Wikipedia at the TREC QA Track. In *The Thirteenth Text Retrieval Conference (TREC 2004)*. Text Retrieval Conference. Available at: http://dare.uva.nl/record/165576.

Davis, C., Nikolic, I. & Dijkema, G.P.J., 2010. Industrial ecology 2.0. *Journal of Industrial Ecology*, 14(5), pp.707–726.

Dede, C., 2008. A Seismic Shift in Epistemology. *Educause*, 43(3), pp.80–81.

Demartini, G. et al., 2010. Why finding entities in Wikipedia is difficult, sometimes. *Information Retrieval*, 13(5), p.534.

Denoyer, L. & Gallinari, P., 2009. Overview of the INEX 2008 XML Mining Track. *Advances in Focused Retrieval*, p.Jaap Kamps
Archives and Information Studies/Humanities, University of Amsterdam, Amsterdam, The Netherlands 1012 XT.

Denoyer, L. & Gallinari, P., 2006. The Wikipedia XML corpus. *SIGIR Forum*, 40(1), pp.64 – 9.

Devereux, B. et al., 2009. Towards Unrestricted, Large-Scale Acquisition of Feature-Based Conceptual Representations from Corpus Data. *Research on Language and Computation*, 7(2-4), pp.137 – 170.

Devgan, L. et al., 2007. Wiki-Surgery? Internal validity of Wikipedia as a medical and surgical reference. *Journal of the American College of Surgeons*, 205(3, supplement), pp.S76–S77.

van Dijk, Z., 2009. Wikipedia and lesser-resourced languages. *Language Problems & Language Planning*, 33(3), pp.234–250.

Dodig-Crnkovic, G., 2002. Scientific methods in computer science. In *Proceedings of the Conference for the Promotion of Research in IT at New Universities and at University Colleges in Sweden*. Conference for the Promotion of Research in IT at New Universities and at University Colleges in Sweden, Skövde, Suecia. Skövde, Suecia, Sweden, pp. 126–130. Available at: http://www.mrtc.mdh.se/~gdc/work/cs_method.pdf [Accessed October 4, 2012].

Dondio, P. & Barrett, S., 2007. Computational trust in Web content quality: a comparative evalutation on the Wikipedia project. *Informatica*, 31(2), pp.151–60.

Dooley, P.L., 2010. Wikipedia and the two-faced professoriate. In *Proceedings of the 6th International Symposium on Wikis and Open Collaboration*. New York, NY, USA: ACM.

Đorđe Stakić, 2009. Wiki technology - origin, development and importance. *Infotheca*, 10(1-2), pp.61–69.

Dorji, T. et al., 2010. Extraction, selection and ranking of Field Association (FA) Terms from domain-specific corpora for building a comprehensive FA terms dictionary. *Knowledge and Information Systems*, pp.1–21.

Duguid, P., 2006. Limits of self-organization: Peer production and "laws of quality. *First Monday*, 11(10), pp.0–0.

Ehmann, K., Large, A. & Beheshti, J., 2008. Collaboration in context: comparing article evolution among subject disciplines in Wikipedia. *First Monday*, 13(10), p.19 pp.

Eijkman, H., 2010. Academics and Wikipedia: Reframing Web 2.0+as a disruptor of traditional academic power-knowledge arrangements. *Campus-Wide Information Systems*, 27(3), pp.173–85.

Eijkman, H., 2008. Web 2.0 as a non-foundational network-centric learning space. *Campus-Wide Information Systems*, 25(2), pp.93–104.

Elia, A., 2009. Quantitative data and graphics on lexical specificity and index of readability: The case of Wikipedia. *RaeL: Revista Electronica de Linguistica Aplicada*, (8), pp.248–271.

Elsas, J.L. et al., 2008. Retrieval and feedback models for blog feed search. In *31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM SIGIR 2008, July 20, 2008 - July 24, 2008*. ACM SIGIR 2008 - 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Proceedings. Singapore, Singapore: Association for Computing Machinery, pp. 347–354. Available at: http://dx.doi.org/10.1145/1390334.1390394 [Accessed November 6, 2010].

Elvebakk, B., 2008. Philosophy democratized? A comparison between Wikipedia and two other Web-based philosophy resources. *First Monday*, 13(2). Available at: http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/2091/1938.

Emigh, W. & Herring, S.C., 2005. Collaborative Authoring on the Web: A Genre Analysis of Online Encyclopedias. In *Proceedings of the Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS'05) - Track 4 - Volume 04*. Washington, DC, USA: IEEE Computer Society, p. 99.1. Available at: http://0-

portal.acm.org.mercury.concordia.ca/citation.cfm?id=1042435.1042895&coll=DL&dl=GUIDE&CFID=112025803&CFTOKEN=32336862&preflayout=flat [Accessed November 21, 2010].

Erdmann, M. et al., 2009. Improving the extraction of bilingual terminology from Wikipedia. *ACM Transactions on Multimedia Computing, Communications and Applications*, 5(4). Available at: http://dx.doi.org/10.1145/1596990.1596995 [Accessed November 5, 2010].

Evgeniy Gabrilovich, 2006. *Feature Generation for Textual Information Retrieval Using World Knowledge*. Doctoral Dissertation. Haifa, Israel: Technion – Israel Institute of Technology. Available at: http://www.cs.technion.ac.il/~gabr/papers/phd-thesis.pdf.

Experiment-Resources.com, 2008. Research Designs - How to construct an experiment or study. *Experiment-Resources.com*. Available at: http://www.experiment-resources.com/research-designs.html [Accessed October 4, 2012].

Fallis, D., 2008. Toward an Epistemology of Wikipedia. *Journal of the American Society for Information Science and Technology*, 59(10), pp.1662–1674.

Famiglietti, A.A., 2011. *Hackers, Cyborgs, and Wikipedians: The Political Economy and Cultural History of Wikipedia*. thesis. Bowling Green State University. Available at: http://etd.ohiolink.edu/view.cgi?acc_num=bgsu1300717552 [Accessed October 9, 2012].

Farhoodi, M., Yari, A. & Mahmoudi, M., 2009. A Persian Web Page Classifier Applying a Combination of Content-Based and Context-Based Features. *International Journal of Information Studies*, 1(4), pp.263–71.

Farrell, H. & Schwartzberg, M., 2008. Norms, Minorities, and Collective Choice Online. *Ethics & International Affairs*, 22(4), pp.357–367.

Ferrandez, S. et al., 2009. Exploiting Wikipedia and EuroWordNet to solve Cross-Lingual Question Answering. *Information Sciences*, 179(20), pp.3473–3488.

Ferriter, M.M., 2009. "Arguably the Greatest": Sport Fans and Communities at Work on Wikipedia. *Sociology of Sport Journal*, 26(1), pp.127–154.

Ferschke, O., Zesch, Torsten & Gurevych, Iryna, 2011. Wikipedia revision toolkit: efficiently accessing Wikipedia's edit history. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Systems Demonstrations*. HLT '11. Stroudsburg, PA, USA: Association for Computational Linguistics, pp. 97–102. Available at: http://dl.acm.org/citation.cfm?id=2002440.2002457 [Accessed October 22, 2012].

Fink, A., 2005. *Conducting Research Literature Reviews: From the Internet to Paper* 2nd ed., Sage Publications, Inc.

Fiore, K., 2011. *APA: Med Students Cram for Exams With Wikipedia*, Available at: http://www.medpagetoday.com/MeetingCoverage/APA/26483.

Fitzpatrick, K., 2009. Peer-to-peer review and the future of scholarly authority. *Cinema Journal*, 48, p.124.

Forte, A., 2009. *Learning in public: Information literacy and participatory media*. United States -- Georgia: Georgia Institute of Technology. Available at: http://proquest.umi.com/pqdweb?did=1879753771&Fmt=7&clientId=10306&RQT=309&VName=PQD.

Forte, A. & Bruckman, A., 2008. Scaling consensus: Increasing decentralization in Wikipedia governance. In *41st Annual Hawaii International Conference on System Sciences 2008, HICSS, January 7, 2008 - January 10, 2008*. Proceedings of the Annual Hawaii International Conference on System Sciences. Big Island, HI, United states: Inst. of Elec. and Elec. Eng. Computer Society. Available at: http://dx.doi.org/10.1109/HICSS.2008.383 [Accessed November 5, 2010].

Forte, A. & Bruckman, A., 2005. Why do people write for Wikipedia? Incentives to contribute to open-content publishing. In *Group 2005 workshop: Sustaining community: The role and design of incentive mechanisms in online systems. Sanibel Island, FL.* Available at: http://www.cc.gatech.edu/ aforte/ForteBruckmanWhyPeopleWrite.pdf.

Forte, A. & Bruckman, A., 2009. Writing, Citing, and Participatory Media: Wikis as Learning Environments in the High School Classroom. *International Journal of Learning*, 1(4), pp.23–44.

Forte, A., Larco, V. & Bruckman, A., 2009. Decentralization in wikipedia governance. *Journal of Management Information Systems*, 26(1), pp.49–72.

Francke, H. & Sundin, O., 2010. An inside view: credibility in Wikipedia from the perspective of editors. *Information Research*, 15(3), p.16 pp.

Friedlin, J. & McDonald, C.J., 2010. An evaluation of medical knowledge contained in Wikipedia and its use in the LOINC database. *Journal of the American Medical Informatics Association: JAMIA*, 17(3), pp.283–287.

Friesen, N. & Hopkins, J., 2008. Wikiversity; or education meets the free culture movement: An etnographic investigation. *First Monday*, 13(10). Available at: http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2234/2031.

Furbach, U. et al., 2010. Logic-Based Question Answering. *KI - Künstliche Intelligenz*, 24(1), pp.51–55.

Gabrilovich, E. & Markovitch, S., 2007. Computing semantic relatedness using Wikipedia-based explicit semantic analysis. In *Proceedings of the 20th international joint conference on Artifical intelligence*. pp. 1606–1611. Available at: http://0-portal.acm.org.mercury.concordia.ca/citation.cfm?id=1625275.1625535&coll=DL&dl=GUIDE&CFID=112059492&CFTOKEN=13405793&preflayout=flat [Accessed November 22, 2010].

Gabrilovich, E. & Markovitch, S., 2006. Overcoming the brittleness bottleneck using Wikipedia: enhancing text categorization with encyclopedic knowledge. In *Proceedings of the Twenty-First National Conference on Artificial Intelligence*. Menlo Park, California: AAAI Press, pp. 1301–1306.

Gabrilovich, E. & Markovitch, S., 2009. Wikipedia-based semantic interpretation for natural language processing. *Journal of Artificial Intelligence Research*, 34, pp.443–498.

Ganter, V. & Strube, M., 2009. Finding hedges by chasing weasels: hedge detection using Wikipedia tags and shallow linguistic features. In *Proceedings of the ACL-IJCNLP 2009 Conference Short*

Papers. pp. 173–176. Available at: http://0-portal.acm.org.mercury.concordia.ca/citation.cfm?id=1667583.1667636&coll=DL&dl=GUIDE&CFID=112025803&CFTOKEN=32336862&preflayout=flat [Accessed November 21, 2010].

Garber, M., 2011. *The contribution conundrum: Why did Wikipedia succeed while other encyclopedias failed?*, Available at: http://www.niemanlab.org/2011/10/the-contribution-conundrum-why-did-wikipedia-succeed-while-other-encyclopedias-failed/.

Gardner, P.P. et al., 2010. Rfam: Wikipedia, clans and the "decimal" release. *Nucleic Acids Res.*, p.gkq1129.

Garud, R., Jain, S. & Tuertscher, P., 2008. Incomplete by design and designing for incompleteness. *ORGANIZATION STUDIES*, 29(3), pp.351–371.

Gehl, R., 2010. *A cultural and political economy of Web 2.0*. United States -- Virginia: George Mason University. Available at: http://proquest.umi.com/pqdweb?did=2035889541&Fmt=7&clientId=10306&RQT=309&VName=PQD.

Geiger, R.S. & Ribes, D., 2010. The work of sustaining order in wikipedia: the banning of a vandal. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*. pp. 117–126. Available at: http://0-portal.acm.org.mercury.concordia.ca/citation.cfm?id=1718918.1718941&coll=DL&dl=GUIDE&CFID=112020990&CFTOKEN=50968312&preflayout=flat [Accessed November 21, 2010].

George, A., 2007. Avoiding tragedy in the wiki-commons. *Virginia Journal of Law and Technology*, 12(8), pp.1–42.

Gever, J., 2007. *Wikipedia Information on Surgical Procedures Generally Accurate*, Available at: http://www.docguide.com/wikipedia-information-surgical-procedures-generally-accurate-presented-acs.

Giles, J., 2005. Internet encyclopaedias go head to head. *Nature*, 438(7070), pp.900–901.

Goldspink, C., 2010. Normative behaviour in Wikipedia. *Information*, 13(5), pp.652–673.

Goldspink, C., 2009. Social self-regulation in computer mediated communities: the case of Wikipedia. *International Journal of Agent Technologies & Systems*, 1(1), pp.19–33.

Gollapudi, S. & Sharma, A., 2009. An axiomatic approach for result diversification. In *Proceedings of the 18th international conference on World wide web*. p. Wolfgang Nejdl L3S and Hannover University. Available at: http://0-portal.acm.org.mercury.concordia.ca/citation.cfm?id=1526709.1526761&coll=DL&dl=GUIDE&CFID=112015986&CFTOKEN=43703661&preflayout=flat [Accessed November 21, 2010].

Greenstein, S., 2007. Wagging Wikipedia's long tail. *IEEE Micro*, 27(2), p.6+79.

Gregor, S., 2006. The nature of theory in information systems. *MIS Q.*, 30(3), pp.611–642.

Grineva, M., Grinev, M. & Lizorkin, D., 2009. Extracting key terms from noisy and multitheme documents. In *Proceedings of the 18th international conference on World wide web*. p. Wolfgang

Nejdl
L3S and Hannover University. Available at: http://0-portal.acm.org.mercury.concordia.ca/citation.cfm?id=1526709.1526798&coll=DL&dl=GUIDE&CFID=112020990&CFTOKEN=50968312&preflayout=flat [Accessed November 21, 2010].

Gunnels, C.B., 2007. Librarians on the verge of an epistemological breakdown. *Community & Junior College Libraries*, 14(2), pp.111–20.

Gunnels, C.B. & Sisson, A., 2009. Confessions of a Librarian or: How I Learned to Stop Worrying and Love Google. *Community & Junior College Libraries*, 15(1), pp.15–21.

Guosong Shao, 2009. Understanding the appeal of user-generated media: a uses and gratification perspective. *Internet Research*, 19(1), pp.7–25.

Gurevych, Iryna & Wolf, E., 2010. Expert-Built and Collaboratively Constructed Lexical Semantic Resources. *Language and Linguistics Compass*, 4(11), pp.1074–1090.

Ha, J.K. & Kim, Y.-H., 2009. An Exploration on On-line Mass Collaboration: focusing on its motivation structure. *International Journal of Social Sciences*, 4(2), pp.138–143.

Hahn, J., 2010. Information seeking with Wikipedia on the iPod Touch. *Reference services review*, 38(2), pp.284–298.

Hahn, J., 2009. On the remediation of Wikipedia to the iPod. *Reference Services Review*, 37(3), pp.272–285.

Haider, J. & Sundin, O., 2010. Beyond the legacy of the Enlightenment? Online encyclopaedias as digital heterotopias. *First Monday*, 15(1), p.16 pp.

Haigh, C.A., 2010. Wikipedia as an evidence source for nursing and healthcare students. *Nurse Education Today*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20646799 [Accessed October 29, 2010].

Halavais, A. & Lackaff, D., 2008. An analysis of topical coverage of Wikipedia. *Journal of Computer Mediated Communication*, 13(2), pp.429–440.

Hansen, S., Berente, N. & Lyytinen, K., 2009. Wikipedia, critical social theory, and the possibility of rational discourse. *The Information society*, 25(1), pp.38–59.

Hara, N., Shachaf, P. & Hew, K.F., 2010. Cross-cultural analysis of the Wikipedia community. *Journal of the American Society for Information Science and Technology*, 61(10), pp.2097–2108.

Hardy, M., 2007. Wiki Goes to War. *AQ - Journal of Contemporary Analysis*, 79(4), pp.17–22.

Harouni, H., 2009. High School Research and Critical Literacy: Social Studies with and Despite Wikipedia. *Harvard Educational Review*, 79(3), pp.473–493.

Hartelius, E.J., 2010. Wikipedia and the Emergence of Dialogic Expertise. *Southern Communication Journal*, 75(5), pp.505–526.

Head, A.J. & Eisenberg, M.B., 2010. How today's college students use Wikipedia for course-related research. *First Monday*, 15(3), p.15 pp.

Heilman, J.M. et al., 2011. Wikipedia: A Key Tool for Global Public Health Promotion. *Journal of Medical Internet Research*, 13(1), p.e14.

Hemphill, C., 2008. Network neutrality and the false promise of zero-price regulation. *Yale Journal on Regulation*, 25(2), p.135.

Hepp, M., Siorpaes, K. & Bachlechner, D., 2007. Harvesting Wiki consensus: using Wikipedia entries as vocabulary for knowledge management. *IEEE Internet Computing*, 11(5), pp.54–65.

Hickerson, C.A. & Thompson, S.R., 2009. Dialogue through wikis: a pilot exploration of dialogic public relations and wiki websites. *PRism*, 6(1). Available at: http://www.doaj.org/doaj?func=abstract&id=613862 [Accessed October 10, 2012].

Hilbert, M., 2009. The Maturing Concept of E-Democracy: From E-Voting and Online Consultations to Democratic Value Out of Jumbled Online Chatter. *Journal of Information Technology & Politics*, 6(2), pp.87–110.

Hochstotter, N. & Lewandowski, D., 2009. What users see - Structures in search engine results pages. *Information Sciences*, 179(12), pp.1796–1812.

Hoffman, D.A. & Mehra, S.K., 2009. Wikitruth through wikiorder. *Emory Law Journal*, 59(1), pp.151–209.

Holley, R., 2010. Crowdsourcing: How and Why Should Libraries Do It? *D-Lib Magazine*, 16(3-4), p.15 pp.

Holloway, T., Bozicevic, M. & Borner, K., 2007. Analyzing and visualizing the semantic coverage of Wikipedia and its authors. *Complexity*, 12(3), pp.30–40.

Hu, B., 2010. WiKi'mantics: interpreting ontologies with WikipediA. *Knowledge and Information Systems*, 25(3), p.445.

Hu, J. et al., 2008. Enhancing text clustering by leveraging wikipedia semantics. In *31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM SIGIR 2008, July 20, 2008 - July 24, 2008*. ACM SIGIR 2008 - 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Proceedings. Singapore, Singapore: Association for Computing Machinery, pp. 179–186. Available at: http://dx.doi.org/10.1145/1390334.1390367 [Accessed November 5, 2010].

Hu, J. et al., 2009. Understanding user's query intent with wikipedia. In *Proceedings of the 18th international conference on World wide web*. p. Wolfgang Nejdl L3S and Hannover University. Available at: http://0-portal.acm.org.mercury.concordia.ca/citation.cfm?id=1526709.1526773&coll=DL&dl=GUIDE&CFID=112020990&CFTOKEN=50968312&preflayout=flat [Accessed November 21, 2010].

Hu, M. et al., 2007. Measuring article quality in wikipedia: Models and evaluation. In *16th ACM Conference on Information and Knowledge Management, CIKM 2007, November 6, 2007 - November 9, 2007*. International Conference on Information and Knowledge Management, Proceedings. Lisboa, Portugal: Association for Computing Machinery, pp. 243–252. Available at: http://dx.doi.org/10.1145/1321440.1321476 [Accessed November 5, 2010].

Hughes, B. et al., 2009. Junior physician's use of Web 2.0 for information seeking and medical education: A qualitative study. *International Journal of Medical Informatics*, 78(10), pp.645–655.

Huss, J.W. et al., 2008. A Gene Wiki for Community Annotation of Gene Function. *PLoS Biology*, 6(7), p.e175.

Huss, J.W. et al., 2010. The Gene Wiki: community intelligence applied to human gene annotation. *Nucleic Acids Research*, 38(Database issue), pp.D633–639.

Huvila, I., 2010. Where does the information come from? Information source use patterns in Wikipedia. *Information Research*, 15(3), p.24 pp.

Hwang, H. et al., 2010. BinRank: Scaling dynamic authority-based search using materialized subgraphs. *IEEE Transactions on Knowledge and Data Engineering*, 22(8), pp.1176–1190.

Iba, T. et al., 2010. Analyzing the Creative Editing Behavior of Wikipedia Editors: Through Dynamic Social Network Analysis. *Procedia - Social and Behavioral Sciences*, 2(4), pp.6441–6456.

Jacobs, M.L., 2009. Libraries and the mobile revolution: remediation = relevance. *Reference services review*, 37(3), pp.286–290.

Jancarik, A. & Jancarikova, K., 2010. Wiki Tools in the Preparation and Support of e-Learning Courses. *Electronic Journal of e-Learning*, 8(2), pp.123–32.

Jankowski, J., 2008a. Copernicus Adding the Third Dimension to Wikipedia. In *Wikimania*. Available at: http://wm08reg.wikimedia.org/schedule/events/21.en.html.

Jankowski, J., 2008b. Copernicus: 3D Wikipedia. In *International Conference on Computer Graphics and Interactive Techniques*. ACM.

Jankowski, J. & Kruk, S.R., 2008. 2LIP: The Step Towards The Web3D. In *Proceedings of the 17th International World Wide Web Conference (WWW2008). April 21-25, 2008. Beijing, China*. pp. 1137–1138. Available at: http://www2008.org/papers/pdf/p1137-jankowskiA.pdf.

Järvinen, P., 2008. Mapping Research Questions to Research Methods. In D. Avison et al., eds. *Advances in Information Systems Research, Education and Practice*. IFIP Advances in Information and Communication Technology. Springer Boston, pp. 29–41. Available at: http://www.springerlink.com/content/8746437n5065516h/abstract/ [Accessed October 4, 2012].

Javanmardi, S., 2011. *Measuring Content Quality in User Generated Content Systems: a Machine Learning Approach*. UNIVERSITY OF CALIFORNIA, IRVINE. Available at: http://gradworks.umi.com/34/73/3473551.html [Accessed October 12, 2012].

Javanmardi, S., Lopes, C. & Baldi, P., 2010. Modeling user reputation in wikis. *Statistical Analysis and Data Mining*, 3(2), pp.126–139.

Jennings, E., 2008. Using Wikipedia to Teach Information Literacy. *College & Undergraduate Libraries*, 15(4), pp.432–437.

Jijkoun, V. & Rijke, M., 2007. Overview of the WiQA Task at CLEF 2006. *Evaluation of Multilingual and Multi-modal Information Retrieval*, p.Springer–Verlag Berlin, Heidelberg ©2007.

Johnson, P.T. et al., 2008. A Comparison of World Wide Web Resources for Identifying Medical Information. *Academic Radiology*, 15(9), pp.1165–1172.

Jones, J., 2008. Patterns of revision in online writing: A study of Wikipedia's featured articles. *Written Communication*, 25(2), pp.262–289.

Jordan, C., 2009. *Contextual Retrieval of Single Wikipedia Articles to Support the Reading of Academic Abstracts*. PhD dissertation. Dalhousie University (Canada). Available at: http://proquest.umi.com/pqdlink?Ver=1&Exp=10-14-2017&FMT=7&DID=1630234261&RQT=309&attempt=1&cfc=1.

Jordan, C. & Watters, C., 2009. Addressing gaps in knowledge while reading. *Journal of the American Society for Information Science and Technology*, 60(11), pp.2255–2268.

Judd, T. & Kennedy, G., 2010. A five-year study of on-campus Internet use by undergraduate biomedical students. *Computers and Education*, 55(4), pp.1564–1571.

Judd, T. & Kennedy, G., 2009. Expediency-based practice? Medical students' reliance on Google and Wikipedia for biomedical inquiries. *British Journal of Educational Technology*, p.no.

Jullien, N., 2012. What we know about Wikipedia. A review of the literature analyzing the project(s). *Available at SSRN 2053597*. Available at: https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID2053597_code728676.pdf?abstractid=2053597&mirid=2 [Accessed October 4, 2012].

Kalantidis, Y. et al., 2010. VIRaL: Visual Image Retrieval and Localization. *Multimedia Tools and Applications*, pp.1–38.

Kane, G.C. & Fichman, R.G., 2009. The Shoemaker's Children: Using Wikis for Information Systems Teaching, Research, and Publication. *MIS Quarterly*, 33(1), pp.1–17.

Kangpyo Lee et al., 2008. FolksoViz: A subsumption-based folksonomy visualization using the wikipedia. *Journal of KISS: Computing Practices*, 14(4), pp.401–11.

Kaplan, A.M. & Haenlein, M., 2010. Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, 53(1), pp.59–68.

Kaptein, R., Sedyukov, P. & Kamps, J., 2010. Linking Wikipedia to the Web. In *Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval*. New York, NY, USA: ACM.

Kasneci, G. et al., 2008. The YAGO-NAGA approach to knowledge discovery. *SIGMOD Record*, 37(4), pp.41–47.

Kazama, J. & Torisawa, K., 2007. Exploiting Wikipedia as external knowledge for named entity recognition. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*. Stroudsburg, PA, USA:

Association for Computational Linguistics, pp. 698–707. Available at: http://acl.ldc.upenn.edu/D/D07/D07-1073.pdf.

Kennedy, K., 2009. *Textual curators and writing machines: Authorial agency in encyclopedias, print to digital*. United States -- Minnesota: University of Minnesota. Available at: http://proquest.umi.com/pqdweb?did=1850115751&Fmt=7&clientId=10306&RQT=309&VName=PQD.

Kim, J.-M., Shin, H. & Kim, H.-J., 2007. Schema and constraints-based matching and merging of Topic Maps. *Information processing & management*, 43(4), pp.930–945.

Kim, J.Y. et al., 2010. The pathology informatics curriculum wiki: Harnessing the power of user-generated content. *Journal of Pathology Informatics*, 1. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20805963 [Accessed October 29, 2010].

Kimmerle, J. et al., 2010. VISUALIZING CO-EVOLUTION OF INDIVIDUAL AND COLLECTIVE KNOWLEDGE. *Information, Communication & Society*. Available at: http://www.informaworld.com/10.1080/13691180903521547.

Kimmons, R.M., 2011. Understanding collaboration in wikipedia. *First Monday*, 16(12). Available at: http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/3613 [Accessed October 22, 2012].

Kinsella, S. et al., 2008. Applications of Semantic Web Methodologies and Techniques to Social Networks and Social Websites. *Reasoning Web*, p.Springer–Verlag Berlin, Heidelberg ©2008.

Kinzler, D., 2008. *Automatischer Aufbau eines multilingualen Thesaurus durch Extraktion semantischer und lexikalischer Relationen aus der Wikipedia*. Diplomarbeit an der Abteilung fÃ¼r Automatische Sprachverarbeitung. Institut für Informatik, Universität Leipzig.

Kinzler, D., 2009. WikiWord: Multilingual Image Search and More. In *Wikimania*.

Kitchenham, B.A. & Charters, S., 2007. *Guidelines for performing Systematic Literature Reviews in Software Engineering*,

Kittur, Aniket, Suh, B., et al., 2007. He says, she says: Conflict and coordination in Wikipedia. In *25th SIGCHI Conference on Human Factors in Computing Systems 2007, CHI 2007, April 28, 2007 - May 3, 2007*. Conference on Human Factors in Computing Systems - Proceedings. San Jose, CA, United states: Association for Computing Machinery, pp. 453–462. Available at: http://dx.doi.org/10.1145/1240624.1240698 [Accessed November 5, 2010].

Kittur, Aniket, Chi, E. & Bryan A. Pendleton, Bongwon Suh, T.M., 2007. Power of the Few vs. Wisdom of the Crowd: Wikipedia and the Rise of the Bourgeoisie. In *Computer/Human Interaction 2007*. Available at: http://www.viktoria.se/altchi/submissions/submission_edchi_1.pdf.

Kittur, Aniket, Chi, Ed H. & Suh, B., 2009. What's in Wikipedia?: mapping topics and conflict using socially annotated category structure. In *Proceedings of the 27th international conference on Human factors in computing systems*. New York, NY, USA: ACM, pp. 1509–1512.

Kittur, Aniket & Kraut, R.E., 2008. Harnessing the wisdom of crowds in wikipedia: Quality through coordination. In *2008 ACM Conference on Computer Supported Cooperative Work, CSCW 08, November 8, 2008 - November 12, 2008*. Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW. San Diego, CA, United states: Association for Computing Machinery, pp. 37–46. Available at: http://dx.doi.org/10.1145/1460563.1460572 [Accessed November 5, 2010].

Klemp, N.J. & Forcehimes, A.T., 2010. From Town-Halls to Wikis: Exploring Wikipedia's Implications for Deliberative Democracy. *Journal of Public Deliberation*, 6(2), p.4.

Knapp, M.M., 2008. eBay, Wikipedia, and the Future of the Footnote. *Theatre History Studies*, 28(1), pp.36–41.

Kobilarov, G. et al., 2009. Media Meets Semantic Web — How the BBC Uses DBpedia and Linked Data to Make Connections. In L. Aroyo et al., eds. *The Semantic Web: Research and Applications*. Lecture Notes in Computer Science. Berlin/Heidelberg: Springer, pp. 723–737. Available at: http://www.georgikobilarov.com/publications/2009/eswc2009-bbc-dbpedia.pdf.

Kohn, R.S., 2010. Of Descartes and of train schedules: Evaluating the Encyclopedia Judaica, Wikipedia, and other general and Jewish Studies encyclopedias. *Library review (Glasgow)*, 59(4), pp.249–260.

Kolbitsch, J. & Maurer, H., 2004. Community Building around Encyclopaedic Knowledge. *Journal of Computing and Information Technology*, 14(3), p.175.

Konieczny, P., 2010. Adhocratic Governance in the Internet Age: A Case of Wikipedia. *Journal of Information Technology & Politics*, 7(4), pp.263 – 283.

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

Konieczny, P., 2007. Wikis and Wikipedia as a Teaching Tool. *International Journal of Instructional Technology & Distance Learning*, 4(1), pp.15–34.

Koolen, M., Kazai, G. & Craswell, N., 2009. Wikipedia pages as entry points for book search. In *WSDM '09: Proceedings of the Second ACM International Conference on Web Search and Data Mining*. New York, NY, USA: ACM, pp. 44–53.

Korfiatis, N.T. et al., 2006. Evaluating authoritative sources using social networks : an insight from Wikipedia. *Online information review (Print)*, 30(3), pp.252–262.

Korosec, L. et al., 2010. Chemical Information Media in the Chemistry Lecture Hall: A Comparative Assessment of Two Online Encyclopedias. *CHIMIA International Journal for Chemistry*, 64(5), pp.309–314.

Korsgaard, T.R. & Jensen, C.D., 2009. Reengineering the Wikipedia for Reputation. *Electronic Notes in Theoretical Computer Science (ENTCS)*, Volume(me 244), pp.81–94.

Kostakis, V., 2010. Identifying and understanding the problems of Wikipedia's peer governance: The case of inclusionists versus deletionists. *First Monday*, 15(3), p.14 pp.

Kriplean, T. et al., 2007. Community, consensus, coercion, control: CS*W or how policy mediates mass participation. In *2007 International ACM Conference on Supporting Group Work, GROUP'07, November 4, 2007 - November 7, 2007*. GROUP'07 - Proceedings of the 2007 International ACM Conference on Supporting Group Work. Sanibel Island, FL, United states: Association for Computing Machinery, pp. 167–176. Available at: http://dx.doi.org/10.1145/1316624.1316648 [Accessed November 6, 2010].

Kriplean, T., Beschastnikh, I. & McDonald, D.W., 2008. Articulations of wikiwork: uncovering valued work in wikipedia through barnstars. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*. pp. 47–56. Available at: http://0-portal.acm.org.mercury.concordia.ca/citation.cfm?id=1460563.1460573&coll=DL&dl=GUIDE&CFID=112015986&CFTOKEN=43703661&preflayout=flat [Accessed November 21, 2010].

Krizhanovsky, A.A. & Smirnov, A.V., 2009. On the problem of wiki texts indexing. *Journal of Computer and Systems Sciences International*, 48(4), pp.616–624.

Krötzsch, M. et al., 2007. Semantic Wikipedia. *Web Semantics*, 5(4), pp.251–261.

Kubiszewski, I., Noordewier, T. & Costanza, R., 2011. Perceived credibility of Internet encyclopedias. *Computers & Education*, 56(3), pp.659–667.

Kupiainen, R., Suoranta, Juha & Vaden, Tere, 2007. Fire Next Time: Or Revisioning Higher Education in the Context of Digital Social Creativity. *E-Learning*, 4.

de Laat, P.B., 2010. How can contributors to open-source communities be trusted? On the assumption, inference, and substitution of trust. *Ethics and Information Technology*, 12(4), pp.1–15.

Lam, S.T.K. et al., 2011. WP:clubhouse?: an exploration of Wikipedia's gender imbalance. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration*. WikiSym '11. New York, NY, USA: ACM, pp. 1–10. Available at: http://doi.acm.org/10.1145/2038558.2038560 [Accessed October 22, 2012].

Lam, S.T.K. & Riedl, J., 2009. Is Wikipedia growing a longer tail? In *2009 ACM SIGCHI International Conference on Supporting Group Work, GROUP'09, May 10, 2009 - May 13, 2009*. GROUP'09 - Proceedings of the 2009 ACM SIGCHI International Conference on Supporting Group Work. Sanibel Island, FL, United states: Association for Computing Machinery, pp. 105–114. Available at: http://dx.doi.org/10.1145/1531674.1531690 [Accessed November 5, 2010].

Lanamäki, A. et al., 2011. Protocol for Systematic Mapping of Wikipedia Studies. In *Proceedings of IRIS 2011 – The 34th Information Systems Research Seminar in Scandinavia*. Information Systems Research Seminar in Scandinavia. Turku, Finland.

Langlois, G., 2008. *The TechnoCultural dimensions of meaning [microform]: towards a mixed semiotics of the World Wide Web*. Thesis (Ph.D.). York University.

Langlois, G. & Elmer, G., 2009. Wikipedia leeches? The promotion of traffic through a collaborative web format. *New Media & Society*, 11(5), pp.773–794.

Laurent, Michael R & Vickers, T.J., 2009. Seeking health information online: does Wikipedia matter? *Journal of the American Medical Informatics Association*, 16(4), pp.471–479.

Lavsa, S.M. et al., 2011. Reliability of Wikipedia as a medication information source for pharmacy students. *Currents in Pharmacy Teaching and Learning*, 3(2), pp.154–158.

Lawler, C., 2006. A "resource review" of Wikipedia. *Counselling & Psychotherapy Research*, 6(3), pp.149–150.

Leinonen, T., Vaden, T. & Suoranta, J., 2009. Learning in and with an open Wiki project: Wikiversity's potential in global capacity building. *First Monday*, 14(2), p.11 pp.

Leithner, A. et al., 2010. Wikipedia and osteosarcoma: a trustworthy patients' information? *Journal of the American Medical Informatics Association*, 17(4), pp.373–374.

Leskovec, J., Huttenlocher, D. & Kleinberg, J., 2010a. Governance in Social Media: A case study of the Wikipedia promotion process. In *ICWSM 2010 - International AAAI Conference on Weblogs and Social Media*. ICWSM 2010 - International AAAI Conference on Weblogs and Social Media. Available at: http://arxiv.org/abs/1004.3547 [Accessed October 22, 2012].

Leskovec, J., Huttenlocher, D. & Kleinberg, J., 2010b. Predicting positive and negative links in online social networks. In *Proceedings of the 19th international conference on World wide web*. pp. 641–650. Available at: http://0-portal.acm.org.mercury.concordia.ca/citation.cfm?id=1772690.1772756&coll=DL&dl=GUIDE&CFID=112015986&CFTOKEN=43703661&preflayout=flat [Accessed November 21, 2010].

Letia, M., Preguica, N. & Shapiro, M., 2010. Consistency without concurrency control in large, dynamic systems. , 44, pp.29–34.

Lewandowski, D. & Spree, U., 2011. Ranking of Wikipedia articles in search engines revisited: Fair ranking for reasonable quality? *Journal of the American Society for Information Science and Technology*, 62(1), pp.117–132.

Li, D. et al., 2010. Keyphrase extraction based on topic relevance and term association. *Journal of Information and Computational Science*, 7(1), pp.293–299.

Li, Yun et al., 2008. Searching and computing for vocabularies with semantic correlations from Chinese Wikipedia. In *China-Ireland International Conference on Information and Communications Technologies, CIICT 2008, September 26, 2008 - September 28, 2008*. IET Conference Publications. Beijing, China: Institution of Engineering and Technology, pp. 58–63. Available at: http://dx.doi.org/10.1049/cp:20080760 [Accessed November 5, 2010].

Liao, H.-T., 2009. Conflict and consensus in the Chinese version of Wikipedia. *IEEE Technology and Society Magazine*, 28(2), pp.49–56.

Lichtenstein, S. & Parker, C.M., 2009. Wikipedia model for collective intelligence: a review of information quality. *International Journal of Knowledge and Learning*, 5(3-4), pp.254–72.

Lih, A., 2004. Wikipedia as participatory Journalism: Reliable Sources? Metrics for evaluating collaborative media as a news resource. In *5th International Symposium on Online Journalism*. Available at: http://jmsc.hku.hk\-/faculty\-/alih\-/publications\-/utaustin-2004-wikipedia-rc2.pdf.

Lim, S., 2009. How and why do college students use Wikipedia? *Journal of the American Society for Information Science and Technology (Print)*, 60(11), pp.2189–2202.

Lim, S. & Kwon, N., 2010. Gender differences in information behavior concerning Wikipedia, an unorthodox information source? *Library & information science research*, 32(3), pp.212–220.

Lin, C.-C. et al., 2009. Learning weights for translation candidates in Japanese-Chinese information retrieval. *Expert Systems with Applications*, 36(4), pp.7695–7699.

Lin, C.-C., Wang, Y.-C. & Tsai, R.T.-H., 2010. Japanese-Chinese Information Retrieval With an Iterative Weighting Scheme. *Journal of information science and engineering*, 26(2), pp.685–697.

Lin, M., 2006. *Sharing knowledge and building communities: A narrative of the formation, development and sustainability of OOPS*. United States -- Texas: University of Houston. Available at: http://proquest.umi.com/pqdweb?did=1251888971&Fmt=7&clientId=10306&RQT=309&VName=PQD.

Lindsey, D., 2010. Evaluating quality control of Wikipedia's feature articIes. *First Monday*, 15(4), p.7 pp.

Liu, S., 2006. *Improve text retrieval effectiveness and robustness*. United States -- Illinois: University of Illinois at Chicago. Available at: http://proquest.umi.com/pqdweb?did=1221734581&Fmt=7&clientId=10306&RQT=309&VName=PQD.

Lizorkin, D. et al., 2010. Accuracy estimate and optimization techniques for SimRank computation. *VLDB Journal*, 19(1), pp.45–66.

Logan, D.W. et al., 2010. Ten simple rules for editing Wikipedia. *PLoS Computational Biology*, 6(9), p.e1000941.

Lopes, R. & Carriço, L., 2008. On the Credibility of Wikipedia: an Accessibility Perspective. In *Second Workshop on Information Credibility on the Web (WICOW 2008)*. New York: ACM.

Lorenzen, M., 2006. Vandals, Administrators, and Sockpuppets, Oh My! An Ethnographic Study of Wikipedia's Handling of Problem Behavior. *MLA forum*, 5.

Luyt, B., Tay, C.H.A., et al., 2008. Improving wikipedia's accuracy: Is edit age a solution? *Journal of the American Society for Information Science and Technology*, 59(2), pp.318–330.

Luyt, B. et al., 2010. Librarian Perception of Wikipedia: Threats or Opportunities for Librarianship? *Libri (Copenhagen)*, 60(1), pp.57–64.

Luyt, B., 2011. The nature of historical representation on Wikipedia: Dominant or alterative historiography? *Journal of the American Society for Information Science and Technology*, 62(6), pp.1058–1065.

Luyt, B., Zainal, C., et al., 2008. Young people's perceptions and usage of Wikipedia. *Information Research*, 13(4). Available at: http://apps.isiknowledge.com/full_record.do?product=WOS&search_mode=GeneralSearch&qid=4&SID=3D9@K9HgCKnBM85J1eH&page=2&doc=82 [Accessed October 29, 2010].

Luyt, B. & Tan, D., 2010. Improving Wikipedia's Credibility: References and Citations in a Sample of History Articles. *Journal of the American Society for Information Science and Technology (Print)*, 61(4), pp.715–722.

Madison, M., Frischmann, B. & Strandburg, K., 2010. Constructing commons in the cultural environment. *Cornell Law Review*, 95(4), pp.657–709.

Maehre, J., 2009. What It Means to Ban Wikipedia: An Exploration of the Pedagogical Principles at Stake. *College Teaching*, 57(4), pp.229–36.

Magnus, P.D., 2008. Early response to false claims in Wikipedia. *First Monday*, 13(9), p.4 pp.

Magnus, P.D., 2009. On Trusting Wikipedia. *Episteme - Edinburgh*, 6(1).

Magrassi, P., 2010. *Free and open-source software is not an emerging property but rather the result of studied design*, Available at: http://arxiv.org/pdf/1012.5625.

Malone, T., Laubacher, R. & Dellarocas, C., 2010. The Collective Intelligence Genome. *MIT SLOAN MANAGEMENT REVIEW*, 51(3), p.21–+.

Maracke, C., 2010. Creative Commons International The International License Porting Project. *jipitec*, 1(1). Available at: http://www.jipitec.eu/issues/jipitec-1-1-2010/2417 [Accessed January 18, 2011].

Martin, O.S., 2012. *A Dynamic Competing Risk Model for Filtering Reliability and Tracking Survivability*. THE GEORGE WASHINGTON UNIVERSITY. Available at: http://gradworks.umi.com/34/90/3490816.html [Accessed October 4, 2012].

Martin, O.S., 2010. A Wikipedia Literature Review. *arXiv:1110.5863*. Available at: http://arxiv.org/abs/1110.5863 [Accessed October 4, 2012].

Mateos-Garcia, J. & Steinmueller, W.E., 2008. Open, But How Much? Growth,Conflict, and Institutional Evolution in OpenSource Communities. *Community, Economic Creativity, and Organization*, 1, pp.254–283.

Mattus, M., 2008. Wikipedia – Free and Reliable? : Aspects of a Collaboratively Shaped Encyclopaedia. *Nordicom Review*, 30(1), pp.183–199.

Matychak, X., 2008. Knowledge Architecture That Facilitates Trust and Collaboration. *Interactions*, 15(4), p.26.

MBA Knowledge Base, 2011. Research Methodology | MBA Knowledge Base. Available at: http://www.mbaknol.com/category/research-methodology/ [Accessed October 4, 2012].

McCrae, J. & Collier, N., 2008. Synonym set extraction from the biomedical literature by lexical pattern discovery. *BMC Bioinformatics*, 9(1), p.159.

McGrady, R., 2009. Gaming against the greater good. *First Monday*, 14(2). Available at: http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2215/2091.

McGuinness, D.L. et al., 2006. Investigations into Trust for Collaborative Information Repositories: A Wikipedia Case Study. In *Proceedings of the Workshop on Models of Trust for the Web*. Edinburgh, United Kingdom. Available at: http://ebiquity.umbc.edu\-/get\-/a\-/publication/274.pdf.

Medelyan, O. et al., 2009. Mining meaning from Wikipedia. *International Journal of Human Computer Studies*, 67(9), pp.716–754.

Mehler, A., Pustylnikov, O. & Diewald, N., 2010. Geography of social ontologies: Testing a variant of the Sapir-Whorf Hypothesis in the context of Wikipedia. Available at: http://dx.doi.org/10.1016/j.csl.2010.05.006.

Meishar-Tal, H. & Tal-Elhasid, E., 2008. Measuring collaboration in educational wikis - a methodological discussion. *International Journal of Emerging Technologies in Learning*, pp.46–9.

Mendoza, H.R., 2009. The WikiID: An Alternative Approach to the Body of Knowledge. *Journal of Interior Design*, 34(2), pp.1–18.

Mercer, J., 2007. Wikipedia and "open source" mental health information. *Scientific Review of Mental Health Practice*, 5(1), pp.88–92.

Messner, M. & South, J., 2010. Legitimizing Wikipedia: How US national newspapers frame and use the online encyclopedia in their coverage. *Journalism Practice*. Available at: http://www.informaworld.com/10.1080/17512786.2010.506060.

Meyer, M., Rensing, C. & Steinmetz, R., 2008. Using community-generated contents as a substitute corpus for metadata generation. *International Journal of Advanced Media and Communication*, 2(1), pp.59–72.

Mihalcea, Rada, 2007. Using Wikipedia for Automatic Word Sense Disambiguation. In *Proceedings of NAACL HLT 2007*. NAACL HLT. Rochester, New York, USA: Association for Computational Linguistics, pp. 196–203.

Mihalcea, Rada & Csomai, Andras, 2007. Wikify!: linking documents to encyclopedic knowledge. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*. pp. 233–242. Available at: http://0-portal.acm.org.mercury.concordia.ca/citation.cfm?id=1321440.1321475&coll=DL&dl=GUIDE&CFID=112020990&CFTOKEN=50968312&preflayout=flat [Accessed November 21, 2010].

Mika, P. et al., 2008. Learning to tag and tagging to learn: a case study on Wikipedia. *IEEE Intelligent Systems*, 23(5), pp.26–33.

Miller, N., 2005. Wikipedia and the disappearing "author." *ETC.: A Review of General Semantics*, 62(1), p.37.

Milne, D., Medelyan, O. & Witten, I.H., 2006. Mining Domain-Specific Thesauri from Wikipedia: A Case Study. In *Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence*. pp. 442–448. Available at: http://0-portal.acm.org.mercury.concordia.ca/citation.cfm?id=1248823.1249168&coll=DL&dl=GUIDE&CFID=112025803&CFTOKEN=32336862&preflayout=flat [Accessed November 21, 2010].

Milne, D. & Witten, I.H., 2008. Learning to link with wikipedia. In *17th ACM Conference on Information and Knowledge Management, CIKM'08, October 26, 2008 - October 30, 2008*. International Conference on Information and Knowledge Management, Proceedings. Napa Valley, CA, United

states: Association for Computing Machinery, pp. 509–518. Available at: http://dx.doi.org/10.1145/1458082.1458150 [Accessed November 5, 2010].

Milne, D., Witten, I.H. & Nichols, D.M., 2007. A knowledge-based search engine powered by Wikipedia. In *16th ACM Conference on Information and Knowledge Management, CIKM 2007, November 6, 2007 - November 9, 2007*. International Conference on Information and Knowledge Management, Proceedings. Lisboa, Portugal: Association for Computing Machinery, pp. 445–454. Available at: http://dx.doi.org/10.1145/1321440.1321504 [Accessed November 5, 2010].

Miquel Ribé, M. & Rodríguez, H., 2011. Cultural Configuration of Wikipedia: measuring Autoreferentiality in Different Languages. In *Proceedings of the International Conference Recent Advances in Natural Language Processing 2011*. Recent Advances in Natural Language Processing. Hissar, Bulgaria: RANLP 2011 Organising Committee, pp. 316–322. Available at: http://aclweb.org/anthology/R11-1044 [Accessed October 22, 2012].

Moeller, E., 2009. *Wikipedia Scholarly Survey Results*, Wikimedia Foundation.

Morse, G., 2008. A conversation with Jimmy Wales. *HARVARD BUSINESS REVIEW*, 86(4), p.26–+.

Moy, C. et al., 2010. Improving Science Education and Understanding through Editing Wikipedia. *Journal of Chemical Education*, 87(11), pp.1159–1162.

Muchnik, L. et al., 2007. Self-emergence of knowledge trees: Extraction of the Wikipedia hierarchies. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 76(1). Available at: http://dx.doi.org/10.1103/PhysRevE.76.016106 [Accessed November 5, 2010].

Mühlhauser, I. & Oser, F., 2008. Does Wikipedia provide evidence-based health care information? A content analysis. *Zeitschrift Für Evidenz, Fortbildung Und Qualität Im Gesundheitswesen*, 102(7), pp.441–448.

Müller-Birn, C. et al., 2010. Seeing similarity in the face of difference: enabling comparison of online production systems. *Social Network Analysis and Mining*. Available at: http://www.springerlink.com/content/u664v208279w7447/ [Accessed November 29, 2010].

Muller-Seitz, G. & Reger, G., 2009. Is open source software living up to its promises? Insights for open innovation management from two open source software-inspired projects. *R & D Management*, 39(4), pp.372–81.

Muller-Seitz, G. & Reger, G., 2010. "Wikipedia, the free encyclopedia" as a role model? Lessons for open innovation from an exploratory examination of the supposedly democratic-anarchic nature of Wikipedia. *International Journal of Technology Management*, 52(3-4), pp.457–476.

Murray, K.K., 2007. *Mass Spectrometry on Wikipedia: Open Source and Peer Review*, Available at: http://mass-spec.lsu.edu/presentations/ASMS_2007/ASMS_2007_ThP161.pdf.

Murugeshan, M.S., Lakshmi, K. & Mukherjee, S., 2010. A negative category based approach for Wikipedia document classification. *International Journal of Knowledge Engineering and Data Mining*, 1, pp.84–97.

Myers, M.D., 2008. *Qualitative Research in Business & Management*, Sage Publications Ltd.

N. Tkacz, 2007. Power, Visibility, Wikipedia. *SOUTHERN REVIEW-ADELAIDE-*, 40(2), p.5.

Nadamoto, A. et al., 2010. Extracting content holes by comparing community-type content with Wikipedia. *International Journal of Web Information Systems*, 6, pp.248–260.

Nature, 2011. *Survey results: Best face forward*, Available at: http://www.nature.com/nature/newspdf/reputation_survey.pdf.

Niederer, S. & Dijck, J. van, 2010. Wisdom of the crowd or technicity of content? Wikipedia as a sociotechnical system. *New Media & Society*, 12(8), pp.1368–1387.

Nielsen, F.Å., 2008a. Clustering of scientific citations in Wikipedia. In *Wikimania*. Available at: http://arxiv.org/abs/0805.1154.

Nielsen, F.Å., 2007. Scientific citations in Wikipedia. *First Monday*, 12(8). Available at: http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1997/1872.

Nielsen, F.Å., 2008b. Wikipedia - nørdernes sejr over vandalerne? Available at: http://www2.imm.dtu.dk/pubdb/views/edoc_download.php/5652/pdf/imm5652.pdf.

Nielsen, F.Å., 2012. Wikipedia Research and Tools: Review and Comments. Available at: http://ssrn.com/abstract=2129874.

Nix, E.M., 2010. Wikipedia: How It Works and How It Can Work for You. *History Teacher*, 43(2), pp.259–64.

Nov, O., 2009. Chapter 1 Information Sharing and Social Computing: Why, What, and Where? In Marvin V. Zelkowitz, ed. *Advances in Computers*. Elsevier, pp. 1–18. Available at: http://www.sciencedirect.com/science/article/pii/S0065245809010018 [Accessed October 9, 2012].

Nov, O., 2007. What motivates Wikipedians? *Communications of the ACM*, 50(11), pp.60–64.

Nov, O. & Kuk, G., 2008. Open source content contributors' response to free-riding : The effect of personality and context. *Computers in human behavior*, 24(6), pp.2848–2861.

Noveck, B., 2007. Wikipedia and the future of legal education. *Journal Of Legal Education*, 57(1), pp.3–9.

O'Neil, M., 2011. The sociology of critique in Wikipedia. *Critical studies in peer production*, (RS 1.2), pp.1–11.

Oboler, A., Steinberg, G. & Stern, R., 2010. The Framing of Political NGOs in Wikipedia through Criticism Elimination. *Journal of Information Technology & Politics*, 7(4), pp.284 – 299.

Okoli, C., 2009. A brief review of studies of Wikipedia in peer-reviewed journals. In *Digital Society, 2009. ICDS'09. Third International Conference on*. pp. 155–160. Available at: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4782868 [Accessed September 17, 2012].

Okoli, C. & Oh, W., 2007. Investigating recognition-based performance in an open content community: A social capital perspective. *Information and Management*, 44(3), pp.240–252.

Okoli, C. & Schabram, K., 2010. A guide to conducting a systematic literature review of information systems research. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1954824 [Accessed September 17, 2012].

Okoli, C. & Schabram, K., 2009a. Protocol for a systematic literature review of research on the Wikipedia. *Sprouts: Working Papers in Information Systems*, 9(65). Available at: http://sprouts.aisnet.org/9-65.

Okoli, C. & Schabram, K., 2009b. Protocol for a systematic literature review of research on the Wikipedia. In *Proceedings of the International Conference on Management of Emergent Digital EcoSystems (MEDES)*. International ACM Conference on Management of Emergent Digital EcoSystems (MEDES). Lyon, France: Association for Computing Machinery, p. 73. Available at: http://dl.acm.org/citation.cfm?id=1643912 [Accessed September 17, 2012].

Okoli, C., Schabram, K. & Kader, B.A., 2009. From the Academy to the Wiki: Practical Applications of Scholarly Research on Wikipedia. In *Proceedings of Wikimania*. Wikimania. Buenos Aires: Wik. Available at: http://chitu.okoli.org/images/stories/bios/pro/research/open/Okolietal2009Wikimania.pdf [Accessed September 17, 2012].

Oreg, S. & Nov, O., 2008. Exploring Motivations for Contributing to Open Source: Initiatives: The Roles of Contribution Context and Personal Values. *Computers in Human Behavior*, 24(5), pp.2055–2073.

Ortega, Felipe & Gonzalez-Barahona, J.M., 2007. Quantitative analysis of the wikipedia community of users. In *Proceedings of the 2007 international symposium on Wikis*. pp. 75 – 86. Available at: http://0-portal.acm.org.mercury.concordia.ca/citation.cfm?id=1296951.1296960&coll=DL&dl=GUIDE&CFID=112020990&CFTOKEN=50968312&preflayout=flat [Accessed November 21, 2010].

Ortega, Felipe, Gonzalez-Barahona, J.M. & Robles, G., 2008. On the inequality of contributions to wikipedia. In *41st Annual Hawaii International Conference on System Sciences 2008, HICSS, January 7, 2008 - January 10, 2008*. Proceedings of the Annual Hawaii International Conference on System Sciences. Big Island, HI, United states: Inst. of Elec. and Elec. Eng. Computer Society. Available at: http://dx.doi.org/10.1109/HICSS.2008.333 [Accessed November 5, 2010].

Ortega, Felix, 2009. *Wikipedia. A quantitative analysis*. Doctoral thesis. Madrid, Spain: Universidad Rey Juan Carlos. Available at: http://libresoft.es/Members/jfelipe/thesis-wkp-quantanalysis/view.

Otto, P. & Simon, M., 2008. Dynamic perspectives on social characteristics and sustainability in online community networks. *System Dynamics Review*, 24(3), pp.321–47.

Overell, S. & Ruger, S., 2008. Using co-occurrence models for place name disambiguation. *International Journal of Geographical Information Science*, 22(3), pp.265–87.

Overell, Simon, Sigurbjornsson, B. & Van Zwol, R., 2009. Classifying tags using open content resources. In *2nd ACM International Conference on Web Search and Data Mining, WSDM'09, February 9, 2009 - February 12, 2009*. Proceedings of the 2nd ACM International Conference on Web Search and Data Mining, WSDM'09. Barcelona, Spain: Association for Computing Machinery, pp. 64–73. Available at: http://dx.doi.org/10.1145/1498759.1498810 [Accessed November 6, 2010].

Page, R., 2010. Wikipedia as an encyclopaedia of life. *Organisms Diversity & Evolution*, 10(4), pp.343–349.

Pak, A.N. & Chung, C.-W., 2010. A wikipedia matching approach to contextual advertising. *World Wide Web*, 13(3), pp.251–274.

Pamkowska, M., 2008. Autopoiesis in virtual organizations. *Informatica Economica*, 12(1), pp.33–9.

Pantel, P. et al., 2009. Web-scale distributional similarity and entity set expansion. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2 - Volume 2*. pp. 938–947. Available at: http://0-portal.acm.org.mercury.concordia.ca/citation.cfm?id=1699571.1699635&coll=DL&dl=GUIDE&CFID=112015986&CFTOKEN=43703661&preflayout=flat [Accessed November 21, 2010].

Park, T.K., 2011. The visibility of Wikipedia in scholarly publications. *First Monday*, 16(8-1).

Patch, P., 2010. Meeting Student Writers Where They Are: Using Wikipedia to Teach Responsible Scholarship. , 37(3), pp.278–85.

Pehcevski, J. et al., 2010. Entity ranking in Wikipedia: utilising categories, links and topic difficulty prediction. *Information Retrieval*, 13(5), p.568.

Pekárek, M. & Pötzsch, S., 2009. A comparison of privacy issues in collaborative workspaces and social networks. *Identity in the Information Society*, 2(1), pp.81–93.

Pender, M.P. et al., 2008. Putting Wikipedia to the test: a case study. In *The Special Libraries Association Annual Conference*. Available at: http://espace.library.uq.edu.au/eserv/UQ:193433/SLA_Paper.pdf.

Pentzold, C., 2009. Fixing the floating gap: The online encyclopaedia Wikipedia as a global memory place. *Memory Studies*, 2(2), pp.255–272.

Pentzold, C., 2011. Imagining the Wikipedia community: what do Wikipedia authors mean when they write about their "community"? Available at: http://nms.sagepub.com/content/early/2010/11/11/1461444810378364.

Pentzold, C. & Seidenglanz, S., 2006. Foucault@Wiki: first steps towards a conceptual framework for the analysis of Wiki discourses. In *Proceedings of the 2006 international symposium on Wikis*. WikiSym '06. New York, NY, USA: ACM, pp. 59–68. Available at: http://doi.acm.org/10.1145/1149453.1149468 [Accessed October 22, 2012].

Peoples, L.F., 2009. The Citation of Wikipedia in Judicial Opinions. *Yale Journal of Law & Technology*. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1272437.

Perea-Ortega, J.M. et al., 2010. Using web sources for improving video categorization. , pp.1–14.

Perona, P., 2010. Vision Of A Visipedia. *Proceedings of the IEEE*, 98(8), pp.1526–34.

Petticrew, M. & Roberts, H., 2006. *Systematic Reviews in the Social Sciences: A Practical Guide* 1st ed., Wiley-Blackwell.

Pfeil, U., Zaphiris, P. & Ang, C.S., 2006. Cultural differences in collaborative authoring of Wikipedia. *Journal of Computer-Mediated Communication*, 12(1), pp.88–113.

Plaza, B., 2011. Google Analytics for measuring website performance. *Tourism Management*, 32(3), pp.477–481.

Poderi, G., 2009. Comparing featured article groups and revision patterns correlations in Wikipedia. *First Monday*, 14(5), p.9 pp.

Pöllä, M. & Honkela, T., 2010. Negative Selection of Written Language Using Character Multiset Statistics. *Journal of Computer Science and Technology*, 25(6), pp.1256–1266.

Pollard, E.A., 2008. Raising the Stakes: Writing about Witchcraft on Wikipedia. *History Teacher*, 42(1), pp.9–24.

Polukarova, N.A., 2007. The concept of open editing from the copyright viewpoint. *Automatic Documentation and Mathematical Linguistics*, 41(3), pp.104–7.

Ponzetto, S.P. & Strube, M., 2007a. Deriving a large scale taxonomy from Wikipedia. In *AAAI-07/IAAI-07 Proceedings: 22nd AAAI Conference on Artificial Intelligence and the 19th Innovative Applications of Artificial Intelligence Conference, July 22, 2007 - July 26, 2007*. Proceedings of the National Conference on Artificial Intelligence. Vancouver, BC, Canada: American Association for Artificial Intelligence, pp. 1440–1445.

Ponzetto, S.P. & Strube, M., 2007b. Knowledge derived from wikipedia for computing semantic relatedness. *Journal of Artificial Intelligence Research*, 30, pp.181–212.

Potthast, M. et al., 2010. Cross-language plagiarism detection. *Language Resources and Evaluation*, pp.1–18.

Potthast, M., Stein, B. & Gerling, R., 2008. Automatic Vandalism Detection in Wikipedia. In *Advances in Information Retrieval*. Lecture Notes in Computer Science. Berlin/Heidelberg: Springer, pp. 663–668.

Prasarnphanich, P. & Wagner, C., 2009a. Explaining the Sustainability of Digital Ecosystems based on the Wiki Model through Critical Mass Theory. *Industrial Electronics, IEEE Transactions on*, PP(99), p.1.

Prasarnphanich, P. & Wagner, C., 2009b. The role of wiki technology and altruism in collaborative knowledge creation. *Journal of Computer Information Systems*, 49(4), pp.33–41.

Preece, J. & Shneiderman, B., 2009. The Reader-to-Leader Framework: Motivating Technology-Mediated Social Participation. *AIS Transactions on Human-Computer Interaction*, 1(1), pp.13–32.

Priedhorsky, R. et al., 2007. Creating, destroying, and restoring value in wikipedia. In *2007 International ACM Conference on Supporting Group Work, GROUP'07, November 4, 2007 - November 7, 2007*. GROUP'07 - Proceedings of the 2007 International ACM Conference on Supporting Group Work. Sanibel Island, FL, United states: Association for Computing Machinery, pp. 259–268. Available at: http://dx.doi.org/10.1145/1316624.1316663 [Accessed November 5, 2010].

Priedhorsky, R., 2010. *The value of geographic wikis*. United States -- Minnesota: University of Minnesota. Available at: http://proquest.umi.com/pqdweb?did=2159254821&Fmt=7&clientId=10306&RQT=309&VName=PQD.

Purdy, J.P., 2006. *Digital archives and the turn to design*. United States -- Illinois: University of Illinois at Urbana-Champaign. Available at: http://proquest.umi.com/pqdweb?did=1253507701&Fmt=7&clientId=10306&RQT=309&VName=PQD.

Purdy, J.P., 2010. The Changing Space of Research: Web 2.0 and the Integration of Research and Writing Environments. *Computers and Composition*, 27(1), pp.48–58.

Purdy, J.P., 2009. When the Tenets of Composition Go Public: A Study of Writing in Wikipedia. *College Composition and Communication*, 61.

Quack, T., Leibe, B. & Gool, L.V., 2008. World-scale mining of objects and events from community photo collections. In *Proceedings of the 2008 international conference on Content-based image and video retrieval*. pp. 47–56. Available at: http://0-portal.acm.org.mercury.concordia.ca/citation.cfm?id=1386352.1386363&coll=DL&dl=GUIDE&CFID=112025803&CFTOKEN=32336862&preflayout=flat [Accessed November 21, 2010].

Radtke, P.J. & Munsell, J.F., 2010. Wikipedia as a tool for forestry outreach. *Journal of Forestry*, 108(7), pp.354–359.

Rahman, M.M., 2007. An Analysis of Wikipedia. *Journal of Information Technology Theory and Application (JITTA)*, 9(3), p.81.

Rahman, M.M., 2006. *Essays analyzing blogs and Wikipedia*. United States -- Kansas: The University of Kansas. Available at: http://proquest.umi.com/pqdweb?did=1126778281&Fmt=7&clientId=10306&RQT=309&VName=PQD.

Rahurkar, M. et al., 2010. Image Interpretation Using Large Corpus: Wikipedia. *Proceedings of the IEEE*, 98(8), pp.1509–25.

Rajagopalan, M.S. et al., 2010. Accuracy of cancer information on the Internet: A comparison of a Wiki with a professionally maintained database. *Bodine Journal*, 3(1), p.7s.

Rajagopalan, M.S. et al., 2011. Patient-oriented cancer information on the Internet: a comparison of Wikipedia and a professionally maintained database. *Journal of Oncology Practice*, 7(5), pp.319–323.

Rand, A.D., 2010. Mediating at the Student-Wikipedia Intersection. *Journal of Library Administration*, 50(7/8), pp.923–932.

Ransbotham, S. & Kane, G.C., 2011. Membership Turnover and Collaboration Success in Online Communities: Explaining Rises and Falls from Grace in Wikipedia. *MIS Quarterly*, 35(3), pp.613–627.

Rask, M., 2008. The reach and richness of Wikipedia: Is Wikinomics only for rich countries. *First Monday*, 13(6), p.10 pp.

Ratcliff, D., 2004. 15 Methods of Data Analysis in Qualitative Research. Available at: http://fycs.ifas.ufl.edu/swisher/6802_12/15methods_Qual_An.pdf [Accessed October 4, 2012].

Ratkiewicz, J. et al., 2010. Characterizing and modeling the dynamics of online popularity. *Physical Review Letters*, 105(15). Available at: http://dx.doi.org/10.1103/PhysRevLett.105.158701 [Accessed November 5, 2010].

Ray, S.K., Singh, S. & Joshi, B.P., 2010. A semantic approach for question classification using WordNet and Wikipedia. *Pattern Recognition Letters*, 31(13), pp.1935–1943.

Read, B., 2007. Middlebury College History Department Limits Students' Use of Wikipedia. *The Chronicle of Higher Education*, 53(24), p.A39.

Reagle, J.M., 2010a. "Be Nice": Wikipedia norms for supportive communication. *New Review of Hypermedia and Multimedia*, 16(1-2), pp.161–180.

Reagle, J.M., 2010b. *Good Faith Collaboration: The Culture of Wikipedia*, Cambridge, Mass: MIT Press.

Reagle, J.M., 2008. *In good faith: Wikipedia collaboration and the pursuit of the universal encyclopedia*. United States -- New York: New York University. Available at: http://proquest.umi.com/pqdweb?did=1510560441&Fmt=7&clientId=10306&RQT=309&VName=PQD.

Rector, L.H., 2008. Comparison of Wikipedia and other encyclopedias for accuracy, breadth, and depth in historical articles. *Reference Services Review*, 36(1), pp.7–22.

Řehůřek, R., 2010. Fast and Faster: A Comparison of Two Streamed Matrix Decomposition Algorithms. In *NIPS 2010 Workshop on Low-rank Methods for Large-scale Machine Learning*. NIPS 2010 Workshop on Low-rank Methods for Large-scale Machine Learning. Vancouver, Canada. Available at: http://www.muni.cz/research/publications/914342 [Accessed October 15, 2012].

Reich, E.S., 2011. Online reputations: best face forward. *Nature*, 473, pp.138–139.

Rodríguez, R., 2007. Liberating Epistemology: Wikipedia and the Social Construction of Knowledge. *Religious Studies and Theology*, 26(2), p.173.

Rosenzweig, R., 2006. Can History Be Open Source? Wikipedia and the Future of the Past. *Journal of American History*, 93(1), pp.117–146.

Roth, C., 2007. Viable wikis: struggle for life in the wikisphere. In *Proceedings of the 2007 international symposium on Wikis*. WikiSym '07. New York, NY, USA: ACM, pp. 119–124. Available at: http://doi.acm.org/10.1145/1296951.1296964 [Accessed October 22, 2012].

Royal, C. & Kapila, D., 2009. What's on Wikipedia, and what's not ... ? Assessing completeness of information. *Social Science Computer Review*, 27(1), pp.138–148.

Rubin, A. & Rubin, E., 2010. Informed Investors and the Internet. *Journal of Business Finance & Accounting*, 37(7-8), pp.841–865.

Ruiz-Casado, M., Alfonseca, E. & Castells, P., 2007. Automatising the learning of lexical patterns: An application to the enrichment of WordNet by extracting semantic relationships from Wikipedia. *Data and Knowledge Engineering*, 61(3), pp.484–499.

Rush, E.K. & Tracy, S.J., 2010. Wikipedia as Public Scholarship: Communicating Our Impact Online. *Journal of Applied Communication Research*, 38(3), pp.309–315.

Sanger, L.M., 2009. The Fate of Expertise after Wikipedia. *Episteme - Edinburgh*, 6(1).

Santana, A. & Wood, D.J., 2009. Transparency and social responsibility issues for Wikipedia. *Ethics and Information Technology*, 11(2), pp.133–144.

Santos, M., 2009. *Toward another rhetoric: Web 2.0, Levinas, and taking responsibility for response ability*. United States -- Indiana: Purdue University. Available at: http://proquest.umi.com/pqdweb?did=2056028761&Fmt=7&clientId=10306&RQT=309&VName=PQD.

Sara Monaci, 2009. Quality assessment process in Wikipedia's Vetrina: the role of the community's policies and rules. *Observatorio (OBS*)*, 3(1). Available at: http://obs.obercom.pt/index.php/obs/article/viewArticle/240 [Accessed January 18, 2011].

Schenkel, R., Suchanek, F.M. & Kasneci, G., 2007. YAWN: A semantically annotated Wikipedia XML corpus. In *12. Symposium on Database Systems for Business, Technology and the Web of the German Society for Computer Science*. 12th Symposium on Database Systems for Business, Technology and the Web of the German Society for Computer Science. pp. 277–291. Available at: http://suchanek.name/work/publications/btw2007.pdf [Accessed October 4, 2012].

Schiltz, M., Truyen, F. & Coppens, H., 2007. Cutting the Trees of Knowledge: Social Software, Information Architecture and Their Epistemic Consequences. *Thesis Eleven*, 89(1), pp.94–114.

Schroer, J. & Hertel, G., 2009. Voluntary engagement in an open Web-based encyclopedia: Wikipedians and why they do it. *Media Psychology*, 12(1), pp.96–120.

Schweitzer, N.J., 2008. Wikipedia and Psychology: Coverage of Concepts and Its Use by Undergraduate Students. *Teaching of Psychology*, 35(2), pp.81–85.

Shachaf, P., 2009. The paradox of expertise: is the Wikipedia Reference Desk as good as your library? *Journal of Documentation*, 65(6), pp.977–996.

Shachaf, P. & Hara, N., 2010. Beyond vandalism: Wikipedia trolls. *Journal of Information Science*, 36(3), pp.357–370.

Shaw, D., 2008. Wikipedia in the newsroom. *American Journalism Review*, 30(1), pp.40–45.

Shim, J.P. & Yang, J., 2009. Why is Wikipedia not more widely accepted in Korea and China? factors affecting knowledge-sharing adoption. *Decision Line*, 40(2), pp.12–15.

Sigurdsson, M. & Halling, S.C., 2007. *Zeeker: A topic-based search engine*. Kongens Lyngby, Denmark: Informatics and Mathematical Modelling, Technical University of Denmark. Available at: http://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=5502.

Silva, F.N. et al., 2010. Identifying the borders of mathematical knowledge. *Journal of Physics A: Mathematical and Theoretical*, 43(32), p.325202 (7 pp.).

Simma, A., 2010. *Modeling events in time using cascades of Poisson processes*. United States -- California: University of California, Berkeley. Available at: http://proquest.umi.com/pqdweb?did=2128789941&Fmt=7&clientId=10306&RQT=309&VName=PQD.

Sipos, R. et al., 2009. Demo: HistoryViz — Visualizing Events and Relations Extracted from Wikipedia. In L. Aroyo et al., eds. *The Semantic Web: Research and Applications: 6th European Semantic Web Conference, ESWC 2009 Heraklion, Crete, Greece, May 31â€"June 4, 2009 Proceedings*. Lecture Notes in Computer Science. Berlin/Heidelberg: Springer, pp. 903–907.

Smets, K., Goethals, B. & Verdonk, B., 2008. *Automatic Vandalism Detection in Wikipedia: Towards a Machine Learning Approach*, Available at: http://www.adrem.ua.ac.be/ksmets/talks/Benelearn08_abstract.pdf.

Spinellis, D. & Louridas, P., 2008. The collaborative organization of knowledge. *Communications of the ACM*, 51(8), pp.68–73.

Spoerri, A., 2007a. Visualizing the overlap between the 100 most visited pages on Wikipedia for September 2006 to January 2007. *First Monday*, 12(4), pp.0–0.

Spoerri, A., 2007b. What is popular on Wikipedia and why? *First Monday*, 12(4), pp.0–0.

Stankus, T. & Spiegel, S.E., 2010. Wikipedia, scholarpedia, and references to books in the brain and behavioral sciences: A comparison of cited sources and recommended readings in matching free online encyclopedia entries. *Science and Technology Libraries*, 29(1-2), pp.144–164.

Stettler, R., 2008. Reframing semiotic telematic knowledge spaces, and the anthropological challenge to designing interhuman relations. *Technoetic Arts: a journal of speculative research*, 6(2), pp.163–170.

Stoddard, M.M., 2009. *Judicial Citation to Wikipedia in Published Federal Court Opinions*. School of Information and Library Science of the University of North Carolina at Chapel Hill.

Stokes, N. et al., 2008. An empirical study of the effects of NLP components on Geographic IR performance. *INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE*, 22(3), pp.247–264.

Stone, B., Dennis, S. & Kwantes, P.J., 2010. Comparing Methods for Single Paragraph Similarity Analysis. *Topics in Cognitive Science*, p.no.

Strube, M. & Ponzetto, S.P., 2006. WikiRelate! Computing Semantic Relatedness Using Wikipedia. In *Proceedings of the Twenty-First AAAI Conference on Artificial Intelligence*. Menlo Park, Califonia: AAAI Press, pp. 1419–1424. Available at: http://www.eml-research.de/strube/papers/aaai06.pdf.

Stvilia, B. et al., 2007. A framework for information quality assessment. *Journal of the American Society for Information Science and Technology*, 58(12), pp.1720–1733.

Stvilia, B. et al., 2008. Information quality work organization in Wikipedia. *Journal of the American Society for Information Science and Technology*, 59(6), pp.983–1001.

Stvilia, B., Al-faraj, A. & Yi, Y.J., 2009. Issues of cross-contextual information quality evaluation—The case of Arabic, English, and Korean Wikipedias. *Library & information science research*, 31(4), pp.232–239.

Stvilia, B. & Gasser, L., 2008. An activity theoretic model for information quality change. *First Monday*, 13(4), p.12 pp.

Su, A.I. et al., 2004. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proceedings of the National Academy of Sciences of the United States of America*, 101(16), pp.6062–6067.

Suchanek, F.M., Kasneci, G. & Weikum, G., 2007. YAGO: A Core of Semantic Knowledge Unifying WordNet and Wikipedia. In *Proceedings of the 16th international conference on World Wide Web*. ACM, pp. 697 – 706.

Suchanek, F.M., Kasneci, G. & Weikum, G., 2008. YAGO: A Large Ontology from Wikipedia and WordNet. *Web Semantics*, 6(3), pp.203–217.

Suh, B. et al., 2008. Lifting the veil: Improving accountability and social transparency in Wikipedia with WikiDashboard. In *26th Annual CHI Conference on Human Factors in Computing Systems, CHI 2008, April 5, 2008 - April 10, 2008*. Conference on Human Factors in Computing Systems - Proceedings. Florence, Italy: Association for Computing Machinery, pp. 1037–1040. Available at: http://dx.doi.org/10.1145/1357054.1357214 [Accessed November 6, 2010].

Suh, B. et al., 2007. Us vs. Them: Understanding Social Dynamics in Wikipedia with Revert Graph Visualizations. In *IEEE Symposium on Visual Analytics Science and Technology, 2007. VAST 2007*. IEEE Symposium on Visual Analytics Science and Technology, 2007. VAST 2007. pp. 163 –170.

Sundin, O., 2011. Janitors of Knowledge: Constructing Knowledge in the Everyday Life of Wikipedia Editors. *Journal of Documentation*, 67(5).

Sundin, O. & Francke, H., 2009. In search of credibility: pupils' information practices in learning environments. *Information Research: An International Electronic Journal*, 14(4), p.19 pp.

Syed, Z., 2010. *Wikitology: A novel hybrid knowledge base derived from wikipedia*. United States -- Maryland: University of Maryland, Baltimore County. Available at: http://proquest.umi.com/pqdweb?did=2157352461&Fmt=7&clientId=10306&RQT=309&VName=PQD.

Tan, B. & Peng, F., 2008. Unsupervised query segmentation using generative language models and wikipedia. In *Proceeding of the 17th international conference on World Wide Web*. pp. 347–356. Available at: http://0-portal.acm.org.mercury.concordia.ca/citation.cfm?id=1367497.1367545&coll=DL&dl=GUIDE&CFID=112015986&CFTOKEN=43703661&preflayout=flat [Accessed November 21, 2010].

Tann, C. & Sanderson, M., 2009. Are web-based informational queries changing? *Journal of the American Society for Information Science and Technology*, 60(6), pp.1290–1293.

Tao Guo et al., 2009. Codifying collaborative knowledge: using Wikipedia as a basis for automated ontology learning. *Knowledge Management Research & Practice*, 7(3), pp.206–17.

Thelwall, M. & Stuart, D., 2007. RUOK? Blogging Communication Technologies During Crises. *Journal of Computer-Mediated Communication*, 12(2007), pp.523–548.

Theobald, M. et al., 2008. TopX: efficient and versatile top-k query processing for semistructured data. *The VLDB Journal — The International Journal on Very Large Data Bases*, 17(1), pp.81 – 115.

Thom-Santelli, J., 2010. *Expressing territoriality in online collaborative environments*. United States -- New York: Cornell University. Available at: http://proquest.umi.com/pqdweb?did=1975863181&Fmt=7&clientId=10306&RQT=309&VName=PQD.

Timme Bisgaard Munk, 2009. Why wikipedia: Self-efficacy and self-esteem in a knowledge-political battle for an egalitarian epistemology. *Observatorio (OBS*)*, 3(4). Available at: http://obs.obercom.pt/index.php/obs/article/view/248 [Accessed January 18, 2011].

Tollefsen, D.P., 2009. Wikipedia and the Epistemology of Testimony. *Episteme - Edinburgh*, 6(1).

Tseng, S.-M. & Huang, J.-S., 2011. The correlation between Wikipedia and knowledge sharing on job performance. *Expert Systems with Applications*, 38(5), pp.6118–6124.

Tumlin, M. et al., 2007. Collectivism vs. individualism in a wiki world: Librarians respond to Jaron Lanier's essay "Digital Maoism: The Hazards of the New Online Collectivism." *SERIALS REVIEW*, 33(1), pp.45–53.

Turdakov, D. & Velikhov, P., 2008. Semantic Relatedness Metric for Wikipedia Concepts Based on Link Analysis and its Application to Word Sense Disambiguation. In *Spring Young Researcher's Colloquium On Database and Information Systems*. Spring Young Researcher's Colloquium On Database and Information Systems. St.-Petersburg, Russia. Available at: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.143.864.

Turdakov, D.Y. & Kuznetsov, S.D., 2010. Automatic word sense disambiguation based on document networks. *Programming and Computer Software*, 36(1), pp.11–18.

Urdaneta, G., Pierre, G. & Steen, M. van, 2009. Wikipedia workload analysis for decentralized hosting. *Computer Networks: The International Journal of Computer and Telecommunications Networking*, 53(11), pp.1830–1845.

Vassia Atanassova, 2011. Topics of Bioengineering in Wikipedia. *Bioautomation*, 13. Available at: http://www.biomed.bas.bg/bioautomation/2009/vol_13.3/files/13.3_4.1.pdf.

Vechtomova, O., 2010. Facet-based opinion retrieval from blogs. *Information Processing and Management*, 46(1), pp.71–88.

Veltman, K.H., 2005. Access, claims and quality on the internet - Future challenges. *Progress in Informatics*, (2), pp.17–40.

Viégas, F.B., Wattenberg, M., Kriss, J., et al., 2007. Talk before you type: Coordination in Wikipedia. In *40th Annual Hawaii International Conference on System Sciences 2007, HICSS'07, January 3,*

2007 - January 6, 2007. Proceedings of the Annual Hawaii International Conference on System Sciences. 40th Annual Hawaii International Conference on System Sciences, 2007. HICSS 2007. Big Island, HI, United states: Inst. of Elec. and Elec. Eng. Computer Society, p. 78. Available at: http://dx.doi.org/10.1109/HICSS.2007.511 [Accessed November 5, 2010].

Viégas, F.B., Wattenberg, M. & Dave, K., 2004. Studying cooperation and conflict between authors with history flow visualizations. In E. Dykstra-Erickson & M. Tscheligi, eds. *Proceedings of the SIGCHI conference on Human factors in computing systems*. New York, NY, USA: ACM, pp. 575–582. Available at: http://0-portal.acm.org.mercury.concordia.ca/citation.cfm?id=985692.985765&coll=DL&dl=GUIDE&CFID=112025803&CFTOKEN=32336862&preflayout=flat.

Viégas, F.B., Wattenberg, M. & McKeon, M.M., 2007. The Hidden Order of Wikipedia. In *Lecture Notes in Computer Science*. Berlin/Heidelberg: Springer, pp. 445–454.

Vivienne Waller, 2011. The search queries that took Australian Internet users to Wikipedia. *Information Research*, 16(2). Available at: http://informationr.net/ir/16-2/paper476.html [Accessed July 28, 2011].

Voss, J., 2005. Measuring Wikipedia. In *Proceedings of the 10th International Conference of the International Society for Scientometrics and Informetrics*. International Conference of the International Society for Scientometrics and Informetrics. Stockholm. Available at: http://eprints.rclis.org/handle/10760/6207#.TvH7B9Tj52A.

Wagner, C., 2005. Breaking the knowledge acquisition bottleneck through conversational knowledge management. *Information Resources Management Journal*, 19(1), pp.70–83.

Wales, J., 2004. Wikipedia Founder Jimmy Wales Responds - Slashdot. *Slashdot*. Available at: http://slashdot.org/story/04/07/28/1351230/wikipedia-founder-jimmy-wales-responds [Accessed October 4, 2012].

Wang, K. et al., 2008. The Adoption of Wikipedia: A Community- and Information Quality-Based View. *PACIS 2008 Proceedings*. Available at: http://aisel.aisnet.org/pacis2008/50 [Accessed December 24, 2010].

Wang, P. et al., 2009. Using Wikipedia knowledge to improve text classification. *Knowledge and Information Systems*, 19(3), pp.265–281.

Wang, P. & Domeniconi, C., 2008. Building semantic kernels for text classification using wikipedia. In *14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2008, August 24, 2008 - August 27, 2008*. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. International Conference on Knowledge Discovery and Data Mining. Las Vegas, NV, United states: Association for Computing Machinery, pp. 713–721. Available at: http://dx.doi.org/10.1145/1401890.1401976 [Accessed November 5, 2010].

Wannemacher, K., 2011. Experiences and perspectives of Wikipedia use in higher education. *International Journal of Management in Education*, 5(1), pp.79–92.

Waters, N.L., 2007. Why you can't cite Wikipedia in my class. *Communications of the ACM*, 50(9), pp.15–17.

Webster, J. & Watson, R.T., 2002. Analyzing the past to prepare for the future: writing a literature review. *MIS Q.*, 26(2), p.xiii–xxiii.

Wedemeyer, B. et al., 2008. Quality of the science articles on the English Wikipedia: Preliminary results. In *Wikimania 2008*. Wikimania. Alexandria, Egypt. Available at: http://www.youtube.com/watch?v=B7bCZbHHeZI.

Weiss, S., Urso, P. & Molli, P., 2010. Logoot-undo: Distributed collaborative editing system on P2P networks. *IEEE Transactions on Parallel and Distributed Systems*, 21(8), pp.1162–1174.

Weld, D.S. et al., 2008. Intelligence in wikipedia. In *Proceedings of the 23rd national conference on Artificial intelligence - Volume 3*. pp. 1609–1614. Available at: http://0-portal.acm.org.mercury.concordia.ca/citation.cfm?id=1620270.1620344&coll=DL&dl=GUIDE&CFID=112059492&CFTOKEN=13405793&preflayout=flat [Accessed November 22, 2010].

West, A.G., Kannan, S. & Lee, I., 2010. Spatio-temporal analysis of Wikipedia metadata and the STiki anti-vandalism tool. In *Proceedings of the 6th International Symposium on Wikis and Open Collaboration*. New York, NY, USA: ACM, p. 18.

West, K. & Williamson, J., 2009. Wikipedia: friend or foe? *Reference services review*, 37(3), pp.260–271.

Wielsch, D., 2010. Governance of Massive Multiauthor Collaboration – Linux, Wikipedia, and Other Networks: Governed by Bilateral Contracts, Partnerships, or Something in Between? *jipitec*, 1(2). Available at: http://www.jipitec.eu/issues/jipitec-1-2-2010/2618 [Accessed January 18, 2011].

Wikipedia contributors, 2012. Epistemology. *Wikipedia, the free encyclopedia*. Available at: http://en.wikipedia.org/w/index.php?title=Epistemology&oldid=515500356 [Accessed October 4, 2012].

Wilkinson, D.M. & Huberman, B.A., 2007a. Assessing the value of cooperation in Wikipedia. *First Monday*, 12(4), p.13 pp.

Wilkinson, D.M. & Huberman, B.A., 2007b. Cooperation and quality in Wikipedia. In *2007 ACM SIGPLAN International Conference on Object-Oriented Programming, Systems, Languages and Applications, OOPSLA - 2007 International Symposium on Wikis, WikiSym, October 21, 2007 - October 25, 2007*. Proceedings of the Conference on Object-Oriented Programming Systems, Languages, and Applications, OOPSLA. Montreal, QC, Canada: Association for Computing Machinery, pp. 157–164. Available at: http://dx.doi.org/10.1145/1296951.1296968 [Accessed November 6, 2010].

Willinsky, J., 2008. Socrates Back on the Street: Wikipedia's Citing of the "Stanford Encyclopedia of Philosophy." , 2, pp.1269–88.

Willinsky, J., 2007. What open access research can do for Wikipedia. *First Monday*, 12(3), pp.0–0.

Witzleb, N., 2009. Engaging with the world: Students of comparative law write for Wikipedia. *Legal Education Review*, 19(1/2), pp.83–97.

Wong, W., Liu, W. & Bennamoun, M., 2007. Tree-traversing ant algorithm for term clustering based on featureless similarities. *Data Mining and Knowledge Discovery*, 15(3), pp.349–381.

Wray, K.B., 2009. The Epistemic Cultures of Science and Wikipedia: A Comparison. *Episteme*, 6(1), pp.38–51.

Wu, F., Hoffmann, R. & Weld, D.S., 2008. Information extraction from Wikipedia: Moving down the long tail. In *14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2008, August 24, 2008 - August 27, 2008*. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Las Vegas, NV, United states: Association for Computing Machinery, pp. 731–739. Available at: http://dx.doi.org/10.1145/1401890.1401978 [Accessed November 5, 2010].

Wu, F. & Weld, D.S., 2007. Autonomously semantifying wikipedia. In *16th ACM Conference on Information and Knowledge Management, CIKM 2007, November 6, 2007 - November 9, 2007*. International Conference on Information and Knowledge Management, Proceedings. Lisboa, Portugal: Association for Computing Machinery, pp. 41–50. Available at: http://dx.doi.org/10.1145/1321440.1321449 [Accessed November 5, 2010].

Xiang, E.W. et al., 2010. Bridging domains using world wide knowledge for transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(6), pp.770–783.

Yang, H.-L. & Lai, C.-Y., 2010a. Motivations of Wikipedia content contributors. *Computers in human behavior*, 26(6), pp.1377–1383.

Yang, H.-L. & Lai, C.-Y., 2010b. Understanding Knowledge Sharing Behaviour in Wikipedia. *Behaviour & Information Technology*. Available at: http://www.informaworld.com/10.1080/0144929X.2010.516019.

Yasseri, T. & Kertész, J., 2012. Value production in a collaborative environment. *arXiv:1208.5130*. Available at: http://arxiv.org/abs/1208.5130 [Accessed October 15, 2012].

Yermilov, I. et al., 2008. What Is the Quality of Surgery-Related Information on the Internet? Lessons Learned from a Standardized Evaluation of 10 Common Operations. *Journal of the American College of Surgeons*, 207(4), pp.580–586.

Younger, P., 2010. Using wikis as an online health information resource. *Nursing Standard*, 24(36), pp.49–56.

Yu, J., Thom, J.A. & Tam, A., 2009. Requirements-oriented methodology for evaluating ontologies. *Information Systems*, 34(8), pp.686–711.

Yuan, Y. et al., 2009. The Diffusion of a Task Recommendation System to Facilitate Contributions to an Online Community. *JOURNAL OF COMPUTER-MEDIATED COMMUNICATION*, 15(1), pp.32–59.

Zaragoza, Hugo et al., 2007. Ranking very many typed entities on wikipedia. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*. Lisbon, Portugal: ACM, pp. 1015–1018. Available at: http://0-portal.acm.org.mercury.concordia.ca/citation.cfm?id=1321440.1321599&coll=DL&dl=GUIDE&CFID=112015986&CFTOKEN=43703661&preflayout=flat [Accessed November 21, 2010].

Zeng, H. et al., 2006. Computing trust from revision history. In *STAR. Vol. 44*. Available at: http://ebiquity.umbc.edu\-/get\-/a\-/publication/302.pdf.

Zesch, T., Mueller, C. & Gurevych, I., 2008. Extracting Lexical Semantic Knowledge from Wikipedia and Wiktionary. In *Proceedings of the Conference on Language Resources and Evaluation (LREC)*.

Zesch, Torsten & Gurevych, Iryna, 2009. Wisdom of crowds versus wisdom of linguists – measuring the semantic relatedness of words. *Natural Language Engineering*, 16(01), p.25.

Zesch, Torsten, Müller, C. & Gurevych, Iryna, 2008. Using wiktionary for computing semantic relatedness. In *Proceedings of the 23rd national conference on Artificial intelligence - Volume 2*. pp. 861–866. Available at: http://0-portal.acm.org.mercury.concordia.ca/citation.cfm?id=1620163.1620206&coll=DL&dl=GUIDE&CFID=112015986&CFTOKEN=43703661&preflayout=flat [Accessed November 21, 2010].

Zhang, W. & Kramarae, C., 2008. Feminist invitational collaboration in a digital age: Looking over disciplinary and national borders. *Women and Language*, 31(2), pp.8–19.

Zhang, X., 2009. *Exploiting external/domain knowledge to enhance traditional text mining using graph-based methods*. United States -- Pennsylvania: Drexel University. Available at: http://proquest.umi.com/pqdweb?did=1818331311&Fmt=7&clientId=10306&RQT=309&VName=PQD.

Zhang, X. (Michael) & Zhu, Feng, 2011. Group Size and Incentives to Contribute: A Natural Experiment at Chinese Wikipedia. *American Economic Review*, Forthcoming. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1021450 [Accessed January 18, 2011].

Zhirov, A.O., Zhirov, O.V. & Shepelyansky, D.L., 2010. Two-dimensional ranking of Wikipedia articles. *The European Physical Journal B - Condensed Matter and Complex Systems*, pp.1–9.

Zhou, A. et al., 2008. Adaptive indexing for content-based search in P2P systems. *Data and Knowledge Engineering*, 67(3), pp.381–398.

Zhu, F., 2008. *Dynamics of platform-based markets*. United States -- Massachusetts: Harvard University. Available at: http://proquest.umi.com/pqdweb?did=1534034141&Fmt=7&clientId=10306&RQT=309&VName=PQD.

Zlatic, V. et al., 2006. Wikipedias: Collaborative web-based encyclopedias as complex networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 74(1), p.016115.

Appendix: Resources for Wikipedia Researchers

The large amount of data available from Wikipedia encouraged researchers as well as practitioners to create tools to extract various types of data from Wikipedia. The table below provides a comprehensive list of datasets identified during the data extraction phase of our study. Other datasets were also suggested by researchers from the Wiki-research-l mailing list.

Resources from the Wikimedia Foundation

It is odd to write a Science 1.0 article about a Web 2.0 phenomenon. An interested researcher may already find good collaborative written articles about Wikipedia research on Wikipedia itself, as listed in Table 3. These articles may have more complete and updated lists of published scientific work on Wikipedia, and much research-like reporting on Wikipedia of relatively good quality occurs outside ordinary academic channels—on webpages and blogs.

Table 3. Wikimedia articles related to Wikipedia research

Some of these articles are in the main namespace, while others require the `Wikipedia:` namespace prefix, while others (m: prefixed) are on the meta wiki (meta.wikimedia.org).

Wikipedia article	Description
m:Research:Index	Primary entry point for Wikimedia research
Wikipedia	Main article about the encyclopedia
en:Reliability of Wikipedia	Wikipedia article
en:Criticism of Wikipedia	Wikipedia article
en:Academic studies about Wikipedia	Wikipedia article
en:User:Moudy83/conference papers	Long list of Wikipedia conference papers
en:User:NoSeptember/The_NoSeptember_Admin_Project	Various statistics on Wikipedia administrators
en:Wikipedia:Academic studies of Wikipedia	Comprehensive list of studies on Wikipedia
en:Wikipedia:Ethically researching Wikipedia	Essay
en:Wikipedia:Modelling Wikipedia's growth	Specific results on the growth of Wikipedia
en:Wikipedia:Notability (academics)	Notability guideline for academics
en:Wikipedia:Researching Wikipedia	Discusses quantitatively measures and links to various statistics
en:Wikipedia:Survey_(disambiguation)	List of surveys on Wikipedia
en:Wikipedia:Wikipedia as an academic source	List of papers
en:Wikipedia:Wikipedia in research	Essay
en:Wikipedia:WikiProject Wikidemia	Page to "design, implement, and discuss academic research about Wikipedia"
en:Wikipedia:WikiProject Vandalism studies	Studies of damaging edits
m:Research	List resources for wiki research and researchers
m:Wiki Research Bibliography	Bibliography of scholar and science articles
m:Wikimedia Foundation Research Goals	Draft listing of research goals for the foundation
en.wikiversity.org/wiki/Portal:Wikimedia Studies	Portal to Wikimedia studies
strategy:Wikimedia-pedia	Overview of research questions

Datasets

Name	**Host/Developer**	**URL**	**Description**
Wikimedia Data Hub	Wikimedia	http://thedatahub.org/en/group/wikimedia	Official source for Wikimedia Foundation data dumps
Wikimedia Downloads	Wikimedia	http://download.wikimedia.org/	Compressed XML files of Wikipedia from its official database dumps.

Wikimedia Downloads Historical Archives	Wikimedia	http://dumps.wikimedia.org/archive/	Historical Archives.
Wikimedia Foundation Image Dump (November 2005)	Wikimedia	http://archive.org/details/wikimedia-image-dump-2005-11	About 296,000 archived images in use on Wikipedia and its related projects.
Picture of the Year (POTY)	Wikimedia	http://dumps.wikimedia.org/other/poty/	Wikimedia picture of the year archive (2006, 2007, 2009, and 2010).
Pagecount	Domas Mituzas	http://dumps.wikimedia.org/other/pagecounts-raw/	Page view statistics for Wikimedia projects including Wikipedia.
poty	Domas Mituzas	http://dumps.wikimedia.org/other/poty/	Picture of the Year archives.
DBpedia	DBpedia	http://wiki.dbpedia.org/Datasets	This is a large domain ontology derived from Wikipedia.
Koblenz Network Collection	Konect	http://konect.uni-koblenz.de/	Large network datasets of all types.
Wiki10+	Arkaitz Zubiaga	http://nlp.uned.es/social-tagging/wiki10+/	English Wikipedia articles with at least 10 annotations on Delicious.
Wikipedia Page Traffic Statistics	Peter N. Skomoroch	http://aws.amazon.com/datasets/2596?_encoding=UTF8andjiveRedirect=1	7 months of hourly page traffic statistics for over 2.5 million Wikipedia articles.
Wikipedia Traffic Statistics V2	Peter N. Skomoroch	http://aws.amazon.com/datasets/4182?_encoding=UTF8andqueryArg=searchQueryandx=0andfromSearch=1andy=0andsearchPath=datasetsandsearchQuery=Wikipedia	This is the second version of the Wikipedia Page traffic statistics dataset. It contains 16 months of hourly page traffic statistics for over 2.5 million Wikipedia articles.
Wikipedia Page Traffic Statistic V3	Scott C. Frase	http://aws.amazon.com/datasets/6025882142118545?_encoding=UTF8andqueryArg=searchQueryandx=0andfromSearch=1andy=0andsearchPath=datasetsandsearchQuery=Wikipedia	This dataset contains 3 months of hourly page traffic statistics from Wikipedia between 1/1/2011 and 3/31/2011.
Freebase Wikipedia Extraction (WEX)	Freebase	http://wiki.freebase.com/wiki/WEX	Processed dump of the English Wikipedia.

page-to-page link	Henry Haselgrove	http://haselgrove.id.au/wikipedia.htm	Downloadable files that contain all links between 5,716,808 Wikipedia pages.
Wikipedia3	systemone	http://labs.systemone.at/wikipedia3	A monthly updated dataset that contains a conversion of the English Wikipedia into RDF.
Wikipedia edit history	Stanford University	http://snap.stanford.edu/data/wiki-meta.html	A complete Wikipedia edit history until January 2008.
Coordinates in Wikipedia articles	Toolserver.org	http://wikipapers.referata.com/wiki/Coordinates_in_Wikipedia_articles	This contains all the coordinates added to Wikipedia.
Deletionpedia	Deletionpedia	http://wikipapers.referata.com/wiki/Deletionpedia	62,471 pages deleted from Wikipedia.
PAN 2012	PAN 2012	http://www.uni-weimar.de/medien/webis/research/events/pan-12/pan12-web/iq-wikipedia.html#corpus	This corpus contains ten quality flaws (unreferenced, orphan, refimprove, empty section, etc...) the Wikipedia articles that are tagged with the respective cleanup tag.
PAN Wikipedia vandalism corpus (WVC) 2011	PAN 2011	http://www.uni-weimar.de/cms/medien/webis/research/corpora/pan-wvc-11.html	This is a corpus for the evaluation of automatic vandalism detectors for Wikipedia. It supplements the PAN Wikipedia vandalism corpus 2010.
PAN Wikipedia vandalism corpus (WVC) 2010	PAN 2010	http://www.uni-weimar.de/cms/medien/webis/research/corpora/pan-wvc-10.html	This is a corpus for the evaluation of automatic vandalism detectors for Wikipedia.
Wikipedia Vandalism Corpus	Andrew G. West	http://wikipapers.referata.com/wiki/Wikipedia_Vandalism_Corpus_(Andrew_G._West)	This is a corpus of 5.7 million automatically tagged and 5,000 manually-confirmed incidents of vandalism in English Wikipedia.
SWEETpedia	Michael K. Bergman	http://www.mkbergman.com/sweetpedia/	A periodic update of Semantic Web-related Research using Wikipedia.
Social networks of Wikipedia	Paolo Massa	http://sonetlab.fbk.eu/data/social_networks_of_wikipedia/	Network extracted from user talk pages of Venetian Wikipedia.
Tamil Wikipedia word list	Tshrinivasan Wikimedia	https://github.com/tshrinivasan/tamil-wikipedia-word-list	A word list extracted from Tamil Wikipedia dump.
Wikipediadoc	Searchdaimon	http://www.searchdaimon.com/community/dataset/	67,537 Wikipedia articles converted to Microsoft Word 2002 .doc format.

Wikicorpus	Samuel Reese Gemma Boleda Montse Cuadros Lluís Padró German Rigau	http://www.lsi.upc.edu/~nlp/wikicorpus/	This is a corpus that contains large portions of Catalan, Spanish and English Wikipedia (based on 2006 dump) enriched with linguistic information.
WikiBiography	Dr. Michael Strube and his (NLP) research group	http://www.h-its.org/english/research/nlp/download/wikibiography.php	This is a corpus of about 1200 annotated biographies from the German Wikipedia.
WikiTaxonomy	Simone Paolo Ponzetto and Michael Strube NLP research group	http://www.h-its.org/english/research/nlp/download/wikitaxonomy.php	A taxonomy extracted from Wikipedia categories network
WikiNet	Dr. Michael Strube and his (NLP) research group	http://www.h-its.org/english/research/nlp/download/wikinet.php	A multi-language ontology developed by exploiting various aspects of Wikipedia.
WikiRelations	Dr. Michael Strube and his (NLP) research group	http://www.h-its.org/english/research/nlp/download/wikirelations.php	A dataset that contains binary relations obtained from processing Wikipedia category names and the category and page network.
Cite journal miner	Finn Årup Nielsen	http://neuro.imm.dtu.dk/services/wikipedia/citejournalminer.html	A dataset with a matrix of scientific journal citations from Wikipedia
Wikitrends	Ed Summers	http://inkdroid.org/wikitrends/	Top 25 Wikipedia Page Views

Tools

Type	Name	Host/ Developer	URL	Description
Tools for information access or extraction (text and images)	wikipedia2text	Evan Jones	http://www.evanjones.ca/software/wikipedia2text.html	A tool to extract text from Wikipedia. It consists of a command-line program that downloads a specified Wikipedia article and formats it for display on the command-line.
	WikipediaFS	Mathieu Blondel	http://en.wikipedia.org/wiki/WikipediaFS	This tool makes raw text Wikipedia articles available under the Linux file system so a Wikipedia article can be viewed and edited as real files that exist on the local hard drive.
	SONIVIS	Claudia Müller et al.	http://sonivis.org/wiki/index.php/Prepared_Database#Wikipedia_based_data_sets	This software is used to extract information from various wikis including Wikipedia based on social networks analysis.
	infobox2rdf	Tomy and Jimmy	http://code.google.com/p/infobox2rdf/	This is a tool to generate RDF datasets from the infobox data in Wikipedia dump files.
	Java Wikipedia Library	Torsten Zesch	http://code.google.com/p/jwpl/	This is a Java-based application programming interface that allows accessing all information in Wikipedia.
	WikiXRay	Wikimedia	http://meta.wikimedia.org/wiki/WikiXRay	This is another software tool written in Python and R which may download and process data from the Wikimedia sites for generating graphics and data files with quantitative results.
	Catdown	Toolserver.org	http://toolserver.org/~platonides/catdown/catdown.php	A tool to download images by category.
	SIOC	Fabrizio Orlandi and Alexandre Passant	http://ws.sioc-project.org/mediawiki/	This is a RDF exporter for MediaWiki wikis.
	Images for biographies	Emilio J. Rodríguez-Posada	http://wikipapers.referata.com/wiki/Images_for_biographies	A tool that suggests images for biographies in several Wikipedias.

	WikiExtractor	Medialab (University of Pisa, Italy)	http://medialab.di.unipi.it/wiki/index.php/Wikipedia_Extractor	This tool is implemented in python and used to extract cleaned text from Wikipedia dumps.
Bots and APIs	Pywikipediabot	MediaWiki	http://www.mediawiki.org/wiki/Manual:Pywikipediabot	A Python-based collection of tools for bot programming on Wikipedia and other MediaWikis. An example of a bot developed using this tool is one which creates interlanguage links.
	perlwikipedia	MediaWiki	https://github.com/mikelifeguard/MediaWiki-Bot	A MediaWiki bot framework written in Perl. It has been lately used by a Recent changes patrolling program.
	MediaWiki API	MediaWiki	http://www.mediawiki.org/wiki/API:Main_page	A web service API used to monitor a MediaWiki installation, or create a bot to automatically maintain one.
	MediaWiki extensions	MediaWiki	http://wikipapers.referata.com/wiki/MediaWiki_extensions	Extended features for MediaWiki wiki.
	Python-wikitools	Wikipedia User:Mr.Z-man	http://code.google.com/p/python-wikitools/	These are scripts written in Python used to interact with the MediaWiki API and source code for some Wikipedia bots.
	StatMediWiki	Emilio J. Rodríguez-Posada	http://wikipapers.referata.com/wiki/StatMediaWiki	A project to create a tool to collect information available in a MediaWiki installation.
Visualization	Wikistream		http://wikistream.inkdroid.org/	Wikistream is a Node Web Application that enables the visualization of current edits in Wikipedia.
	Wiki Trip	Federico "fox" Scrinzi with contributions by Paolo Massa and Maurizio Napolitano of SoNet@FBK	http://sonetlab.fbk.eu/wikitrip/#|en|	Wiki Trip provides visualization about the edits on a specific page in a temporal context. Statistics about the type and gender of users (editors) as well as the distribution of edits per country are available through Wiki trip.

	History Flow	Joan DiMicco and her research team at IBM	http://www.research.ibm.com/visual/projects/history_flow/index.htm	A tool used to analyze the evolutionary history of Wikipedia pages.
	wmcharts	Emilio J. Rodríguez-Posada	http://toolserver.org/~emijrp/wmcharts/	"wmcharts is a compilation of charts about Wikimedia projects."
	wikitweets	Ed Summer	http://wikitweets.herokuapp.com/	"wikitweets is an experimental visualization of how Wikipedia is cited on twitter"
	Wikiswarm	Jamie Wilkinson	https://github.com/jamiew/wikiswarm	This tool provides a visualization of Wikipedia page edits.
	WikiVis	Pui I Leong, Choi Nga Lou, Weng Kit Tong	http://sourceforge.net/projects/wikivis-um/	"This tool provides an interactive visualization of the Wikipedia information space, primarily as a means of navigating the category hierarchy as well as the article network. The project is implemented in Java, utilizing the Java 3D package."
	Wikipedia JS	Rufus Pollock	http://okfnlabs.org/wikipediajs/	JavaScript library to fetch information from Wikipedia via DBpedia
Tracking and analyzing users' behavior	clicktracking	MediaWiki/ Nimish Gautam, Trevor Parscal	http://www.mediawiki.org/wiki/Extension:ClickTracking	Clicktracking is a MediaWiki extension which enabled tracking users' navigation around the wiki. During Wikipedia Usability Initiative, the Wikimedia Foundation enabled the extension for beta testing.
	Chromogram	Joan DiMicco and her IBM research team.	http://www.research.ibm.com/visual/projects/chromogram.html	A tool for visualizing long sequences of text. It is used to analyze the behavior of Wikipedia users.
	WikiDashboard	Bongwon Suh and Ed Chi	http://wikidashboard.appspot.com	Embeds a user-time matrix in each Wikipedia
	WikiEvent	Ulrik Brandes, Patrick Kenis, Jürgen Lerner, and Denise van	http://www.inf.uni-konstanz.de/algo/software/wikievent/	"WikiEvent is a small graphical java software with which the edit network associated with the history of Wikipedia pages can be computed."

		Raaij		
Semantic and Linguistic tools	Wikipedia-Similarity	Michael Strube and Simone Paolo Ponzetto	http://www.h-its.org/english/research/nlp/download/wikipediasimilarity.php	This tool is used to compute semantic similarity using Wikipedia.
	Manypedia	Paolo Massa, Federico Scrinzi	http://www.manypedia.com/	This tool provides a comparison of Linguistic Points Of View (LPOV) of different language Wikipedias.
	WikipediaMiner	University of Waikato	http://wikipedia-miner.cms.waikato.ac.nz/	"WikipediaMiner is a toolkit for tapping the rich semantics encoded within Wikipedia."
Anti-vandalism tools	AVBOT	Emilio J. Rodríguez-Posada	http://wikipapers.referata.com/wiki/AVBOT	An anti-vandalism bot in Spanish Wikipedia
	ClueBot NG	Christopher Breneman and Cobi Carter	http://en.wikipedia.org/wiki/User:CluebBot_NG	Anti-vandalism bot
	Huggle	Huggle Gurch	http://code.google.com/p/huggle/	Anti-vandalism tool for use on Wikipedia and other Wikimedia projects.
	Igloo	Wikipedia User:Ale jrb	http://en.wikipedia.org/wiki/Wikipedia:Igloo	This is a browser-based, JavaScript tool for handling vandalism on Wikipedia.
	STiki	Andrew G. West	http://en.wikipedia.org/wiki/Wikipedia:STiki	STiki is a tool used to detect and revert vandalism on Wikipedia
	Salebot	Wikimedia	http://fr.wikipedia.org/wiki/Aide:Salebot	Salebot is an anti-vandalism bot in French Wikipedia.
	Twinkle	Carl Fürstenberg Wikipedia User:AzaToth	http://en.wikipedia.org/wiki/Wikipedia:Twinkle	Twinkle is a set of JavaScript functions that gives registered users many extra options to assist them in common Wikipedia maintenance tasks, and to help them deal with acts of vandalism.
	WikiTrust	de Alfaro and others	http://www.wikitrust.net	System with Firefox add-on for online reputation for Wikipedia authors and content.
Miscellaneous Tools	Wikibu	Wikibu	http://www.wikibu.ch	A tool that indicates the reliability of a Wikipedia article by embedding the Wikipedia articles on a page that also

				shows the number of visitors, number of editors, number of links and sources of the Wikipedia article.
	Link Suggester	User:Nickj	http://can-we-link-it.nickj.org/	A link suggestion tool
	Indywiki	Markos Gogoulos Serafeim Zanikolas	http://indywiki.sourceforge.net/index.html	"Indywiki is an open source project that aims to explore different ways of visually browsing wikipedia pages."
	Qwiki	Doug Imbruce	www.qwiki.com/	A system that displays images from Wikipedia and other sources in a multimedia environment.
	WikiDashboard	Bongwon Suh Ed Chi	http://wikidashboard.appspot.com/	A tool that generates a visualization of the edit activity of each Wikipedia page.
	Wikiblame	flominator	http://wikipedia.ramselehof.de/wikiblame.php	"An online browser-based tool for searching the revision history of a MediaWiki based wiki for a text string to identify the author of a particular change to the page."
	WikEd	User:Cacycle	http://en.wikipedia.org/wiki/User:Cacycle/wikEd	"A full-featured Wikipedia-integrated advanced text editor for regular to advanced wiki users."
	Wikipanion	Robert Chin	http://www.wikipanion.net/	A Wikipedia browsing application for iOS.
	Wikipedia Diver		http://download.cnet.com/Wikipedia-Diver/3000-11745_4-75029595.html	A Firefox plugin that logs clicks between Wikipedia pages to a database and displays the browsing path graphically.
	Kiwix	Kiwix.org	http://www.kiwix.org/index.php/Main_Page/en	An offline multimedia reader that makes Wikipedia available offline.
	Wikireader	Wikimedia Foundation	http://en.wikipedia.org/wiki/WikiReader	"A project to deliver an offline, text-only version of Wikipedia on a mobile device."
	Wikirage	Craig Wood	http://www.wikirage.com/	"It tracks the pages in Wikipedia which are receiving the most edits over various periods of time."
	Wikichecker	MediaWiki	http://en.wikichecker.com/	It generates statistics over users and individual articles as well as lists of highly edited pages.

	WikiChanges	Sérgio Nunes	http://sergionunes.com/p/wikichanges/	A web-based tool that creates on-the-fly graph of the edit history of one or two Wikipedia articles.
	Wikipedia article traffic statistics	User:Henrik	http://stats.grok.se/	A useful interactive web-service that renders the statistics in monthly histograms.
	Wikitrends	Johan Gunnarsson	http://toolserver.org/~johang/wikitrends/english-uptrends-this-week.html	A web service from the Toolserver with up- and down-trends based on day, week or month and across the language versions of Wikipedia.

Lists of Datasets and Tools

Name	Host/Developer	URL	Description
Datamob datasets	Datamob	http://datamob.org/datasets/tag/wikipedia	Datamob collects a list of public datasets; the ones here are tagged "Wikipedia"
Amazon web services	Amazon	http://aws.amazon.com/search?searchQuery=Wikipediaandsearch Path=datasetsandx=0andy=0	Keyword search on "Wikipedia" for the Amazon Web Services category "Public datasets"
IBM data	IBM	http://www-958.ibm.com/software/data/cognos/manyeyes/datasets?q=wikipedia	List of public datasets including articles from Wikipedia and statistics about some of its content.
Wikipapers tools	Emilio (Wikipedia User: emijrp)	http://wikipapers.referata.com/wiki/List_of_tools	List of tools to extract data from various wikis including Wikipedia.
Wikipapers datasets	Emilio (Wikipedia User: emijrp)	http://wikipapers.referata.com/wiki/List_of_datasets	List of datasets including data extracted from Wikipedia and other sources.
Wikimedia toolserver	Wikimedia	https://wiki.toolserver.org/view/Main_Page	This is a platform for hosting various software tools written and used by Wikimedia editors.

Books about Wikipedia

Title	Authors/Editors	Year	Publisher
The Wisdom of Crowds: why the Many Are Smarter Than the Few and How	James Surowiecki	2004	Doubleday, Anchor

Collective Wisdom Shapes Business, Economies, Societies and Nations			
Understanding Knowledge as a Commons. From Theory to Practice	Hess, Charlotte; Ostrom, Elinor, editors.	2006	MIT Press
La Révolution Wikipédia (English: The Wikipedia Revolution)	Pierre Gourdain, Florence O'Kelly, Béatrice Roman-Amat, Delphine Soulas, Tassilo von Droste zu Hülshoff	2007	*Les Mille et Une Nuits*
The Cult of the Amateur	Andrew Keen	2007	Crown Business, Doubleday, Random House
How Wikipedia works	Phoebe Ayers, Charles Matthews and Ben Yates	2008	No Starch Press
MediaWiki (Wikipedia and Beyond)	Daniel J. Barrett	2008	O'Reilly
Wikipedia: The Missing Manual	Johan Broughton	2008	O'Reilly
The Wikipedia Revolution	Andrew Lih	2009	Hyperion (US version); Aurum Press (UK version)
The World and Wikipedia	Andrew Dalby	2009	Siduri Books
Lazy Virtues: Teaching Writing in the Age of Wikipedia	Robert E. Cummings	2009	Vanderbilt University Press
Critical Point of View: A Wikipedia Reader	Geert Lovink and Nathaniel Tkacz (eds)	2011	Institute of Network Cultures
Good Faith Collaboration: The Culture of Wikipedia	Joseph Michael Reagle Jr.	2011	MIT Press

Scientific Meetings

Several dedicated scientific meeting centers around wikis and Wikipedia.

Name	**Description**
WikiSym	An ACM affiliated meeting which presents results in all areas of wiki research.
SemWiki	This is a workshop that started in 2006 and focused on presenting results in semantic wiki research. The research community around that meeting also interacts at the semanticweb.org site,—itself a semantic wiki.
WikiAI	WikiAI is a workshop that is concentrated on the interface between artificial intelligence, machine learning and computational linguistics on one side and wiki and other collaborative-built knowledge bases on the other side.
Workshop on Collaboratively Constructed	This is a workshop focused on social semantics and social natural language processing with several contributions centered on the Wikipedia corpus.

Semantic Resources	
Wikimania	This is a community meeting that focuses on Wikipedia and its sister Wikimedia foundation operated projects. Apart from community-related topics the meeting usually has a good deal of research-oriented material presented.

Other Communication Channels

Name	URL	Description
WikiSym mailing lists	http://www.wikisym.org/cgi-bin/mailman/listinfo	A listing of all the public mailing lists on www.wikisym.org.
wiki-general	http://www.wikisym.org/cgi-bin/mailman/listinfo/wiki-general	General discussion of all things wiki.
wiki-research	http://www.wikisym.org/cgi-bin/mailman/listinfo/wiki-research	Discussion of wiki research and practice.
wiki-standards	http://www.wikisym.org/cgi-bin/mailman/listinfo/wiki-standards	The discussion list for wiki standards.
wikisym-announce	http://www.wikisym.org/cgi-bin/mailman/listinfo/wikisym-announce	Announcements about WikiSym, the Wiki Symposium.
Wikimedia mailing lists	https://lists.wikimedia.org/mailman/listinfo	A listing of all the public mailing lists on lists.wikimedia.org.
Wiki-research-l	https://lists.wikimedia.org/mailman/listinfo/wiki-research-l	Research into Wikimedia content and communities
Wikipedia Signpost	http://en.wikipedia.org/wiki/Wikipedia:Wikipedia_Signpost/Archives/Years http://www.wikipediasignpost.com/blog/	A newsletter that discusses different matters of Wikimedia projects.
Wikipedia Review	www.wikipedia-watch.org	A forum with critical commentaries
Wikipedia Weekly	http://en.wikipedia.org/wiki/Wikipedia:WikipediaWeekly	A podcast with episodes from 2006 to presently 2009
Wikimedia Research Newsletter	www.meta.wikimedia.org/wiki/Research:Newsletter	A monthly newsletter that is focused on new Wikimedia-related research.