

An Analysis of Topical Coverage of Wikipedia

Alexander Halavais

School of Communications
Quinnipiac University

Derek Lackaff

Department of Communication
State University of New York at Buffalo

Many have questioned the reliability and accuracy of Wikipedia. Here a different issue, but one closely related: how broad is the coverage of Wikipedia? Differences in the interests and attention of Wikipedia's editors mean that some areas, in the traditional sciences, for example, are better covered than others. Two approaches to measuring this coverage are presented. The first maps the distribution of topics on Wikipedia to the distribution of books published. The second compares the distribution of topics in three established, field-specific academic encyclopedias to the articles found in Wikipedia. Unlike the top-down construction of traditional encyclopedias, Wikipedia's topical coverage is driven by the interests of its users, and as a result, the reliability and completeness of Wikipedia is likely to be different depending on the subject-area of the article.

doi:10.1111/j.1083-6101.2008.00403.x

Introduction

We are stronger in science than in many other areas... we come from geek culture, we come from the free software movement, we have a lot of technologists involved... that's where our major strengths are. We know we have systemic biases. - James Wales, Wikipedia founder, Wikimania 2006 keynote address

Satirist Stephen Colbert recently noted that he was a fan of “any encyclopedia that has a longer article on ‘truthiness’ [a term he coined] than on Lutheranism” (July 31, 2006). The focus of the encyclopedia away from the appraisal of experts and to the interests of its contributors stands out as one of the significant advantages of Wikipedia, “the free encyclopedia that anyone can edit,” but also places it in sharp contrast with what many expect of an encyclopedia.

There have been a number of recent attempts to measure the accuracy and reliability of Wikipedia as a resource. The processes by which traditional encyclopedias were constructed contributed to their trustworthiness, and Wikipedia is constructed in a radically different way. Little sustained effort has been made in measuring the

diversity of content of Wikipedia. Here we present two efforts to map the diversity of content on Wikipedia, to better understand not whether it is accurate in its content, but to what degree it represents collected knowledge. The degree to which Wikipedia is a useful resource depends not only on its accuracy, but on the degree to which it is, indeed, encyclopedic in its breadth. The metrics presented suggest an applicability beyond Wikipedia, and the possibility of better mapping and comparing other collections of knowledge.

Measuring Wikipedia

An investigation published in *Nature* (Giles, 2006), and other attempts to measure Wikipedia's accuracy and reputation (Lih, 2004; Chesney, 2006), bring questions about Wikipedia's quality and authority to the forefront of discussions among librarians, educators, and other knowledge workers. Many of these attempts have tried to move beyond anecdotal examples of article success or failure to provide more thorough metrics of quality. Voss (2005) examines the number of articles, division of the language-specific sites, growth of the site, editing behavior of authors, sizes of articles, and other formal elements of the Wikipedia site. Emigh and Herring (2005) performed genre analysis of Wikipedia and Everything2 (another popular wiki-based collection of general knowledge), and concluded that the socio-technical processes of article creation shaped the style of the content presented. Other work has focused mainly on the process of collaboration (Bryant, Forte, and Bruckman, 2005), the content of articles (Braendle, 2005; Pfeil, Zapheris, & Ang, 2006), the ways in which it is interlinked (Bellomi, & Bonato, 2005; Stvilia et al., 2005), or the dynamics of how it evolves (Hassine, 2005; Kittur et al., 2007; Viégas, Wattenburg, & Dave, 2004; Viégas et al., 2007; Wilkinson & Huberman, 2007; Zlatić et al., 2006). As a result, we are beginning to see a number of potential ways of measuring the content of Wikipedia, and the evolutionary process by which that content changes.

The work here attempts to measure a particular aspect of Wikipedia: its topical scope and coverage. Similar attempts have been made in other knowledge domains. In the field of information retrieval, for example, it is necessary to compare two items to measure how similar they might be. For images, one way to do this is to generate a color histogram for each of two images, and determine the percentage of difference between these two histograms (Swain & Ballard, 1991). In other words, in order to determine similarity, some feature is extracted (in this case color), and compared across examples, to create some measure of similarity. The same approach has long been used to provide an indication of similarity between texts, at the heart of information retrieval (Luhn, 1958).

Similar approaches have been taken to compare the variety of material available. The economic literature has examined the scope and variety of products in a market (Lancaster, 1990), and this has found its way into examinations of media goods as well. Compaine & Smith (2001), for example, attempt to determine whether internet radio represents a broader set of content than traditional over-the-air broadcast radio, and George (2001) looks at changes in the distribution of content in

newspapers over time. The only major attempt to examine the diversity of content on Wikipedia is undertaken by Holloway, Bozicevic, and Börner (2006), who provide an indication of how different parts of Wikipedia relate to one another in terms of content, currency, and authorship. As part of this work, they also compared the category structures of Wikipedia, Britannica, and Encarta.

Our analysis examines the distribution of Wikipedia articles at two levels, first at the overall level of all full articles in the English-language Wikipedia, and then at the level of articles within three particular academic fields. Broadly, the exploration presented here demonstrates what many who are involved in Wikipedia already have a sense of: Wikipedia remains particularly strong in some of the sciences, among other areas, but not as strong in the humanities or social sciences. If an encyclopedia is only as good as its weakest areas, it is important to identify these weaknesses.

Study 1: Understanding the topical diversity of Wikipedia

While not intended to provide a measure of the distribution of topics, classification systems like the Dewey Decimal System and the Library of Congress Classification provide a structure through which we can measure one collection of knowledge against another. Since all measurement is essentially a comparison, we make use of Bowkers Books in Print to determine the degree to which it relates to Wikipedia's diversity of content. This does not imply that books in print are an ideal distribution, merely that they represent an established distribution against which we are able to measure. While no subject categorization system is expected to provide an even distribution of items, the nature of such a distribution tells us something about what it considers to be more worthy of elaboration.

Method

A sample of 3,000 articles was drawn randomly in the spring of 2006 from a downloaded list of article headwords in the English-language Wikipedia, excluding articles with less than 30 words of text. These articles were classified by Library of Congress category (at the broadest level) by two coders familiar with both Wikipedia and the LC system. A subset of 500 of these articles was categorized by two coders, and intercoder reliability, as measured by Cohen's Kappa, was 0.92.

When collected, the length (in kilobytes) of the HTML of each page was recorded, providing another indication of the amount of content in particular categories. Note that this measurement is imperfect, as it does not include, for example, the number of images. In addition, the number of edits made by contributors to each page was recorded.

Results and discussion

Figure 1 presents the distribution of topics on Wikipedia when compared with Books in Print. Categories on the left side represent those in which Wikipedia has

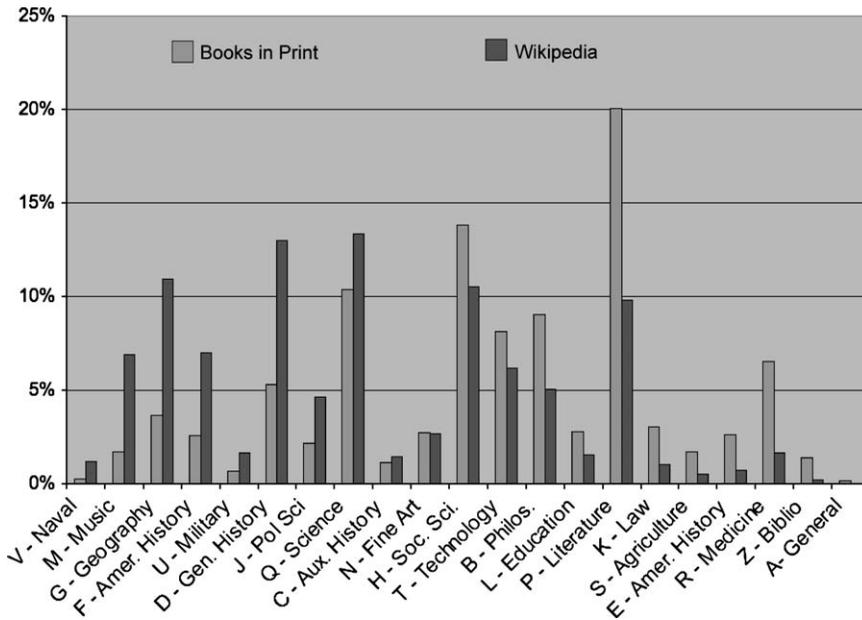


Figure 1 Wikipedia vs. Books in Print, percentage of total in each category, ranked by percentage difference between the two collections.

a relatively large number of articles, compared with the world of books, and those on the right side represent areas in which Wikipedia has far less emphasis.

While we certainly would not expect the two distributions to be identical, the universe of books available to consumers should represent some indication of general interest and general knowledge. Some variations in topical coverage can be attributed to Wikipedia’s technical attributes. For example, Wikipedia has more articles on naval sciences and the military than is found in Books in Print. Because it is easy to quickly generate headwords in both areas—many of the entries represent individual ships in the US and British fleet, and the military category has extensive lists of arms—they have tended to expand quickly. Such a listing bias is even more pronounced in the history categories. While some of the extensive lists articles in categories D, E, and F probably reflect the interest of Wikipedians in the subject, the number of articles was artificially boosted as a result of the automatic importation of data from a public data sources such as the 2000 United States Census. Often, such entries have remained untouched by subsequent edits, and provide only the most basic facts about a town or city. Since this local information is often classified as “history,” these categories are disproportionately represented.

One of the most marked differences, that in language and literature (P), is to be expected. An encyclopedia is unlikely to map the publishing industry in every regard, and since nearly 15% of new books published each year are fictional (Bowker, 2007), and fiction is not appropriate for an encyclopedia, there is a discrepancy. In practice,

there is actually a substantial number of articles that represent literary criticism on Wikipedia, otherwise the disparity would be even greater. The documentation of the *Lord of the Rings* trilogy, for example, or commentary on the Harry Potter series, is voluminous.

The heavy emphasis on music, and relative parity of fine arts, may be surprising to those—like Wales (2006)—who believe that Wikipedia turns a cold shoulder to the humanities. The nature of the articles within these groups, however, is enlightening. The vast majority of the articles in the music category are not on theory or performance, but pages for particular popular music bands. Fans drive the creation and development of articles not just in music, but to a lesser extent, in the fine arts (e.g., comics) and literature.

Perhaps not surprisingly to those familiar with Wikipedia, the sciences (Q) are well represented, while the social sciences, particularly present in H and J, are not nearly as well covered. These differences, however, are not as stark as might be expected. Indeed, it appears that the issue is more nuanced. Perhaps surprising is that while science is generally well-covered, it is not universally the case. Articles in medicine (R) are particularly sparse, and the technology category (T) is not particularly prominent.

This categorization gives an indication of the distribution of Wikipedia in terms of individual articles, but for a more complete view, we should examine the length and “churn” of these articles—that is, the number of times they have been edited—as presented in Figures 2 and 3. The number of articles within a particular topic is an

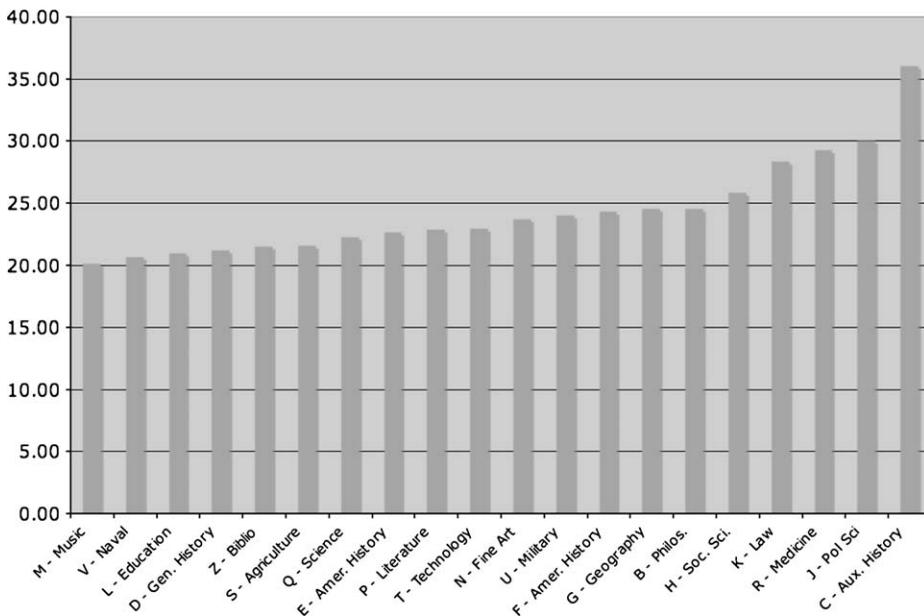


Figure 2 Average size (in total page HTML kilobytes) of articles in each category.

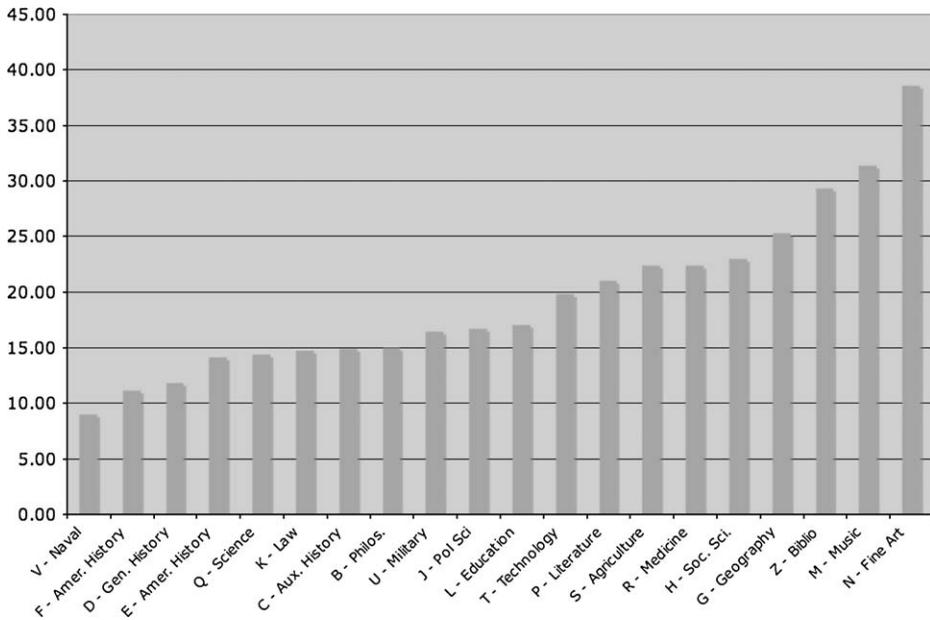


Figure 3 Average number of edits of articles in each category.

important metric, but as noted above, some of these articles are little more than “stubs,” waiting to be filled in by others. An indication of length of the article (in kilobytes on the page, not including images) provides an indication of how extensive the articles are on average. Likewise, how many edits an article has sustained is sometimes used as a proxy for quality, since articles that have been more frequently modified are expected to be of higher quality. There can certainly be other reasons for high numbers of modifications (especially controversial content, for example), but this gives us an indication of the character of the articles in each subject area, as a whole.

These data should be considered with a caveat: many of the top-level LC categories had very few articles. Small numbers of articles in any given category can provide outliers particular influence. For example, in these two figures, the “Superman” article has been removed from Fine Arts (N), as it had an extremely large number of edits (4,867) and was by far the longest article (149K).

As noted above there are an anomalous number of articles related to geographical locations, but these are generally very short articles with little or no history of revision. While they may exist in large number, they remain on the periphery of Wikipedia. On the other hand, a “fan effect” is visible in Fine Arts and in Music. While articles in these two areas are not exceptionally long, they are frequently edited.

The longest articles appear in some of the areas that appear otherwise to be “light”: medicine, law, and the social sciences. It may be that these topics lend

themselves to longer articles than do the more factual articles in science. For whatever reason, this mitigates somewhat the observation that Wikipedia is lacking in these areas. While it may still be far from parity with Books in Print, those articles that do appear tend to be among the lengthier on Wikipedia.

Study 2: Wikipedia and expert taxonomies

Our second study was designed to provide a more comprehensive comparison between Wikipedia's topical coverage and other reference sources. Academic or scholarly knowledge domains are a good starting point for this type of study, as the established disciplines already contain authoritative experts. These experts, alone or collectively, attempt to define, explain, and bound their domains by publishing textbooks and encyclopedias. We determined that one method of evaluating Wikipedia's topical coverage would be to compare the content of printed scholarly encyclopedias with Wikipedia. The encyclopedias used in comparison were *Encyclopedia of Linguistics* (Strazny, 2005), *New Princeton Encyclopedia of Poetry and Poetics* (Preminger & Brogan, 1993), and *Encyclopedia of Physics, 2nd Ed.* (Lerner & Trigg, 1991). Each encyclopedia is widely available, widely cited, and edited by highly-qualified academic experts. These particular encyclopedias were chosen as they represent knowledge domains with limited scholarly overlap and contained a comparable number of individual articles, hopefully covering their topics at a similar level of specificity.

Method

The encyclopedias were compared on the basis of article titles, or headwords, found in each. While such headwords might be open to interpretation, generally article topics refer to terms of art and key terms that represent the existing organization of the disciplines. (Alternative approaches that might draw on the content of the articles are possible; see Ruiz-Casado, Alfonseca, & Castells, 2005). Naming conventions for topically-related Wikipedia articles tend to be developed over time, and there are no formal naming conventions on Wikipedia in the fields considered here. Because there is variability in the ways in which headwords are decided, and the topical coverage of articles, there are difficulties in doing direct comparisons. Nonetheless, since the organization of a field is in some part related to the knowledge of that field, we would expect to find some coherence among encyclopedias.

In early Spring of 2006, each headword from the printed encyclopedias was used as the search phrase in a Google search of the English Wikipedia. This was a complete mapping of the topical space of each printed encyclopedia onto Wikipedia. The headword "Burmese poetry" from the poetry encyclopedia, for example, was searched via Google as "Burmese poetry site:en.Wikipedia.org". Of the top five results, the best match (if any) was chosen by a human coder as the corresponding Wikipedia article. Rather than using only exact matches, we made use of near matches provided by Google, and when ambiguities arose in the headwords, we

presumed similarity. This strategy overestimates, somewhat, congruence between the two sources, but allows for term matches (e.g. “Burmese poetry” and Wikipedia’s “Myanmar literature”) that would be inaccessible using automated matching procedures such as word stemming.

We were not only interested in mapping the topical coverage of the printed encyclopedias to Wikipedia, but also mapping the Wikipedia’s topical coverage of the same knowledge domains back to the printed encyclopedias. Due to the decentralized nature of Wikipedia, generating a headword list for each of the three knowledge domains was challenging. One useful Wikipedia organizational convention is the “category,” a software container into which articles and other categories (subcategories) can be placed. The “Physics” category, for example, contains physics-related articles, as well as subcategories such as “Thermodynamics” and “Physicists.” In the interest of systematic headword list creation, articles from the Wikipedia categories “Linguistics,” “Physics,” and “Poetry” were sampled to a depth of three subcategories using Daniel Kinzler’s CatScan script, an external software tool that allows categories to be easily searched and analyzed. This method resulted in a partial mapping of the Wikipedia’s total topical space for each domain. Since categories can be nested recursively and to an arbitrary depth, we assume that a relevant “core” of the topical space can be sampled in this way.

Results and discussion

While the total number of articles in the traditional encyclopedias may have been relatively small (Table 1), a considerable number of these in each field could not be matched with articles in Wikipedia. The number of orphan articles ranged from 89 (18%) of the articles on physics, to 330 (37%) of the poetry articles. This appears to indicate that Wikipedia’s topical coverage is more limited than that of the printed, expert-created encyclopedia. As articles on Wikipedia are created and develop according to the interest of contributors, some topics expand rapidly (popular culture and physical science) while other topics are developed more slowly (national poetries and prosodies).

The hierarchical category system, one of several organizational structures on Wikipedia, provides an expedient, accessible headword list organized hierarchically by topic. Under linguistics, for example, there are 292 subcategories within a nested depth of three levels. These categories include linguistic topics ranging from “Linguistic morphology” to “Finnish profanity.” While these categories are far from

Table 1 Coverage comparison between printed encyclopedias and Wikipedia.

	Coincidental Articles	Print Only Articles	Total Articles
Physics	399 (81.76%)	89 (18.24%)	488
Linguistics	424 (79.10%)	112 (20.90%)	536
Poetry	551 (62.54%)	330 (37.46%)	881

comprehensive, even at local levels (there are not categories for profanity in many languages, for example) they cover a broader range of subtopics than any printed encyclopedia could reasonably approach. Further, Wikipedia is able to break complex topics into an arbitrary number of articles, where a print encyclopedia might be forced to consolidate such topics. As of this writing, 12,554 individual articles are listed within Wikipedia's linguistics subcategories. Wikipedia's physics category contains 7,916 articles, while the poetry category contains 2,735 articles. Despite these large numbers of articles, there appear to be blind spots.

There may be as much variation among different printed encyclopedias as has been found between Wikipedia and the encyclopedias used here. Within our small sample of encyclopedias, differences in specificity and breadth are apparent. A substantial part of why these fail to overlap is related to different editors' organizational approaches to the knowledge in their fields. Approximately one quarter (22) of the unmatched linguistics terms were personal names, for example. The top-down approach of encyclopedia-editing may not apply to Wikipedia, but a communal agreement about what belongs within individual articles (and whether articles should themselves be included) is seen to emerge there. That this policy decision is more distributed does not change the force of editorial control, and in comparison with the bound encyclopedias, Wikipedia has a fairly conservative boundary for article inclusion. This shared limitation to specifically topical issues may be the factor that leads to such strong congruence between it and the *Encyclopedia of Physics*.

Conclusion

The traditional printed encyclopedia is subject to physical and structural constraints of the paper medium. Any encyclopedia contains articles dealing with only a subset of all possible topics, whether it is a source of general knowledge (*Encyclopedia Britannica* with over 65,000 articles) or domain-specific knowledge (*Encyclopedia of Physics* with 488 articles). Online encyclopedias, unrestricted by weight, volume, and time spent flipping pages, hold out the promise of being truly comprehensive.

Wikipedia presents a new model of encyclopedic knowledge creation and maintenance. While Wikipedia lacks the structures of authority that support the popular faith in printed encyclopedias, proponents argue that its model of populist participation provides an equally valid and useful organizing structure. Current research is examining the ability of Wikipedia to maintain high-quality and factually-accurate articles. We maintain that topical coverage is of equal importance in Wikipedia's quest for mainstream and academic acceptance.

Overall, we found that the degree to which Wikipedia is lacking depends heavily on one's perspective. Even in the least covered areas, because of its sheer size, Wikipedia does well, but since a collection that is meant to represent general knowledge is likely to be judged by the areas in which it is weakest, it is important to identify these areas and determine why they are not more fully elaborated. It cannot be a coincidence that two areas that are particularly lacking on Wikipedia—law and

medicine—are also the purview of licensed experts. Many attorneys have taken up blogging with open arms and medical research is now frequently published in open access journals, both suggesting that there is not always an impediment to these groups contributing to online resources.

Despite the noted difficulties of partitioning Wikipedia into topical domains, the sheer number of articles presented by Wikipedia far outstrips the bound encyclopedias we investigated. Can you have too much of a good thing? There may be some question as to whether an article on “Finnish Profanity” rises to the same level of importance as “Finnish Grammar”—someone seeking out the most important topics in any sub-domain of human knowledge might have difficulty finding them in Wikipedia. But assuming the most important topics are covered well, there is no reason that other topics that may be considered somewhat more marginal should not also be available.

At present, several projects are underway to ensure that important topics receive appropriate coverage. WikiProject Physics, for example, has several dozen participants who are actively contributing to the breadth, quality, and organization of physics-related articles on Wikipedia. The project maintains a list of missing and inadequate articles, as well as a list of articles awaiting expert review. Several of the orphan articles located by our comparison were actually listed on various “missing topics” pages, indicating that if this study were replicated in the future, the correlation between the printed encyclopedias and Wikipedia would increase.

Both approaches taken here provide some indication of the kinds of topics that Wikipedia emphasizes. We have provided some initial observations as to why these differences exist, but there is still much to be done in this regard. Wikipedia remains a surprise in many ways, in part because it is difficult to gauge the motivations of its contributors. By understanding why and how people contribute to Wikipedia, particularly within various knowledge sub-domains, we may be able to encourage work in areas that are, relatively speaking, in need of more contributions.

References

- Bellomi, F., & Bonato, R. (2005). Network analysis for Wikipedia. Proceedings of Wikimania 2005: The First International Wikimedia Conference. Frankfurt, Germany. Retrieved November 1, 2007 from <http://www.fran.it/blog/2005/08/network-analysis-for-wikipedia.html>
- Bowker, R.R., Inc. (2007). Bowker reports US book production rebounded slightly in 2006. Retrieved November 1, 2007 from http://www.bowker.com/press/bowker/2007_0531_bowker.htm
- Braendle, A. (2005). Many cooks don't spoil the broth. Proceedings of Wikimania 2005—The First International Wikimedia Conference. Frankfurt, Germany. Retrieved November 1, 2007 from <http://meta.wikimedia.org/wiki/Transwiki:Wikimania05/Paper-AB1>
- Bryant, S., Forte, A., & Bruckman, A. (2005). Becoming Wikipedia: Transformation of participation in a collaborative online encyclopedia. In *Proceedings of GROUP 2005*.

- Chesney, T. (2006). An empirical examination of Wikipedia's credibility. *FirstMonday*, 11(11). Retrieved November 1, 2007 from http://www.firstmonday.org/issues/issue11_11/chesney/
- Colbert, S. (July 31, 2006). Report on "Wikiality." Colbert Report. Comedy Central.
- Comaigne, B., & Smith, E. (2001). Internet radio: A new engine for content diversity? 29th Telecommunications Policy Research Conference. Arlington, VA.
- Emigh, W., & Herring, S. C. (2005). Collaborative authoring on the web: A genre analysis of online encyclopedias. *Proceedings of the 38th Hawai'i International Conference on System Sciences (HICSS-38)*. Los Alamitos: IEEE Press. Retrieved May 21, 2007 from <http://ella.slis.indiana.edu/~herring/wiki.pdf>
- Giles, J. (2006). Internet encyclopaedias go head to head. *Nature*, 438, 900–901.
- George, L. (2001). What's fit to print: The effect of ownership concentration on product variety in daily newspapers. 29th Telecommunications Policy Research Conference. Arlington, VA.
- Hassine, T. (2005). The dynamics of NPOV disputes. *Proceedings of Wikimania 2005—The First International Wikimedia Conference*. Frankfurt, Germany. Retrieved November 1, 2007 from <http://meta.wikimedia.org/wiki/Transwiki:Wikimania05/Paper-TH1>
- Holloway, T., Bozicevic, M., & Börner, K. (2006). Analyzing and visualizing the semantic coverage of Wikipedia and its authors.. Submitted to *Complexity*, preprint available from ArXiv.org: <http://arxiv.org/abs/cs.IR/0512085>.
- Kinzler, D. (undated). CatScan. Retrieved May 21, 2007 from <http://tools.wikimedia.de/~daniel/WikiSense/CategoryIntersect.php?wikilang=de&wikifam=.wikipedia.org&userlang=de>
- Kittur, A., Chi, E. H., Pendelton, B. A., Suh, B., & Mytkowicz, T. (2007). Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. 25th Annual ACM Conference on Human Factors in Computing Systems. Retrieved November 1, 2007 from http://www.viktoria.se/altchi/submissions/submission_edchi_1.pdf
- Lancaster, K. (1990). The Economics of Product Variety: A Survey. *Marketing Science*, 9(3), 189–206.
- Lerner, R. G., & Trigg, G. L. (eds.). (1991). *Encyclopedia of Physics*, 2nd ed. New York: VCH.
- Lih, A. (2004). Wikipedia as participatory journalism: Reliable sources? Presented at the Fifth International Symposium on Online Journalism, April 16-17, Austin. Retrieved November 1, 2007 from <http://jmsc.hku.hk/faculty/alih/publications/utaustin-2004-wikipedia-rc2.pdf>
- Luhn, H. P. (1958). The automatic creation of literature abstracts. *IBM Journal of Research and Development*, 2, 159–165.
- Pfeil, U., Zaphiris, P., & Ang, C. S. (2006). Cultural differences in collaborative authoring of Wikipedia. *Journal of Computer-Mediated Communication*, 12(1). November 1, 2007 from <http://jcmc.indiana.edu/vol12/issue1/pfeil.html>
- Preminger, A., & Brogan, T. V. F. (eds.). (1993). *The New Princeton Encyclopedia of Poetry and Poetics*. Princeton: Princeton University Press.
- Ruiz-Casado, M., Alfonseca, E., & Castells, P. (2005). Automatic assignment of Wikipedia encyclopedic entries to WordNet synsets. *Lecture Notes in Computer Science*, v. 3528.
- Strazny, P. (ed.). (2005). *Encyclopedia of Linguistics*, 2 vols. New York: Fitzroy Dearborn.
- Stvilia, B., Twidale, M. B., Gasser, L., & Smith, L. (2005). Information quality discussions in Wikipedia. Proceedings of the 2005 International Conference on Information Quality. Retrieved November 1, 2007 from <http://www.isrl.uiuc.edu/~stvilia/papers/qualWiki.pdf>

- Swain, M. J., & Ballard, D. H. (1991). Color indexing. *International Journal of Computer Vision*, 7(1), 11–32.
- 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 from ACM CHI 2004 Conference on Human Factors in Computing Systems* (pp. 575–582). Vienna, Austria. Retrieved May 21, 2007 from http://web.media.mit.edu/~fviegas/papers/history_flow.pdf
- Viégas, F. B., Wattenberg, M., Kriss, J., & van Ham, F. (2007). Talk before you type: Coordination in Wikipedia. In *Proceedings of the 40th Hawaii International Conference on System Sciences*.
- Voss, J. (2005). Measuring Wikipedia. In *Proceedings of the Tenth International Conference of the International Society for Scientometrics and Informetrics: Stockholm*. Retrieved November 1, 2007 from <http://eprints.rclis.org/archive/00003610/01/MeasuringWikipedia2005.pdf>
- Wales, J. (2006). Keynote talk. Wikimania. Cambridge, Mass. August.
- Wilkinson, D., & Huberman, B. (2007). Assessing the value of cooperation in Wikipedia. arXiv preprint [cs/0702140v1](http://arxiv.org/abs/cs/0702140v1). Retrieved November 1, 2007 from <http://arxiv.org/abs/cs.DL/0702140v1>
- Zlatić, V., Božičević, M., Štefančić, H., & Domazet, M. (2006). Wikipedias: Collaborative web-based encyclopedias as complex networks. *Physical Review E*, 74, 016115.

About the Authors

Alexander Halavais [alex@halavais.net] is assistant professor of interactive communication at Quinnipiac University. His research addresses the relationship of networked communication to social change in politics, journalism, and education. He blogs at <http://alex.halavais.net>

Address: School of Communications, Quinnipiac University, 275 Mount Carmel Avenue, Hamden, Connecticut 06518

Derek Lackaff [lackaff@lackaff.net] is a doctoral candidate in Department of Communication at the State University of New York at Buffalo. His research interests include online collaboration, the psychology of social media, and community informatics.

Address: Department of Communication, State University of New York at Buffalo, Buffalo, New York 14260-1020

The authors wish to express their gratitude to Brenda Battleson and Catherine Munro for their assistance on this project.