

Protocol for a systematic literature review of research on the Wikipedia

Chitu Okoli
Decision Sciences and MIS
Concordia University
(514) 848-2424
Chitu.Okoli@concordia.ca

Kira Schabram
Management
Concordia University
(514) 848-2424
k_scha@jmsb.concordia.ca

ABSTRACT

Context: 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 attracted over 1,000 scholarly studies that treat Wikipedia as a major topic or data source. *Objectives:* This article presents a protocol for conducting a systematic mapping (a broad-based literature review) of research on Wikipedia. It identifies what research has been conducted; what research questions have been asked, which have been answered; and what theories and methodologies have been employed to study Wikipedia. *Methods:* This protocol follows the rigorous methodology of evidence-based software engineering to conduct a systematic mapping study. *Results and conclusions:* This protocol reports a study in progress.

Categories and Subject Descriptors

D.2.9 [Software Engineering]: Management – *copyrights, software process models.*

General Terms

Management, Human Factors, Theory.

Keywords

Wikipedia, literature review, open content, open source

1. INTRODUCTION

Since its inception, Wikipedia has become the world's leading source of Web reference information. It comes as no surprise that the phenomenon is a prime target for academic inquiry. In our preliminary searches, we have identified as many as 1,000 scholarly studies that use Wikipedia as a major topic or data source. Research goals have varied widely, ranging from attempting to understand open content creation mechanisms, to investigating the concept of a web-based encyclopaedia, to using Wikipedia's abundant data as a source for various kinds of research. To date, only three literature reviews of the

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MEDES 2009, October 27-30, 2009, Lyon, France.
Copyright 2009 ACM 978-1-60558-829-2/09/0010.\$10.00.

scholarly work have been published [1,2,3]; none of these has attempted an in-depth critical analysis of Wikipedia.

Although there is an abundance of literature reviews in software engineering and in management information systems, there are few that follow a systematic, rigorous methodology. In response to this shortage, the Evidence-Based Software Engineering Group led by Barbara Kitchenham, has developed detailed guidelines on writing systematic reviews in software engineering [5]; we closely follow these specifications to describe here a systematic literature review, defined as "a form of secondary study that uses a well-defined methodology to identify, analyse and interpret all available evidence related to a specific research question in a way that is unbiased and (to a degree) repeatable" [5]. Such a "systematic mapping study" is a broad outline of the field, the precursor to reviews with more focused research questions. Following Kitchenham's requirement for an explicit research protocol before the actual study is conducted, this article presents a specification of our intended review. Disseminating the protocol for public review provides in itself an exemplar for the execution of similar studies.

Our study investigates the following questions:

1. What high-quality research has been conducted with Wikipedia as a major topic or data source?
2. What research questions have been asked by various sources, both academic and practitioner?
3. What theoretical frameworks and reference theories have been used to study the topic?
4. What methodological approaches have been employed to answer research questions?
5. What conclusions have been made from existing research? That is, what questions from RQ2 have been answered, and what are these answers? (These constitute lessons learned.)
6. What questions from RQ2 are left unanswered? (These present directions for future research.) Unanswered questions fall into two categories:
 - No quality research has been conducted to address the question.
 - Research has been conducted, but the results are inconclusive, or they are suspect because of the unsatisfactory quality of the research.

2. METHODOLOGY

The first step of our literature review was to *retrieve the subject material*. For our Wikipedia review, the recent nature of the subject matter does not lend itself to physical browsing in a library. Instead, we reviewed all 484

databases available to researchers at Concordia University spanning all areas of inquiry: business/commerce, the fine arts, humanities, science and engineering, and the social sciences. Although keywords are normally a critical choice in a literature review, in the case of our systematic mapping, the choice was very simple: we searched *only* for the words “Wikipedia”, “Wikipedian” and “Wikipedians”. To validate the exhaustiveness of our databases, we will verify our results by consulting with subject “experts” [4], which for the purpose of this study constitute the Wikimedia Foundation’s compilation of bibliographies of research on Wikimedia projects (including non-Wikipedia studies) [6,7], as well as the Wikimedia researchers’ mailing list (Wiki-research-l) hosted by the Wikimedia Foundation [8]. By matching our own database search to these independent bibliographies, we were able to establish with certainty that our databases would tag all pertinent work and we were not missing any important search mechanisms. For managing our references, we use Zotero, an open source bibliography manager.

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. Based on Fink’s [4] criteria for the *practical screen*, we restrict ourselves to content limitations set by RQ1: refereed articles that bring up one of the key terms in the title, abstract or subject only- and to English-language articles. In March 2009 we conducted a preliminary search in the original 484 databases. We determined a final list of 34 databases that retrieved a total of 2,115 articles. Eliminating duplicates and those that do not fit our practical screen, we anticipate our final article collection to comprise roughly 1,000 articles. We are presently refining this initial search, and thus do not have an exact number at this time.

It is essential that all reviewers involved in a systematic literature review study be *thoroughly trained* in both note-taking and reviewing techniques. In the case of our Wikipedia study, we have three reviewers. Two selected the articles for inclusion while a third acts as a “gold standard” [4] to reconcile disagreements in article selection and scoring. To establish the rigour of our work, we also randomly exchange articles among reviewers and calculate inter-rater reliability in inclusion and exclusion decision. Once all potentially eligible articles have been collected, the next step would normally be to screen them for inclusion criteria based on the methodology followed for carrying out the studies, that is, based on the quality of the study execution. However, because of the broad nature of this systematic mapping review, we will not rate or exclude studies based on their quality and instead move directly to the next step, *data extraction*.

In order to answer research questions 2 to 6, we need to read the papers more carefully than was done in our original scan for inclusion. Because of the sheer quantity of the studies, it will not be practical to read each paper thoroughly. However, with a focused data extraction protocol, the necessary answers to the research questions can be systematically obtained and recorded. The final requirement will be the *synthesis of data*. A systematic

mapping study is too high-level to try to combine studies with a quantitative meta-analysis. For the purpose of answering our research questions, our analysis will consist primarily of tabulation and qualitative synthesis methods. A prerequisite to synthesizing the data will be to code the extracted data into combinable categories whenever this makes sense. For each category, we will calculate the number and percentage of studies. In this section, we discuss the necessary synthesis of data to answer each research question. When the data is synthesized, we will write up the systematic mapping review following Kitchenham’s guidelines [5].

3. CONCLUSION

We believe that scholarly research is a critical contributor to thoroughly understanding the workings of Wikipedia, an important and widely used global resource. This mapping study will make sense of the varied research that has been done to date, outline research conclusions drawn, as well as highlight implications for policy and administration for Wikipedia contributors, administrators, and the Wikimedia Foundation.

4. ACKNOWLEDGEMENTS

This study is funded by the Social Sciences and Humanities Research Council of Canada. We thank Bilal Abdul Kader for his assistance in searching the literature databases.

5. REFERENCES

- [1] C. Okoli, “A Brief Review of Studies of Wikipedia in Peer-Reviewed Journals,” *Proceedings of the Third International Conference on Digital Society, 2009.*, Cancun, Mexico: 2009, pp. 155-160.
- [2] P. Ayers, “Researching wikipedia - current approaches and new directions,” *Proceedings of the American Society for Information Science and Technology*, vol. 43, 2006, pp. 1-14.
- [3] N. Miller, “Wikipedia revisited,” *ETC: A Review of General Semantics*, vol. 64, Apr. 2007, pp. 147-150.
- [4] A. Fink, *Conducting Research Literature Reviews: From the Internet to Paper*, Thousand Oaks, California: Sage Publications, 2005.
- [5] Barbara Kitchenham, *Guidelines for performing Systematic Literature Reviews in Software Engineering*, Keele, UK: Keele University, 2007.
- [6] Wikimedia contributors, “Research - Meta,” *Wikimedia Foundation*, Jan. 2009.
- [7] Wikipedia contributors, “Wikipedia:Wikipedia in academic studies - Wikipedia, the free encyclopedia,” *Wikipedia*, Jan. 2009.
- [8] Wikimedia Foundation, “Wiki-research-l Info Page,” *Wikimedia Mailing Lists*, Jan. 2009.