

WikiLit: Collecting the Wiki and Wikipedia Literature

Phoebe Ayers
UC Davis
Phys. Sci. & Eng. Library
Davis, CA 95616
(530) 752-9948
psayers@ucdavis.edu

Reid Priedhorsky
IBM T.J. Watson Research Center
1 Rogers St.
Cambridge, MA 02142
reid@reidster.net

ABSTRACT

This workshop has three key goals. First, we will examine existing and proposed systems for collecting and analyzing the research literature about wikis. Second, we will discuss the challenges in building such a system and will engage participants to design a sustainable collaborative system to achieve this goal. Finally, we will provide a forum to build upon ongoing wiki community discussions about problems and opportunities in finding and sharing the wiki research literature.

Categories and Subject Descriptors

H.3.7 [Information Storage and Retrieval]: Digital Libraries

General Terms

Documentation

Keywords

wikis, Wikipedia, research literature, literature databases

1. INTRODUCTION

In the 16 years since wikis were first introduced by Ward Cunningham, and in the 10 years since Wikipedia was founded, wiki technology and projects have been a source of fascination for the research community. Thousands of research papers explore the technical and social aspects of various wikis and wiki communities. Making sense of such a large body of work is a problem in many areas of research, not just those that study topics related to wikis. However, there are four important characteristics of our research area that make this a uniquely difficult and interesting problem.

First, the wiki literature is highly cross-disciplinary, ranging from sociology to theoretical physics to education, and papers may be unintelligible for researchers attempting to cross field boundaries. This broadness in scope also means that the wiki research literature is not collected in a single traditional library database (for example, mathematical literature collections contain few social science papers, and vice versa) and different fields have different standards of which publications are considered archival (notably,

conference publications in computer science typically pass a rigorous gauntlet of peer review, while conferences in other fields have high acceptance rates and focus on preliminary, non-archival results). Researchers are left to depend on complicated, error-prone literature searches or general-purpose tools like Google Scholar and can thus overlook important publications.

Second, literature about Wikipedia dominates the field, which can make it difficult to locate research about wikis in general or other wiki projects. Third, much of the research literature is of interest to wiki practitioners in addition to researchers, but paywalls and quirks of academic publishers often make this literature difficult for practitioners in industry or wiki communities to access. This means that research results of practical benefit to the wiki community are again easily overlooked. Finally, research is available in a wide variety of both traditional and non-traditional venues. There are many hundreds of white papers, in-wiki community documents, and news articles adding to the huge number of traditionally peer-reviewed academic publications. For the new (or even the experienced) wiki researcher, the sheer quantity and variety of information is often overwhelming.

As a response to this problem, the same collaborative impulse that made wikis a successful and widespread technology has also led to several proposals and projects within the wiki community to collaboratively collect, summarize, and share this rapidly growing body of research literature. Tools for managing the wiki research literature have ranged from dedicated wikis and databases (such as the early Wikindx-based bibliography, originally hosted by Wikimedia Germany [1] or the general AcaWiki, which is used by some wiki researchers [2]) to Zotero groups [3] and in-wiki lists (such as those on Wikipedia [4]), to comprehensive literature review projects [5, 6]. However, these tools have suffered from several problems, notably lacking stable long-term maintenance and standards agreed upon by the community. Some early projects have disappeared from the Internet, while others are simply hopelessly incomplete. Traditional one-time literature reviews suffer from the impracticality of summarizing such a massive and diverse body of knowledge in one paper, and rapidly become dated as new research is published in great volume.

Despite these difficulties, there is deep and active interest in the notion of building a shared and open literature organizing system, as has been evidenced by continued efforts to build such a system and recent discussions on the wiki-research-l mailing list [7]. Such projects are often influenced by interest in the appropriateness of wikis themselves to build an editable annotated literature database of useful scale, an approach that has been to date explored only partially.

Copyright is held by the author/owner(s).

WikiSym '11, October 3–5, 2011, Mountain View, CA, USA.
ACM 978-1-4503-0909-7/11/10.

2. AIMS AND OBJECTIVES

While the notion of a centralized database of wiki and Wikipedia research has been much discussed in diverse places, the research community will benefit from a dedicated, in-person session at WikiSym 2011 to discuss this problem and focus toward a solution. This workshop will (a) explore efforts to collect the wiki and Wikipedia literature to date, (b) discuss further ideas for moving forward, and (c) engage participants to enumerate requirements for a sustainable system to collect, preserve, disseminate, and summarize the wiki and Wikipedia research literature. Ideally, this session will provide a roadmap for those interested in the topic to move forward in building such a collaborative literature system. The workshop organizers will actively solicit participation from community members who have been active in or who have expressed interest in past efforts to address this problem.

3. WORKSHOP STRUCTURE

This will be a working session. We will divide the workshop into three parts:

- First, we will outline the impetus behind building a wiki literature system and give a brief history of prior and ongoing efforts to do so.
- The bulk of the workshop will be participatory. We will discuss requirements and specifications for a system for collecting the wiki literature, including specifically sustainability-related issues. We will also analyze possible tools matching these specifications.
- Finally, we will conclude by collaboratively building a roadmap for the future, including recommendations for improvements to existing systems and/or design of new ones.

The specifications, roadmap, and other artifacts we produce will be published.

At the end of the session, participants will have a concrete idea of what sorts of tool(s) are needed and how they can be created, joined, or acquired.

As this is a participatory workshop, workshop leaders will frame the problem and facilitate discussion with working outcomes. We ask all workshop participants to arrive prepared to actively engage in the subject. Other leaders may be added.

4. ORGANIZER BIOGRAPHIES

Phoebe Ayers is the Computer Science and Electrical Engineering Librarian at the University of California, Davis and a Wikimedia Foundation Trustee. She has been involved in Wikimedia projects for many years, and holds a long-standing interest in bringing the wiki practitioner and research communities closer through shared conferences and projects.

She brings a thorough knowledge of the Wikipedia research literature and of previous projects to collect and disseminate it as well as deep knowledge of traditional library databases and other literature-gathering projects.

Reid Priedhorsky is a Research Staff Member at the IBM T.J. Watson Research Center in Cambridge, Massachusetts, USA. He has been a wiki researcher for several years, with a special focus on geographic wikis such as Cyclopath, a wiki for bicyclists which he founded and whose construction he led. He has also published research related to Wikipedia and brings a long-standing interest in building better tools for managing the scholarly literature.

5. REFERENCES

- [1] Wikibibliographie Encyclen. http://wikindx.inrp.fr/biblio_encyclen/index.php
- [2] Acawiki. <http://acawiki.org/Home>
- [3] Zotero Wikipedia research group. <http://www.zotero.org/groups/6693>
- [4] Academic studies of Wikipedia. http://en.wikipedia.org/wiki/Wikipedia:Academic_studies_of_Wikipedia
- [5] Okoli, C. et al. 2011. In progress.
- [6] Nielsen, F Å. 2011. Wikipedia research and tools: Review and comments. http://www2.imm.dtu.dk/pubdb/views/edoc_download.php/6012/pdf/imm6012.pdf
- [7] Wiki-research-l mailing list, March 2011. <http://lists.wikimedia.org/pipermail/wiki-research-l/2011-March/thread.html>