

The Democratization of Semantic Properties: An Analysis of Semantic Wikis

Yolanda Gil

Information Sciences Institute
University of Southern California
4676 Admiralty Way
Marina del Rey, CA 90292
+1-310-822-1511
gil@isi.edu

Angela Knight

Kevin Zhang
Larry Zhang
Harvard-Westlake School
700 North Faring Road
Los Angeles, CA 90077

Varun Ratnakar

Ricky Sethi
Information Sciences Institute
University of Southern California
4676 Admiralty Way
Marina del Rey CA, 90292
varunr@isi.edu, ricky@sethi

Abstract—Semantic wikis augment wikis with semantic properties that can be used to aggregate and query data through reasoning. Semantic wikis are used by many communities, for widely varying purposes such as organizing genomic knowledge, coding software, and tracking environmental data. Although wikis have been analyzed extensively, there has been no published analysis of the use of semantic wikis. In this paper, we analyze twenty semantic wikis selected for their diverse characteristics and content. We analyze the property edits and compare to the total number of edits in the wiki. We also show how semantic properties are created over the lifetime of the wiki.

Keywords- *Semantic wikis, social knowledge collection, Semantic Web, knowledge capture*

I. INTRODUCTION

Wikis are one of the most popular frameworks for collaboration on the Web. Wikis are easy to use, track the provenance and the history of all user changes, and scale well to thousands of users. Semantic wikis augment wikis with the ability to structure information through semantic annotations, and with reasoning capabilities that exploit that structure to organize the wiki's knowledge [3]. Semantic wikis are wikis with extensions that support the creation of structured content. A semantic wiki allows users to organize topic page categories as classes in an ontology, and to define properties that apply to each class. As content is added using these structured properties, the semantic wiki can use reasoning and inference. Users can then query the content to generate dynamic content for wiki pages. Semantic wikis can be seen as a microcosm of the Semantic Web, since users exploit semantic technologies while retaining a very accessible Web collaboration interface.

There are many implementations of semantic wikis. [3] give a detailed overview of semantic wikis and a thorough comparison of semantic wiki frameworks. Semantic MediaWiki [10] is a diverse set of extensions for the popular MediaWiki wiki platform, and allows users to easily create new structured properties. OntoWiki [1] is another semantic wiki that requires that a schema be defined before users enter content to populate it. It is not a proper wiki system but rather

a form-based web-interface. AceWiki [11] provides a more powerful knowledge representation formalism than most other semantic wikis, with the cost of requiring the contributors to learn and use a semi-formal logical language.

There are hundreds of semantic wikis in use. Some semantic wikis have a serious use, such as scientific knowledge organization. Others have practical use, for example gardening or restaurant finding. Some wikis have users that are on the younger side, such as teens who are card traders and want to organize all the information about the different characters in the cards.

In this paper, we present an analysis of twenty semantic wikis. This represents an initial formative analysis that can provide indications of how semantic wikis are used. This kind of analysis can be the basis to understand communities of semantic wiki users and improve the support and usability of semantic wikis.

There is no prior work, to our knowledge, on analyzing user behaviors in semantic wikis. There are many published analyses of wikis. Many studies are focused on Wikipedia, aimed at understanding social content collection systems, with topics as varied as the editorial process [19], incentives to contributors [13], critical mass of contributors [18], coordination across contributions [8], group composition [12], conflict [9], trust [17], and user interaction design [15]. Other work focuses on extracting structured content from wikis [2; 20]. Structured knowledge collection from volunteers has been studied in prior research, including OpenMind [14], the Cyc FACTory [16], and Learner [4]. [5] has analyzed several sources in order to identify knowledge gaps in wikis, and particularly in semantic wikis.

The contributions of this paper are twofold. First, we present a compilation of data about twenty semantic wikis that focus on a variety of topics and with diverse communities. Second, we present a quantitative analysis of the twenty semantic wikis regarding the creation of structured properties and the editors that create them. We discuss patterns of use that are common across wikis. These patterns could be the

basis to understand communities of semantic wiki users and improve the support and usability of semantic wikis.

II. BACKGROUND: SEMANTIC WIKIS

Semantic wikis are wikis with extensions that support the creation of structured content. Traditional wikis support some ways to structure content, for example by assigning categories to topic pages. In Wikipedia, for example, infoboxes are defined for countries, soccer players, etc. A semantic wiki allows users to organize topic page categories as classes in an ontology, and to define properties that apply to each class. Semantic wikis allow users to constrain properties by the range of values that they can take, which are called structured properties. As content is added using these structured properties, the semantic wiki can use reasoning and inference. Users can then query the content to generate dynamic content for wiki pages.

[3] give a detailed overview of semantic wikis and a thorough comparison of semantic wiki frameworks.

III. GOALS OF THE ANALYSIS

The goal of our work is to understand the use of structured properties in existing semantic wikis. For this work, we use Semantic MediaWiki (SMW)¹, as it is a very popular semantic wiki framework. There is a site that compiles a list of 489 existing Semantic MediaWikis².

Table 1 has a brief description of the wikis analyzed. We selected a sampling of wikis that had interesting topics, diverse communities, and different sizes. Our analysis focuses on 20 SMWs. It is meant to be informative of the kinds of behaviors that users make of the semantic aspects of the wikis. The wikis were chosen for the diversity of topics covered, communities forming them, and size of their contents.

As in any wiki, the contributor data is publicly accessible through the history pages of the wikis. SMWs have APIs that allow external access of aggregate data about user activities as well as the structured content.

The use of semantic properties leads to a distinction among semantic wiki users. We use the term users to refer to all people who access the wiki. Viewers are users who simply access information in the wiki without editing or contributing content. Editors are users who change any information stored within the wiki. We refer to property editors as the subset of editors who edit structured properties.

Table 2 shows overall statistics for the wikis. The wikis vary widely in terms of age, size, proportion of editors to users, and amount of structured properties.

Table 1: Overview of the semantic wikis analyzed.

Semantic Wiki	Description
Artwiki	Artist profiles
Foodfinds	Restaurants
Biodiversityofindia	Biology of life in India
Beachapedia	Data on coastal areas
Rosettacode	Collaboration on coding problems
Mozilla	Mozilla software development
Wikidevi	Computer hardware
Deepskypedia	Observations by astronomers
Roadsignmath	Road signs with interesting properties
Gardenology	Plants/gardening
Dnd	Dungeons and dragons
Stowiki	Reference for online Star Trek game
Farmafrikipedia	Farming practices in Africa
Dexid	Product encyclopedia
Neurolex	Neuroscience terms
Navi	Fan site of the Navi in the movie Avatar
Genewikiplus	Genes and proteins
Practicalplants	Plant cultivation info
Scientolipedia	Scientology
Enipedia	Applications of wikis to energy issues

IV. ANALYSIS OF SEMANTIC WIKIS

We asked a series of specific questions about the use of the semantic aspects of the wikis.

Are users creating semantic properties?

Users are creating hundreds of properties in some wikis, and dozens of properties in most wikis. The number of property edits is larger than the number of properties, showing that the properties are being changed after their creation. But that number is not much larger, so there is on average a small number of edits per property.

How many users of the wiki create semantic properties?

The number of property editors tends to be small, and typically much smaller than the number of general editors.

Five wikis have one user making 100% of the property edits: Practicalplants, Farmafrikipedia, Roadsignmath, Scientolipedia, and Wikidevi. Most wikis have a small group of property editors (under 10), though many of them are quite large in terms of the total number of pages and the total number of editors.

Mozilla shows by far the largest number of property editors, with more than 320. Neurolex has 36 property editors, which is a sizeable group. Enipedia has 14 editors. Since the set of wikis that we analyzed is a sample, there must be other wikis with similarly large groups of editors creating semantic properties. Further analysis would be required to determine whether there is overlap in the properties that these users edit,

¹ <http://www.semantic-mediawiki.org>

² <http://smw.referata.com/wiki/Special:BrowseData/Sites>

and how they coordinate their edits so that wiki contents have overall consistency.

Does the number of semantic property editors grow in proportion to the number of general editors?

Figure 1 shows the editors for each wiki. The graph shows that the number of property editors does not increase in proportion with the number of editors.

How many semantic properties do different editors create?

Figure 2 shows the proportion of property edits done by each property editor in descending order. We group the wikis into four subsets that seem to follow distinct patterns. In one pattern, shown in Figure 2(a), one, two, or three editors are responsible for all the edits to the structured properties. We note that the plots for Navi and Dexid are superimposed with one another and not clearly visible, and that it is also hard to see the five wikis mentioned with one single editor. In these wikis, these few editors keep a close lid on the structure of the wiki, sometimes reversing the edits of other users. We refer to this behavior as a monarchy, as a very small set of users controls the structure of the wiki. We observed a second pattern, shown in Figure 2(b), where more than 3 but less than 10 editors are responsible for the edits to the semantic properties. We refer to these wikis as an oligarchy. Two additional patterns are shown in Figure 2(c) and 2(d), with 10-40 editors and >40 editors respectively. We describe these patterns as a republic and a democracy respectively.

Do overall edits follow similar patterns to semantic property edits?

We wondered if the patterns we identified for semantic property edits are also followed for general edits. Figure 3 shows the proportion of all page edits done by each of the top

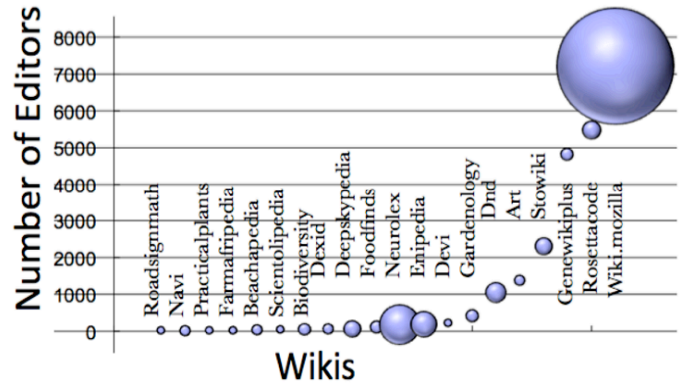


Figure 1: The number of editors versus the number of property editors shown as a bubble chart. The location of the sphere is the number of editors, and the area of the sphere indicates the number of property editors.

20-40 editors in descending order. We observed similar kinds of patterns as we group the wikis in the different graphs, shown in Figures 3(a), (b), (c), and (d). Figure 3(a) shows a monarchy pattern, where almost the entirety of the proportion of edits (over 95% in all cases) is done by just 1-3 editors. Figure 3(b) shows an oligarchy pattern, where the first few editors do a large proportion of edits but even the top editor is below 77% of the edits in most cases. Figure 3(c) shows a republic pattern, where the top editor accounts for at most 14% of the edits.

Figure 3(d) shows a democracy pattern, where individual editors account for a maximum of 3.8% of the edits. In both the republic and democracy patterns there is a much greater number of editors accounting for the majority of the proportion of edits.

Table 2: Statistics for the semantic wikis analyzed.

Semantic Wiki	Number of pages	Number of properties	Number of editors	Number of property editors	Number of page edits	Number of property edits	Start date
Artwiki	5197	53	1365	2	24182	76	Dec 2011
Foodfinds	521	13	107	3	1292	40	Mar 2008
Biodiversityofindia	1460	189	39	3	4238	360	May 2005
Beachapedia	1246	26	31	2	15484	44	May 2009
Rosettacode	2705	53	5461	7	80114	99	Jan 2007
Mozilla	27381	657	7202	322	371009	5905	Jul 2004
Wikidevi	7526	141	207	1	49977	178	May 2009
Deepskypedia	29175	690	48	6	65538	1394	Oct 2008
Roadsignmath	241	21	6	1	5159	26	May 2010
Gardenology	97595	42	398	3	114918	79	Jun 2005
Dnd	11907	113	1035	9	65403	208	Nov 2005
Stowiki	8050	135	2292	6	53618	217	May 2009
Farmafrikipedia	244	4	22	1	405	8	Mar 2011
Dexid	24656	58	47	2	60449	187	Jan 2010
Neurolex	27279	163	159	36	34095	373	Jul 2008
Navi	5299	63	8	2	19804	207	Jan 2011
Genewikiplus	45274	25	4806	3	129942	36	Feb 2002
Practicalplants	13936	36	10	1	60754	86	Jan 2012
Scientolipedia	292	50	31	1	2280	89	Dec 2011
Enipedia	95121	463	171	14	223703	549	Feb 2010

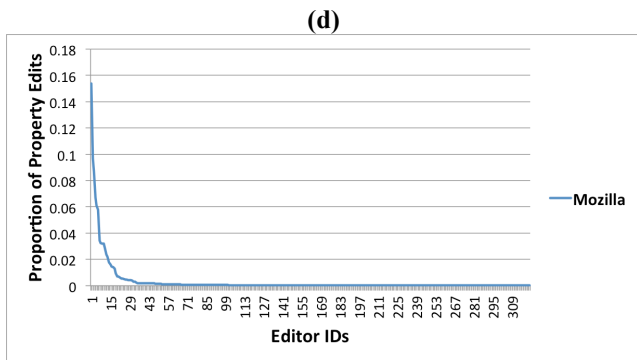
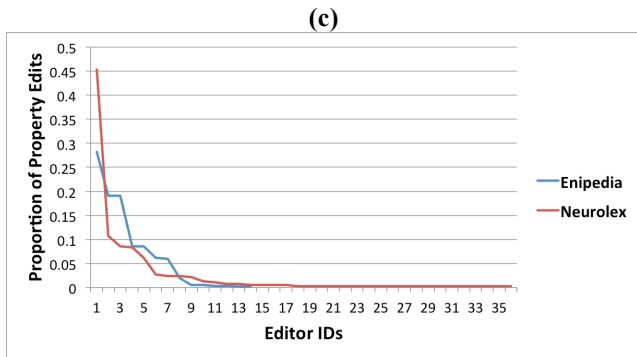
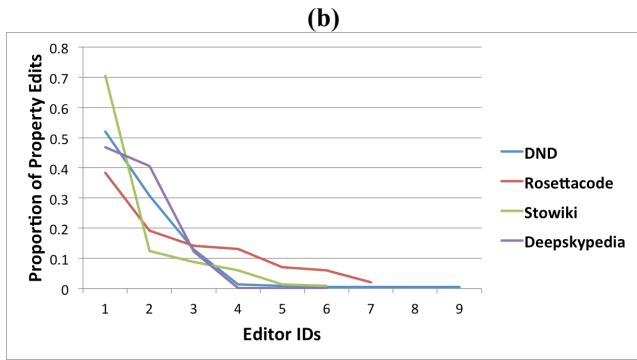
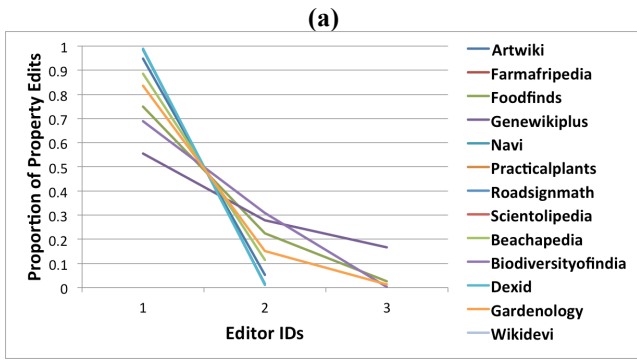


Figure 2: The proportion of property edits made by editors varies widely across semantic wikis: (a) 2-3 editors, (b) 4-9 editors, (c) 10-40 editors, and (d) >40 editors.

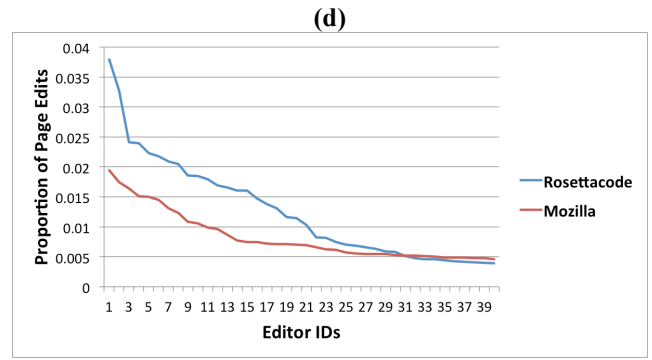
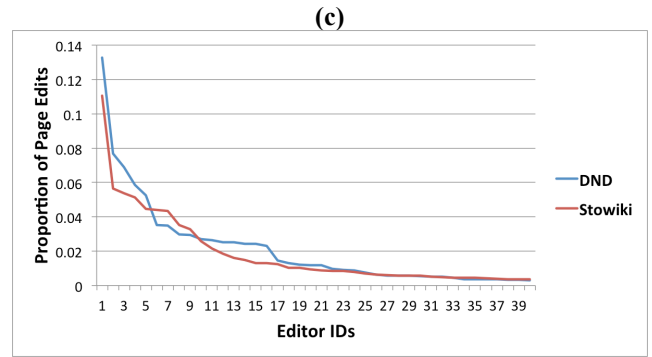
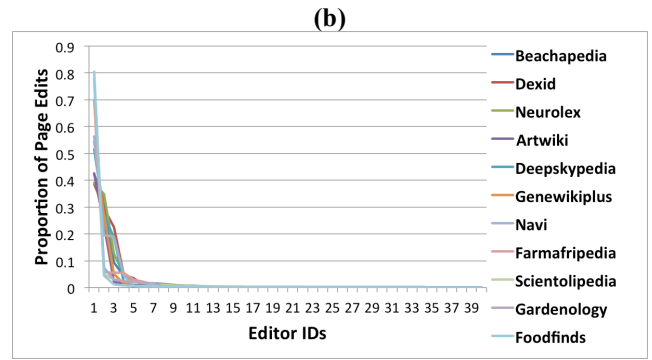
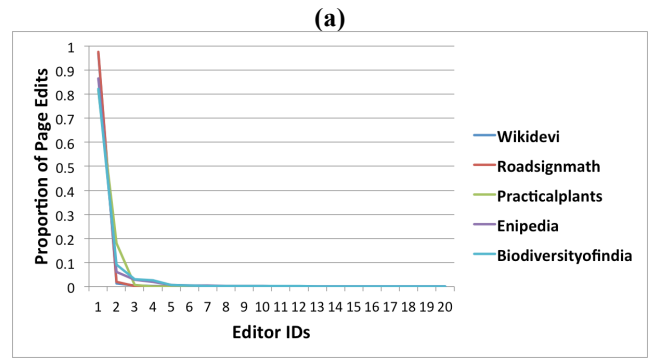


Figure 3: The pattern of proportion of page edits made by editors varies widely across semantic wikis: from a monarchy pattern (a) to a democracy pattern (d).

Do semantic property edits and general edits follow the same pattern for a given wiki community?

Interestingly, the pattern that a semantic wiki community follows for semantic property edits does not necessarily correspond to the pattern of all edits, as can be seen by comparing the subgroups of wikis in Figures 2 and 3. There are a few exceptions. There are two wikis that have a monarchy for both semantic properties and regular edits: Roadsignmath and Practicalplants. There is one wiki where there is an oligarchy for both edit categories, which is Deepskypedia. No wikis had the same republic behaviors in both categories. Mozilla is a democracy for both semantic property edits and general edits.

Do the same users that edit most of the semantic properties also do most of the general edits?

We looked at whether there was overlap in the top editors for semantic properties and for general edits.

- The wikis that had one user doing all the semantic property edits (Practicalplants, Farmafripedia, Roadsignmath, Scientolipedia, and Wikidevi) had the same user do the vast majority of the edits. As we mentioned, for Practicalplants and Roadsignmath only that user did the vast majority of the regular edits. However, in Farmafripedia, Scientolipedia, and Wikidevi there were several other users (an oligarchy) who contributed the vast majority of regular edits.
- The wikis that had a democracy in regular edits, Rosettacode and Mozilla, had different top users for semantic properties than the top users for regular edits.
- For most wikis, the same user was the top editor for property edits and for regular edits, including Artwiki, Foodfinds, Biodiversity, Beachapedia, Deepskypedia, Gardenology, Navi, and Practicalplants.

We noted that some wikis use bots to make edits, and there is a convention of how to name bots to make it transparent. In some cases, a bot extracts content from other wikis and places that content in the semantic wiki. These wikis are referred to as “meta-wikis.” An example is GeneWiki+, where the content comes from Wikipedia’s GeneWiki portal. When Wikipedia users edit information about genes, then that information is imported by bots into GeneWiki+. Users are not allowed to directly edit GeneWiki+, only indirectly by editing Wikipedia.

However, it is not easy to identify bots automatically. According to the Mediawiki Manual, the "bot" user right and the "bot" user group are not reliable indicators to determine if a revision is created by a bot. And the "bot" flag, which is the only reliable indicator, is only set for recent revisions and not for all revisions in most wikis.

V. DISCUSSION

Further analysis is needed to understand the behaviors of the semantic wiki user communities. This will require obtaining finer grained data about the nature of the edits.

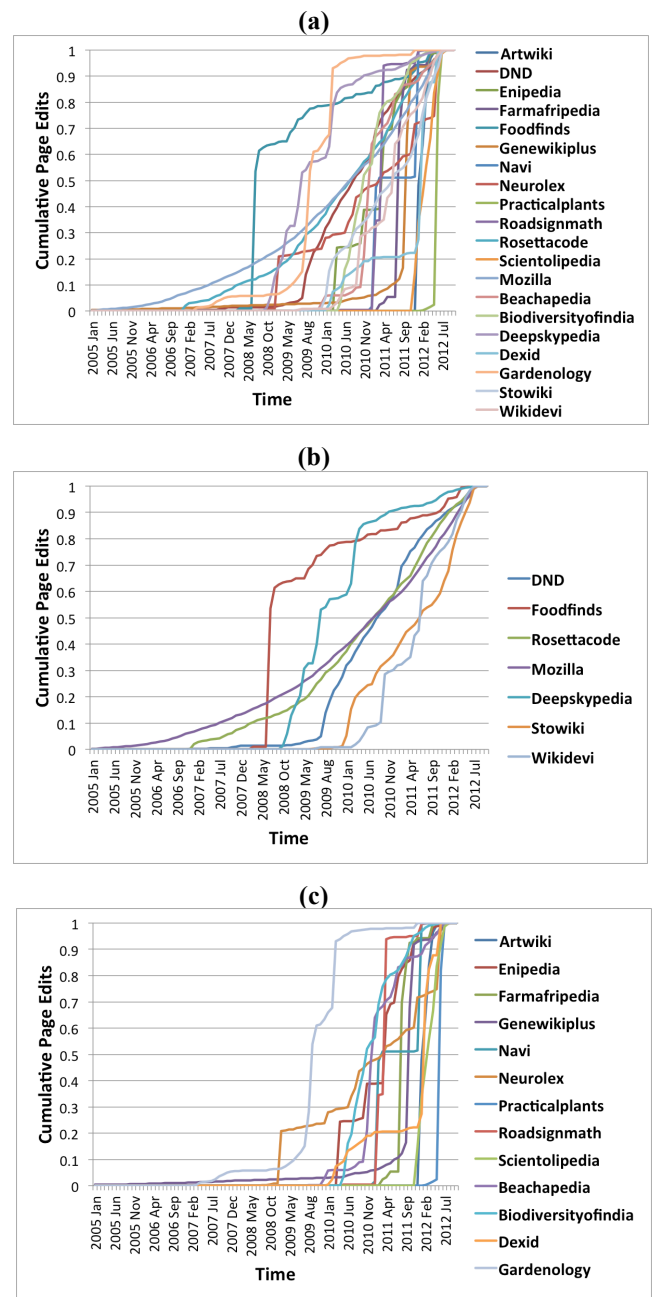


Figure 4: Cumulative page edits over time.

As an example, consider the cumulative page edits over time shown in Figure 4. Figure 4(a) shows all the wikis. We found a distinction between wikis that had a sloped growth, shown in Figure 4(b), and wikis that had spiked growth, shown in Figure 4(c). An interesting question is whether the spiked growth corresponds to the creation or edit to a structured property in the wiki. In order to answer this question, we would have to compile data about the nature of the individual edits to the wiki, which involves parsing history pages and extracting whether each edit is related to a structured property or not. We believe that further analysis with additional data will be required to understand the impact of semantic properties in the addition of content to semantic wikis.

An interesting question is what proportion of structured properties is automatically created by bots. We discussed an example earlier: GeneWiki+. We noticed that the use of bots is prevalent in many semantic wikis. These may be global bots for the whole wiki, or automated scripts created by individual users. But bot or script usage is hard to quantify from the data that can be obtained through the API. Parsing of user names might reveal bots if users follow the naming conventions.

VI. CONCLUSION

We presented an analysis of semantic wikis, looking at the creation of semantic properties that help structure the contents of the wikis. We identified four patterns that those semantic wiki communities follow, from very few users defining all the semantic properties to a large set of users defining them. We found that the amount of editors does not correlate with the amount of property editors in semantic wikis. We plan to compile additional data for further analysis about the effect of structured property edits in the growth of wiki content, the use of bots to create structured properties, and uncover further patterns of behavior in semantic wikis. These kinds of analyses will enable a better understanding of the use of semantic technologies in these collaboration platforms, and improve their effective use and dissemination.

ACKNOWLEDGMENT

This research was supported in part by a grant from the US National Science Foundation (NSF) through award number IIS-1117281.

REFERENCES

- [1] Soren Auer, Sebastian Dietzold, and Thomas Riechert. "OntoWiki - A Tool for Social, Semantic Collaboration." 5th International Semantic Web Conference, 2006.
- [2] Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, Zachary Ives. "DBpedia: a nucleus for a web of open data." Proceedings of the 6th international semantic web conference, 2007.
- [3] François Bry, Sebastian Schaffert, Denny Vrandečić and Klara Weiland. "Semantic Wikis: Approaches, Applications, and Perspectives." Lecture Notes in Computer Science, Reasoning Web. Semantic Technologies for Advanced Query Answering, Volume 7487, 2012.
- [4] Tim Chklovski and Yolanda Gil. "Towards Managing Knowledge Collection from Volunteer Contributors," Proceedings of the 2005 AAAI Spring Symposium on Knowledge Collection from Volunteer Contributors (KCVC), Stanford, CA, March 2005.
- [5] Hans-Jörg Happel. "Woogle4MediaWiki: from searchers to contributors." Proceedings of the 6th International Symposium on Wikis and Open Collaboration, 2010.
- [6] Aniket Kittur, Bongwon Suh, Ed H. Chi. "Can you ever trust a wiki? Impacting perceived trustworthiness in Wikipedia." Proceedings of the ACM conference on Computer supported cooperative work, 2008.
- [7] Aniket Kittur, Robert E. Kraut. "Harnessing the wisdom of crowds in Wikipedia: Quality through coordination." Proceedings of the ACM conference on Computer supported cooperative work, 2008.
- [8] Aniket Kittur, Bryant Lee, Robert E. Kraut. "Coordination in collective intelligence: the role of team structure and task interdependence." Proceedings of the 27th international conference on Human factors in computing systems, 2009.
- [9] Aniket Kittur, Robert E. Kraut. "Beyond Wikipedia: coordination and conflict in online production groups." Proceedings of the 2010 ACM conference on Computer supported cooperative work.
- [10] Markus Krötzsch, Denny Vrandečić, Max Völkel, Heiko Haller, Rudi Studer. "Semantic Wikipedia." Journal of Web Semantics, 5(4), pages 251-261, December 2007.
- [11] Tobias Kuhn. "AceWiki: A Natural and Expressive Semantic Wiki." Proceedings of the Fifth International Workshop on Semantic Web User Interaction (SWUI 2008), CEUR Workshop Proceedings, Volume 543, 2009.
- [12] Shyong (Tony) K. Lam, Jawed Karim, John Riedl. "The effects of group composition on decision quality in a social production community." Proceedings of the 16th ACM international conference on Supporting group work, 2010.
- [13] J. Leskovec, D. Huttenlocher, J. Kleinberg. Governance in Social Media: A case study of the Wikipedia promotion process. Proceedings of the AAAI International Conference on Weblogs and Social Media (ICWSM), 2010.
- [14] Henry Lieberman, Hugo Liu, Push Singh, and Barbara Barry. "Beating some common sense into interactive applications." AI Magazine, 2004.
- [15] Raphael Hoffmann, Saleema Amershi, Kayur Patel, Fei Wu, James Fogarty, Daniel S. Weld. "Amplifying Community Content Creation Using Mixed-Initiative Information Extraction." Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI), 2009.
- [16] Cynthia Matuszek, Michael J. Witbrock, Robert C. Kahlert, John Cabral, David Schneider, Purvesh Shah, Douglas B. Lenat. "Searching for Common Sense: Populating Cyc from the Web." Proceedings of the 20th National Conference on Artificial Intelligence (AAAI), 2005.
- [17] Deborah L. McGuinness, Honglei Zeng, Paulo Pinheiro da Silva, Li Ding, Dhyanesh Narayanan, and Mayukh Bhaowal. "Investigations into Trust for Collaborative Information Repositories: A Wikipedia Case Study." Proceedings of the Workshop on Models of Trust for the Web, 2006.
- [18] Daphne R. Raban, Mihai Moldovan, Quentin Jones. "An empirical study of critical mass and online community survival." Proceedings of the ACM conference on Computer supported cooperative work, 2010.
- [19] Diomidis Spinellis and Panagiotis Louridas. "The Collaborative Organization of Knowledge." CACM, August 2008.
- [20] Daniel S. Weld, Fei Wu, Eytan Adar, Saleema Amershi, James Fogarty, Raphael Hoffmann, Kayur Patel, Michael Skinner. "Intelligence in Wikipedia." Proceedings of the 23rd national conference on Artificial intelligence, 2008.