

# Knowledge Capture in the Wild: A Perspective from Semantic Wiki Communities

Yolanda Gil and Varun Ratnakar  
Information Sciences Institute  
University of Southern California  
4676 Admiralty Way  
Marina del Rey CA, 90292  
gil@isi.edu, varunr@isi.edu

## ABSTRACT

Semantic wikis augment wikis with semantic properties that can be used to structure content that can therefore be aggregated and queried through reasoning. Semantic wikis have been adopted by many communities for very diverse purposes, such as organizing genomic knowledge, coding software, learn about hobbies, and tracking environmental data. Although wikis have been analyzed extensively, there has been little analysis of the use of semantic wikis. In this paper, we analyze the formalization of knowledge in 230 semantic wiki communities. We report our findings in terms of the edits of semantic concepts and properties, as well as the communities of editors for these semantic features of the wikis.

## Author Keywords

Semantic wikis; semantic web; social knowledge collection.

## ACM Classification Keywords

H.5.2 [Information Interfaces and Presentation]: User Interfaces - Interaction styles.

## INTRODUCTION

Wikis are one of the most popular frameworks for collaborative knowledge collection on the Web. Wikis are easy to use, track the history of all edits and their provenance, and scale well to thousands of users. Wikis sometimes have conventions that provide structure to their content. For example, Wikipedia has infoboxes to organize basic information for some types of wiki pages such as athletes, politicians, and countries. But even though infoboxes can be the basis for automatic extraction of knowledge bases from Wikipedia [Auer et al 2007; Weld et al 2008], the wiki itself cannot understand their structure or reason about their content. In contrast, semantic wikis augment wikis with the ability to structure information

through semantic annotations in RDF, for example through defining classes and properties [Bry et al 2012]. As a result, the contents of a semantic wiki can be aggregated and queried. Visualizations can be created by overlaying semantic information in maps or charts. While wikis can be seen as sites for social content collection, where people come together to share information in a virtual community, semantic wikis can be seen as frameworks for *social knowledge collection* where contributors structure content and organize it into knowledge structures [Gil 2013].

Semantic wikis are becoming very popular, as they offer the simplicity of a wiki with additional capabilities to organize content. There are hundreds of installations of semantic wikis used for very different purposes and communities. Some semantic wikis have a serious use, such as scientific knowledge organization. Others focus on hobbies, for example gardening or restaurant directories. Some wikis have users that are on the younger side, such as teens who are Yu-Gi-Oh card traders and want to organize all the information about the different characters in the cards.

Semantic wikis are an example of knowledge capture in the wild, where non-experts design and use knowledge structures in an organic way. Semantic wikis can be seen as a microcosm of the Semantic Web, since users exploit semantic web technologies while retaining a very accessible Web collaboration interface. While there have been many analysis of wiki communities (e.g. [Kittur et al 2009; Leskovec et al 2010]), there are no longitudinal studies about social knowledge collection in semantic wikis.

In prior work, we carried out an analysis of 20 semantic wikis regarding the creation of semantic properties [Gil et al 2013]. The wikis had a varying amount of users editing semantic properties, and we found some patterns that were common across wikis. That initial analysis was small and limited to property editing, but already raised the question of whether the patterns would occur in a larger set of wikis.

This paper presents an analysis of 230 semantic wikis, focusing on how users edit wikis to add structure to the contents. We analyze the concepts and properties created, the amount of editors involved in creating them, and the amount of edits for concepts and for properties. This

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.

K-CAP'13, June 23-26, 2013, Banff, Canada.

Copyright (c) 2013 ACM 978-1-45-03-2102-0/13/06... \$15.00

analysis contributes to the understanding of semantic wikis as a platform for social knowledge collection, and suggests improvements for the design of semantic wikis.

### RELATED WORK

[Bry et al 2012] gives a detailed overview of semantic wikis and a thorough comparison of two dozen alternative implementations, which include Semantic MediaWiki [Krötzsch et al 2007], OntoWiki [Auer et al 2006], and AceWiki [Kuhn et al 2009].

While there are many published analyses of wikis, semantic wiki communities have not been studied in depth. Many studies are focused on Wikipedia, with topics as varied as the editorial process [Spinellis and Louridas 2008], incentives to contributors [Leskovec et al 2010], critical mass of contributors [Raban et al 2010], coordination across contributions [Kittur et al 2009], group composition [Lam et al 2010], conflict resolution [Kittur and Kraut 2010], and user interaction design [Hoffman et al 2009]. Other work focuses on automatic extraction of structured content from wikis [Auer et al 2007; Weld et al 2008].

Structured knowledge collection from volunteers has been studied in prior research, including OpenMind [Lieberman et al 2004], the Cyc FACTory [Matuszek et al 2005], and Learner [Chklovski and Gil 2005].

### SEMANTIC WIKI DATA

We used data from 230 semantic wikis, all installations of Semantic MediaWiki so the data would be comparable. There is a list of Semantic MediaWiki installations<sup>1</sup>, with 489 sites at the time of this analysis. The wikis vary widely in topic, community, scope, size, and language. Of those, we selected those listed as active, which was a total of 362 wikis. From those, we selected those that had an easily accessible API and allowed anonymous read access which were a total of 255. Some wikis were unresponsive, which left us with 238 wikis. Of those, we removed several that had been clearly created by semantic web researchers, since we wanted to focus our analysis on the use by non-experts. Our final corpus contains 230 wikis.

Table 1 shows the metrics that we compiled for each wiki. The data was obtained through the Semantic MediaWiki API. For each wiki, we extracted the structured properties defined, noting each of the edits made to those properties and by what editors. We also extracted the concepts defined, the edits to those concepts and the editors. We also extracted counts of total number of pages, page edits, and total number of editors. The API does not provide some data that we would have wanted to collect, such as the number of properties or assertions in a given page. Although contributor data is publicly accessible through wiki history pages, that structure would need to be extracted for further analysis.

**Table 1. Metrics of editor activities in semantic wikis.**

Metric	Description
SC	Semantic concepts defined
SCC	Concept creators who edited concepts
SCE	Concept edits that changed concepts
SP	Structured properties defined
SPC	Property creators who edited structured properties
SPE	Property edits that changed structured properties
SPA	Property assertions
PP	Pages in the wiki
PPC	Contributors to the wikis that made any edits to pages
PPE	Page edits

**Table 2. Semantic wiki indices indicative of the proportion of semantic-related metrics to the overall size of a wiki.**

Index	Description
SC-Index	SC/PP
SP-Index	SP/PP
SPA-Index	SPA/PP
SPC-Index	SPC/PPC
SPE-Index	SPE/PPE
SPE-Density	SPE/SP
SPA-Density	SPA/SP

We removed the data for anonymous users. Although some of them might be making legitimate contributions, most anonymous users create spams in wikis and we wanted to eliminate that effect. Legitimate contributors normally want to be part of the community and be credited for their contributions, so they would not typically be anonymous. We also removed data for users blocked by the admins.

We separated the data from users identified as bots. Bots are sometimes set up in wikis to create content automatically. For example, a user may create a bot to import objects from an existing external site and create a page for each object. In the case of some wikis, their contents are almost entirely created by bots. This enabled us to differentiate contributions from bots and from users.

To compare the metrics collected across wikis of different sizes we defined several indices, summarized in Table 2, that assessed each wiki relative to its overall size:

- *SC-Index*: is the proportion of concepts defined in the wiki over the total amount of pages
- *SP-Index*: is the proportion of semantic properties defined in the wiki over the total pages
- *SPA-Index*: the proportion of property assertions over the total amount of pages
- *SPC-Index*: the proportion of semantic property contributors over the total contributors to the wiki
- *SPE-Index*: the proportion of semantic property edits over the total amount of edits to the wiki
- *SPE-Density*: the proportion of semantic property edits to the total amount of properties
- *SPA-Density*: the proportion of property assertions to the total amount of properties

<sup>1</sup> <http://smw.referata.com/>

**Table 3: Total pages, concepts defined, concept creators, and concept edits are shown for all users and for human users (no bots).**

Wiki Name	All Users (Humans + Bots)					Humans				
	PP	SC	SCC	SCE	SC-Index	PP-H	SC-H	SCC-H	SCE-H	SC-H-Index
VroniPlag Wiki	10492	148	14	291	0.01	10380	138	13	281	0.01
WikiTranslate	3147	109	1	160	0.03	3147	109	1	160	0.03
SaveMLAK	21493	104	13	183	0.00	3887	104	13	183	0.03
Yu-Gi-Oh! Wikia	18318	53	11	186	0.00	18318	53	9	160	0.00
SKYbrary	5577	49	2	158	0.01	5577	49	2	158	0.01
Free Software Directory	7181	38	3	65	0.01	131	6	2	15	0.05
Säsongsmat	1766	31	1	59	0.02	76	0	0	0	0.00
Narutopedia	4247	22	4	43	0.01	4247	22	4	43	0.01
Complex Operations Wiki	6872	18	1	42	0.00	6872	18	1	42	0.00
OpenEnergyInfo	78282	15	6	51	0.00	6622	15	6	51	0.00
EntropiaPlanets.com	2118	14	1	36	0.01	2118	14	1	36	0.01
Read Write Book Club	206	14	1	38	0.07	206	14	1	38	0.07
Salem Wiki	859	12	2	61	0.01	859	12	2	61	0.01
Virtual Skipper Wiki	189	12	1	32	0.06	188	12	1	32	0.06
Detective Conan Wiki	287	10	1	21	0.03	286	10	1	21	0.03
Links thing	5050	9	1	9	0.00	5050	9	1	9	0.00
Navi	1987	8	1	14	0.00	1955	8	1	14	0.00
NeuroLex	253	7	3	27	0.03	138	7	3	27	0.05
Digitabulum	17	4	1	8	0.24	17	4	1	8	0.24
Dungeons and Dragons Wiki	10444	4	1	5	0.00	10421	4	1	5	0.00
EMSwiki.com	3465	4	1	7	0.00	560	4	1	7	0.01
Natural History of Southeast Alaska	6983	4	1	70	0.00	6983	4	1	70	0.00
Practical Plants	11041	4	1	16	0.00	11041	4	1	16	0.00
Stargate Wiki	172	4	1	12	0.02	172	4	1	12	0.02
The Business Model Project	17949	4	1	7	0.00	17949	4	1	7	0.00
Mount Wiki	702	3	1	16	0.00	702	3	1	16	0.00
Van Hamel wiki	10824	3	1	4	0.00	10824	3	1	4	0.00
Web Community Wiki	10749	3	1	12	0.00	10127	3	1	12	0.00
WikiAplary	7273	3	1	5	0.00	691	3	1	5	0.00

All the metrics and indices have a corresponding one for human edits, named with a “-H” suffix.

### USE OF CONCEPTS

Defining concepts is one way for wiki users to create semantic structures in a wiki. One could view a wiki as being more semantic when more concepts are created and edited by more users.

Table 3 shows the data for wikis with more than 2 concepts defined (SC>2), 29 in total. It shows total pages, concepts defined, concept creators, and concept edits. We show these totals for all users (humans and bots) and human users only. Of the 230 wikis that we analyzed, only 53 wikis had concepts defined.

#### Concept Definitions

Table 3 shows that only 3 wikis had more than 100 concepts defined (SC>100), and only 15 had between 99 and 5 concepts defined (99>SC>5).

Only 10 wikis had an SC-Index over 0.02 (Digitabulum, Read Write Book Club, Virtual Skipper, Eroge, Detective Conan, WikiTranslate, NeuroLex, Stargate, Bioinformatics Core, and Säsongsmat). 9 other wikis had a value of 0.01, the rest had a value of 0.

Only 11 wikis had an SC-H-Index over 0.02 (Digitabulum, Read Write Book Club, Virtual Skipper, NeuroLex, Eroge, Free Software Directory, Detective Conan, WikiTranslate, SaveMLAK, Stargate, and Bioinformatics Core). 9 wikis had a value of 0.01, the rest had a value of 0.

#### Concept Contributors

Only 3 wikis had more than 10 concept contributors (shown in the table with SCC>10). Only 4 wikis had between 9 and 3 concept contributors. 4 wikis had 2, and 42 wikis had 1 concept contributor. The rest had 0.

Most contributors were humans, and many wikis had 1 or 2 bots contributing concepts. There was 1 wiki (Säsongsmat) with all concepts (31) created by 1 bot, with 59 edits (total pages: 1766).

### Concept Edits

5 wikis had more than 100 edits (shown in the table). 19 wikis had between 99 and 10 edits.

### USE OF STRUCTURED PROPERTIES

The use of structured properties defined in a wiki is also correlated with the use of semantic aspects of the wiki. Ultimately, the utility of the properties is to create assertions about objects in the wiki. The more assertions made in the wiki, the more its semantic features are being used.

#### Property Definitions

Table 4 shows a list of wikis with more than 200 properties defined (SP>200), 18 wikis in total. Their SP-Index ranges from a value of 0.01 in Enipedia to 0.49 in CSDMS.

All 230 wikis had some property defined. 50 wikis had more than 100 properties defined (SP>100), and an additional 53 wikis had more than 50 properties (100>SP>50). Only 19 wikis had less than 10 properties defined.

Table 5 shows the wikis with the highest SP-Index. Several wikis have an SP-Index greater than 1. For most wikis, the index is smaller than 1, which means that some pages have no properties defined for them.

#### Property Creators

Table 6 shows the wikis with more than 10 property creators (SPC>10), with 12 wikis total. Only 1 wiki (TermWiki) had more than two dozen creators. The SPC-Index is 0.00 for many of the wikis. It is non-zero for 4 of the wikis, with 0.06 in NeuroLex and even 2.22 in TermWiki.

Table 7 shows the wikis with the highest SPC-Index. TermWiki and Neurolex also appear in this list, but many other wikis appear here that were not in Table 6.

#### Property Edits

Table 8 shows the wikis with more than 400 property edits (SPE>400), with 17 wikis total. All wikis but 1 have less than 25 property creators. The exception is TermWiki with 745 property creators. The SPE-Index ranges from 0.01 in Enipedia to 0.79 in CSDMS and 0.92 in WikiDeb, with the highest value as 6.93 in TermWiki.

Table 9 shows the wikis with the highest SPE-Index. TermWiki appears in this list. Two other wikis have a higher value for this index, but they are both quite small in size (PP).



**Table 9: Semantic wikis with the highest SPE-Index.**

Wiki Name	All Users (Humans + Bots)									Humans						
	PP	SP	SPC	SPE	SPA	SP-Index	SPC-Index	SPE-Index	SPA-Index	PP-H	SP-H	SPC-H	SPE-H	SP-H-Index	SPC-H-Index	SPE-H-Index
Digitabulum	17	67	1	332	2492	3.94	0.06	19.53	146.59	17	67	1	332	3.94	0.06	19.53
CPP Ouest-1	7	51	2	67		7.29	0.29	9.57	0.00	7	51	2	67	7.29	0.29	9.57
TermWiki	336	173	745	2330		0.51	2.22	6.93	0.00	336	173	745	2330	0.51	2.22	6.93
Biostatistics Annual Report	42	160	4	202		3.81	0.10	4.81	0.00	42	160	4	202	3.81	0.10	4.81
Marshruta.net	10	22	1	45		2.20	0.10	4.50	0.00	10	22	1	45	2.20	0.10	4.50
Bioinformatics Core CRG	45	83	3	175	27936	1.84	0.07	3.89	620.80	45	83	3	175	1.84	0.07	3.89
HTML5 Video Wiki	42	48	3	77	836	1.14	0.07	1.83	19.90	42	48	3	77	1.14	0.07	1.83
CAcert in Berlin	50	46	1	77	1728	0.92	0.02	1.54	34.56	50	46	1	77	0.92	0.02	1.54
FORSYS - FP00804	254	190	4	376		0.75	0.02	1.48	0.00	254	190	4	376	0.75	0.02	1.48

**Table 10. Semantic wikis with the highest SPE-Density.**

Wiki Name	All Users (Humans + Bots)								Humans				
	PP	SP	SPE	SPA	SP-Index	SPC-Index	SPA-Density	SPE-Density	PP-H	SP-H	SPE-H	SP-H-Index	SPE-H-Density
TermWiki	336	173	2330		0.51	0.00	0.00	13.47	336	173	96	0.51	0.55
SpieleWiki	247	8	48	4276	0.03	17.31	534.50	6.00	246	8	107	0.03	13.38
Cultural Heritage Connections	3088	33	164	34518	0.01	11.18	1046.00	4.97	3088	33	205	0.01	6.21
Digitabulum	17	67	332	2492	3.94	146.59	37.19	4.96	17	67	66	3.94	0.99
Psychiatrienet	724	13	63		0.02	0.00	0.00	4.85	724	13	40	0.02	3.08
SILO	1118	33	151	20286	0.03	18.14	614.73	4.58	1118	33	173	0.03	5.24
Traditional Tune Archive	23702	54	232	58447	0.00	2.47	1082.35	4.30	23702	54	68	0.00	1.26
AutoBiblia	71	7	30	1898	0.10	26.73	271.14	4.29	71	7	154	0.10	22.00
WikiUniversite	178	29	118	2281	0.16	12.81	78.66	4.07	178	29	237	0.16	8.17

**Table 11. Semantic wikis with more than 400,000 structured property assertions.**

Wiki Name	All Users (Humans + Bots)									Humans						
	PP	SP	SPC	SPE	SPA	SP-Index	SPC-Index	SPE-Index	SPA-Index	PP-H	SP-H	SPC-H	SPE-H	SP-H-Index	SPC-H-Index	SPE-H-Index
Säsongsmat	1766	312	3	494	3081494	0.18	0.00	0.28	1744.90	76	33	2	63	0.43	0.03	0.83
SNPedia	60194	73	2	115	2593173	0.00	0.00	0.00	43.08	8841	73	2	115	0.01	0.00	0.01
Enipedia	65858	369	8	470	2346068	0.01	0.00	0.01	35.62	65015	369	8	470	0.01	0.00	0.01
911datasets.org	118	16	1	17	2145185	0.14	0.01	0.14	18179.53	115	16	1	17	0.14	0.01	0.15
The Business Model Project	17949	97	1	179	2026201	0.01	0.00	0.01	112.89	17949	97	1	179	0.01	0.00	0.01
Practical Plants	11041	35	1	86	1342266	0.00	0.00	0.01	121.57	11041	35	1	86	0.00	0.00	0.01
Van Hamel wiki	10824	244	1	553	1065713	0.02	0.00	0.05	98.46	10824	244	1	553	0.02	0.00	0.05
Yu-Gi-Oh! Wikia	18318	610	16	1563	1015246	0.03	0.00	0.09	55.42	18318	592	14	1420	0.03	0.00	0.08
Wikinosh	21540	66	2	105	981167	0.00	0.00	0.00	45.55	21540	66	2	105	0.00	0.00	0.00
WoWWiki	76706	123	5	193	804175	0.00	0.00	0.00	10.48	73729	123	5	193	0.00	0.00	0.00
Free Software Directory	7181	134	4	305	767462	0.02	0.00	0.04	106.87	131	43	3	176	0.33	0.02	1.34
HPhysics	4787	20	1	25	756481	0.00	0.00	0.01	158.03	3	20	1	25	6.67	0.33	8.33
Tieteen kansallinen termipankki	139	38	4	84	717664	0.27	0.03	0.60	5163.05	139	38	4	84	0.27	0.03	0.60
GeneWikiPlus	38207	25	3	37	637975	0.00	0.00	0.00	16.70	34746	11	2	16	0.00	0.00	0.00
15Mpedia	4477	150	5	219	622029	0.03	0.00	0.05	138.94	3505	150	5	219	0.04	0.00	0.06
WikiApiary	7273	109	2	127	589972	0.01	0.00	0.02	81.12	691	109	2	127	0.16	0.00	0.18
The Sock Book	12646	136	2	392	496429	0.01	0.00	0.03	39.26	12323	127	1	360	0.01	0.00	0.03
GaretienWiki	527	125	4	173	485060	0.24	0.01	0.33	920.42	527	125	4	173	0.24	0.01	0.33

**Table 12: Semantic wikis with an SPA-Index over 100.**

Wiki Name	All Users (Humans + Bots)									Humans						
	PP	SP	SPC	SPE	SPA	SP-Index	SPC-Index	SPE-Index	SPA-Index	PP-H	SP-H	SPC-H	SPE-H	SP-H-Index	SPC-H-Index	SPE-H-Index
911datasets.org	118	16	1	17	2145185	0.14	0.01	0.14	18179.53	115	16	1	17	0.14	0.01	0.15
Tieteen kansallinen termipankki	139	38	4	84	717664	0.27	0.03	0.60	5163.05	139	38	4	84	0.27	0.03	0.60
Säsongsmat	1766	312	3	494	3081494	0.18	0.00	0.28	1744.90	76	33	2	63	0.43	0.03	0.83
Marvel: War of Heroes Wiki	155	25	1	31	165021	0.16	0.01	0.20	1064.65	154	25	1	31	0.16	0.01	0.20
Planet Kubb Wiki	118	115	2	150	124580	0.97	0.02	1.27	1055.76	118	115	2	150	0.97	0.02	1.27
GaretienWiki	527	125	4	173	485060	0.24	0.01	0.33	920.42	527	125	4	173	0.24	0.01	0.33
Bioinformatics Core CRG	45	83	3	175	27936	1.84	0.07	3.89	620.80	45	83	3	175	1.84	0.07	3.89
Translatewiki.net	298	8	4	13	98550	0.03	0.01	0.04	330.70	297	8	4	13	0.03	0.01	0.04
Community Surface Dynamics Modeling System	526	256	3	416	155649	0.49	0.01	0.79	295.91	525	256	3	416	0.49	0.01	0.79
Milieuhulp	1199	170	3	173	335038	0.14	0.00	0.14	279.43	1199	170	3	173	0.14	0.00	0.14
Read Write Book Club	206	86	2	114	42121	0.42	0.01	0.55	204.47	206	86	2	114	0.42	0.01	0.55
SporeWiki	1159	60	5	94	209498	0.05	0.00	0.08	180.76	941	60	4	93	0.06	0.00	0.10
HPhysics	4787	20	1	25	756481	0.00	0.00	0.01	158.03	3	20	1	25	6.67	0.33	8.33
Digitabulum	17	67	1	332	2492	3.94	0.06	19.53	146.59	17	67	1	332	3.94	0.06	19.53
Wikimedia Labs - labsconsole	103	67	5	119	14749	0.65	0.05	1.16	143.19	103	67	5	119	0.65	0.05	1.16
15Mpedia	4477	150	5	219	622029	0.03	0.00	0.05	138.94	3505	150	5	219	0.04	0.00	0.06
Practical Plants	11041	35	1	86	1342266	0.00	0.00	0.01	121.57	11041	35	1	86	0.00	0.00	0.01
The Business Model Project	17949	97	1	179	2026201	0.01	0.00	0.01	112.89	17949	97	1	179	0.01	0.00	0.01
Free Software Directory	7181	134	4	305	767462	0.02	0.00	0.04	106.87	131	43	3	176	0.33	0.02	1.34
EntropiaPlanets.com	2118	465	4	660	215599	0.22	0.00	0.31	101.79	2118	465	4	660	0.22	0.00	0.31

Table 10 shows the wikis with the highest SPE-Density. The highest is 13.47 for TermWiki. 8 wikis have a value between 4 and 10. 60 wikis have values between 2 and 4.

**Property Assertions**

The real test of the use of properties is whether property assertions are added to the wikis. We were able to extract

this information only 278 wikis, as they were 52 wikis that not responsive to these API queries.

Table 11 shows the wikis with more than 400,000 semantic property assertions (SPA>400,000). There are 18 wikis in this list, 8 of them with more than a million assertions. 251 of the rest of the wikis have thousands of assertions. 8 had less than 1,000 property assertions, and 53 had none.

**Table 13: The 100 wikis with the highest SPA-Density.**

Wiki Name	PP	SP1	All Users (Humans + Bots)		SPA-Index	SPA-Density	Humans	
			SPA	SPA-Index				
911datasets.org	118	16	2145185	0.14	18179.53	134074.06	115	16
Practical Plants	11041	35	1342266	0.00	121.57	38350.46	11041	35
HPhysics	4787	20	756481	0.00	158.03	37824.05	3	20
SNPedia	60194	73	2593173	0.00	43.08	35252.92	8841	73
Grazwiki	4420	1	32460	0.00	7.34	32460.00	4420	1
SaveMLAK	21493	10	285840	0.00	13.30	28584.00	3887	10
GeneWikiPlus	38207	25	637975	0.00	16.70	25519.00	6246	11
Wikindex	16850	14	333620	0.00	16.80	23830.00	6324	14
The Business Model Project	17949	97	2026201	0.01	112.89	20888.67	17949	97
Tieteen kansallinen termipankki	139	38	717664	0.27	5163.05	18885.89	139	38
Wikinosh	21540	66	981167	0.00	45.55	14866.17	21540	66
Translatewiki.net	298	8	98550	0.03	330.70	12318.75	297	8
Såsongsmat	1766	312	3081494	0.18	1744.90	9876.58	76	33
Links thing	5050	30	277573	0.01	54.96	9254.23	5050	30
Grand Theft Wiki	6829	9	76751	0.00	11.24	8527.89	5317	9
Botnets	1302	3	24803	0.00	19.05	8267.67	1302	3
Perrypedia	26804	18	147089	0.00	5.49	8171.61	26804	17
Scoutpedia.nl	2833	28	212142	0.01	74.88	7576.50	2325	28
Marvel: War of Heroes Wiki	155	25	165021	0.16	1064.65	6600.84	154	25
WoWWiki	76706	123	804175	0.00	10.48	65380.01	73729	123
Dati e gestione dell'acqua in Italia	8278	6	39055	0.00	4.72	6509.17	8278	6
Enipedia	65858	39	234068	0.01	35.62	6357.91	65858	39
Free Software Directory	7181	134	767462	0.02	106.87	5727.33	131	43
WikiApiary	7273	109	589972	0.01	81.12	5412.59	691	109
VroniPlag Wiki	10492	40	209345	0.00	19.95	5233.63	10380	40
Appropedia	6218	5	25728	0.00	4.14	5145.60	6142	5
OrigamiWiki	2523	19	87829	0.01	34.81	4622.58	2523	19
Van Hamel wiki	10824	244	1065713	0.02	98.46	4367.68	10824	244
GnWiki	12980	3	151502	0.00	11.67	4328.63	12980	3
15Mpedia	4477	150	622029	0.03	138.94	4146.36	256	150
Animanga Wiki	1098	3	12226	0.00	11.13	4075.33	1097	3
GartienWiki	527	125	485060	0.24	920.42	3880.48	527	125
The Sock Book	12646	136	496429	0.01	39.26	3650.21	12323	127
Transit Unlimited	14067	46	166417	0.00	11.83	3617.76	14067	46
SporeWiki	1159	60	209498	0.05	180.76	3491.63	941	60
Poképédia	7772	73	232876	0.01	29.96	3190.08	7772	73
Broadcast for Schools	2044	39	246694	0.01	82.80	2921.29	2044	39
Narutopedia	4247	85	187270	0.02	44.09	2203.18	4247	85
NuestraAgenda	6628	136	272917	0.02	41.18	2006.74	1611	123
Milieuhelp	1199	170	335038	0.14	279.43	1970.81	1199	170
Dungeons and Dragons Wiki	10444	113	208952	0.01	20.01	1849.13	10421	113
Wiki Law School	9774	43	78324	0.00	8.01	1821.49	541	43
Rosetta Code	1371	52	92767	0.04	67.66	1783.98	1370	52
ArtWiki	5168	7	80894	0.00	15.45	1717.81	5168	7
Navi	1987	61	104075	0.03	52.38	1706.15	1955	61
Yu-Gi-Oh! Wikia	18318	610	1015246	0.03	55.42	1664.34	18318	592
Wikiteater	5319	53	86166	0.01	16.20	1625.77	5319	53
Natural History of Southeast Alaska	6983	192	283740	0.03	40.63	1477.81	6983	192
Culture.si	3033	109	152819	0.04	50.39	1400.01	729	109
WikiPapers	3452	65	80827	0.02	23.41	1243.49	3115	65
DeurneWiki	5687	35	30768	0.00	5.41	1230.72	5687	35
Combat Arms Wiki	1954	23	27707	0.01	14.18	1204.65	1950	23
Open Resource Bank for Interactive Teaching (ORBIT)	1565	13	15300	0.01	9.78	1176.92	1565	13
Offene Naturführer	4615	117	136517	0.03	29.58	1166.81	4615	117
Giki	11969	125	142433	0.01	11.90	1139.46	1164	125
Planet Kubby Wiki	118	115	124580	0.97	1055.76	1083.30	118	115
Traditional Tune Archive	23702	54	58447	0.00	2.47	1082.35	23702	54
Shufi	551	3	8495	0.01	15.42	1061.88	551	3
Cultural Heritage Connections	3088	33	34518	0.01	11.18	1046.50	3088	33
Beachapedia	1157	25	25583	0.02	22.11	1023.32	1157	25
Casiopea	7999	113	112457	0.01	14.06	995.19	7999	113
The Features Wiki	1046	45	44255	0.04	42.31	983.44	1046	45
IEEE Global History Network	2260	34	32355	0.02	14.32	951.62	2260	34
WikiTranslate	3147	44	41218	0.01	13.10	936.77	3147	44
Wikiteanuar	772	18	16263	0.02	21.07	903.50	772	18
Antique Jewelry University	2156	26	22701	0.01	10.53	873.12	2156	26
MozillaWiki	31571	141	116976	0.00	3.71	829.62	31570	141
Discourse DB	3438	85	69506	0.02	20.22	817.72	3438	85
Star Trek Online Wiki	6705	136	101659	0.02	15.16	747.49	6705	136
Muyscubun	2568	52	38767	0.02	15.10	745.52	2325	52
Web Community Wiki	10749	618	394633	0.06	36.71	638.56	10127	618
SILO	1118	33	20286	0.03	18.14	614.73	1118	33
Frank	1393	40	24552	0.01	8.17	542.78	1393	40
Community Surface Dynamics Modeling System	526	256	155649	0.49	295.91	608.00	526	256
Complex Operations Wiki	6872	133	80858	0.02	11.77	607.95	6872	133
DCPedia	993	21	12598	0.02	12.69	599.90	993	21
Verwaltungskooperation	705	15	8857	0.02	12.56	590.47	705	15
Zombie Jombie Wiki	381	10	5795	0.03	15.21	579.50	380	10
Microcosm Aquarium Explorer	4343	77	44522	0.02	10.25	578.21	4343	77
Hackerspace Brussels	601	49	4912	0.01	8.17	542.78	601	49
SpieleWiki	247	8	4276	0.03	17.31	534.50	246	8
HackerspaceWiki	1849	96	49368	0.05	26.70	514.25	1848	96
Rage of Bahamut Wiki	474	40	19677	0.08	41.51	491.93	474	40
Read Write Book Club	206	86	42121	0.42	204.47	489.78	206	86
Mikomos	727	39	18669	0.05	25.68	478.69	727	39
EntropiaPlanets.com	2118	465	215599	0.22	101.79	463.65	2118	465
Wiki:step-project.com	430	74	33670	0.17	78.30	455.00	430	74
Keroro Wiki	927	29	12083	0.03	14.11	451.14	926	29
Turf Wars Wiki	1045	11	4924	0.01	4.71	447.64	1045	11
GRF Wiki	1323	67	29911	0.05	22.61	446.43	1323	67
Leibowitz's Candle	250	13	5653	0.05	22.61	434.85	250	13
Global Women's Network	2121	45	18596	0.02	8.77	413.24	2121	45
World Directory of Churches	180	29	11600	0.16	64.44	400.00	180	29
AliceSoftWiki	1153	27	10766	0.02	9.34	396.74	1153	27
ArthropodBase Wiki	7203	178	66895	0.02	9.29	375.81	7203	178
Das Verdamnte Wiki	1137	28	9910	0.02	8.72	353.93	1137	28
Bioinformatics Core CRG	45	83	27936	1.84	620.80	336.58	45	83
Road Sign Math	248	33	10971	0.13	44.24	332.45	248	33
MetaBase	2073	41	13274	0.02	6.40	323.76	2073	41
OpenCongress wiki	7341	339	106116	0.05	14.46	313.03	7341	339
SKYbrary	5577	254	77455	0.05	13.89	304.94	5577	254

Table 12 shows wikis with a SPA-Index over 100, a total of 20 wikis. 127 wikis have a value above 10, 49 wikis had a value between 1 and 10. Table 13 shows the 100 wikis with the highest SPA-Density, all have a value above 300. The highest is over 134,000 for 911datasets.org. There are 12 wikis with higher than 12,000. It is over 100 for 150 wikis, and is between 10 and 99 for 27 wikis.

**DISCUSSION**

Of the 230 semantic wikis analyzed, 25% had no semantic concepts defined. Users in a large proportion of wikis do not seem interested in using this feature. It is possible that wiki page “categories” are being used in lieu of concepts.

We collected data about the number of categories and the number of category assertions in the wikis. As suspected, categories are far more popular than concepts, with 43 wikis having more than 1,000 categories and only 15 wikis with no categories. Further investigation is required to determine the reasons for the limited use of concepts.

The use of structured properties was much more widespread. All wikis had some property defined. 50 wikis had more than 100 properties defined.

The creation of property assertions is done at large volumes across all wikis.

The number of property creators was below two dozen for all but one wiki. This suggests that adoption and normalization of properties as well as resolution of conflicts among property definitions may be confined to a small subset of users. In fact, some wikis have established editorial processes for semantic properties. For example, the Battlemaster wiki states the following policy:

*“Do not create new properties without approval. Properties listed in the table below can be added to pages or templates as appropriate. If you need a property that is not yet defined, make a suggestion on the talk page. Any and all properties added to the wiki without approval will be deleted and all changes to your pages that use them will be reversed.”*

It is possible that many more wikis have similar policies. Further investigation of the dynamics of the creation of semantic properties would reveal whether potential conflicts are collaboratively detected and resolved.

Many semantic edits are done by bots. However, the vast majority is done by humans. We did not have data about whether property assertions were done by bots (i.e., an SPA-H metric), since that data is not available from the API and would have to be extracted from the wiki history pages.

From our analysis, it appears that the adoption of semantic wikis is in large part due to the ability to create semantic properties and use them to make assertions.

It would be useful to understand the interplay between property edits and growth of the wiki. Figure 1(a) shows edits over time for all the wikis. We separated wikis that had editing spikes, defined as having more than 30% of edits in any one month. Wikis with no editing spikes are shown in Figure 1(b). The spikes of growth for the wikis shown in Figure 1(c) could be correlated to the definition of semantic properties. Investigating this would require extracting details of what edits were done to each page in a finer granularity than what is available through the API.

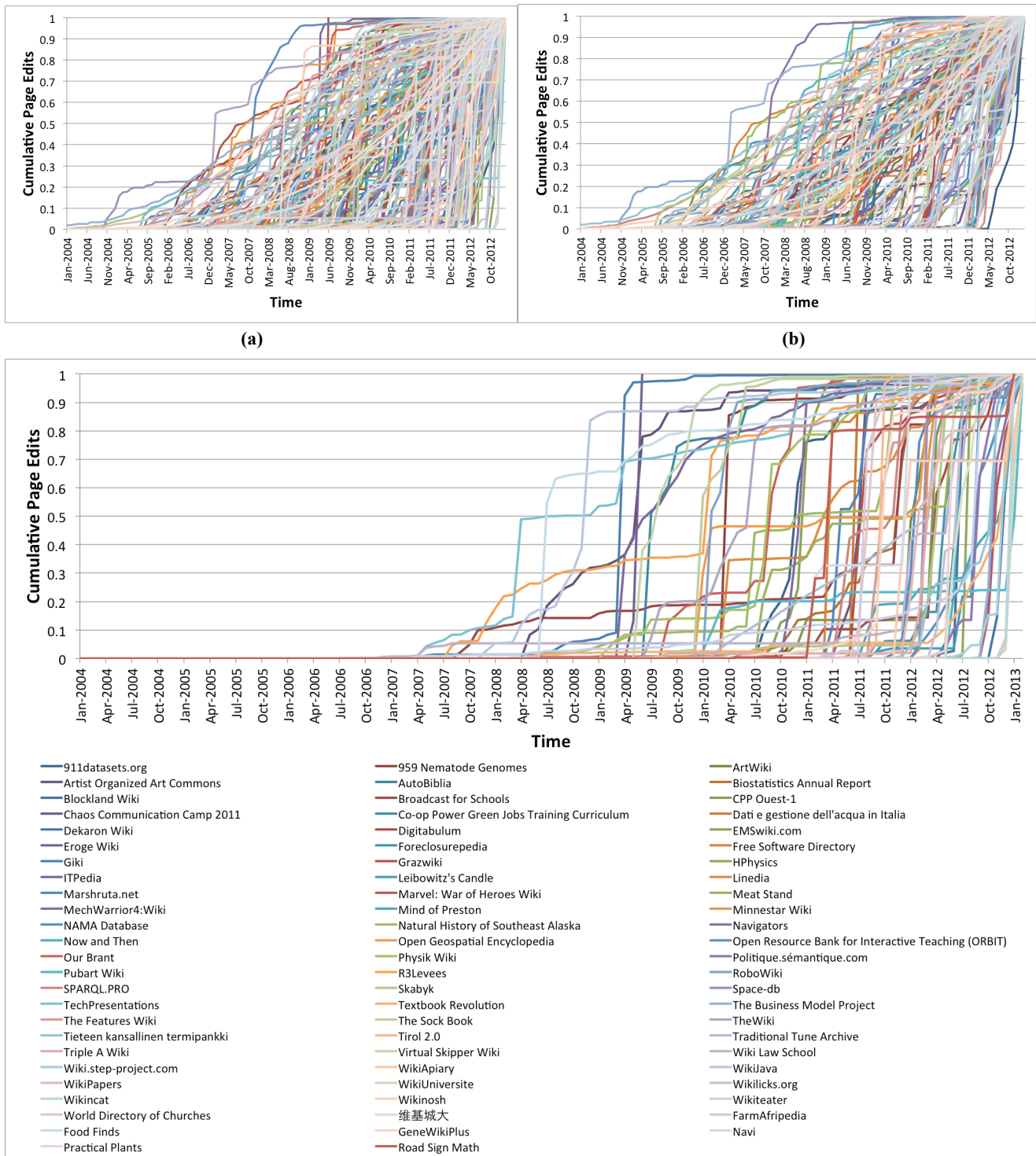


Figure 1. Wiki edits over time, showing (a) edits over time for all wikis, (b) the subset of wikis with no editing spikes, (c) the subset of wikis with editing spikes.

## CONCLUSIONS

In this paper, we report on an analysis of 230 semantic wikis regarding the creation of semantic concepts and properties as well as assertions. We collected data about their creation and edits, as well as the wiki users who edited

them. We found that concepts are not used very often, and not used at all in many wikis. Properties were used in all the wikis, although very small numbers of users edit them. Very large numbers of property assertions are used in almost every wiki.

An important challenge is to understand the limited use of some semantic features of the wiki, such as defining concepts, as well as the small number of users who make property definitions. The extent of use and enforcement of restrictive editing policies should be further investigated. It is possible that additional users would be involved in the creation of properties if there were facilities in the wiki to detect and resolve conflicts collaboratively. A future research direction is to investigate extensions to semantic wikis that proactively prompt users to define concepts or properties, detect inconsistent definitions, and mediate their resolution.

There are hundreds of communities that are using semantic wikis for social knowledge collection. They are collaboratively creating structured content. Improving the ability of all users to contribute to the semantic aspects of the wiki will help extend the use of knowledge-rich systems and broaden the adoption of semantic web technologies.

#### ACKNOWLEDGMENTS

We gratefully acknowledge the support from the National Science Foundation with grant IIS-1117281.

#### REFERENCES

- [Auer et al 2006] Soren Auer, Sebastian Dietzold, and Thomas Riechert. "OntoWiki - A Tool for Social, Semantic Collaboration." 5th International Semantic Web Conference, 2006.
- [Auer et al 2007] 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.
- [Bry et al 2012] 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.
- [Chklovski and Gil 2005] Tim Chklovski and Yolanda Gil. "An Analysis of Knowledge Collected from Volunteer Contributors." Proceedings of the Twentieth National Conference on Artificial Intelligence (AAAI), Pittsburgh, PA, July 2005.
- [Gil 2013] Yolanda Gil. "Social Knowledge Collection." To appear in the Handbook for Human Computation, P. Michelucci (Ed). Springer, 2013.
- [Gil et al 2013] Yolanda Gil, Angela Knight, Kevin Zhang, Larry Zhang, and Ricky Sethi. "An Initial Analysis of Semantic Wikis." Proceedings of the ACM International Conference on Intelligent User Interfaces, Santa Monica, CA. March 2013.
- [Kittur et al 2008] 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.
- [Kittur et al 2009] 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.
- [Kittur and Kraut 2010] 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.
- [Krötzsch et al 2007] Markus Krötzsch, Denny Vrandečić, Max Völkel, Heiko Haller, Rudi Studer. "Semantic Wikipedia." Journal of Web Semantics, 5(4), pages 251-261, December 2007.
- [Kuhn 2009] 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.
- [Lam et al 2010] 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.
- [Leskovec et al 2010] 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.
- [Lieberman et al 2004] Henry Lieberman, Hugo Liu, Push Singh, and Barbara Barry. "Beating some common sense into interactive applications." AI Magazine, 2004.
- [Hoffman et al 2009] 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.
- [Matuszek et al 05] 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 AAAI 2005.
- [Raban et al 2010] 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.
- [Spinellis and Louridas 2008] Diomidis Spinellis and Panagiotis Louridas. "The Collaborative Organization of Knowledge." CACM, August 2008.
- [Weld et al 2008] 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 (AAAI), 2008.