

Mediawiki Report – wiki 2

Wiki Explorer*

August 26, 2009

1 Introduction

This is an automated report of your wiki. Section 2 provides some basic statistics on the global level of the wiki, on the local level of wiki users (Subsection 2.1) and on the local level of wiki pages (Subsection 2.2). This gives you a first and general impression of what is going on in your wiki. Section 3 introduces a number of cumulative distribution functions. These functions allow for statements such as “20 % of the users do 80 % of the work.” Section 4 delves into the revision history and with it takes a first look at the dynamics of your wiki. Section 5 continues the investigation of dynamics in terms of author participation. Section 6 offers various visualizations of your wiki. In particular, it shows the hyperlink network and the co-authorship network.

2 Statistics

For those unfamiliar with wikis, we briefly introduce some of the most common terminology. A wiki consists of *pages* or *articles* which, in turn, are written by *users* or *authors*. Each time a user edits (i. e., creates or changes) a page, a new *revision* of the page is added to the *revision history*.

Pages are grouped in *namespaces*. Ordinary pages (e. g., articles) belong to the main namespace (namespace 0). There are namespaces for personal user pages (indicated by prefix “User:”), help pages (prefix “Help:”), and others, and each page may have a corresponding discussion page (prefix “Talk:” for the main namespace, “User_talk:” for user discussion pages and so on).

Since each revision is the result of a single edit by a single user, we know *who* contributed to *which* page—at least if the user is logged in. In case the wiki allows for anonymous edits, such as user mapping not possible. Anonymous edit are clearly visible in the revision history with the IP address of the computer used to edit. Unfortunately,

*<http://wiki-explorer.rubyforge.org>

Table 1: Global Wiki Statistics

Namespace		#pages	#revisions
<i>main</i>	(0)	1350	11522
Talk	(1)	6	6
User	(2)	13	56
User_talk	(3)	1	2
wiki	(4)	4	11
Image	(6)	172	172
Image_talk	(7)	2	2
MediaWiki	(8)	7	10
Template	(10)	7	24
	(106)	22	189
	(108)	2	46
total		1586	12040

Table 2: Global User Statistics

	mean	s. d.	median	min.	max.
Edits	405.97	766.70	76	0	3059
Edited Pages	90.38	146.75	17	0	605
Edits/Page	3.47	2.17	3.00	1.00	9.77
Self-Edits	271.76	566.78	35	0	2417
Foreign-Edits	79.90	127.03	21	0	452
Category-Edits	0.00	0.00	0	0	0
Image-Edits	0.00	0.00	0	0	0

private Internet Service Providers (ISPs) assign a different IP address to a computer every time a user connects to the Internet. Therefore, we exclude anonymous users into user-dependent statistics. In similar vein, we exclude the system user, who is the owner of default and automatically generated pages

Now, your wiki has 1586 pages with 12040 altogether revisions, and there are 29 non-anonymous users. Table 1 gives a more detailed overview.

2.1 User Statistics

Table 2 shows the global user statistics. Most of these aggregated statistics are self-explanatory. For example, the mean number of edits points out how many edits, on average, a single user makes. The standard deviation (s. d.), median, minimum (min.), and maximum (max.) are likewise statistics for the number of edits, edited pages,

Table 3: Top 20 Users

#edits	#pages	edits/page	#self edits	#foreign edits
3059	313	9.77	2417	394
2524	445	5.67	1875	452
1978	605	3.27	1163	343
944	266	3.55	561	276
749	246	3.04	386	233
521	79	6.59	349	110
350	125	2.80	186	61
337	41	8.22	265	38
312	171	1.82	110	136
219	90	2.43	108	53
194	42	4.62	138	28
114	17	6.71	82	23
103	57	1.81	35	64
89	28	3.18	50	21
76	22	3.45	40	21
47	17	2.76	29	18
40	8	5.00	26	8
33	9	3.67	24	2
24	7	3.43	11	8
19	7	2.71	12	6

edits per page. The statistics for self-edits and foreign-edits, however, require some explanation. Self-edits are two consecutive edits by one and the same user on a single page. Therefore, they indicate phenomena such as page savings and uninterrupted work on a page. In contrast, foreign-edits are two consecutive edits by two different users on a single page. These, then, indicate collaboration on a page. Last but not least, category-edits and image-edits pertain to the number of edits that add a category or an image to a page.

Table 3 shows the top 20 users ranked by their number of edits. In addition, the number of edited pages, edits per page, self-edits, and foreign-edits are given. In general, users with many edits are likely to edit many pages, too. This is even more true for many foreign-edits, that is, collaboration is more likely to spread across many pages, too.



Figure 1: Cumulative Distribution Functions

2.2 Page Statistics

Table 4 shows the top 5 pages ranked by the number of edits, users, outgoing links (links→, a. k. a. outdegree), incoming links (→links, a. k. a. indegree), and text size (in bytes, i. e., characters). Again, these statistics are self-explanatory.

3 Cumulative Distribution Functions

Figure 1 shows three cumulative distribution functions (a. k. a. Lorenz curves), one for each combination of authors, pages, and revisions. For example, plotting authors versus revisions allows for statements such as “80 % of all revisions are made by 20 % of the authors.” This “80-20 rule” or, as it is more commonly known, the *Pareto principle* states that, for many events, roughly 80 % of the effects come from 20 % of the causes (e. g., 80 % of a nation’s wealth is held by 20 % of its citizens). A distribution far less than the *Pareto principle* suggests that your wiki is driven by very few people. In all likelihood, this may be an undesirable state, although the overall number of authors puts this finding into perspective. In the German Wikipedia, for example, approximately 80 % of the revisions are made by 2 % (!) of the 132,238 authors.

4 Revision History

Figures 2 and 3 show the absolute and cumulative number of revisions per month. On the one hand, this gives a first impression of the work that is being done in your wiki on a monthly basis, while this impression translates to the growth of the wiki, on the other hand. Note that the number of revisions does not necessarily point to work in terms of content, that is, content may easily come about few revisions. However, many revisions generally account for an active use of the wiki so that its content is up to date. In both figures, you should be able to identify holidays (Christmas, summer vacation,

Table 4: Top 5 Pages

title	#users	#edits	#links→	#→links	size
<i>by edits</i>					
p995	17	780	61	10	6316
p1	15	318	24	0	2776
p2101	4	289	10	1	94365
p943	11	206	65	1	7094
p1088	5	175	7	2	221
<i>by users</i>					
p995	17	780	61	10	6316
p1	15	318	24	0	2776
p1766	14	41	0	3	1887
p943	11	206	65	1	7094
p940	9	67	18	1	1296
<i>by outgoing links</i>					
p2514	1	4	163	1	10461
p2231	1	1	72	1	4219
p2140	3	9	65	1	9864
p943	11	206	65	1	7094
p995	17	780	61	10	6316
<i>by incoming links</i>					
p1395	1	47	0	71	4102
p1380	3	33	0	19	2715
p1972	3	14	0	13	3637
p2501	1	9	7	12	3485
p995	17	780	61	10	6316
<i>by text size</i>					
p1417	1	5	0	1	133015
p2101	4	289	10	1	94365
p1023	1	3	0	4	30402
p1498	1	11	0	1	28630
p1881	2	32	0	1	23289

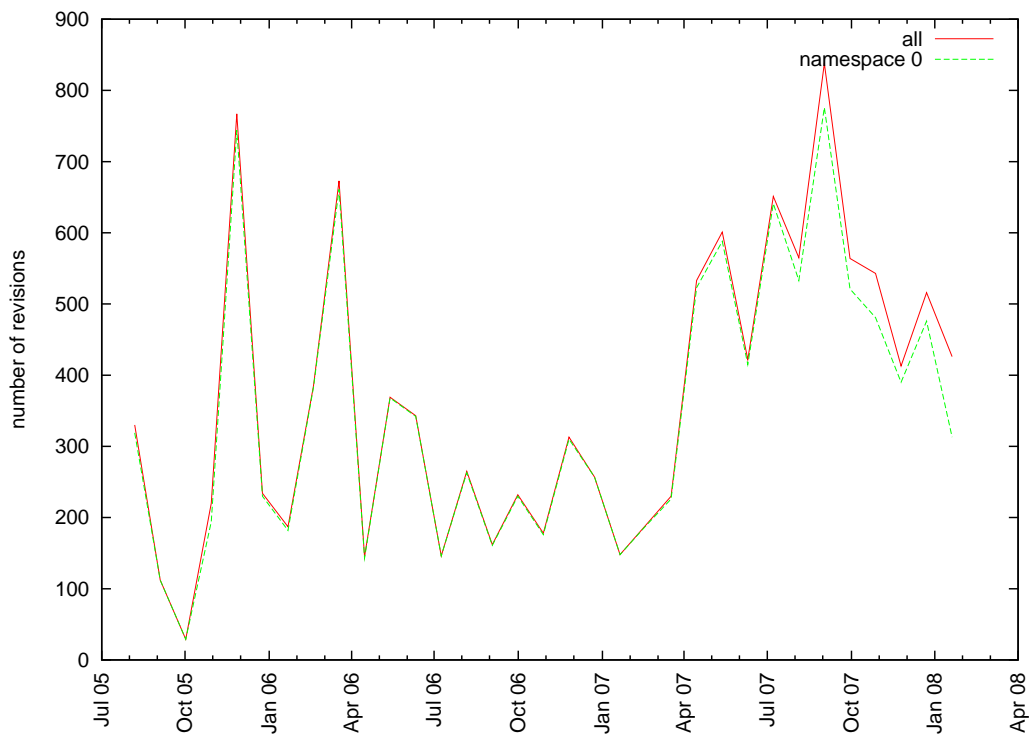


Figure 2: Revisions per Month

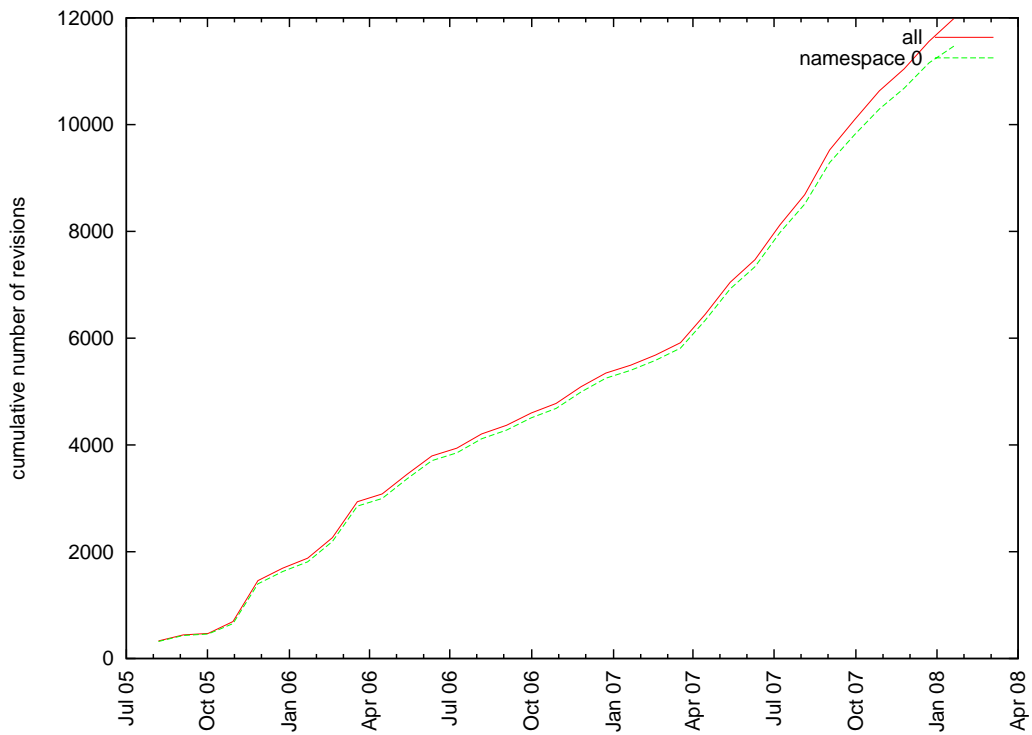


Figure 3: Cumulative Revisions per Month

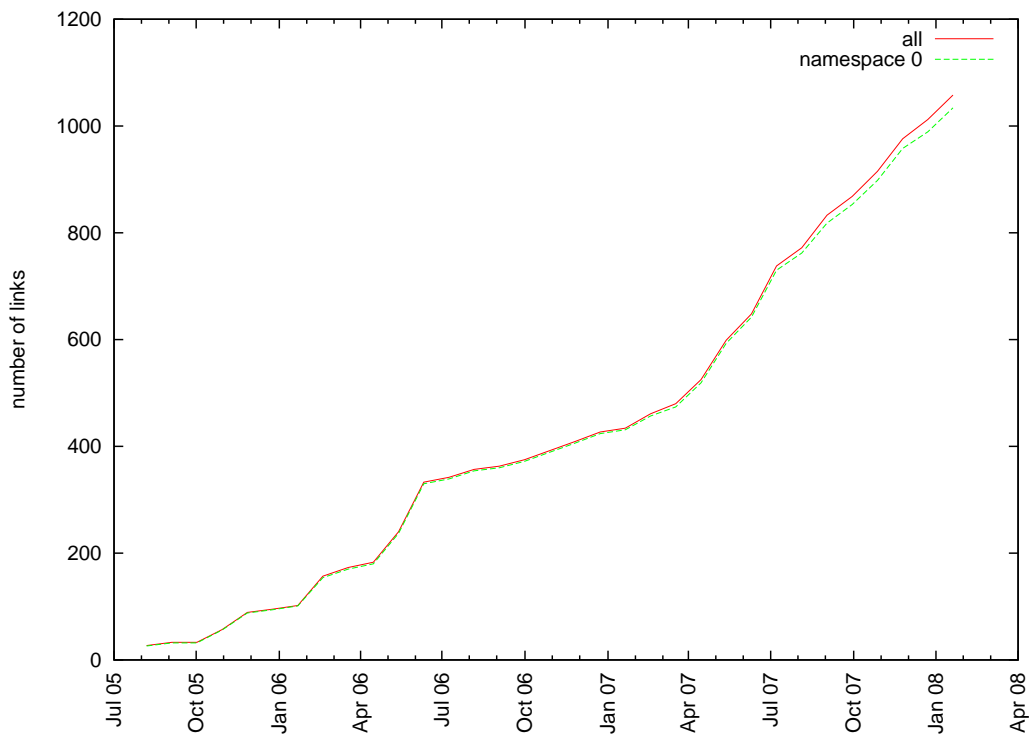


Figure 4: Number of Links

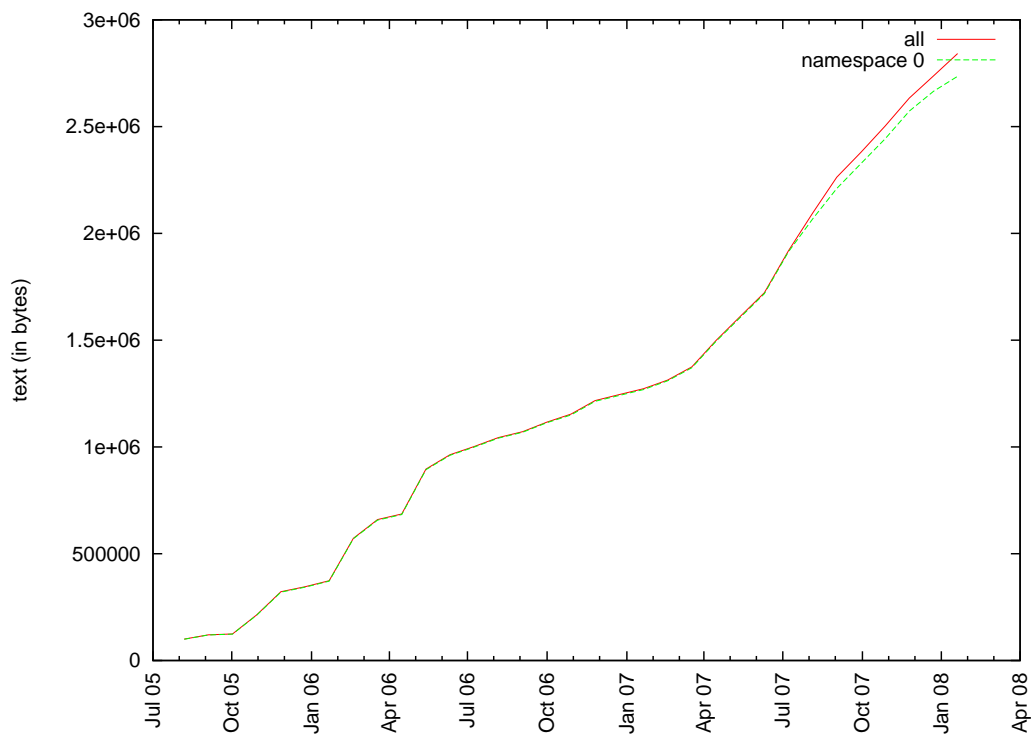


Figure 5: Amount of Text (in Bytes)

etc.) when the number of revisions is presumably low. If you ever tried to “kick off” or “jump start” your wiki, you should also be able to identify these events in the face of rather high numbers of revisions.

5 Author Participation

Figure 6 shows the number of authors who edited at least one document at any given month. While the absolute numbers are likely to fluctuate, the trend line gives an idea of just how many authors participate in the wiki.

Figure 7 shows the percentage of authors who edit at least one document at any given week. For example, a value of 50 % indicates that half of all active authors participate in the wiki. Note that we define an author as active for the time between his first and his last edit. Therefore, and in contrast to Figure 6 , relative author participation adjusts for growth of the wiki in question. Again, while the absolute percentages are likely to fluctuate, the trend line gives an idea of just the percentage of authors who participate in the wiki.

6 Network Visualization

Figure 8 shows the co-authorship network in which the nodes are authors, and a link between any two nodes represents authorship of one or more pages. There are 30 nodes connected by 300 links in the network. The co-authorship network gives an idea of who is working with whom. Single nodes with above average links to others identify authors who work with many others, usually on many pages. Rather than content, these authors provide most of the structure to the wiki. Clusters of nodes identify authors working in close concert with each other. These authors provide most of the content to the wiki.

Figure 9 shows the page network in which the nodes are pages, and a link between any two nodes represents a hyperlink from one to the other page. There are 807 nodes connected by 959 links in the network. The page network gives an idea of which page links to which other. Single nodes with above average links to others identify pages which serve as portals. These pages provide most of the structure of the wiki. Clusters of nodes identify pages which revolve around a common theme or topic. These pages provide most of the content of the wiki.

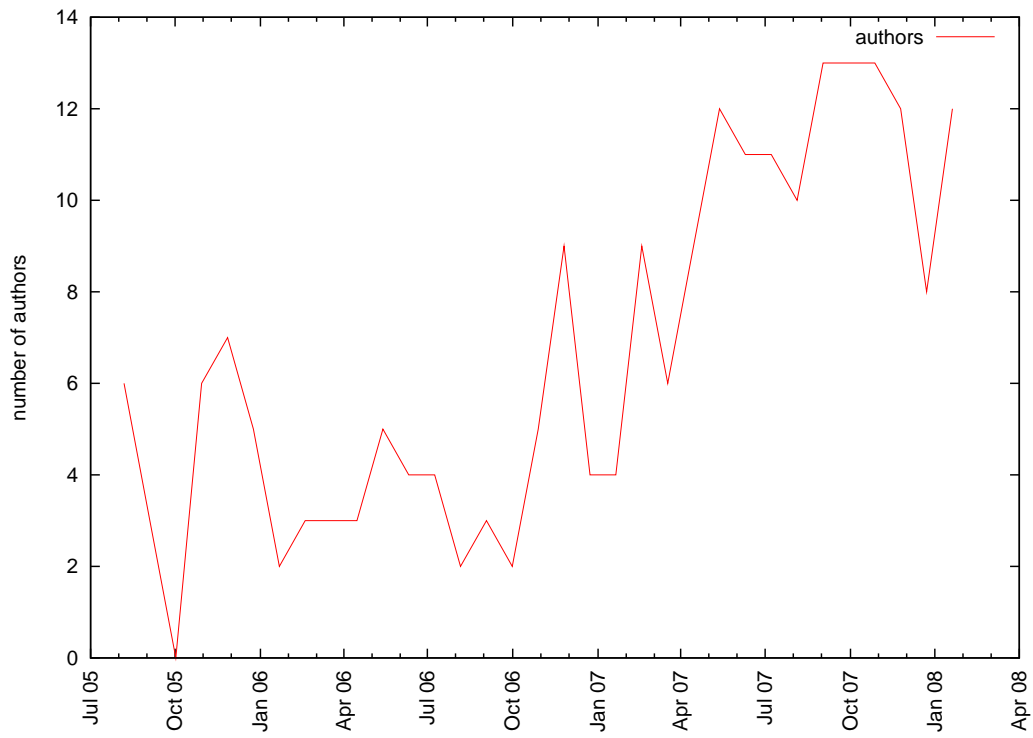


Figure 6: Author Participation

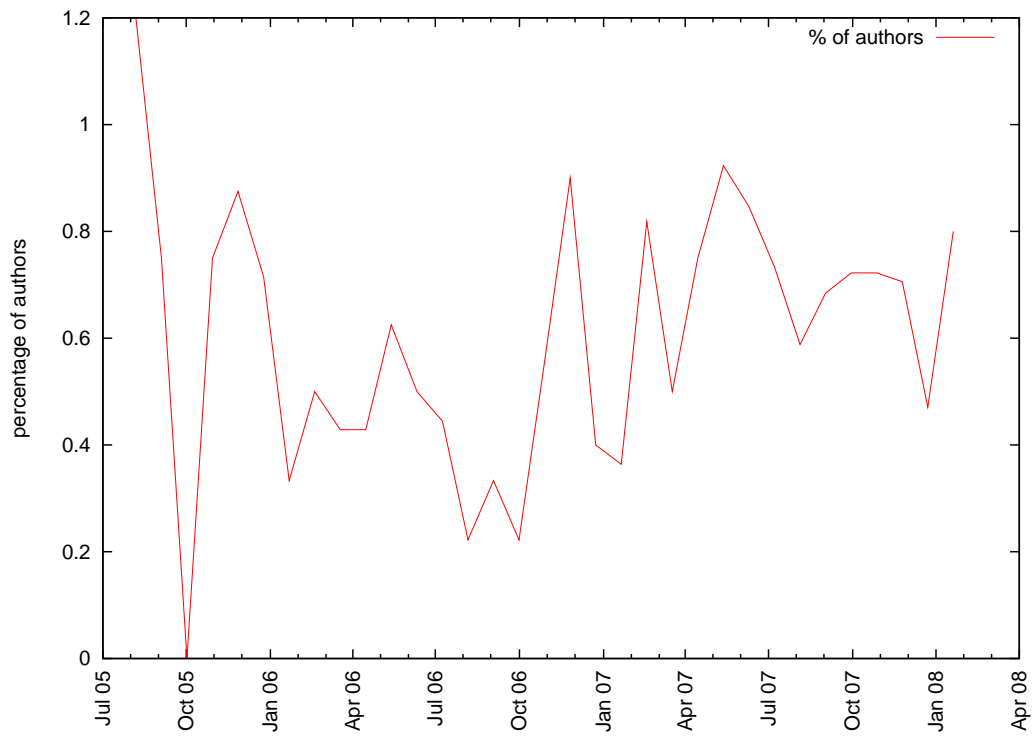


Figure 7: Relative Author Participation

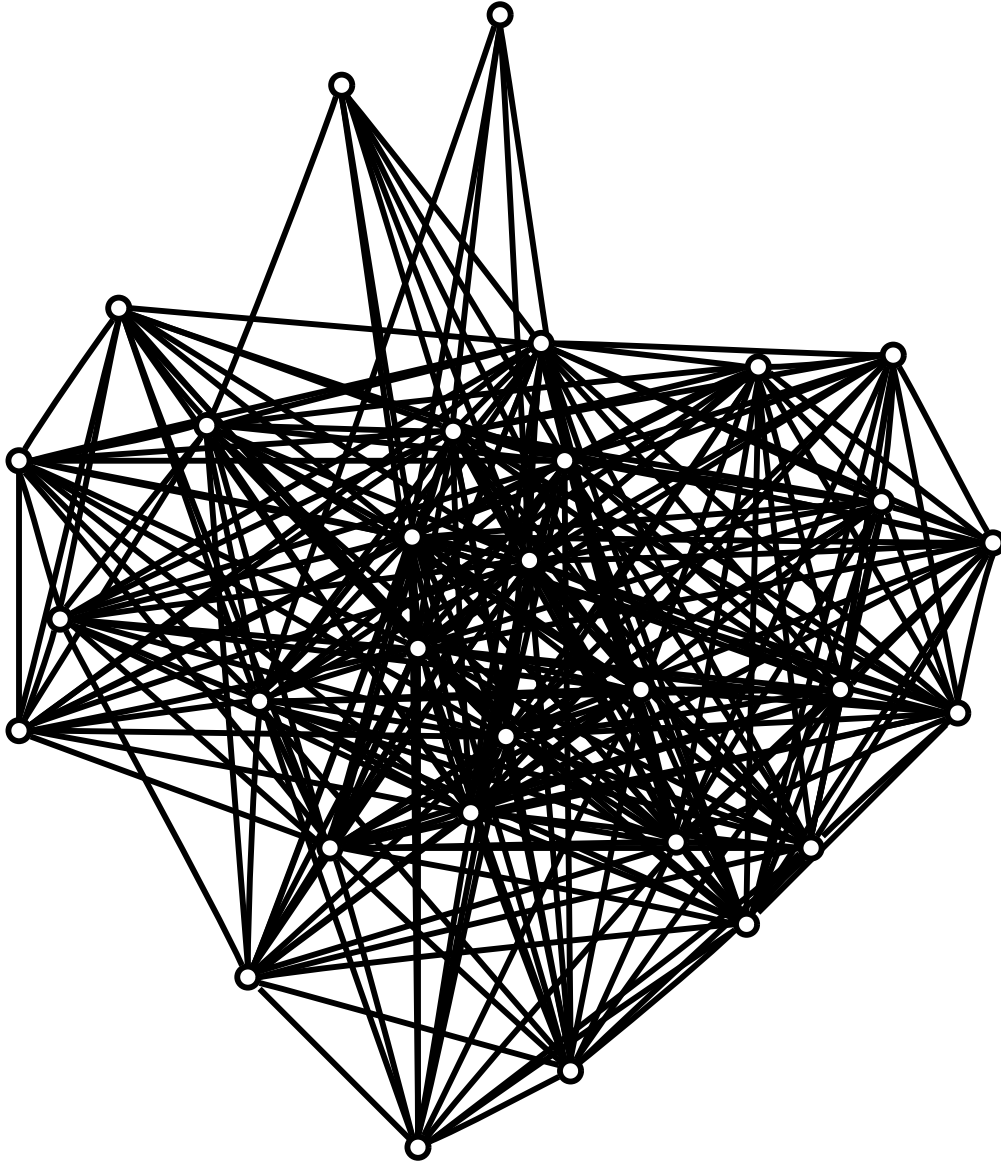


Figure 8: Coauthorship Network

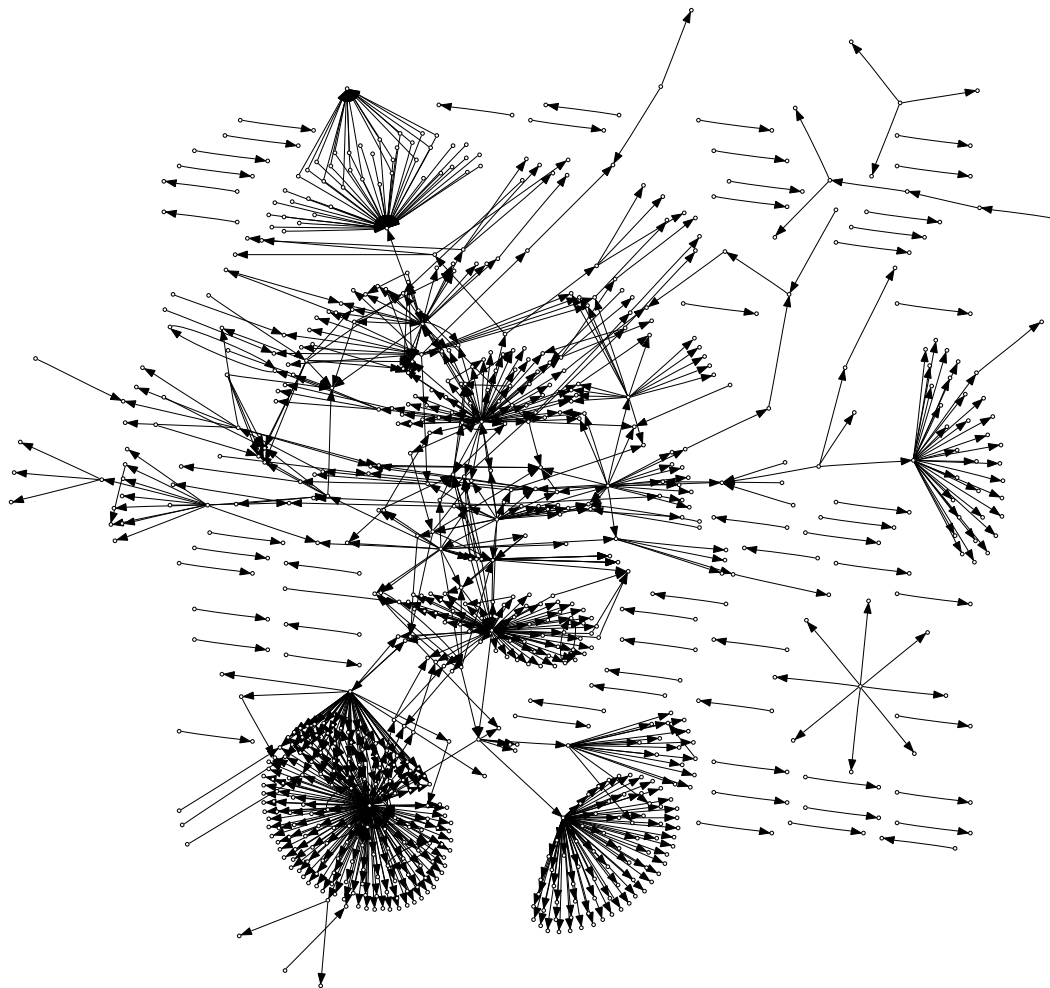


Figure 9: Page Network

Acknowledgements

This report was created using Wiki Explorer¹, a ruby library for scientific research on wikis (and other CMS, focus: Mediawiki) for interactive exploration, statistics and visualization of (network) data.

Wiki Explorer was developed² by

Dr. Klaus Stein
Laboratory for Semantic Information Technology
Chair of Computing in the Cultural Sciences
Otto-Friedrich University Bamberg

and is available as open source.

¹<http://wiki-explorer.rubyforge.org>

²in the WiO project supported by a grant from the Volkswagen Stiftung.