

An Outsider's View on "Topic-oriented" Blogging

Judit Bar-Ilan

School of Library, Archive and Information Studies
The Hebrew University of Jerusalem, Jerusalem, 91904, Israel
and

Department of Information Science, Bar-Ilan University
Ramat Gan, 52900, Israel
judit@cc.huji.ac.il

ABSTRACT

The number of Web blogs is growing extremely fast, thus this phenomenon cannot be ignored. This paper discusses the issue through monitoring a set of blogs for a two months period in September-October 2003 and characterizing these blogs based on descriptive statistics and content analysis.

Categories and Subject Descriptors

K.4.2 [Computers and Society]: Social Issues

General Terms

Human Factors

Keywords

Blogs, bloggers

1. INTRODUCTION

What is a blog? There seems to be a rather wide agreement on the definition of a *blog*. They are Web pages with several dated entries usually arranged in reverse chronological order [15, 2] (for slightly different definitions, see [16, 17]). Most blogs or Weblogs contain many links to Web sites or blogs and their main purpose to discuss the contents of these links or simply to inform the readers of the blog about the existence of these sites. Such blogs resemble NCSA's old "What's new pages" (for an archived list of these pages, see <http://archive.ncsa.uiuc.edu/SDG/Software/Mosaic/Docs/old-whats-new/>). Probably the earliest such page news page, announcing "Here's the latest (that we know) about W3" was created by Tim Berners-Lee in January 1992 (archived at <http://www.w3.org/History/19921103-hypertext/hypertext/WWW/News/9201.html>). Thus Tim Berners-Lee can be seen as the "father of blogs". For additional historical information on weblogs see [16, 5].

Nowadays there are blogs with no links at all (or almost no links) and they are simply personal diaries on the Web, "nothing more than an outbreak of self-expression" [5]. However almost all blogs, even those whose primary purpose is to list "interesting/new" websites, have a personal side as well. First of all the chosen sites and their evaluations already reflect on the author's personality, and they often add pieces of personal information as well.

Sometimes the blog is not a "one-man show": there are *community blogs* where anyone can become member (e.g. MetaFilter, <http://www.metafilter.com>, currently has more than 17,000 members) or *group blogs* (sometimes called *distributed blogs*, with a small number of pre-invited regular members (e.g.

[misbehaving.net](http://www.misbehaving.net/), <http://www.misbehaving.net/> – discussed later in the paper) or a mixture of group and community blog, where a small number of authors review and moderate the submissions from a very large community (see Slashdot, <http://slashdot.org/>).

There are a number of possible categorizations of weblogs: the *content* may be mainly associative, personal and self-expressive, or topic oriented where the topic can be related to a hobby or to the author's profession or business (business promotion). There are several possible *formats* as well: the entries can be lists of sites of interest to the author, or essays that may or may not include links to other sites or blogs; or they can be simply announcements to the prospective readers of the blog (e.g. used as alternates to school newsletters). The postings can be either monologues or have a more interactive nature, encouraging comments from readers and even asking them questions. Weblogs can also be classified according to the *authorship*, they can be maintained by a single person or by a small group or by a large community. The blogs can be set up at the individual's or the company's own initiative, or it can serve educational purposes: in some places children are encouraged to maintain blogs where they can express themselves and improve their writing skills [6]. The above categorizations are along different dimensions (content, format, and authorship) and sometimes it is not easy to decide what category the specific blog belongs to, mainly in the content and format dimensions. The authors of the blog often do not state what is its exact purpose, and even if they do the topics and format may drift over time. See [13, 10] for further discussions of categorization. The categorization becomes more accurate when it is carried out on a per posting basis as we propose in this paper.

The *blogspace* (blogs on the Web) is growing very fast. According to Cyberatlas [8], based on data from Blogcount [3] there were an estimated 2.4-2.9 million blogs as of June 2003, but only about half the registered blogs are reported to be active (where active probably means updated during the last eight weeks). According to the data in this article, 2% of the online community set up blogs and 4% of the community read blogs. Perseus [14] estimated that there are 4.12 million hosted weblogs (not including blogs that individuals maintain on their own servers using their own tools), but only 33% of the blogs were active (updated during the last two months), with 26% of the inactive weblogs having just a single posting (written at set-up time). The Perseus White Paper carries no date, but according to Blogcount [4] the paper was released in September 2003. Perseus projects that by the end of 2004 there may be as many as 10 million blogs on the Web.

Another interesting question is how many Web pages and links are produced by these blogs. There is no simple answer to this question, since it depends on the how the blog is set up. Each posting has a *permalink*, a permanent URL address for each

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

WWW 2004, May 17–22, 2004, New York, New York, USA.

ACM 1-58113-912-8/04/0005.

posting, but this does not necessarily mean that each posting is on a separate web page, sometimes a number of postings have the same initial permanent URL and only the part after the # sign (the anchor inside the webpage) differentiates between them. Halavais [9] based on data from weblogs.com charted the update frequencies of active blogs, the median was somewhere around three days. A survey of 311 20six UK blog users [1] found that about 32% of the bloggers updated their blogs 3 to 4 times a week. The growth of the number of blogs has a large impact on the growth of the number of Web pages, because unlike regular webpages which are often removed from the Web or replaced by newer versions, the basic philosophy of blogging is to store every posting permanently on the Web for possible future access (this is only the philosophy behind blogging, however in practice blogpost may very well disappear or become permanently inaccessible). In addition to the links appearing in the postings, there are usually a large number of links on the sidebars of the blog pages, and these sidebars appear on each and every archived page as well as the current page. Bloggers usually publish *blogrolls* – lists of blogs they read and additional links of interest on the sidebars. Also sometimes there are links to different parts of the blog archive from the sidebar. In some cases the comments of the readers are an integral part of the blog page, where each comment contains at least one additional link (to the author of the comment).

One of the reasons for starting this research was to try to figure out why people write blogs, and why other people read these blogs. Graham [7] lists reasons why people blog (the primary reason is the need to publish), while one of the reasons mentioned for reading blogs that it is a new source of news [12]. Some blogs provide useful information, and personal touch of well-written blogs capture the audience, once they start reading them. In most blogs readers can *comment* on the postings, thus engage in active discussions and become part of the blogspace without actually setting up a blog. A nice feature of blogs is that the owner can moderate the comments, thus nasty comments are not published and do not spoil the atmosphere of the blog (this of course means that blogs are far from being objective). An additional feature is *trackback* that keeps account of the Web pages and blogs that link to the specific posting. Blogdex (<http://blogdex.net/>), Daypop (<http://www.daypop.com/blogstats>) and Technorati (<http://www.technorati.com/>) allow users to track the number of sites/blogs that link to a specific blog.

Currently the estimates claim [8] that only 4% of the Web users read blogs, however I believe that more than 4% of the online community will become blog readers as the phenomenon becomes more widely exposed. The size of the readership will probably rise as more and more articles on blogging appear in the press, and the number of blogs reaches a critical mass.

2. METHODOLOGY

After browsing the blogspace for a while, a convenience sample of 15 weblogs were chosen. I picked weblogs based on my professional interests and knowledge, only “topic oriented” blogs were chosen. The categorization below was based on the initial information that was available about the authors and the blogs (in the “about” sections, where existing or by looking at the first entry in the blog). These blogs were monitored for a period of two months, in September-October 2003. The chosen blogs were (in italics is the abbreviated name of each blog as used in this paper):

Library and Information Science

1. Information Research weblog - <http://www.free-conversant.com/irweblog/index> . Tom Wilson, Professor of Information Science. Members of the “we-blog group” can post messages directly to the blog. *InfRes*
2. ResourceShelf <http://www.resourceshelf.com/> Resources and News for Information Professionals by Gary Price. *ResShelf*
3. Basil's Blog - <http://basilsblog.blogspot.com/> News from the information world. *Basil*

Networking, social science perspective

4. Networks, complexity and relatedness <http://www.byeday.net/weblog/networkblog.html> Inquiry and learning into social networks, network analysis and the relationships between people and systems. *Networks*
5. A weblog about weblogging - <http://royby.com/research/weblog.php> . A research and fact gathering blog about the phenomenon that is weblogging. *Blogging*
6. Seb's Open Research - <http://radio.weblogs.com/0110772/> Pointers and thoughts on the evolution of knowledge sharing and scholarly communication, collected by Sébastien Paquet. *Seb*

Computer Science

7. My Computational Complexity Web Log <http://fortnow.com/lance/complog/> - Computational complexity and other fun stuff in math and computer science as viewed by Lance Fortnow, U. Chicago. *Fortnow*
8. Security Weblog - <http://blog.qaddisin.com/> by Qaddisin A network and security company. *Qaddisin*
9. Michael Nielsen <http://www.qinfo.org/people/nielsen/blog/> Occasional thoughts and how to do effective research in physics. The author is a member of the Quantum information science group at the University of Queensland. *MNielsen*

Usability, HCI, and social aspects of technology

10. Sylvie's HCI weblog - <http://dagwood.dgrc.crc.ca/~sylvie/weblog/myblog.html>. A research scientist in Human-Computer Interactions (HCI) for the government of Canada. *Sylvie*
11. misbehaving.net <http://www.misbehaving.net/> A weblog about women and technology. *Misbehaving*
12. WebWord.com <http://webword.com/index.html> Usability and Human Factors for the Internet. *WebWord*

Web

13. Google Weblog <http://google.blogspot.com/> - Aaron Swartz is a teenage writer, coder and hacker. *Google*
14. Onlineblog.com - Guardian Online's weblog <http://www.onlineblog.com/> A weblog covering internet and technology news produced daily by the Guardian Online team. *Online*
15. ResearchBuzz - <http://www.researchbuzz.com/> ResearchBuzz is designed to cover the world of Internet research. Maintained by Tara Calishain. *RBuzz*

For each weblog, we present the average, daily maximum and minimum number of postings and links per postings, average and maximum number of comments per posting, number of links on the sidebar of the main page and archived pages, authorship, and present backlink info retrieved from Technorati and DayPop. The

contents of these blogs were characterized by analyzing the content of the individual postings [11] and by categorizing the links within the postings.

For the characterization of the individual postings, we defined the following attributes:

Posting topic – theme (posting closely related to the defined purpose of the blog), personal, general/other

Length of posting - short, extensive

Links to – no link, news item, weblog, blog posting, article (popular or research), press release, report, reference item, newsletter/e-zine/magazine, book/scientific journal, project, search engine/directory, search result, application (downloadable)/service, event, entity (person, company, university, organization, library, archive, museum etc.), database (“invisible Web”source), list, content site/page, audio/video/image, other

Relation between posting and links in posting – description, quotation, comment (short), discussion (description + extensive commenting), source (the posting is based on this source), mentioning (just the link without any further info), no links in posting, other

Placement of links in posting – embedded, list (a single link is embedded if additional text surrounds it, otherwise it is a list)

The values assigned to the different attributes were developed inductively, i.e. they were defined through browsing actual blog postings.

3. FINDINGS

3.1 Basic statistics related to the blogs

All the blogs we examined based most or all of their postings on information appearing elsewhere and they linked to these sources.

All except four blogs were single authored, three had two authors (InfRes, Basil and Qaddisin) and one had multiple authors - eight during the period of examination (Misbehaving). The number of links on the sidebars of the blogs ranged between 9 (Qaddisin) and 300 (Seb) on the main page of the blog, on archived pages the number of links on sidebars was between 3 (MNielsen) and 300 (Seb – same links as on main page). Table 1 provides details on the posting and links in postings distributions.

Table 1 shows that all the examined blogs took days off without posting any new information. WebWord for example, had almost daily postings (multiple) in September, but none in October (none until mid-November), without any notification. Other authors sometimes put up personal notes saying they’ll be away for a while or provide other reasons for not posting for a period (server down, health problems, etc). The most prolific blog in terms of postings was ResourceShelf, where postings appeared almost every day. We see peaks in the number of postings per day. There are similar peaks for the number of links per posting as well. Four blogs had at least one hypertext link in each and every one of the postings.

Table 2 provides details on the popularity of the examined blogs. We present the number of blogs and links pointing to the specific blog as retrieved from Technorati on November 12, 2003. Technorati, at that time, monitored about 1.2 million blogs and 100 million links. Another tool that tries to measure blog popularity is BlogStats of DayPop, which ranks the blogs by score and by the number of links pointing to it (on November 12, 2003 the ranking were out of 39452 blogs). This ranking is an indication of the relative popularity of the blog among all other blogs monitored by DayPop. An additional indication of popularity is the number of comments on blog postings, when comments are enabled on the blog.

Table 1: Statistics on postings and links per blog during the 61 day period

Blog	average postings/day	number of posting days out of 61	total postings in 61 days	max postings/day	average links/posting	total links in 61 days	min links/posting	max links/posting
InfRes	0.61	37	21	5	2.76	102	0	11
ResShelf	4.85	58	296	11	3.65	1081	1	19
Basil	0.46	15	28	5	1.25	35	0	2
Networks	0.28	14	17	2	1.53	26	0	4
Blogging	0.11	6	7	2	1.86	13	1	3
Seb	2.74	37	167	11	3.34	558	0	31
Fortnow	0.41	24	25	2	1.52	38	0	4
Qaddisin	0.30	13	18	3	0.89	16	0	1
MNielsen	0.33	11	20	3	2.10	42	0	9
Sylvie	1.84	28	120	10	0.54	61	0	4
Misbehaving	0.54	14	33	8	2.06	68	0	8
WebWord	1.05	23	63	8	1.03	66	0	4
Google	0.33	10	20	7	2.80	56	1	8
Online	1.95	50	119	6	1.61	192	0	7
Rbuzz	1.79	37	109	5	1.43	156	1	7

Table 2: Popularity indicators – DayPop: 39352 blogs, Technorati: 1.2 million blogs, 100 million links

Blog	DayPop rank by score	DayPop rank by citation	Technorati - # incoming blogs	Technorati -# incoming links	total comments in 61 days	average comments /posting	max comment s/posting
InfRes	892	1457	1	1	not enabled		
ResShelf	353	272	123	398	not enabled		
Basil	not listed		0	0	not enabled		
Networks	19916	13502	6	6	10	0.59	3
Blogging	2206	2487	22	23	9	1.29	4
Seb	N/A	N/A	335	551	122	0.73	
Fortnow	not listed		21	24	19	0.76	6
Qaddisin	6480	2300	25	26	not enabled		
MNielsen	not listed		3	3	29	1.53	8
Sylvie	9658	7289	4	4	not enabled		
Misbehaving	not listed		125	162	128	4.00	18
WebWord	not listed		6	6	212	3.31	11
Google	184	105	325	408	not enabled		
Online	227	279	154	185	369	3.13	23
Rbuzz	412	336	183	238	not enabled		

Among the examined blogs, the Google blog seems to be most popular (first among the examined blogs on DayPop and second on Technorati), even though during the period we monitored it, it published a relatively few postings and 10 out of the 20 postings overlapped with postings on the same issue by other blogs in our set as well. Misbehaving appeared on the Web only recently, the first posting is from September 24, 2003 – but has already succeeded in gaining a lot of attention.

3.2 Characterization of postings

The *topics* of most postings in all the blogs were characterized as closely related to the declared topic (except for Fortnow, where only 44% were considered as such). In several cases (7 blogs) all the postings were “on theme”.

In the *length of the postings* there were considerable stylistic differences. Only four blogs (InfRes, Fortnow, Misbehaving and ResarchBuzz) published mainly extensive postings, while for the others the majority of the postings was categorized as short (Blogging published only short items during the whole period).

Only in a single blog (Sylvie) the majority of the postings were without links. In postings including hypertext links, usually the links are embedded, although lists of links are published occasionally. An exception was ResourceShelf, where most of the postings were in a “telegraphic style” with very short descriptions of a list of items in each posting. Altogether 1085 postings were published on all fifteen blogs during the monitoring period. Out of this, in 11.8% there were no links, in 15.9% of the postings the links appeared in lists, while for the large majority of cases (72.3%) the links were embedded within the text.

Tables 3a and 3b display the characterizations of the link targets per blog. This characterization provides us with a clear view of the types of information sources these blogs utilize. The categories are arranged in decreasing order of the total targets mentioned in all the weblogs during the whole period (2510 links).

The tables show that the most often occurring categories are news items, content/site page and blog posting, when all the links from all the blogs are considered as one multi set of links (links may occur with multiplicity, when the same target is mentioned several times in a given blog or is mentioned in several blogs). If the categories *blog posting* and *weblog* are combined, then they become the most frequently occurring target type (18.4%). Thus bloggers rely very heavily on information published in the blogspace. This definitely is a change from the original “what’s new” pages (the ancestor”of weblogs), where the links did not point to other “what’s news pages” but to other types of resources on the Web. However a closer look at the distribution of link types show that the most frequently occurring target type was blogs (postings and weblogs combined) only for two blogs (Seb and Sylvie), news items were most frequently occurring in four blogs (InfRes, Basil, Qaddisin, and Online), content sites were most popular targets in four blogs as well (ResourceShelf, Basil, Blogging and WebWord), the two blogs that mentioned most frequently search tools - search engines or directories - were Google and ResearchBuzz). Another two blogs linked most frequently to entities (Misbehaving and Fortnow), and one blog linked most often to applications/services (Blogging) and to event/presentation (MNielsen). Thus the linking behavior is far from uniform even if we consider this small subset of blogs, where the theme of many of them was rather closely related.

Table 3a: Link targets of the examined blogs (in percentages)

	Total links/ category (2510 links)	% of links in category out of total links	InfRes (102 links)	ResShelf (1081 links)	Basil (35 links)	Networks (26 links)	Blogging (13 links)	Seb (558 links)	Fortnow (38 links)	Qaddisin (16 links)
news item	336	13.4%	13.7%	14.0%	20.0%	7.7%	7.7%	2.3%	5.3%	62.5%
content site/page (incl. documentation)	324	12.9%	8.8%	15.6%	20.0%	11.5%	23.1%	9.9%	5.3%	6.3%
blog posting	301	12.0%	6.9%	6.0%	5.7%	11.5%	7.7%	26.3%	10.5%	12.5%
weblog	160	6.4%	5.9%	0.6%	2.9%	7.7%	0.0%	21.0%	2.6%	0.0%
article (popular and scientific)	152	6.1%	11.8%	6.6%	11.4%	0.0%	15.4%	2.7%	18.4%	12.5%
se/directory	136	5.4%	2.9%	6.2%	14.3%	0.0%	0.0%	1.4%	0.0%	0.0%
press release	135	5.4%	0.0%	11.2%	2.9%	3.8%	0.0%	0.0%	0.0%	0.0%
entity (e.g. person, company, organization)	128	5.1%	2.0%	4.4%	2.9%	30.8%	0.0%	5.2%	21.1%	0.0%
application/service	120	4.8%	4.9%	1.9%	2.9%	3.8%	23.1%	10.2%	2.6%	0.0%
list	107	4.3%	0.0%	7.6%	0.0%	7.7%	0.0%	2.2%	0.0%	0.0%
report	107	4.3%	1.0%	9.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
database ("invisible web")	89	3.5%	7.8%	5.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
event/presentation	75	3.0%	4.9%	1.1%	0.0%	7.7%	0.0%	1.6%	18.4%	0.0%
search results	68	2.7%	1.0%	3.4%	0.0%	0.0%	0.0%	3.2%	0.0%	0.0%
other	62	2.5%	4.9%	0.9%	0.0%	0.0%	0.0%	3.9%	0.0%	6.3%
reference item	53	2.1%	1.0%	0.7%	2.9%	0.0%	23.1%	5.0%	10.5%	0.0%
newsletter/e-zine/magazine	51	2.0%	4.9%	1.9%	5.7%	0.0%	0.0%	0.4%	0.0%	0.0%
book/scientific journal	45	1.8%	11.8%	1.9%	0.0%	7.7%	0.0%	0.7%	5.3%	0.0%
audio/video/image	32	1.3%	2.0%	0.6%	0.0%	0.0%	0.0%	2.7%	0.0%	0.0%
project	29	1.2%	3.9%	0.8%	8.6%	0.0%	0.0%	1.3%	0.0%	0.0%

Table 3b: Link targets of the examined blogs (in percentages)

	Total links/ category (2510 links)	% of links in category out of total links	MNielsen (42 links)	Sylvie (61 links)	Misbehav- ing (68 links)	Webword (66 links)	Google (56 links)	Online (192 links)	Rbuzz (156 links)
news item	336	13.4%	0.0%	4.9%	10.3%	22.7%	10.7%	52.6%	2.6%
content site/page (incl. documentation)	324	12.9%	9.5%	18.0%	14.7%	28.8%	16.1%	3.6%	9.6%
blog posting	301	12.0%	11.9%	31.1%	17.6%	10.6%	12.5%	9.9%	0.6%
weblog	160	6.4%	4.8%	3.3%	1.5%	0.0%	3.6%	5.2%	5.8%
article (popular and scientific)	152	6.1%	9.5%	4.9%	7.4%	13.6%	1.8%	6.3%	3.2%
se/directory	136	5.4%	0.0%	1.6%	0.0%	1.5%	17.9%	0.5%	25.6%
press release	135	5.4%	0.0%	0.0%	0.0%	3.0%	0.0%	3.1%	2.6%
entity (e.g. person, company, organization)	128	5.1%	7.1%	0.0%	20.6%	1.5%	1.8%	3.1%	4.5%
application/service	120	4.8%	0.0%	14.8%	2.9%	3.0%	5.4%	2.6%	6.4%
list	107	4.3%	0.0%	0.0%	1.5%	0.0%	0.0%	0.5%	5.8%
report	107	4.3%	0.0%	0.0%	0.0%	3.0%	1.8%	1.0%	0.0%
database ("invisible web")	89	3.5%	0.0%	1.6%	0.0%	0.0%	1.8%	0.0%	15.4%
event/presentation	75	3.0%	54.8%	11.5%	8.8%	0.0%	0.0%	2.1%	0.0%
search results	68	2.7%	0.0%	0.0%	0.0%	0.0%	16.1%	1.0%	0.6%
other	62	2.5%	0.0%	3.3%	1.5%	3.0%	1.8%	7.3%	2.6%
reference item	53	2.1%	0.0%	1.6%	0.0%	7.6%	3.6%	0.0%	0.0%
newsletter/e-zine/magazine	51	2.0%	0.0%	1.6%	0.0%	0.0%	0.0%	0.0%	13.5%
book/scientific journal	45	1.8%	0.0%	1.6%	4.4%	0.0%	0.0%	0.5%	0.0%
audio/video/image	32	1.3%	2.4%	0.0%	4.4%	0.0%	5.4%	0.5%	0.0%
project	29	1.2%	0.0%	0.0%	4.4%	1.5%	0.0%	0.0%	1.3%

Table 4: Relationship between source and target

	quotation		description		comment		discussion		no link		mentioning		other		total assignments
InfRes	8	14.55%	12	21.82%	7	12.73%	18	32.73%	5	9.09%	0	0.00%	3	5.45%	55
ResShelf	163	39.85%	120	29.34%	23	5.62%	24	5.87%	0	0.00%	60	14.67%	4	0.98%	409
Basil	0	0.00%	17	58.62%	6	20.69%	0	0.00%	1	3.45%	5	17.24%	0	0.00%	29
Networks	0	0.00%	5	26.32%	3	15.79%	8	42.11%	1	5.26%	2	10.53%	0	0.00%	19
Blogging	1	10.00%	4	40.00%	4	40.00%	0	0.00%	0	0.00%	1	10.00%	0	0.00%	10
Seb	80	35.87%	18	8.07%	36	16.14%	45	20.18%	4	1.79%	23	10.31%	7	3.14%	223
Fortnow	0	0.00%	9	31.03%	6	20.69%	4	13.79%	6	20.69%	2	6.90%	2	6.90%	29
Qaddisin	0	0.00%	7	33.33%	8	38.10%	3	14.29%	2	9.52%	1	4.76%	0	0.00%	21
MNielsen	1	4.76%	1	4.76%	2	9.52%	3	14.29%	8	38.10%	6	28.57%	0	0.00%	21
Sylvie	5	3.79%	8	6.06%	19	14.39%	11	8.33%	78	59.09%	6	4.55%	2	1.52%	132
Misbehaving	7	17.50%	2	5.00%	6	15.00%	13	32.50%	12	30.00%	0	0.00%	0	0.00%	40
Webword	43	63.24%	0	0.00%	12	17.65%	1	1.47%	8	11.76%	2	2.94%	2	2.94%	68
Google	5	18.52%	10	37.04%	4	14.81%	7	25.93%	0	0.00%	1	3.70%	0	0.00%	27
Online	76	46.06%	34	20.61%	25	15.15%	12	7.27%	2	1.21%	7	4.24%	1	0.61%	165
Rbuzz	0	0.00%	96	60.00%	42	26.25%	11	6.88%	1	0.63%	2	1.25%	1	0.63%	160
Total	389	27.63%	343	24.36%	203	14.42%	160	11.36%	128	9.09%	118	8.38%	22	1.56%	1408

Thus we see large variability in the distribution of the target type. Somewhat to our surprise it was far easier to categorize the relation between the postings and the targets, seven values were sufficient for the characterization. The results appear in Table 4. Multiple assignments were allowed per posting, but if in a posting there were links to several targets that were in similar relation to the source, this relation type was noted only once per posting. The values for this attribute and short explanations appear in section 2 of this paper. The percentages are calculated out of the total number of assignments for this category for each blog.

We see that blogs most often quote from other resources, or describe the target or its content, sometimes the bloggers add their short comments and only rarely engage in extensive discussion of the target. For general Webpages there are many more relation types between the source and the target; however these types did not appear often enough in the postings we characterized.

4. CONCLUSIONS

“Professional” blogs are excellent sources of secondary or tertiary information. Most information (apart from occasional extensive discussions of topics) can be easily found elsewhere, but these blogs concentrate and filter this info, and they can be viewed as one-stop information kiosks. The postings are based mainly on information appearing in other weblogs, news items and press releases.

Seemingly, the aim of the authors of these blogs is to disseminate information and to express their opinion related to the disseminated information. Starting discussions with readers and providing details on personal life seem to be of secondary importance. The blog authors disseminate information, but for whom? Who reads blogs, what blogs, why, and how often? These are intriguing questions beyond the scope of this paper, however these questions have to be answered as well if we want to gain an understanding of the blogosphere.

The relatively large number of tools listing and tracking weblogs and blog postings (for a list of such tools consult FaganFinder, <http://www.faganfinder.com/blogs/>) indicates that this community is very much aware of itself and is interested to learn about the popularity of specific topics and blogs. The next step for them will be to try to reach more Web users that may become active bloggers after reading blogs for a while.

Analyzing the content of weblog postings provides us with a picture of what these blogs are about. The findings are based on results of a well-accepted research methodology and not on impressions alone. Of course further research along these lines is needed in order to gain a more thorough understanding of this emerging phenomenon – blogging.

5. REFERENCES

- [1] Azeem, A. Blogger demographics and behaviour. <http://www.20six.co.uk/weblogCategory/1mspx188uju6?p=1>
- [2] Bausch, P., Haughey, M., and Hourihan, M. We blog: Publishing online with Weblogs. Wiley Publishing, Indianapolis, IN, 2002.
- [3] Blogcount. The Blogcount estimate. http://dijest.com/bc/2003_06_23_bc.html#105638688729256217
- [4] Blogcount. Perseus: 1.4 million hosted blogs. http://dijest.com/bc/2003_10_03_bc.html#106521436105333320
- [5] Blood, R. Weblogs: A history and perspective. In We've got blog: How weblogs are changing our culture, John Rodzvilla (ed). Perseus Publishing, Cambridge, MA, 2002. Also http://www.rebeccablood.net/essays/weblog_history.html
- [6] Embrey, T. R. You blog, we blog: Guide to How Teacher-Librarians Can Use Weblogs to Build Communication and Research Skills. Teacher Librarian 30(2), 2002. http://teacherlibrarian.com/pages/30_2_feature.html

- [7] Graham, B. L. Why I weblog: A rumination on where the hell I'm going with this website. In *We've got blog: How weblogs are changing our culture*, John Rodzvilla (ed). Perseus Publishing, Cambridge, MA, 2002. Also <http://www.bradlands.com/words/maybe/maybe02.html>
- [8] Greenspan, R. Blogging by the numbers. http://cyberatlas.internet.com/big_picture/applications/article/0,,1301_2238831,00.html
- [9] Halavais, A. Weblog update frequency. <http://alex.halavais.net/news/archives/000369.html>
- [10] Hargittai, E. Blog types. <http://campuscgi.princeton.edu/~eszter/weblog/archives/0000191.html>
- [11] Krippendorff, K. *Content analysis – An introduction to its methodology*. Sage Publications, Beverly Hills, CA, 1980.
- [12] Lasica, J. D. Weblogs: A new source of information. In *We've got blog: How weblogs are changing our culture*, John Rodzvilla (ed). Perseus Publishing, Cambridge, MA, 2002. Also <http://www.ojr.org/ojr/lasica/p1019165278.php>
- [13] Mernit, S. Blogger classifications : Some thoughts, http://susanmernit.blogspot.com/archives/2003_10_05_susanmernit_archive.html#106554742635066352
- [14] Perseus. The Blogging Iceberg. <http://www.perseus.com/blogsurvey/>
- [15] Rodzvilla, J. (ed). *We've got blog: How weblogs are changing our culture*. Perseus Publishing, Cambridge, MA, 2002.
- [16] Webopedia. Blog. <http://www.webopedia.com/TERM/b/blog.html>
- [17] Weil, D. Top 20 definitions of blogging. <http://www.wordbiz.com/archive/20blogdefs.shtml>
- [18] Wiener, D. The history of weblogs. <http://www.userland.com/theHistoryOfWeblogs>