

The Meaning of Links in Blogs

Outi Nivakoski

outi.nivakoski@abo.fi

Department of Information Studies, Åbo Akademi University, Tavastgatan 13, FIN-20500 Åbo (Finland)

Abstract

This study aims to increase the knowledge of linking on a distinct Web space, the blogosphere. It investigates the meaning of hyperlinks in blogs. The material consists of 160 links from 80 Finnish blogs. Reasons for creating links, visible attitudes towards target pages and contexts of links were analyzed. The results show that links in the blogosphere are means for showing appreciation and valuation. They connect similar topics in the blogosphere as well as on the Web generally. Links in blogs can also reveal negative attitudes towards target pages, contrary to what is usually the case on other Web environments. In diary blogs, of which the sample was largely formed, links are more used to tell about interests and valuations of the blogger than to add something to the content of the blog. The study confirms the picture of links as an essential element of the blogosphere, without which it would lose much of its specific nature.

Introduction

A blog (weblog) can be described as a Web page, where content is added relatively often and where the most recent content is at the top of the page (Herring et al., 2005b; Cohen & Krishnamurthy, 2006). Contents appear in postings that include also permalink and other additional information (Miller & Shepherd, 2004). Content types are manifold: text, images, video clips, hyperlinks, and various mixtures of these. One of the central features is that readers can comment blogs through the comment feature (Miller & Shepherd, 2004; Herring et al., 2005a; Kilpi, 2006). A blog without the comment feature would lose much of its nature as a forum for interaction. It could even be said that it wouldn't be a blog anymore, only a publishing platform.

Typically blogs are categorized according to the forms of their content (e.g. Blood, 2000; Herring et al., 2005b). A filter blog contains mostly hyperlinks pointing to other Web documents, usually with blogger's comments (Blood, 2000). A diary blog presents blogger's personal life and topics include everyday life, hobbies, etc. (Miller & Shepherd, 2004; Wei, 2004). A k-log (knowledge blog) offers information, observations and experiences of some particular topic (Herring et al., 2005b). Picture blogs present some kind of visual material: photos (photo blogs), art, etc.

The blogosphere is the community of all blogs on the Web (Kilpi, 2006). Blogs form a distinct category among other Web documents. Compared to other content on the Web, the blogosphere forms a new kind of time-space and a social network where movement both chronologically backwards and horizontally between blogs or other Web sites is possible by following hyperlinks (Sirkkunen, 2006).

Hyperlinks are an essential part of the blogosphere. Linking is an important means for communication and with links bloggers can form and maintain communities (Blood, 2000; Kumar et al., 2004; Miller & Shepherd, 2004; Chin & Chignell, 2006; Sirkkunen, 2006). Links have been found to represent bloggers' opinions and social appreciation (Miller & Shepherd, 2004). However, some blogs contain very few links or do not contain them at all (Clyde, 2004; Herring et al., 2005a; Herring et al., 2005b). This can be explained with different purposes and content types of blogs: for example, diary blogs contain less links than filter blogs (Herring et al., 2005a; Herring et al., 2005b).

The goal of this research is to study what links in the blogosphere really represent. The goal will be fulfilled by studying the motivations behind linking and the contexts where links occur. In general, the lack of knowledge concerning the meaning of links restrains a deeper

understanding of the Web, as well as the development of webometrics as a science (Wilkinson et al., 2003; Thelwall, 2006). The research concerning meaning of links in blogs has been scarce all around the world and in Finland the subject has not been studied before. Researchers call for the development of theories and methods in webometrics because new theories and tested methods can help to better understand topology, functions, and possibilities for use and development of the Web (Björneborn & Ingwersen, 2001). This study will contribute to the research field of webometrics by studying the meaning of links in the blogosphere and with that, to increase our knowledge about the Web.

Linking on the Web

Web documents gain link citations for various reasons (Thelwall, 2006). For example, sites gaining lots of in-links are not necessarily semantically valuable; they just contain useful information or represent some eminent direction (Thelwall, 2002). The most common hypothesis is that links are created for positive reasons. For example, some search engines use the Web's link structure and the amounts of gained in-links as quality indicators for Web pages (cf. Brin & Page, 1998).

Generally it can be maintained that the creator of a link thinks that the target is somehow worth noticing, either positively or negatively (Berners-Lee, 2000). Link creator wants to direct the reader to somehow interesting page or to an information source (Haas & Grams, 1998; Davison, 2000). This means that links can be classified into two main categories: navigation links and informational links (Haas & Grams, 1998). This classification is, however, too wide to help creation of deeper knowledge of linking and researchers have created more precise categorizations for links.

Haas & Grams (1998) have categorized links to create clearer picture of network structure and contents of the Web. Their material was uncategorized Web sites and their links. Thelwall (2003) has categorized links from academic Web sites and focused especially on those reasons that differ from reasons for academic citation. Bar-Ilan (2005) has presented a few link types in the literature review of her study. Ali-Hasan & Adamic (2007) have studied social ties in the blogosphere and described some general reasons for linking in that environment. The summary of reasons for linking mentioned in these studies is presented in table 1.

Each of above mentioned studies names some kind of conventional linking (following the typical linking patterns on the Web) but definitions for that vary: courtesy links (Haas & Grams, 1998; Thelwall, 2003; Bar-Ilan, 2005), formulaic links (Thelwall, 2003), as well as reciprocal links (Haas & Grams, 1998; Bar-Ilan, 2005; Ali-Hasan & Adamic, 2007) can be conventional links. Navigation links, courtesy links, reciprocal links, links created to show acknowledgement or appreciation, links created for social reasons, and advertisement links are apparently occurring commonly on the Web, because they are mentioned in more than one study. There are also some link types that occur rarely on the Web. For example, Haas & Grams (1998) present persuasion links only as a minor subcategory within advertisement links. Thelwall (2003) presents links that are created for expressing technical competence as a minor type of links that is hard to identify.

Table 1. Reasons for creating links in various types of Web sites.

Reason for creating a link	Description	Authors
Navigation	Links within a single Web page or site. Makes moving within the site easier.	Haas & Grams, 1998; Thelwall, 2003
Courtesy	Linking to a target that is mentioned, but where the link is not actually needed nor expected to be used.	Haas & Grams 1998; Thelwall, 2003; Bar-Ilan, 2005
Formulaic	Following the typical manner of linking of particular Web site type.	Thelwall, 2003
Reciprocal	Links between the source and the target pages are reciprocated, but only for superficial, complimentary reasons.	Haas & Grams, 1998; Bar-Ilan, 2005; Ali-Hasan & Adamic, 2007
Acknowledgement	The link creator wants to recognize a party that has contributed to the creation of the site or some of its contents.	Haas & Grams, 1998; Thelwall, 2003
Appreciation	The link represents creator's true positive valuation of the target.	Bar-Ilan, 2005; Ali-Hasan & Adamic, 2007
Presenting a content entity	The target presents an entity that gives wider or more detailed information on a subject.	Haas & Grams, 1998
Offering information	Target is in some way interesting or useful; the subject is not necessarily the same as in the source page.	Haas & Grams, 1998
Social	Linking represents real social ties within either online or offline world.	Thelwall, 2003; Ali-Hasan & Adamic, 2007
Advertisement	Advertising some other page, site, or a product.	Haas & Grams, 1998; Bar-Ilan, 2005
Persuasion	Advertising the content of the same site and trying to entice more visitors there (often competitions, little games etc.).	Haas & Grams, 1998
Technical	Created for technical reasons, for example downloading or uploading a file.	Bar-Ilan, 2005
Expressing technical competence	The creator of the link wants to show that s/he can create a link.	Thelwall, 2003
Practice	The creator of the link wants to practice link creation.	Thelwall, 2003

Results from citation analysis studies are often used as starting and reference points in link analysis studies (e.g. Thelwall, 2003; Chu, 2005), even though citations and links represent quite different phenomena (Thelwall, 2003; Björneborn & Ingwersen, 2004; Bar-Ilan, 2005). Motivations for creating citations are different from motivations for creating links (Thelwall, 2004; Bar-Ilan, 2005). Some of the motivations for creating links are clearly new and unique products of the Web and they cannot be equated with reasons for references in academic

publications (Thelwall, 2003). For example, several studies focusing on reasons for creating citations have found criticizing citations that are used for negating results, methods, or opinions of other authors (cf. Weinstock, 1971; Oppenheim & Renn, 1978). On the Web this kind of linking seems to be very rare (Chu, 2005).

Linking in the blogosphere

Linking is a central element for the communities and networks of the blogosphere (e.g. Kumar et al., 2004; Herring et al., 2005a; Chin & Chignell, 2006; Cohen & Krishnamurthy, 2006; Ali-Hasan & Adamic, 2007). However, research on why bloggers actually create links and what links mean for blogs has been almost nonexistent. Blog link research thus far has concentrated mostly on social issues that links reveal (e.g. Herring et al., 2005a; Chin & Chignell, 2006; Ali-Hasan & Adamic, 2007).

Along with commenting, linking is a means for social control and a sign of appreciation. Links in blogs are actually expected to reveal opinions and valuations (Miller & Shepherd, 2004). With links bloggers create sub-communities and hierarchies within the blogosphere (Blood, 2000; Kumar et al., 2004; Miller & Shepherd, 2004; Sirkkunen, 2006). Linking has similar implications in the blogosphere as communication has in offline communities (cf. Orlikowski & Yates, 1994). Linking is therefore a part of the ongoing communication in the blogosphere.

On the Web links are assumed to represent the link creator's interests and valuations (Thelwall, 2004; Bar-Ilan, 2005). The same applies also to the blogosphere: a blog that is gaining more in-links is seen to gain more respect and popularity (Cohen & Krishnamurthy, 2006; Ali-Hasan & Adamic, 2007). However, even if linking in the blogosphere reminds of linking on the Web in general, some differences certainly occur.

Links in blogs tend to point to other blogs (Herring et al., 2005b; Cohen & Krishnamurthy, 2006; Ali-Hasan & Adamic, 2007) and links between blogs are often reciprocated (Ali-Hasan & Adamic, 2007). Some researchers argue, however, that reciprocated linking is not automatically a sign of genuine interaction and discussion (Herring et al., 2005a; Ali-Hasan & Adamic, 2007). Even if a group of blogs is connected through links, discussions between bloggers are less likely to occur (Herring et al., 2005a). This implies that links represent more reciprocated politeness and acknowledgement than genuine social ties (Ali-Hasan & Adamic, 2007).

The majority of links in blogs are in blogrolls (lists of other blogs that each blogger frequently reads; usually situated in sidebars (Kumar et al., 2004)) (Ali-Hasan & Adamic, 2007). Most of the blogroll links are reciprocated (Ali-Hasan & Adamic, 2007). However, these links quite seldom represent real social ties: reciprocated commenting occurs rarely and these blogs are not mentioned in postings (Ali-Hasan & Adamic, 2007). The links in blogrolls do still manifest bloggers' values and approval.

Linking patterns vary remarkably within the blogosphere. Some studies have revealed that diary blogs typically contain less links than filter blogs (Herring et al., 2005a; Herring et al., 2005b). Different sub-communities also link differently; variation occurs in amount of links and accumulation patterns of links (Herring et al., 2005a; Ali-Hasan & Adamic, 2007). There are densely interconnected sub-groups as well as individualists, who do not link to other blogs nor Web sites (Herring et al., 2005a). Linking often occurs in bursts, which means that the activity level of linking varies (Kumar et al., 2004).

Research questions

The goal of this study is to find out why are links created in blogs: Do they represent real valuation of the target page or are they created only for a habit? Do links add something to the content of a blog or are they used to define and show characteristics of the bloggers?

There are two things that need to be reviewed when studying the meanings of links. First, *the motivations* for creating links must be studied (Holmberg & Thelwall, 2007). In this study the linking motivation consists of the apparent reason for which the link has been created and the attitude that the link represents (cf. Kim, 2004, p. 87).

Second, *the context* where links occur on a page has to be studied. The context consists of link's concrete location on the page and the possible description of the target page (Haas & Grams, 1998; Bar-Ilan, 2005). The context of the link helps to define why the link was created because the link text (or picture, etc.) and the immediate surroundings of the link can reveal important facts of the creator's intentions (Haas & Grams, 1998). In blogs links placed in different parts of a blog (sidebar, blogroll, or posting) may represent different kinds of relationships between source and target pages and with that, various intentions of the blogger (cf. Ali-Hasan & Adamic, 2007).

The research questions that seek to clarify to the meaning of links in blogs are as follows:

1. What kinds of linking motivations do bloggers have?
2. In what kinds of contexts do links occur in blogs?

Material

The blogs for the study were collected from the Finnish blog listing service called Blogilista (<http://www.blogilista.fi/>). The service lists blogs of various topics maintained by Finnish bloggers or concentrating on Finnish topics (Blogilista, 2008) but it naturally doesn't give an exhaustive list of all Finnish blogs. At the time of the study there were 14,108 blogs listed on Blogilista (Blogilista, 2008).

Using Blogilista's random blog function a sample of 80 blogs was collected for this study. The collecting of the sample took place on March 15-16, 2008. To be taken into the sample, a blog's latest posting had to be posted after December 15, 2007. This condition was used for limiting the amount of abandoned blogs that might include dead links. Two of the blogs were maintained by a group of bloggers, increasing the total amount of bloggers to 90.

A specially designed Web crawler was used to collect the links from the blogs (SocSciBot, 2007). The links studied were directed outside from the source blogs (out-links) and pointed to an external Web document. The total amount of links found in blogs was 915,972. Using the Web crawler's random sample selector 160 links were extracted for content analysis. Thelwall (2004) proposes taking at least this amount of links for analysis to be able to establish an acceptable level of validity. Some blogs did not have any links and hence the sample of 160 links could not be extracted evenly from all the blogs. Two links were randomly chosen from 60 blogs, one link was chosen from 20 blogs and an additional third link was chosen from 20 blogs.

Some links first collected were considered as irrelevant for this study after more specific analysis and hence removed from the sample. A total of 1,510 links were analyzed to get a sample of 160 links for the content analysis. Among the links considered to be irrelevant were 660 control links for the blog maintenance, 129 navigation links, 173 advertisement links added by the blog hosting service, 69 URL-addresses of those who had commented blogs, and 79 embedded links (for example picture URLs). The target page was no longer available in 141 occasions and 99 of the irrelevant links were duplicates of the relevant links.

Methods

Content analysis was used for interpreting the links. Content analysis suits well for studying both linking motivations and link contexts (Thelwall, 2003; Bar-Ilan, 2005). It is the most used method in link analysis studies and it has many features supporting its use. For example, the researcher has access to wide range of material because source and target pages can be reached quite easily (Chu, 2005). The amount of links can be so high that other methods, like

interviewing link creators, are almost impossible to carry out. With content analysis it is possible to handle larger amounts of material (Wilkinson et al., 2003; Chu, 2005).

Content analysis has also some drawbacks. Because of its qualitative nature, it produces results that are based on subjective judgments (Chu, 2005). There are many reasons for creating a link and these overlapping reasons can be difficult to detect only through source and target pages (e.g. Thelwall, 2003; Wilkinson et al., 2003). However, some researchers (e.g. Chu, 2005) argue that typically there is only one clear reason for creating a link.

The general problem with link analysis is that most of the links are irrelevant for the research problem (Björneborn & Ingwersen, 2001; Thelwall, 2004), which may have an impact on the results (Thelwall, 2004). Irrelevant links can be created by someone else than the actual owner of the source site, and often links are created only for navigational reasons. These kinds of links do not express anything about the relationship between the source and target pages, or site owner's attitude towards the target page (Thelwall, 2004). A commonly used relevance criteria for inclusion of links for the analysis was used: every link had to be created individually and manually, and the target had to be chosen through some kind of evaluation (Thelwall, 2004). Link should also be created by the owner of the site (Thelwall, 2004). These criteria exclude for example navigation links and links created by advertisers.

To answer the first research question *types of source and target pages, reasons for linking, and attitudes that links represent* were studied.

Source blogs were classified according to the form of their contents. Types 'combination' and 'other' recommended by Herring et al. (2005b) were also included. The classification scheme included six categories: filter blogs, diary blogs, k-logs, picture blogs, combination blogs (blogs combining several types of content), and other blogs (not fitting in other categories).

The classification scheme for target page types was based on the study of blog links by Herring et al. (2005b). According to Herring et al. (2005b) links from blogs usually point to Web sites (maintained by other people), blogs (maintained by other people), news sites, and other blogs and sites maintained by the blogger. Instead of Herring et al.'s (2005b) 'webring' as a target site, type labeled 'Web community' was used. This type could entail webrings, discussion forums, etc. To these a content entity that could include for example a single article, photo, etc. (cf. Almind & Ingwersen, 1997) was added. The classification scheme for target page types included six categories.

The summary of linking reasons mentioned in four earlier studies (see table 1) was used as a starting point for categorizing the reasons for linking. The navigation links were removed from the classification scheme because they were considered as irrelevant for the goal of this study. The classification scheme for linking reasons therefore included 13 categories.

The attitude of links could be positive, neutral, or negative. This is a typical classification scheme when studying attitudes of people (Pietilä, 1976). Often the attitude can be inferred from the context of the link (Berners-Lee, 2000).

The *context of links* was studied to answer the second research question. Links in blogs can be situated in postings, blogrolls, and link lists. It was hypothesized that the links in postings are connected to the contents and the subject of the blog and the links in blogrolls and link lists tell more about the blogger's interests, characteristics, and relationship with other bloggers and the Web as a whole.

Demographic information about the bloggers was collected from what they told about themselves in sidebars, profile page, or on other page linked to the blog. Information available on the first page of the blog was also examined. Information collected included age, sex, place of residence, whether the blogger revealed his / her name or not, and if there was a recognizable photo of the blogger.

Findings

Demographics

Of the 90 bloggers 57 were women and 20 were men. 13 bloggers didn't reveal their sex. For 57 bloggers age was not definable. Of those 33 bloggers who did reveal their age, five were 20 years old or under, 16 were 21 – 30 years old, eight 31 – 40 years old, and three 41 – 50 years old. The youngest blogger was a 15 year old girl and the oldest was over 60 year old woman. Of the bloggers that revealed their geographic location 52 lived in Finland and three lived abroad.

A recognizable photo was presented only by 28 bloggers. 51 bloggers reported their first name or nickname, both first and last names told 27 bloggers. Both the photo and the whole name was presented by 19 bloggers, of whom nine (that wrote the same group blog) blogged as part of their job for a blog of a women's magazine.

The most popular blog hosting service was Blogger: of 80 blogs it hosted 39. A Finnish blog hosting service Vuodatus hosted 19 blogs. Ten blogs were hosted by the bloggers themselves, Wordpress hosted seven blogs, and other services (Spaces, Blogsome, another Finnish service Blogitus) hosted altogether five blogs.

Motivations for creating links

The majority of blogs (64) were personal diaries. Typical subjects were hobbies, fashion and style, and everyday life. Three blogs could be identified as filters: two of them commented on societal issues and one contagion of infectious diseases in the world. Altogether six blogs were k-logs. They discussed development of technical products; a popular Finnish television show; different colors, pigments, and paints; a religious concept; and societal issues. One k-log was a PR blog of a Finnish company. Three blogs were picture blogs: two contained photos, and one was a cartoon blog. One blog could be defined as a combination of diary and filter blogs. Three blogs fell in the class "other": a blog that published blog layouts for other bloggers to use, a blog for teaching the Finnish language, and a blog that coordinated a handicraft swap of bloggers.

Table 2 depicts targets of links of each source blog type. The sample shows evidence of self-citations: 11.3% of links pointed to bloggers' other Web documents, such as blogs, home sites, and social networking service profiles. There was only one link pointing to a news site. This seems to strongly support the claim of Herring et al. (2005b) that the amount of blogs linking to news sites has been greatly exaggerated. K-logs contain proportionally many links that point to Web sites maintained by others. These links were mostly targeted towards sites that discussed some issue deeply.

Table 2. Types of source and target sites for links.

		Target site type						Total
		Web site	Blog	News site	Own site / blog	Web community	Content	
Source blog type	Filter	4	1				1	6
	Diary	40	55	1	13	5	13	127
	K-log	8	1		2		3	14
	Picture	4			1			5
	Combination	1	1					2
	Other		4		2			6
	Total	57	62	1	18	5	17	160

Bloggers' reasons for creating links are represented in table 3. The three most popular reasons (appreciation, pointing to information, presenting a content entity) accounted for 68.8% of all the links. The remaining 31.2% of links were divided between six reasons: courtesy, acknowledgement, social, advertisement, technical, and social / reciprocal.

Table 3. Reasons for creating links.

Reason for creating a link	Number	%
Courtesy	5	3.1
Formulaic	0	
Reciprocal	0	
Acknowledgement	7	4.4
Appreciation	56	35.0
Presenting a content entity	24	15.0
Pointing to information	30	18.8
Social	6	3.7
Advertisement	16	10.0
Persuasion	0	
Technical	9	5.6
Expressing technical competence	0	
Practice	0	
Social / Reciprocal	7	4.4
Total	160	100.0

The class “social / reciprocal links” was added to the original classification scheme. It includes links that were reciprocal between the source blog and the target page, but for which the evidence of the depth of the relationship did not come out. Links in this category could be placed in either “reciprocal” or “social” links after deep, chronological analysis on postings, comments, and linking patterns, but this was not in the scope of this study. The six genuinely social links could be recognized because bloggers told about their relationships, or there were discussions between them in the comment sections.

The sample did not include advertisement links that were created by others than bloggers themselves. The advertisement links mentioned in table 2 are therefore created by the blogger and they advertised books, HTML- and CSS-validators, magazines, etc.

The conventional courtesy linking visible on the Web occurs also in the blogosphere. However, if courtesy links are strictly determined as links pointing to a direction mentioned in text (Thelwall, 2003), blog links occurring in other contexts than postings are ruled out. This may be too one-sided a condition for the blogosphere.

The absence of formulaic links is understandable because it's not clear what is typical linking in the blogosphere. The lack of persuasion links, links expressing technical competence, and practice links was not a big surprise. Persuasion links and links created for expressing technical competence are only minor types of links on the Web (Haas & Grams, 1998; Thelwall, 2003). Practice links may occur in blogs but they can also be quite difficult to

identify and likely this reason occurs with other reasons that are more visible. The absence of reciprocal linking is in connection with the method of study, which resulted in the creation of the new category, social / reciprocal links.

The appreciation links were investigated further to understand better why the links were created. The appreciation links were classified according to the subjects of source blogs and target pages to find out if the topics of source blogs and target pages were similar.

All the appreciation links were placed on blogrolls and link lists and they naturally showed a positive attitude towards target pages. 28.6% pointed towards other blogs and the rest pointed towards Web sites. Over half (58.9%) of the source blogs and target pages discussed similar topics. Of these, nearly half (48.5%) discussed different kinds of hobbies: handicrafts, reading, cooking, pets, or furnishing. Other topics that were common for both the source and target pages were fashion, everyday life, societal issues, politics, and new media. Blogs concentrating on these kinds of topics have also in earlier studies noticed to form quite closely connected communities (Wei, 2004; Chin & Chignell, 2006).

A clear majority of links (73.8%) expressed a positive attitude towards target pages. 24.4% of links expressed neutral attitude and 1.9% negative attitude. All the reciprocal, appreciation, advertisement, social, and social / reciprocal links were deemed to be positive because of the inherent nature of these linking reasons. When the reason for link creation was either presenting a content or information, the attitudes were most varied: links presenting a content entity or offering more information contained 71.8% of neutral links, two of the three negative links, but only 19.7% of all the positive links.

The three negative links were in two blogs that discussed and criticized societal issues. One link pointed towards a Web site of a Finnish political party and two links pointed towards a Finnish state-owned company. For every link the reason for creation was different: one pointed towards a content entity, one towards additional information, and one was pure courtesy link (even if the attitude was not politely positive). The negative attitude of each link was clearly visible in contexts of links.

Contexts of links

The majority of links situated in link lists (41.2%), and 20.0% of links were in blogrolls. Many blogs did not include both a blogroll and a link list and all the links in sidebars were compiled in one list. This explains why the proportion of blogroll links is low compared to some earlier studies (e.g. Ali-Hasan & Adamic, 2007). The proportion of links in postings was 18.8%. 16.2% of links occurred in side bars outside of a blogroll or a link list, and 3.8% of links occurred in upper and lower edges of pages.

The links situated in postings had slightly different characteristics than other links. Two of the negative links and nearly half of the neutral links (43.6%) occurred in postings, even if the amount of posting links was quite low. On the contrary, only 8.5% of the positive links were in postings. Mostly posting links were targeted towards a content entity (53.3%).

Discussion and conclusions

In general, bloggers in this study did not reveal much of their personalities. There were some bloggers that revealed even their home addresses and cell phone numbers, but they were definitively a minority. The exception was a blog of a women's magazine, in which all the contributors revealed their full names and faces. In the study of Herring et al. (2005b) bloggers were more open about their identities. Finnish bloggers may choose not to publish personal information because of the relatively small population of the country – other people may quite easily recognize a blogger based on information of their location, age, and contents of postings, even if the real name or a recognizable photo was not revealed.

Based on the results, bloggers' motivations for linking are as follows: Bloggers want to show their appreciation to the target page, or they think that its contents are somehow useful or interesting for themselves or blog readers. The attitude towards the target page is strongly positive. Neutral attitude is evident mostly on self-citations or when the link is created for conventional courtesy. Negative links are rare, but it can be assumed that especially socially conscious bloggers are linking also in a critical manner.

Links in blogs are mostly on blogrolls and link lists and links in postings are a minority. Over half of the links in postings cite some additional information. This demonstrates that especially links situated in postings are connected with contents of blogs and add some depth, breadth, or details to them.

Based on the answers for the research questions, the meaning of links in blogs can be defined as follows: Links interweave the contents of the blogosphere into the contents of the Web. Bloggers use linking as a way to keep in touch with the blogosphere and the Web. Links represent mostly bloggers' positive valuations and they are mostly used for providing information about the blogger and his/hers interests and appreciation. Even if links that are directly connected with contents of blogs are not created very often, they are a crucial element in blogging: they make discussions more diverse and enable the commentary nature that is essential for the blogosphere. Links in the blogosphere can be also negative, clearly showing that linking is a means for showing valuation, discussion, and social control.

A specific similarity between citations in academic publications and blog links seems to be that both can be done for negative, criticizing reasons: the negative links in this study had a close resemblance to negative citations in academic publications. Even though the amount of negative links in this study was small, it seems that they are an important part of linking in the blogosphere. It is not surprising that negative links occurred in k-logs that had a critical attitude towards societal issue – bloggers are expected to reveal opinions and valuations with links (Miller & Shepherd, 2004).

In future studies concerning linking in blogs negative linking should be taken into account already in the planning phase. In classification schemes aimed for academic citations this kind of negative category has been typical, whereas in the linking studies this kind of acting seems to be almost absent. Only Chu (2005) has previously identified negative links in the Web.

Even though nearly 80.0% of blogs in this study were diary blogs, only 7.5% of blogs did not include links pointing to other Web documents. In earlier studies about one fifth – one third of blogs have been without links (Henning, 2003; Herring et al., 2005b). The lack of non-linking blogs could be due to the fact that over the years linking has become more usual in diary blogs as well. Another possible reason is that blogs without links cannot be found with the technology that the blog listing service Blogilista uses to find blogs, and they thus were not included in the study sample.

In this study, differences in features and linking patterns between different blog types could not be identified because the proportions of filter blogs, k-logs, picture blogs, mixed blogs, and other blogs were so small. It would be reasonable to study different blog types separately. For example, in filter blogs and k-logs linking patterns can be different. Also company blogs are likely to have different linking patterns.

The reason for lack of news site links in this study can be explained with the great amount of diary blogs in the sample. It is however surprising that k-logs in this sample did not have links to news sites or Web communities, which could give wider perspective to the issue on hand. Aiming research on different link types could be fruitful: for example, studying links in postings could give more varied picture of links as one element of blog content. This could give more diverse picture of linking in the blogosphere and the importance of links as an element of blogs.

There are some reliability issues with the results of this study. If the aim is to study attitudes and reasons of link creation, interviews and/or questionnaires are highly recommended. However, that is seldom possible an option because of the number of links analyzed and the difficulties to find the right person that have created the links. Using multiple classifiers would improve the reliability of studies using link analysis. Researcher-based analysis has been proved to be the best for introductory studies and for creation of a general picture, but in the following studies also insights of link creators should be used (Wilkinson et al., 2003). The validity of the study was evaluated through the suitability of the used classification schemes. The scheme for reasons for creating links was least in congruence with the material of this study because it was the only classification scheme that was not based on earlier studies of the blogosphere. In the future studies concerning links in the blogosphere or other sub-spaces of the Web the classification should be based on the material or the classification scheme should be formed according to studies of that particular sub-space. However, similar results have been gained from earlier studies concerning links in the blogosphere, which supports the reliability and validity of this study, as well as indicates some level of consistency in linking behavior in the blogosphere. The results also widen the general picture of linking on the Web. The Web consists of several different types of documents and sub-spaces and studying them separately is the best way to add some light to them and the Web as a whole.

Acknowledgements

This paper is based on my master's thesis. I want to thank my supervisor PhD Terttu Kortelainen for her comments and support while writing the original text. I also want to thank researcher Kim Holmberg for his helpful comments on this paper.

References

- Ali-Hasan, N. F. & Adamic, L. A. (2007). Expressing social relationships on the blog through links and comments. Retrieved September 24, 2007 from: <http://www.icwsm.org/papers/2--Ali-Hasan--Adamic.pdf>.
- Almind, T. C. & Ingwersen, P. (1997). Informetric analyses on the World Wide Web: Methodological approaches to 'webometrics'. *Journal of Documentation*, 53, 404-426.
- Bar-Ilan, J. (2005). What do we know about links and linking? A framework for studying links in academic environment. *Information Processing and Management*, 41, 973-986.
- Berners-Lee, T. (2000). *Weaving the web: the original design and ultimate destiny of the World Wide Web by its inventor*. New York: Harper Collins.
- Björneborn, L. & Ingwersen, P. (2001). Perspectives of webometrics. *Scientometrics*, 50, 65-82.
- Blogilista.fi (2008). Etusivu [Front page]. Retrieved March 15, 2008 from <http://www.blogilista.fi>.
- Blood, R. (2000). Weblogs – a history and perspective. Retrieved March 17, 2007 from: http://www.rebeccablood.net/essays/weblog_history.html.
- Brin, S. & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30, 107-117.
- Chin, A. & Chignell, M. (2006). Finding evidence of community from blogging co-citations: A social network analytic approach. In P. Kommers, P. Isaías & A. Goikoetxea (Eds.), *Proceedings of the IADIS International Conference on Web Based Communities* (pp.191-200).
- Chu, H. (2005). Taxonomy of inlinked Web entities: What does it imply for webometric research? *Library & Information Science Research*, 27, 8-27.
- Clyde, L. A. (2004). *Weblogs and libraries*. Oxford: Chandos Publishing Limited.
- Cohen, E. & Krishnamurthy, B. (2006). A short walk in the Blogistan. *Computer Networks*, 50, 615-630.
- Davison, B. D. (2000). Topical locality in the web. In N. J. Belkin, P. Ingwersen & M.-K. Leong (Eds.), *Proceedings of the 23rd Annual International Conference on Research and Development in Information Retrieval (SIGIR 2000)*, (pp. 272-279).

- Haas, S. W. & Grams, E. S. (1998). Page and link classifications: Connecting diverse resources. In I. Witten, R. Akscyn & F.M. Shipman, III (Eds.), *Proceedings of the 3rd International ACM Conference on Digital Libraries*, (pp. 99-107). New York: ACM Press.
- Herring, S. C., Kouper, I., Paolillo, J. C., Scheidt, L. A., Tyworth, M., Welsch, P., Wright, E. & Yu, N. (2005a). Conversations in the Blogosphere: An analysis "from the bottom up". In R.H. Sprague, Jr. (Ed.), *Proceedings of the Thirty-Eighth Hawaii International Conference on System Sciences (HICSS-38)* (11 pages). Los Alamitos: The IEEE Computer Society Press.
- Herring, S. C., Scheidt, L. A., Wright, E. & Bonus, S. (2005b). Weblogs as a bridging genre. *Information Technology & People*, 18, 142-171.
- Henning, J. (2003). The blogging iceberg – Of 4.2 million hosted weblogs, most little seen, quickly abandoned. Retrieved February 2, 2007 from: <http://www.perseus.com/blogsurvey/thebloggingiceberg.html>.
- Holmberg, K. & Thelwall, M. (2007). Local government Web sites in Finland: A geographic and webometric analysis. In D. Torres-Salinas & H. F. Moed (Eds.), *Proceedings of ISSI 2007. 11th International Conference of the International Society for Scientometrics and Informetrics*, (pp. 378-386). Madrid: CINDOC & CSIC.
- Kilpi, T. (2006). *Blogit ja bloggaaminen*. Jyväskylä: Gummerus Kirjapaino Oy.
- Kim, K. (2004). The motivation for citing specific references by social scientists in Korea. The phenomenon of co-existing references. *Scientometrics*, 59, 79-93.
- Kumar, R., Novak, J., Raghavan, P. & Tomkins, A. (2004). Structure and evolution of blogspace. *Communications of the ACM*, 47, 35-39.
- Miller, C. R. & Shepherd, D. (2004). Blogging as social action: a genre analysis of the weblog. In L.J. Gurak, S. Antonijeivic, L. Johnson, C. Ratliff & J. Reyman (Eds.), *Into the blogosphere: Rhetoric, community, and culture of weblogs*. Retrieved February 2, 2007 from: http://blog.lib.umn.edu/blogosphere/blogging_as_social_action_a_genre_analysis_of_the_weblog.html.
- Oppenheim, C. & Renn, S. P. (1978). Highly cited old papers and the reasons why they continue to be cited. *Journal of the American Society for Information Science*, 29, 225-231.
- Orlikowski, W. J. & Yates, J. (1994). Genre repertoire: The structuring of communicative practices in organizations. *Administrative Science Quarterly*, 39, 541-574.
- Pietilä, V. (1976) *Sisällön erittely*. Helsinki: Gaudeamus.
- Sirkkunen, E. (2006). Blogit ovat evoluutiota eivät revolutiota. In *Journalismikritiikin vuosikirja 2006* (pp. 51-61). Vammala: Vammalan kirjapaino Oy.
- SocSciBot (2007). SocSciBot 3. Link crawler for the social sciences. Retrieved December 5, 2007 from: <http://socscibot.wlv.ac.uk>.
- Thelwall, M. (2002). The top 100 linked-to pages on UK university web sites: high inlink counts are not usually associated with quality scholarly content. *Journal of Information Science*, 28, 483-491.
- Thelwall, M. (2003). What is this link doing here? Beginning a fine-grained process of identifying reasons for academic hyperlink creation. *Information Research*, 8. Retrieved September 24, 2007 from: <http://informationr.net/ir/8-3/paper151.html>.
- Thelwall, M. (2004). *Link analysis. An information science approach*. San Diego: Elsevier Inc.
- Thelwall, M. (2006). Interpreting social science link analysis research: A theoretical framework. *Journal of the American Society for Information Science and Technology*, 57, 60-68.
- Wei, C. (2004). Formation of norms in a blog community. In L.J. Gurak, S. Antonijeivic, L. Johnson, C. Ratliff & J. Reyman (Eds.), *Into the blogosphere: Rhetoric, community, and culture of weblogs*. Retrieved February 2, 2007 from: http://blog.lib.umn.edu/blogosphere/formation_of_norms.html.
- Weinstock, M. (1971). Citation indexes. In A. Kent & H. Lancour (Eds.), *Encyclopedia of library and information science vol. 5* (pp. 16-40). New York: Marcel Dekker.
- Wilkinson, D., Harries, G., Thelwall, M. & Price, L. (2003). Motivations for academic web site interlinking: evidence for the Web as a novel source of information on informal scholarly communication. *Journal of Information Science*, 29, 49-56.