BLOGGING FOR ENGINES

Blogs under the Influence of Software-Engine Relations

Name: Anne Helmond
Student number: 0449458
E-mail: anne@annehelmond.nl
Blog: http://www.annehelmond.nl

Date: January 28, 2008
Supervisor: Geert Lovink
Secondary reader: Richard Rogers
Institution: University of Amsterdam
Department: Media Studies (New Media)

Keywords

Summary

This thesis proposes to add the study of software-engine relations to the emerging field of software studies, which may open up a new avenue in the field by accounting for the increasing entanglement of the engines with software thus further shaping the field.

The increasingly symbiotic relationship between the blogger, blog software and blog engines needs to be addressed in order to be able to describe a current account of blogging. The daily blogging routine shows that it is undesirable to exclude the engines from research on the phenomenon of blogs. The practice of blogging cannot be isolated from the software blogs are created with and the engines that index blogs and construct a searchable blogosphere. The software-engine relations should be studied together as they are co-constructed.

In order to describe the software-engine relations the most prevailing standalone blog software, WordPress, has been used for a period of over seventeen months. By looking into the underlying standards and protocols of the canonical features of the blog it becomes clear how the blog software disperses and syndicates the blog and connects it to the engines. Blog standards have also enable the engines to construct a blogosphere in which the bloggers are subject to a software-engine regime. The daily blogging practice brings users directly into the disarray of software-engine politics as illustrated in the case of spam and nofollow.

FOREWORD

For more than ten years now I have maintained a homepage in several shapes and forms. I initially used simple What You See Is What You Get (WYSIWYG)-editors such as HotDog and CoffeeCup but I gradually taught myself HTML. In these ten years I have several times completely changed the form and the looks of my homepage. This was always a major operation which I did not quite look forward to. Not only did I have to determine the looks and the layout but I also had to adjust all the pages manually. Just simply changing of a font could mean hours and hours of work. This aspect has changed thanks to the introduction of Cascading Style Sheets (CSS) which separates content from presentation.

In 1996 I created my first homepages which I remember placing on De Digitale Stad (DDS) and the free webhosting service Geocities. Unfortunately, there are no traces of these homepages left. In 1997 I created my second Geocities homepage which is still online as the "Unofficial Eric's Trip Homepage" which is dedicated to a Canadian band. It was a tribute website to one of my favorite bands at the time (see illustration 13, p. 117).

In 1998 I created a new personal homepage which I hosted at a friend's domain who "rented" a small web space plus e-mail address for only five guilders a year. This homepage contained a collection of my drawings, poetry, photos, sound files, high school essays and contained a typical "about" page and a "link" page (see illustration 14, p. 117). It was a classical personal homepage as described by John Seabrook in 'Home on the Net' for the New Yorker in 1995:

> What exactly is a home page? In the simplest terms, it is like an E-mail address, a place on the Net where people can find you; but whereas an E-mail address is just a mailbox, a home page is a front porch. Although building home pages or Web sites (a collection of pages of which the home page is the natural starting point) is mainly a commercial enterprise, it doesn't have to be. It's also a way to meet people. You want guests to have a good time when they visit your home page, and you hope they will take away a favorable impression of you. You can link your home page to the home pages of friends or family, or to your employer's Web site, or to any other site you like, creating a kind of neighborhood for yourself. You can design your page in any way you wish, and furnish it with anything that can be digitized--your ideas, your voice, your causes, pictures of your scars or your pets or your ancestors. You can have a professional photographer come and take pictures of your apartment, and then upload those pictures onto the Web, or you can just describe your day. (Seabrook, 1995)

Around 2000 my friend's domain ceased to exist and I moved my homepage to the student server of the Utrecht School of Arts. While studying Interaction Design at the Utrecht School of Arts I updated my website only once in these four years to put a small portfolio online (see illustration 15, p. 118).

Although designing websites was part of this study it took too much time and effort to frequently update my own. My website was one of the many sites online that seemed to bear the permanent 'under construction' label. It was a place that was never finished and that constantly needed work or was worked on. Nowadays the 'under construction' banners and labels seem to have vanished which may point to a changing attitude towards websites as objects that contribute to the perceived freshness fetish on the web. (Helmond, 2007b)

Between 1996 and 2005 my homepage was always moving from one (free) server to an other until I registered my own domain in 2005[1]. When I started the New Media Master at the University of Amsterdam (UvA) in 2006 one of our first assignments was to start a collaborative blog. This project resulted in our Masters of Media weblog for which we initially used a WordPress.com account. This was an easy way to quickly set up our blog and start blogging. At the same time I brought my homepage back to life by replacing it with a blog and used it as a research blog for my MA thesis.

I technically set up both blogs, but the Masters of Media blog was originally hosted at WordPress.com while my own blog was hosted at my own web server. The main difference I have experienced is that hosting WordPress on your own server allows you a greater degree of freedom in changing items and adding items to your blog. This eventually led to the decision to move the Masters of Media blog from WordPress.com to a UvA server.

My homepage is now a blog that is updated regularly. It is different from my previous homepages both in content and in form. My first homepage served as an online representation of the self and it was a self-referential closed space with only a few links to other pages on the 'links' page. It did not allow for any feedback and my e-mail address was only present on two pages. My second homepage was a portfolio that showed my work and contained only a few links to project pages. My e-mail address was more prominently visible on the homepage but I do not remember having

---

[1]   In 2003 it became possible for private individuals to register a .nl country code top-level domain (ccTLD). Until then this was restricted to companies or individuals registered with the Kamer van Koophandel (Chamber of Commerce).

ever received any e-mail regarding my website.

My blog is very different from my previous homepages in various aspects. It still contains an 'about' page and a portfolio space consisting of a 'research papers' page and a 'photos' page. Whereas these photos were previously hosted at the same space as my homepage, they are now hosted at an external space, namely Flickr. The photos on my blog are linked from my Flickr account. My homepage consisted of a few pages and a set of sub pages. The content was pretty timeless and updating consisted of adding a new page or subpage. Updates were invisible as I did not use any markers such as the popular 'new' gif to indicate updated content. The "Unofficial Eric's Trip Homepage" used a 'Last Modified' JavaScript to show when the website was last updated. It did not show which specific parts were updated.

Blog software not only enables me to easily update my website but it also adds new content in the form of blog posts which are automatically timestamped and organized in reverse-chronological order. The latest content is always visible at the top. My blog revolves around blog posts that mainly deal with the subject of new media. The blog posts contain a large number of links to other relevant blog posts and websites and are open for commentary.

In contrast to my homepage, my blog links to other blogs and receives links from other blogs. Every blog post is open to feedback from readers in the comment section. My blog is not the closed and self-referential space my homepage once was.

It seems that once you have installed and configured WordPress on your server it hasn't even quite begun. The process of building and shaping your blog never seems to end. After the basics such as choosing a theme to give your blog a certain look there are still hundreds of plugins and widgets to 'accessorize' (VanFossen, 2006a) your blog and every blog post contributes to the blog.

It does not end there. The blog is caught in a web of relations and engines that actively or passively shape the blogging experiences. With the intense process of linking I have also become particularly interested in who links to me and where my visitors come from. I am subscribed to various services and use various tools and plugins in an attempt to grasp a glimp of the network my blog resides in. These services and tools show that the engines form an important part of this network. Technorati and Google have become major actors in the discourse on blogging in the blogosphere itself but not so much in the academic sphere. This has led me to further investigate how the blog has become increasingly entangled in a web of relations and

specifically with the engines.

This thesis was born of my personal blog building and writing experiences over the past year. Even though I am relatively new to blogging and have only been blogging for seventeen months a lot has happened since. By actively participating in the blogosphere by commenting on other blogs I was approached by Tony Hung, former editor of the Blog Herald. I now write a weekly post on blogging, the blogosphere and blog software for a leading blog with a major audience. The technical and social aspects of blogging and their relationship to the engines have fascinated me from the beginning and have led me to explore this issue further in this thesis.

It feels strange to write a thesis when I have gotten so used to the blog format. Writing about blogging in word processing software is a completely different writing experience. Consciously or unconsciously I constantly want to place a link to a blog post or website and simply replacing it with a footnote is simply not the same. A PDF or any other document format is fixed once it has been printed and handed in. It does not allow the direct blog feedback in the form of trackbacks and comments. A thesis as a blog allows for creating internal and external links and would show the workings and idiosyncrasies of the medium.

Instead of handing in my URL[2] I would like to invite you to explore the links in this document and visit my blog to see the workings of the medium.

Acknowledgements

---

2   http://www.annehelmond.nl

TABLE OF CONTENTS

INTRODUCTION

As of January 2008 the blog engine Technorati tracks over 112 million blogs with 175,000 new blogs being created every day[3] and

> While the growth rate of the blogosphere seems to have slowed down recently (Sifry, 2007), these numbers are remarkable even if one considers that a substantial number of blogs might be abandoned after a couple of weeks or months or are created for the sole purpose of search engine optimization. (Schmidt, 2007)

Research conducted by the Pew Internet & American Life Project report showed that in 2006

> Eight percent of internet users, or about 12 million American adults, keep a blog. Thirty-nine percent of internet users, or about 57 million American adults, read blogs – a significant increase since the fall of 2005. (Lenhart & Fox, 2006)

Blogs are especially popular among American teenagers of which "28% have created their own online journal or blog, up from 19% in 2004." (Lenhart et. al., 2007). The number of blogs is still growing and blogs have become a very popular and increasingly important medium over the past decade (Sifry, 2007).

The phenomenon of blogs is often studied from the perspective of use and impact. By looking at blogs from mainly this perspective "we are in danger of always dealing only with effects rather than causes; the output that appears on a computer screen rather than the programs and social cultures that produce these outputs." (Bristow, 2007) There is a need to focus on how blogs and the blogosphere are constructed by looking at blog software and blog engines[4]. This is an often overlooked aspect in current research that needs to be addressed in order to gain more understanding into the phenomenon of blogs.

This thesis wishes to contribute to the growing amount of research on blogs by addressing it from a different angle focussing on the increasing symbiotic relationship between blogger, blog software and blog engines.

---

3   http://www.technorati.com/about/ (accessed 13 September 2007).
4   Previous research in this line has been done by Rebecca Blood in 'How Blogging Software Reshaped the Online Community' (2004) which focussed on how Blogger changed the blogosphere (see paragraph 2.2.3).

1. What Is a Blog?

The coinage of the word weblog is attributed Jorn Barger who started 'logging' interesting 'websites' on his Robot Wisdom WebLog as of December 1997. Peter Merholz shortened weblog to 'blog' when

> Sometime in April or May of 1999 (I can't say for sure when I exactly did it), I posted, in the sidebar of my homepage:
>
> "For What It's Worth
> I've decided to pronounce the word "weblog" as wee'- blog. Or "blog" for short."
>
> I didn't think much of it. I was just being silly, shifting the syllabic break one letter to the left. I started using the word in my posts, and some folks, when emailing me, would use it, too. I enjoyed it's crudeness, it's dissonance...
> (Merholz, 2002)

The word blog has been adapted considerably since and is now the generally accepted term. There is no single definition of a what a blog is and bloggers, academics, blog software and blog engines define blogs differently. Researcher and blogger danah boyd recognizes the problems that arise when trying to settle for a final definition and

> rather than arguing for a definitive definition, this paper invites scholars to conceptualize blogging as a diverse set of practices that result in the production of diverse content on top of a medium that we call blogs. (boyd, 2006)

Instead of seeing blogs as a form of technology or as a form of cultural activity danah boyd calls for seeing blogs as a medium and a diverse social practice:

> By reconceptualizing blogs as a medium and bi-product of practice, it becomes possible to understand the diverse intentions that produce diverse output and analyze blogs even when the output itself is inconsistent in terms of style and content. (boyd, 2006)

This reconceptualization rightly acknowledges the diverse intentions and outputs of blogging practices. In the current discourse blogs are often defined from a genre point of view or through their form or structure. This leads to conceptualizations of the blog as a diary (Nardi et al. 2004, p. 44), a new form of journalism (Bruns, 2005 & Gillmor, 2006 & Hewitt, 2006), a personal publication tool ('Push-Button Publishing' Blogger.com), a reverse-chronological ordered collection of entries (Evans in Turnbull, 2001) or a product of blog software[5] (Conniff, 2005). However, problems

---

5   "We don't have an official definition," says a spokesperson for Technorati, the blog tracking service. "It's something that's created with blog software. I don't know how to answer that question. We don't get that question. (Conniff, 2005)

arise when the output is so diverse that some blogs do not fit in a neatly defined category. Another approach is to look at blogs from a network perspective:

> This includes studies looking at the information flow within the blogosphere (Adar & Adamic, 2005; Adar, Adamic, Zhang, & Lukose, 2004; Gruhl, Guha, Liben-Nowell, & Tompkins, 2004), as well as structural analysis of emerging networks (Herring, Kouper, et al., 2005), sometimes within topical clusters such as the war blogosphere (Tremayne, Zheng, Lee, & Jeong, 2006) or with respect to certain blog providers (Schmidt, forthcoming). While many of these works focus on the English-speaking blogosphere, a growing number of studies examine other country- or language-specific blogospheres (Esmaili, Jamali, Neshati, Abolhassani, & Soltan-Zadeh, 2006; Merelo, Orihuela, Ruiz, & Tricas, 2004; Schlobinski & Siever, 2005; Trammell, Tarkowski, Hofmokl, & Sapp, 2006) or blog networks within certain hosting platforms, such as LiveJournal (Herring et al., 2007), twoday.net (Schuster, 2004), and Wallop (Lento, Welser, Gu, & Smith, 2006). (Schmidt, 2007)

While the network perspective takes into account that blogs are part of a network, the blogosphere, it often lacks criticism of other important actors in the network, in this case the engines. A final problem the analysis of blogs faces, is

> [...] that they are both the product of blogging and the medium through which the blogger produces their expressions. Blogs emerge because bloggers are blogging. And yet, what they are blogging to is the blog itself. (boyd, 2006)

I will adhere to danah boyd's reconceptualization of blogs "as a medium and bi-product of practice" (boyd, 2006) but I wish to elaborate further on the meaning of the terms 'medium' and 'practice'.

1.1 The Blog as Medium

danah boyd defines 'medium' in the context of communication studies in which "a medium is the channel through which people can communicate or extend their expressions to others." (boyd, 2006) Seeing the medium of blogs merely as a channel of communication is not sufficient for our understanding of blogs. Two other meanings of the word medium could add an extra layer to this understanding. Firstly a medium is "an intervening substance through which something else is transmitted or carried on" and secondly, a medium, in computer science, is "an object or device, such as a disk, on which data is stored."[6]

 The first and second added interpretations of a medium points to both the

---

6 http://dictionary.reference.com/browse/medium (accessed November 23, 2007).

browser and the blog software. The blog is usually retrieved, accessed and displayed in the browser.[7] The browser software acts as an intervening substance for the blog as the content of the blog is retrieved from the underlying database. This database is installed on a host server. The blog software itself also acts as an intervening substance for the database.

The third added definition highlights this often neglected aspect of the blog's database. The WordPress interface acts as a mask for the database. Our current understanding of blogs must take into account first the digitally native environment the blog resides in, the browser, second the software most blogs are created with and third the database that stores the content of the blog. This thesis will mainly focus on blog software and how it connects to the engines.

1.2 The Blog as Bi-Product of Practice

The blog as bi-product of practice is further complicated as a blog can no longer be defined as a singlehandedly constructed object. Instead a blog should be seen as co-produced bi-products from different actors with different intentions. The practice of blogging is expressed in various ways and is not acted out in isolation but within a web of relations. The practice of blogging often includes a daily blogging routine. In June 2007 I described my personal daily blogging routine as follows:

> I look at my blog daily, maybe even several times a day. I don't visit it to check if it's still there. I am quite sure it is still where I put it and I have enough confidence in my hosting company that it is there for anyone to see. I do it because it is part of my daily routine. When I turn on my computer my e-mail program and browser are loaded on startup. After reading my e-mail and my RSS subscriptions I look at my blog statistics. After viewing the stats, I check my comments in moderation, view my blog, check MyBlogLog readers in the sidebar, visit unknown readers and scroll back to the top of my page. This is my daily routine which ends with a confrontation with my latest post. (Helmond, 2007b)

The practice of blogging often includes reading other blogs, commenting, checking statistics and references and replying to comments. This part of the blogging practice is often not described in current discourses on blogging. Studying the blogging routine (see paragraph 4.1) reveals that blogs are increasingly a bi-product of practice under the influence of software-engine relations. This may be seen in the practice of using plugins and widgets in the sidebar to display customized data or content from

---

7   Blogs can also be read with a desktop or browser feed reader or even by e-mail.

third-party websites such as Flickr. This has become increasingly popular such a common practice that the well renowned ProBlogger blog actually added the task to 'declutter your sidebar' to its '31 Days to Building a Better Blog Challenge.' (Rowse, 2007)

Statistics (see paragraph 4.1.1) show that blog engines form a large part of our blogging practice as they are often used to search for sources and reference material. They also form an important entry point for blogs because the engines function as major traffic sources. The blogging practice is influenced by internal and external blogging forces in the blogosphere. Internal blogging forces consist of the drive for freshness caused by both the format of the blog and the fixation of keeping the content fresh by updating frequently. External blogging forces include the blog software and the blog engines that enable and constrain certain practices (Helmond, 2007b). Blogs are also the bi-product of a web-native practice.

1.3 The Blog as Native to the Web

Rebecca Blood describes in *the Weblog Handbook* that blogs are native to the web as

> The weblog, updated regularly, is designed to be visited again and again [...]
> The weblog is many things to many people, but it is, above all, a form that is
> native to the web. [...] The Web enables continual publishing, in which updates
> can occur at any time; it is this aspect of the Web that weblogs can capitalize on.
> (Blood, 2002, p. 9)

Blogs are not simply a form of remediation of the diary or the newspaper because

> "Everything about them — their format, their reliance on links, their immediacy,
> their connections to each other — is derived from the medium in which they were
> born. They are of the Web itself." (Blood, 2003, p. xi)

Web pages are at the core of the web and displayed through a web browser. The personal homepage is one of the many types of websites and typically consists of self generated pages containing content of a personal nature (Chandler, 2000.) Blogs are often seen as the dynamic and interactive successor of the static and self-referential personal homepage (Blood, 2004; Lovink, 2007) which blossomed in the mid nineties with the growing popularity of the Mosaic Browser (1993) and the Netscape Navigator browser (1994). However, blogger Cameron Barrett disagrees with blogs "rediscovering the power of a quality home page" as

> Home pages are places where you put pictures of your family and your cats. It's a

place to distribute information to a close circle of family and friends. Weblogs, however, are designed for an audience. They have a voice. They have a personality. Simply put, they are an interactive extension of who you are. (Barrett, 1999)

Personal homepages can be seen as relatively closed objects whereas blogs are increasingly connected to other blogs and the engines as a bi-product of embedded linking practices in the blog software. These links turn the blog into a more interactive extension.

The personal homepage revolved around the unit of the web page and was most popular in the mid 1990s. Meg Hourihan describes how blogs arose at a time of a more maturing web and are native to the web in their distinct way:

When the Web began, the page was the *de facto* unit of measurement, and content was formatted accordingly. Online we don't need to produce content of a certain length to meet physical page-size requirements. And as the Web has matured, we've developed our own native format for writing online, a format that moves beyond the page paradigm: The *weblog*, with its smaller, more concise, unit of measurement; and the *post*, which utilizes the medium to its best advantage by proffering frequent updates and richly hyperlinked text. (Hourihan, 2002)

Because of the networked and hypertextual environment blogs have originated and reside in, they are not static and closed objects. Blogs thrive on the distributed network of the internet and are highly dispersed across this network.

1.4 The Blog as Assemblage

The expanded definition of the blog as a medium and bi-product of practice shows that the blog is no longer an autonomous unit but is entangled in a web of relations. Blogs as native objects to the web with their hyperlinked, distributed, dispersed, and aggregative and syndicative character could be considered a Work as Assemblage:

[...] with the advent of the Web, communication pathways are established through which texts cycle in dynamic intermediation with one another, which leads to what might be called Work as Assemblage (WaA). (Hayles, 2005, p.105)

Seeing the blog as a Work as Assemblage allows for an analysis that takes into account the various 'texts' that constitute the blog. Examples of such texts are your Flickr photos, Upcoming.org calendar, latest Twitter message, MyBlogLog visitors or Google Maps.

[...] the WaA derives its energy from its ability to mutate and transform as it

grows and shrinks, converges and disperses according to the desires of the loosely formed collectives that create it. (Ibidem, p. 107)

The blog is able to transform and cycle different types of content and adjust them to the form of the blog. In its turn the blog is able to cycle its texts to different actors including the engines. In the 'Perceived Freshness Fetish' I argued that blogs "derive their energy from dispersion, aggregation and pinging." (Helmond, 2007b) These blog qualities inject the blog with a certain degree of vitality which in a case study actually and accidentally led to the vitalization of a dead blog.

In the Perceived Freshness Fetish I described how I wrote a blog post on 'I'm sorry blog excuses' and the phenomenon of dead blogs. Dead blogs are blogs that have been 'abandoned' by their authors and thus have not been updated for a while. By linking to these dead blogs I automatically sent out a pingback, a link notification. Some of the blogs had enabled auto-receival of pingbacks thus reinjecting some 'life' into the blog again. In one case the author of the 'dead blog' was notified of my pingback, and returned to blogging by writing a follow-up blog post in which he explained why he stopped blogging. The Work as Assemblage can take into account this dispersed and networked nature of the blog and the web of relations the blog resides in because

Moving fluidly among and across media, its components take forms distinctive to the media in which they flourish, so the specificities of media are essential to understanding its morphing configurations. (Ibidem, p. 107)

Blog elements can smoothly move across various media as the blog can be displayed on a mobile phone and the blog feed can be read in different media environments.

2. The Form of the Blog

The blog as a medium and bi-product of practice is shaped by both the medium and the practice. As blogs and blogging are not restricted to a narrow definition of both medium and practice it allows for a great variety of intentions and outputs. A blog is often defined by its form which is typically "frequently updated, reverse-chronological entries on a single Web page." (Blood, 2004, p. 53)

Over the years blogs have developed a very distinctive format (see illustration 1).
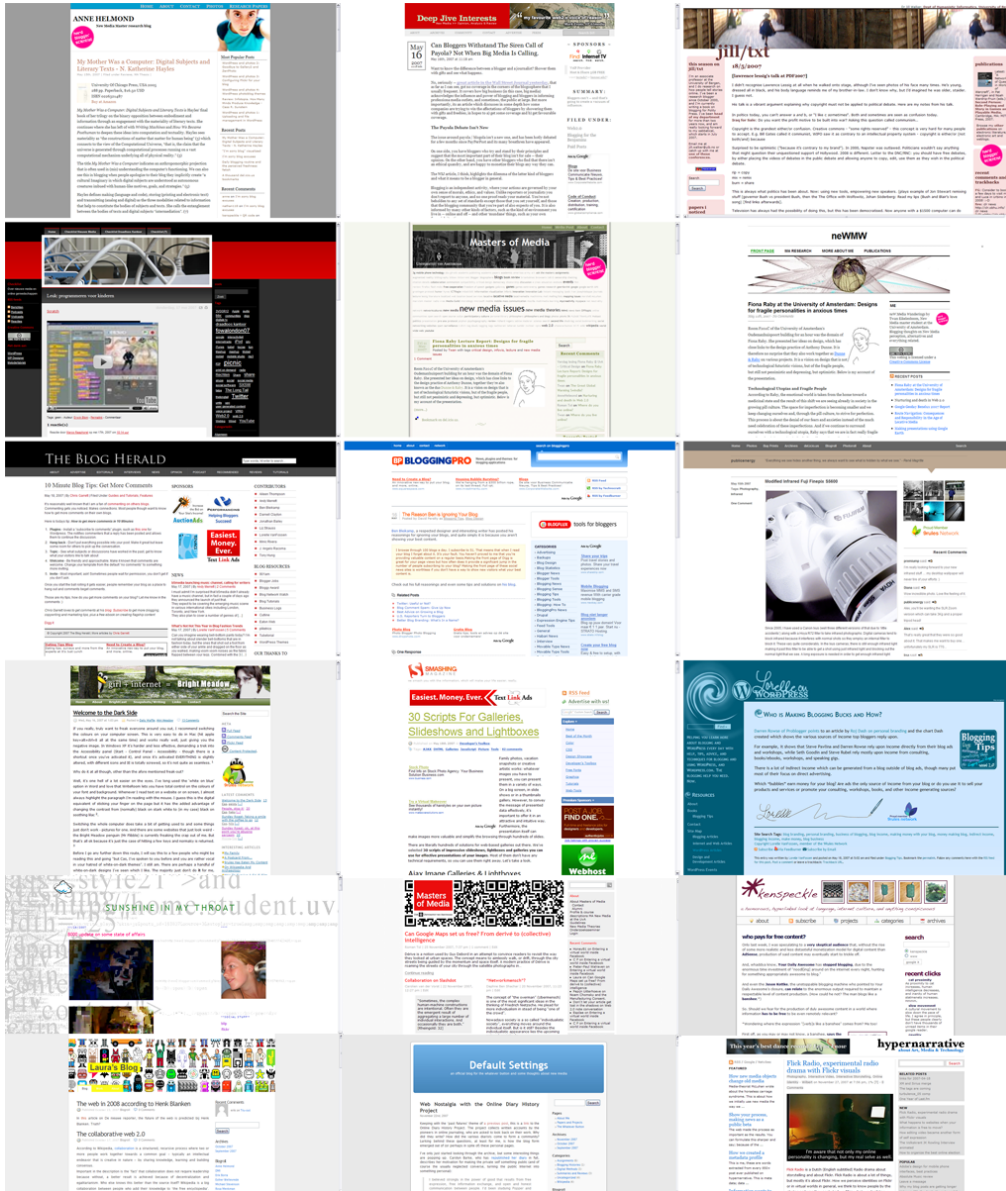
Illustration 1: The anatomy of blogs. (See DVD for detailled illustrations)

Taking the diverse outputs of the blog into account, deconstructing my personal blog reveals the anatomy of a typical blog: the header, navigation, content, sidebar and footer (see illustration 2.)

blog name →

← navigation

ANNE HELMOND

New Media Master research blog

← header

blog description →

post title →

My Mother Was a Computer: Digital Subjects and
Literary Texts - N. Katherine Hayles

← sidebar

← sidebar element

post category

post comments

post content →

post comments →

post title →
post date →
post category →
post comments →

post content

plugin /widget

plugin /widget

plugin /widget

plugin /widget

← sidebar

post comments →

navigation →

tagcloud →

← footer

Illustration 2: Anatomy of my blog, May 2007

The header consists of the top navigation of the different blog pages. It further displays the blog title, tagline and an image (logo). The content area consists of my latest three blog posts in reverse-chronological order. The sidebar shows the most recent comments, popular blog posts, categories, archives, a search function, a blogroll, a tag cloud and my two most recent Flickr photos. The footer displays the Creative Commons License my blog is licensed under. All these different parts collectively form the blog. However, as paragraph 3.2.1 will show, the blog post is considered to be the main item of the blog. The blog post is what really distinguishes the blog from the traditional web page. but it is the blog post that distinguishes it from the traditional website. Hourihan describes how the blog has moved beyond the old paradigm of the page and how it has developed its own natively digital object: the post. In 'What We're Doing When We Blog' she describes the anatomy of a blog post as follows:

> A weblog post can be identified by the following distinguishing characteristics: a date header, a time stamp, and a permalink. Oftentimes the author's name appears beneath each post as well, especially if multiple authors are contributing to one blog. If commenting is enabled (giving the reader a form to respond to a specific post) a link to comment will also appear. (Hourihan, 2002)

Five years later we can add a few new distinguishing characteristics in addition to the date, timestamp, permalink, author and commentlink, namely the category, the tag, the trackback link and the various social bookmarking links.

The comment link and trackback link are part of the blog post itself but the actual comments and trackbacks are not. They form an additional part of the blog post which are displayed on the same page of the blog post but they are not syndicated with the post (see paragraph 3.2.1) The page is another feature that resides outside of the regular blog chronology of reverse ordered posts and is not syndicated.

Pages were introduced by WordPress in version 1.5 on February 17, 2005.[8] The software makes a clear distinction between posts and pages in the interface with one tab for 'Write Post' and one for 'Write Page.' Although both tabs are part of the general 'Write' tab WordPress makes a rigid distinction between the two which is duplicated by the engines (see paragraph 3.2.1).

Posts are usually filed under one or more categories for categorization and archiving. It is a common practice to show the category a post is filed under in order

---

8   http://www.wordpress.org/development/2005/02/strayhorn/ (accessed November 12, 2007)

to supply the reader with a reference, and the opportunity to read older posts dealing with the same subject. Tags are added as meta information to the describe the content of the blog post. Usually either the categories or tags are displayed underneath the blog post title or underneath the post itself. They both serve navigational and search-driven purposes (VanFossen, 2006b) within the blog itself and within the engines.

If a blog is enabled to automatically receive trackbacks or pingbacks whenever someone references a particular blog post, a trackback or pingback will appear underneath the post. These link references typically appear above, in between or below the comments. It could be debated whether blog comments and blog pingbacks belong to the blog post itself or form an object of their own.

The final upcoming post characteristic is the social bookmarking link in the form of a plain link, icons and/or buttons, eg. 'ShareThis Link,' 'Add to del.icio.us,' or 'Submit to Digg,' 'Find related stories on Technorati' (see illustration 7, p. 69)

These additions to the anatomy of the blog (post) reveal an increasing relationship with blog engines. This particular structure of the blog and the anatomy of the blog post has consequences for the blog's relation to the engines.

3. Research Questions

This thesis wishes to address the increasing symbiotic relationship between blogger, blog software and blog engines. What is the relationship between blog software and the machines that index the blogosphere? How are blogs as a medium and bi-product of practice both enabled and constrained by software and engines? How do these software-engine relations relate to the blogger and influence one's blogging behavior?

4. Overview

Chapter one describes the relevance, approach and methodology used to answer these questions in relation to the emerging field of software studies. The next chapter is a history of blogs, blog software and blog engines that aims to describe several milestones in blogs that have contributed to the standardized blog format. It will also deal with the issue of software concentration that has left us with only a few dominant players in both blog software and blog engines. Chapter three deals with the

relationship between the blog software and the engines that index the blogosphere by focusing on the canonical features of the blog. Having used the most prevailing blog software package, WordPress, for a period of seventeen months, this relationship will be described from an expert point of view combined with insights from various bloggers. Chapter four describes the influence of the engines on the daily blogging routine and the practice of blogging under software-engine relations. It describes what happens if the daily blogging routine collides with the software-engines politics as illustrated by spam and nofollow.

1. SOFTWARE STUDIES

This thesis wishes to create a new starting point into the research on blogging by addressing the software-engine relations. An analysis of blogging starts with the blogger and blog software but these two do no longer account for blogging as it stands. Geert Lovink states that:

> The path to understanding blogs lies somewhere between an analysis of software functionalities and the early adopter culture that invented and shaped the blogosphere. (Lovink, 2007)

Lovink's path to understanding blogs deals with the relation between the blogger and the software. Analyzing the path between the functionalities of the software and the early adopter culture is an important first step in understanding blogs. However, as blogs as a medium and bi-product of practice have matured, the relation between the blogger and the software has changed as it now also includes the engines. If we describe Lovink's proposed analysis as the path between the user and the software then the path between the user and the engine and the path between the software and the engine could be added.

Software is more than a tool that constructs an object, the blog, as it is entangled in a web of relations. An understanding of blogs and the blogosphere should start and not end with an analysis of how blog software has helped to shape blogs and the blogosphere and vice versa. The understanding of blogs, blogging and the blogosphere lies in the recognition of the amalgamation of the blogger, the blog software and the blog engines:

> If, contra McLuhan, "A society is defined by its amalgamates, not by its tools"[2] then Office is an attempt to pre-empt this amalgamation by not only providing what rationalist programmers are content to describe merely as tools but also the paths between them, how they intermix, and the boundaries and correlations between their different functions, the objects they work on and the users that they amalgamate with. (Fuller, *Behind the Blip*, p. 138)

The blogosphere, as a society, is defined by its amalgamates and requires an analysis of the paths, boundaries and correlations of the blogger, the blog software and the blog engines. Therefore, I wish to propose to add the blog indexing and search engines to our understanding of blogs and the blogosphere. There is an increasingly symbiotic

relationship between the blogger, blog software and blog engines that needs to be addressed. This thesis focuses on the software and its relations to the engines and how these software-engines relations relate to the blogger.

1.1 Studying Software-Engine Relations

By starting from the native environment of the blog, the web, a new approach to studying blogs must be taken. Starting with the web means to start with Google as "the start page for the Internet" (Dodge, 2007) because "Google has become such a commonly used resource that people are beginning to regard it as synonymous with the Web." (Searls in Gudrais, 2007) Thus, on the web, an analysis of software "begins and ends with the engines." (Stevenson, 2008)

This research addresses the phenomenon of an increasing symbiotic relationship between blog software and blog engines. Studying software-engine relations deepens our understanding of blogs as well ass our understanding of how the blogosphere is constructed through engines. Blog software and blog engines have to be studied together because they are co-constructed. Blog software is often taken for granted in research on blogging but;

> Blog software is an important structural aspect of blogging practices, since the various functionalities built into the code frame, but not determine, actual use. It is also subject to ongoing change, either through innovations in the software itself, or through the (re-)combination of different tools. (Schmidt, 2006)

The broad diffusion of software in our current society has made it nearly inconspicuous:

> Software has quietly penetrated our daily life and in many cases it has also become invisible. The ubiquity and so-called transparency of software not only renders it invisible but at the same time points to the importance of studying it. Software is an understudied object as "Software is often a blind spot in the theorization and study of computational and networked digital media." (Fuller, forthcoming) (Helmond, 2007)

Lev Manovich addresses the importance of studying software in *The Language of New Media* (2001):

> To understand the logic of new media, we need to turn to computer science. It is here that we may expect to find the new terms, categories, and operations that characterize media that became programmable. From media studies, we move to something that can be called 'software studies' -- from media theory to software theory. (Manovich, 2001, p. 48)

New issues and questions arise from this important turn in the culture in which our media have become programmable. The current 'software society' requires a new field of study that addresses software as:

> a layer that permeates all areas of contemporary societies. Therefore, if we want to understand contemporary techniques of control, communication, representation, simulation, analysis, decision-making, memory, vision, writing, and interaction, our analysis can't be complete until we consider this software layer. (Bristow, 2007)

Methods for analysis are currently being developed, as the new field of software studies has not been properly defined yet. It is currently being established and shaped by the Software Studies Initiative at UCSD[9] that "intends to play the key role in establishing this new field." (Bristow, 2007) The Software Studies Initiative is led by Lev Manovich and one of its goals is to develop "methodologies and software tools for the study of software from humanities and social science perspectives." (Ibidem). The roots of software studies, however, can be traced back to various different fields of study:

> The coinage "software studies" belongs to Lev Manovich, but software studies grows out of the work of other provocative theorists as well, such as Matthew Fuller and Geert Lovink, and has antecedents in human-computer interaction (in the work of Donald A. Norman), sociology (Manuel Castells; Steve Woolgar), law (James Boyle; Lawrence Lessig), literary studies (Katherine Hayles), and media theory (Friedrich Kittler). (Truscello, 2005, p. 346)

These different fields of studies have contributed to software studies but criticism on one of its antecedents, Human Computer Interaction (HCI), is that it turns models into users and separates the users from the context and the processes they are involved in. Software is embedded in social practices and relations and there is a need to "inject history and culture into the debate over the production of software and its materiality, a debate that thus far has focused on reception" (Truscello, 2006, p. 183) Software studies seems to be mainly concerned with software-user and software-production relations.

However, in current criticism the role of the software-engine relations is missing. Engines play a central role on the web and have a great impact on both user and producer of blog software. On top of that they construct the blogosphere through the characteristic features of the blog as standardized by blog software. The engines themselves also have a culture, aesthetics and politics embedded that is influenced by

9   The University of California, San Diego.

and in its turn influences software. The politics of the search engines do not only apply to the web itself (Introna and Nissenbaum, 2000 & Rogers, 2004) but also to software on the web. Including the engines into an analysis of (web-based) software would open up a new avenue in the understanding of software and the defining of the emerging field.

Software studies is steadily shaping itself through a growing amount of 'software theory' in which the upcoming lexicon *Software Studies* edited by Matthew Fuller plays a key role. "Fuller previously addressed the topic of software and the culture of software in his collection of essays in *Behind the Blip*. In the opening essay Fuller argues for a 'software criticism' that should move authors away from detailing accounts of the functionality of a particular piece of software. Instead of dealing with the performance of software a more critical approach should be applied, one that is concerned with the question how software consists of different elements that are embedded in a dynamic web of relations." (Helmond, 2007c) However, research into the creation and adaptation of specific functionalities of software within the developer and the user community may contribute to our understanding of the idiosyncrasies of the software. By detailing the core functionalities of WordPress its relationship with the engines can be explored and described.

This thesis strives to contribute to the growing field of software studies by creating a new critical pathway **into and out of** blog software. By criticizing WordPress as the dominant standalone blog software this thesis wishes to contribute to the understanding of blogs by focusing on their relationship with the engines. By describing software-engine relations it wants to create a pathway **out of** software **into** the engines thus further shaping the field of software studies.

1.2 Object of Study

Matthew Fuller is looking for a way forward to study software by the means of a theorized practice and he calls for

> Theorisations of software that are able to operate on the level of a particular version of a program, a particular file structure, protocol, sampling algorithm, colour-scheme, public interface, request for comments, and so on, are necessary. Further, it is essential to understand any such element or event as only one layer or node in a wider set of intersecting and multi-scalar formation. (Fuller, *Behind the Blip*, p. 17-18)

This thesis will study WordPress version 2.3.1 as released on October 26, 2007. Adrian Mackenzie describes software as a very mutable object that is entangled in a web of relations (Mackenzie, 2006, p. 19). WordPress turned out to be a very mutable object as it released several new versions while conducting research. When I started my blog and research on September 25, 2006 I used WordPress version 2.0.4 and thirteen months later no less than sixteen updates have been released[10]. This mutability sometimes makes it a difficult object of study because functions described may have been changed or removed in a recent update.

Updates in the WordPress software do not only affect the interface or functions of the software. These updates occur within a web of relations and may affect other parts of the software such as plugins, WordPress' main requirements. These updates may also affect their relation to the blogger and the blog engines. This is a general problem that the field of software studies is faced with.

To be able to refer to specific notions of the software this thesis describes WordPress version 2.3.1 as released on October 26, 2007 unless noted. I initially settled for version 2.1.3 as released on April 3, 2007 which I ran up till November 18, 2007. Two major new versions, 2.2 and 2.3, did not only include several security fixes but also major updates such as integrated widgets and tagging. Not only did the release of version 2.3 render the development of some of my plugins obsolete it also changed the software's relation to the engines. It turned out to be nearly impossible to study a particular version of a program as the software was entangled in a web of relationships that pushed me towards an upgrade. The forced upgrade culture is at full strength and full speed ahead.

The object of study is not an online hosted blogging service such as WordPress.com or Blogger.com because such blogging services are relatively limited and closed. Generally, they do not allow direct access to the code and they cannot be taken apart like the standalone WordPress software. WordPress 2.3.1 is released under the GNU Public License that allows for changes and additions in the software with the requirement that the derived work is released with the same GNU Public License. WordPress is an Open Source project that allows you to download, view and adjust the to your own needs. This enables you to take it apart and expose the several layers the software consists of showing how they relate, entangle, constrain and

---

10   http://www.wordpress.org/download/release-archive/ (accessed 5 September 2007).

empower.

In order to study the software-engines relations and how they relate to the blogger this thesis focuses on two of the main engines in the blogosphere: Google and Technorati. The engines are even more mutable objects of study as they constantly update their indexes, adjust their algorithms and APIs. Such engine 'updates' are usually fairly invisible but may heavily influence the relation with blog software and blogger.

1.3 Approach and Corpus

I started using WordPress as a part of the MA course 'New Media Practices' in September 2006. I blogged my course assignments and wrote about New Media events and issues. The 'tagline' of my blog *New Media Master research blog* refers to my double position both as a blogger and as a researcher. I am part of and entangled in the web of relations I study.

I also blog at the collaborative Masters of Media blog, a student blog for the New Media Masters at the University of Amsterdam. From September 2006 to September 2007 I was responsible for technically setting up the blog and maintaining it. By actively contributing to and participating in the blogosphere I was able to shape my ideas. My position in the blogosphere has changed in the past year, from September 2006 to December 2007. I was contacted by the editor of the Blog Herald if I wanted to write for them on blogging, blog software and the blogosphere. My active participation in the blogosphere thus has led to a job as a so-called 'pro-blogger' for a major blog in the blogosphere. In October 2007 I published my first post on 'The Blog as Database.' Writing for my personal blog and writing for the Masters of Media and the Blog Herald collaborative blogs were completely different experiences. Not only has my position in the blogosphere allowed me to shape my ideas, but also my position and ideas have changed over the past year. As an active participant in the entangled web of relations I am constantly shaping and adjusting my ideas.

A part of my daily blogging routine is keeping up with my blog subscriptions. Over the past year I have read thousands of blog posts to keep up with the latest news about blogging, blog software, the blogosphere and blog engines (see appendix 1 for a list of my daily subscriptions). I have also subscribed to a number of related

mailinglists such as the Nettime-NL, Nettime-L and The Internet Research Association. I have frequently used The Internet Archive's Way Back Machine, the WordPress Development Blog, the WordPress Trac System[11], the WP Hackers Mailinglist and various WordPress developers' blogs for historical and technical reference purposes. Social networking applications and sites such as Twitter and the 9Rules Community also proved to be interesting resources. A final resource are the numerous conferences on New Media and the New York City WordPress November Meetup I attended. The corpus of this study thus consists of the trac tickets of the WordPress Trac System, the e-mails of the WP Hackers Mailinglist, the previously mentioned websites, the blogs listed in appendix 1 and and a number of WordPress themes and plugins. These plugins and themes are selected from the thousands available on the criteria of popularity and connections to the engines and in the case of the McFarland themes on extraordinary database use.

1.4 Methodology

In *Software Studies* Fuller has not gone as far as to attempt to start a new field of study but instead he calls for new theorizations of software. The entries of the lexicon all bear the marks of the different paradigm the authors reside in and that have their own typical methodologies. My methodology for studying WordPress is very diverse and was roughly developed during practice. In an attempt to define the approach I would propose to call it a 'Fulleresque' software studies approach which consists of a materialistic theorized practice in the line of Matthew Fuller and Alexander R. Galloway.

This thesis focuses on blog software and its relations to the engines and how these software-engines relations relate to the blogger. By starting with a history of blogs, blog software and blog engines several milestones can be described that have contributed to our shaping and perception of both the medium and practice of blogs. The analysis takes the canonical features of the blog as a starting point for a genealogical description. The historical account illustrates the rich history behind the typical features of the blog that have allowed the blog to connect to the engines and

---

11   "Trac is the place to follow along with the development of WordPress. You can track changes in the Timeline section of this site. [...] Bugs reports are also kept on Trac." http://trac.wordpress.org/wiki (accessed January 6, 2008)

vice versa. Looking into these connections starts with a deconstruction of the software exploring the different layers of the software that constitute the medium of the blog. The second step is taking the canonical features of the blog, looking at the underlying standards and protocols and seeing how they connect to the engines. The third step is critically looking into these connections through the eyes of the daily blogging practice. In the final step the impact of the software-engine relations on both the blogger and the blogosphere is analysed.

1.5 Limitations

The field of blogs is very diverse and incredibly large. Blog software has become such a popular web publishing tool that it is often used for other practices than blogging. The line between websites, homepages and blogs is blurring and social networking sites such as Facebook and MySpace now also offer the possibility to create a blog.

This thesis deals specifically with self hosted blogs created with WordPress software. The software, engines and blogs described here are mainly English-language (North American) products and tools. The scope of this research concerns the English language- North American blogosphere, excluding, for example 72.82 million Chinese blogs (CNNIC, 2007).

Even though I am Dutch I do not feel part of the Dutch blogosphere because I choose to blog in English and mainly link to English sources. I do read a few Dutch blogs and recognize Frank Schaap's relation with the Dutch blogosphere:

> I must confess that much of the vibrant Dutch weblog community remained hidden from me for quite a while, even though I am Dutch, have my own weblog, and frequently use the Internet. I was aware of some of the more high profile Dutch weblogs, but as an early adopter and Internet researcher, my attention was mostly trained on the English-speaking part of the Internet. Few native English weblogs link to non-English weblogs in their blogroll and those English language weblogs that do link to non-English weblogs are usually written by non-native English speakers.[1] Conversely, Dutch weblogs almost exclusively link to other Dutch weblogs in their blogroll, which only occasionally includes English language weblogs. In their entries, Dutch weblogs do link to English language web sites and copy or translate their news but the links of affiliation remain overwhelmingly with other Dutch weblogs. The Internet may be transnational but many communities remain bound by barriers of language. (Schaap, 2004)

I have mainly read American blogs and comment on English-language blogs. Being aware of the fact that this is just a tip of the ice berg of the blogosphere it allows me to limit the scope of my research. By focusing on the most popular American blogging

tools and engines, specifically WordPress and Technorati and Google Blog Search, any non-English blog software[12] and engines[13] are excluded from this research.

In order to be able to describe the relationship between the software and the engines I have chosen WordPress version 2.3.1 as the object of study. Detailed research into bloggers under the influence of software-engine describes a particular kind of blogger. A blogger who is aware of both the default settings and the specific settings of WordPress which a standard user may not be aware of.

---

12 In the Netherlands the Dutch blog software Pivot is still quite popular.
13 In China the Chinese search engine Baidu is far more popular than Google.
    http://chinatechstory.blogspot.com/2007/09/cnnic-reports-baidu-with-745-market.html
    (accessed January 20, 2008)

2. THE HISTORY OF BLOG SOFTWARE AND BLOG ENGINES

The history of blogs and blog software is a very contested one. There are numerous debates on where and how blogs originated, what was the first blog and who developed the first blog software[14]. These often heated debates reveal the difficulty of exactly pinning down when and where the medium and bi-practice of blogs originated.

> As Rex Hammock points out there is no single history of blogs and argues that "everyone should write their own version of the history of blogging." As blogging is a practice that has shaped itself over time it is nearly impossible to point to one single blog as "the first blog" in retrospect. Blogs evolved out of a practice that is still developing and shaping itself. (Helmond, 2007a)

Aware of the pitfalls of reification I want to adhere to Peter Cooper's view on the history of blogs:

> I'm going to take an opposing view to everyone else and cite the start of blogging as being when the term was actually invented.
>
> If we didn't call a site a "blog" in 1997, was it actually one? I recall calling them journals, e/n sites, or sometimes even just "homepages". But.. before "blog" became popular, "weblog" was used.
>
> I guess it's like saying how old Web 2.0 is. Sites that offered similar features to Web 2.0 were around 10 years ago too (though not many!) but do we say Web 2.0 is 10+ years old? (Cooper, 2007)

James W. contributes that:

> "Blogging" really began only 6 or so years ago when the term was coined. Before that it was just "writing" or "journaling" (that's why it's called liveJOURNAL and not liveBLOG) and was going on years before these dolts care to conceive. (W. 2007)

The term weblog was coined by by Jorn Barger on 17 December 1997 and later

---

14   The Wall Street Journal article 'Happy Blogiversary' celebrated the 10th anniversary of blogs with Jorn Barger's *Robot Wisdom*. The author of the article, Tunku Varadarajan, was accused of rewriting blog history by Duncan Riley and getting the history of blogs wrong by Robert Scoble and Mike Arrington. A heated debate on the history of blogs was the result on the blogs and in the comments of Riley, Scoble and Arrington blog posts.

abbreviated by Peter Merholz in 1999 (Barger, 1997 & Merholz, 2002). I would like to argue that Barger was the first to name the medium and Merholz the first to name the practice. Merholz' term 'blog' was "quickly adopted as both a noun and a verb ('to blog,' meaning 'to edit one's weblog or to post to one's weblog')" and has stuck ever since. (Merholz 2002 & Wikipedia, 2007[15]). Every practice has its roots and influences which should not be denied:

> As I recall, there were things like blogs before there were blogs. People had websites on which they regularly posted new entries with links to new things on the Internet back in 1993. Even before that, there were things like blogs dating back to pre-historic times. Add an RSS feed to caveman drawings and you have Flickr. In 1994, the Drudge Report started as an email newsletter and later became a website that looked exactly like it still does. It was like a blog as it, well, had links to stories appearing on newspaper websites and a long list of links that is sort of what we'd now call a blogroll. Matt Drudge has steadfastly refused to call the Drudge Report a blog, and, frankly, he shouldn't, because it's not. In the history of the beginning of blogging, this makes Matt Drudge unique, as everyone else who was doing anything like a blog in the middle 1990s, today wants to say it was the first weblog. Matt Drudge, who could lay claim to the title of first blogger with some fairly convincing arguments — at least better than others — is still insulted when his site is referred to as such. (Ratzkinzin, 2007)

Ratzkinzin repeatedly states that "there were things *like* blogs" which exactly nails down the problem of a history of blogs. I want to adhere to a history of blogs that started in the late nineties with the birth of the word 'blog' which coincides with the upcoming of blog software.

I propose to describe the history of blogs, blog software and blog engines as one that is aware of roots, influences and milestones without creating a teleological or progressive account. By connecting these milestones to the engines a current account of blogging can be described. I propose a historical account that acknowledges the phenomenon of blogs as a medium and bi-product of practice that didn't originate with a single blogger, script or piece of software. However, some bloggers, scripts or pieces of software have shaped the medium and practice of blogging significantly to constitute what we now refer to as blogs:

> The politics and aesthetics defined by first users will characterize the medium for decades to come. Blogs appeared during the late 1990s, in the shadow of dot-com mania. (Lovink, 2007)

The late 1990s represent the time in which we saw a transition from Web 1.0 to Web 2.0. Even though the terms 'Web 1.0' and 'Web 2.0' are often described as vague and

---

15   http://en.wikipedia.org/wiki/Blog (accessed November 22, 2007).

suffer the same fate as 'blog' they do offer a good reference point to describe the time blogs originated in. Web 1.0 is a term used to describe the early World Wide Web with its static HTML pages. Tim O'Reilly states that "The bursting of the dot-com bubble in the fall of 2001 marked a turning point for the web" (O'Reilly, 2005) and led to a second generation of the web. In this new perception of 'the web as platform' the web is dynamic and has a focus on syndication, dispersion and aggregation (O'Reilly, 2005.) In the web as platform user participation and contribution gain a new focus and importance which leads to perceptions of the 'Web-as-participation-platform' or the 'participatory Web.'

Blogs are seen as exemplary of Web 2.0 and blog software is often associated with the infrastructure of Web 2.0. This generalization should not neglect the fact that blogs had existed for years and popular blog software such as Blogger had been developed and well used before the bubble burst. Blogger was an important blog service that contributed to the popularity and expansion of blogging but it was by no means the first and only one.

A problem one faces when studying the history of blogs is that the web has a very specific relationship with time and history:

> when one studies the Internet archive, what stands out is not so much that the Internet is archived, but how it is. Unlike a Web search engine, at archive.org's wayback machine, one queries a URL, not a key word. Moreover, one cannot 'surf' or search the Web as it was at some given date. In other words, a series of decisions was taken on how to build the archive, and those decisions constrain the type of research one can perform. (Digital Methods Initiative, 2007)

On the one hand the web is subject to the 'perceived freshness fetish' (Helmond, 2007b) and on the other hand it is oblivious of time. It is possible to look up what WordPress.org looked like in 2004, if it was indexed by the Internet Archive, but it is impossible to search for a chronological mention of 'WordPress' in '2004.'[16] On top of that the web is filled with dead links, websites and blogs that cannot always be recovered by the Internet Archive. This makes it hard to find out when a certain script or function was written and/or implemented.[17]

---

16 Google Search goes offer Timeline View but it is still under development in the Google Labs and does not offer very usable results.
17 I wrote a related blog post on 'Archiving Blogs and the Blogosphere' for the Blog Herald at: http://www.blogherald.com/2007/12/03/archiving-blogs-and-the-blogosphere/ (accessed January 11, 2008)

2.1 Pre-Blog Software History

A historical account of blogging should acknowledge the several scripts and software efforts that have led to the automation of publishing online, some variants of which we now refer to as blogging. An alternative history of blogging should highlight the scripts and functions that were written before what we now know as 'blog software.' It should also address the scripts and functions that were never adopted by the major blog software services.

It is impossible to describe all the single scripts, communication systems and software efforts that have led to our current understanding of blog software. However, this paragraph wishes to describe some milestones in the developments in the pre-web and pre-blog software era that may contribute to our understanding of the relation between blog software and the engines. This historical account focuses on the origins of the now characteristic features of blogs. The standardization of blog features has led to the development of standards and protocols within the software. These standards and protocols have enabled a blogosphere through the manual and automatic interlinking of blogs and the construction of a blogosphere by the engines.

2.2.1 Bulletin Board System (BBS)

The history of blogs is often traced back to Usenet, e-mail lists and Bulletin Board Systems (Broache & McCullagh, 2007 & Wikipedia[18]) where message boards were very popular and discussions flourished. A Bulletin Board System (BBS) is

> [...] a computer system running software that allows users to dial into the system over a phone line (or Telnet) and, using a terminal program, perform functions such as downloading software and data, uploading data, reading news, and exchanging messages with other users. (Wikipedia[19])

BBS was very popular from the late 1970s to the mid 1990 and is often considered "a precursor to the modern form of the World Wide Web and other aspects of the Internet." (Wikipedia) The BSS bulletin boards had a well-known structure with reverse chronological ordering, archive pages and keyword search:

> The most commonly used BBS structure is a two dimensional topic tree. Each leave contains a list of senders and subject titles in a reverse chronological order.

---

18   http://en.wikipedia.org/wiki/Blog (accessed November 22, 2007).
19   http://en.wikipedia.org/wiki/BBS (accessed November 22, 2007).

> The contributions are kept for a certain period of time and may be sometimes
> retrieved in an archive. Browsing of subject keywords may be available. (Alton-
> Scheidl, 1993)

The discussion boards on the BBSes were often local because early BBSes were centralized networks and most BBS networks were not linked. FidoNet changed this with the development of the Fido protocol that

> allowed BBSes to communicate with one another to create a FidoNet for wider
> distribution of messages. Instead of posting a message to just one board, a user
> could post a message and have it fan out to hundreds or thousands of boards at a
> time. (Zetter, 2005)

FidoNet could be considered to act as a distributing engine for the BBS medium. BBS was a medium that thrived on its ability to connect computer users and with the creation of the Fido protocol BBSes connected to FidoNet were linked worldwide. From the late 1970s to the mid 1990s BBS was a very popular decentralized information-sharing network that connected users in discussions on certain topics. Due to the upcoming popularity of the Internet around 1994-1996 the use of BBS declined.[20]

2.2.2 UseNet

UseNet was a very popular communication network in the eighties and nineties and was developed by Duke University graduate students Tom Truscott and Jim Ellis in 1979.

> Usenet (a contraction of user network) is a global, decentralized, distributed
> Internet discussion system that evolved from a general purpose UUCP
> architecture of the same name. (Wikipedia[21])

UseNet resembles BBS with the main difference that BBS is a decentralized network

---

20   Blog pioneer Dave Winer ran a BBS in the early eighties and in 1995 he wrote a "Web-based conferencing system for Macintosh servers called the Userland BBS." He mentions in the installation notes that: "After you install the BBS software, and connect to bbs.welcome.fcgi, you'll be greeted by Don & Marie. Log onto their bbs, you'll see the three main discussion areas, and a "Scratchpad" section. Every BBS should have one, it's the place where all things that don't have a place belong. It's going to be the most active discussion section on every Home Page, I'm sure of it." (Winer, 2005)
With the decline of the popularity of BBS and the increasing popularity of the Internet and the World Wide Web it is not surprising that Winer continued working on community discussion related scripts and software later pioneered some important blog features.

21   http://en.wikipedia.org/wiki/Usenet (accessed November 30, 2007)

and UseNet is a distributed network.

> One notable difference from a BBS or web forum is that there is no central server, nor central system owner. Usenet is distributed among a large, constantly changing conglomeration of servers which store and forward messages to one another. These servers are loosely connected in a variable mesh.

> Individual users usually read from and post messages to a local server operated by their ISP, university or employer. The servers then exchange the messages between one another, so that they are available to readers beyond the original server.

> [...] The articles that users post to Usenet are organized into topical categories called newsgroups, which are themselves logically organized into hierarchies of subjects. For instance, sci.math and sci.physics are within the sci hierarchy, for science. When a user subscribes to a newsgroup, the news client software keeps track of which articles that user has read.

> [...] In most newsgroups, the majority of the articles are responses to some other article. The set of articles which can be traced to one single non-reply article is called a thread. (Wikipedia, UseNet)

As this thesis is not a historical account of BBS or UseNet it is impossible to go into the large and rich history of both communication systems. However, it is important to refer to the distributed nature of UseNet as a communication system. Threads were organized in a hierarchical structure and by subscribing to a newsgroup the news client software would keep track of the threads and comments a user had read. Feed readers for keeping up-to-date with blog subscriptions certainly resemble many aspects of BBS' newsreaders.

2.2.3 .plan files

An important pre-web feature that enabled frequent updating of the self was the Finger protocol that

> would supply information such as whether a user is currently logged-on, e-mail address, full name etc. As well as standard user information, Finger displays the contents of the .project and .plan files in the user's home directory. Often this file (maintained by the user) contains either useful information about the user's current activities, or alternatively all manner of humor. (Wikipedia[22])

Inventor of the protocol, Les Earnest, was interviewed by CNET on the 10-year

---

22   http://en.wikipedia.org/wiki/Finger_protocol (accessed December 1, 2007)

anniversary of the term 'blog' and describes the use of .plan files as follows:

> It was used in much the same way as blogs are now, that is, the .plan file was intended to be just a way to tell people where you were going to be. If you were going off on vacation or a trip or something, or were just going to sleep for a while, you could post that in your .plan file. But then people noticed that it could be used as a statement of personal views on things and they started doing that, and then still later it was used by quite a few people to post their public key for secure communications, so it got used in several ways. (For) expressing your personal views on things, it was very much like a blog, a personal blog. It of course did not allow others to post things on it. (Broache & McCullagh, 2007)

The .plan files were used by many people and created a practice of regularly updating one's personal information and current status. The user culture of the .plan files may resemble the practice of blogging but it arose in a different time with a different user culture.

The .plan files were mostly oblivious to history as an update to a .plan file (basically a text file) overwrote the previous update. Of course it was possible to insert new updates in the .plan file and leave old updates untouched and in 1994 "one undergraduate student at Carnegie Mellon University created a rambling online diary in his .plan file that was hundreds of pages long." (Ibidem) This was, however, an exception to the norm which was more 'Twitteresque'. Twitter constantly asks the question 'What are you doing?' that can only be answered within 140 characters.

The practice of using .plan files may resemble some practices of blogging but .plan files are not a medium. The .plan file is stored in the user's home directory and can be retrieved by using the Finger protocol to display current user information. The .plan files do not share the duality of blogs that are both a medium and a bi-product of practice.

2.2.4 What's New?

As mentioned previously there is a not one single history of blogs. A blog is a fairly new medium and bi-product of practice that is still 'under construction.' As much as blogging itself is a process of constant updating, the medium and practice are also constantly redefining themselves. In retrospect the first (archived) website ever is often credited as the first existing blog:

> The first weblog was the first website, http://info.cern.ch/, the site built by Tim Berners-Lee at CERN. From this page TBL pointed to all the new sites as they came online. Luckily, the content of this site has been archived at the World Wide

Web Consortium. (Thanks to Karl Dubost for the link.)

NCSA's What's New page took the cursor for a while, then Netscape's What's New page was the big blog in the sky in 1993-96. (Winer, 2001)[23][24]

These 'What's New?' pages knew the same reversed chronological order as the BBS and UseNet discussion boards. On August 6, 1991, Tim Berners Lee posted a summary of the World Wide Web project on the alt.hypertext newsgroup with the announcement to "try it" which marked the debut of the Web as a publicly available service on the Internet (Wikipedia[25] & Berners-Lee, 1991). It wasn't until April 30, 2003 that

CERN announced[7] that the World Wide Web would be free to anyone, with no fees due. Coming two months after the announcement that the Gopher protocol was no longer free to use, this produced a rapid shift away from Gopher and towards the Web. An early popular web browser was ViolaWWW, which was based upon HyperCard.

Scholars generally agree, however, that the turning point for the World Wide Web began with the introduction[8] of the Mosaic web browser[9] in 1993, a graphical browser developed by a team at the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign (NCSA-UIUC), led by Marc Andreessen. (Wikipedia, World Wide Web)

With the World Wide Web going public, the release of the first graphical browser and growing internet access we saw the rise of the personal homepage in the mid 1990s and the rise of the blog in the late 1990s. On the web there seems to be an increasing drive to constantly update and there is a major focus on freshness. In the mid 1990s this was visible in the use of under construction signs, 'new' gifs and last updated scripts on homepages:

The last updated scripts were used to inform the website's audience if and when the web page had been updated. The script only indicates the freshness of the page it is placed in and not the underlying pages. This could lead to an index page that had been last updated a year ago while the content of the underlying pages had been updated only a few days ago. The "new" image was often used in combination with the JavaScript on the main page to show which sections of the website had been updated.

Blogs would change this by always showing the lastest item first. (Helmond, 2007b)

---

23   http://www.w3.org/History/19921103-hypertext/hypertext/WWW/News/9201.html (accessed November 29, 2007).

24   http://wp.netscape.com/home/whatsnew/whats_new_0693.html (accessed November 29, 2007).

25   http://en.wikipedia.org/wiki/World_Wide_Web (accessed November 29, 2007).

The first weblogs were hand-coded HTML pages that were frequently updated manually. New items would be added to the top of the page or linked to from the front page. Such early weblogs include Justin Hall's links.net and Dave Winer (1996), Jorn Barger at Robot Wisdom and Cameron Barrett at CamWorld (1997). Blog software automated incorporated scripts to automate the practice of updating and displayed fresh content on top.

2.2 The Rise of Blog Software

This historical account wishes to address the shift from the early hand-coded 'weblogs' to the automatically updated 'blogs' with a grey area in between. Blog software enabled those without any HTML or other scripting knowledge to easily update their site and Geert Lovink states that "the motor behind the expansion of the blogosphere is the move away from code towards content" (2007). Blog software does not require any knowledge of code to use the default features and settings. Popular blog services such as Blogger and WordPress.com only require to create an account and a blog title in order to start blogging. Blog software reduced the barrier of starting a blog and simplified the process of creating new entries.

The practice of blogging changed with the introduction of blog software as coding knowledge is no longer required. Easy-to-use blog software introduced the practice of blogging —that previously belonged to (semi) tech-savvy web users— to the 'average' web user. However, customizing the blog requires a move back to code as for example Greymatter, Movable Type and WordPress show. Different blog software serves different kinds of users with different needs and level of expertise.

The shift from manually updated webblogs to automatically updated blogs entails a shift from offline HTML editing software[26] to online web-based blog software. Blog software has become browser based as an extension of our web practices.

The following depiction is an illustration of several scripts and services that have contributed to the automation of publishing online, some forms we now refer to as blogging. It describes several people and (start-up) companies that have played a major role in the development of blog software, blog standards and blog protocols. It

---

26   Or plain text editors.

aims to focus on the history of the characteristic features of the blog.

2.2.1 Scripting News (1997)

Dave Winer started publishing on DaveNet in 1994 and on Scripting News in 1997. He wrote various scripts to update his Frontier News and Scripting News pages and decided to release the software he used to update his weblogs. Winer helped popularize blogs as Jorn Barger was among the many to use Dave Winer's software to update his blog and Cameron Barrett modeled his blog after Scripting News (Riley, 2007).

Dave Winer started a company named UserLand and developed blog software and web publishing tools such as Radio Userland and Manila. Winer developed several technical standards that have become very important in blogging, such as XML-RPC is "a remote procedure call protocol which uses XML to encode its calls and HTTP as a transport mechanism"[27]. The protocol was created in 1998 in cooperation with Microsoft. The protocol allows any scripting language like Perl[28] or PHP[29] to be encoded or decoded in XML for transfer. This opens up the blogging across various platforms and platform for various ways of blogging. XML-RPC allows you to build your own Weblog Client to post to your blog (WordPress or not). XML-RPC plays a very important role in blogging as it is the technical background for ping, the pingback, trackback and RSS. Winer has also contributed to the development and popularizing of RSS.

In 2000 Winer started offering "a free blog-starter service and community offered by UserLand" at Weblogs.com. The service unexpectedly closed down in 2004 and "explaining his reasons for the shutdown, Winer cited the financial costs of hosting the sites, technical difficulties in moving the blogs to a new server, stress and personal health issues as the reasons for the sudden shutdown." (Delio, 2004) Weblogs.com did not only serve as a service and community but it was also the first ping server and "considered a key part of the blog world" (Naraine, 2005) and will be

---

27 http://en.wikipedia.org/wiki/XML-RPC (accessed November 30, 2007).
28 "Perl is a dynamic programming language." http://en.wikipedia.org/wiki/Perl (accessed January 27, 2008)
29 "PHP is a computer scripting language originally designed for producing dynamic web pages.[1] The name PHP is a recursive initialism for PHP: Hypertext Preprocessor." http://en.wikipedia.org/wiki/PHP (accessed January 27, 2008)

31

further addressed in paragraph 2.4.2.

2.2.2 Open Diary (1998)

Open Diary was founded in October 1998 by Bruce Ableson who was fascinated by an online journal written by a traveling teenager. He became a vivid reader and found himself coming back regularly. He searched the web for more of these kinds of journals and eventually found a few hundred and

> It seemed to me that if I found these journals compelling - that if I wanted to come back and read them every day - that other people would feel the same way. However, it was hard to find them, and hard to keep track of them - if there was a central location where they could be gathered together in an easy-to-read and easy-to-search format, there could be a larger audience for them.
>
> Also, it was hard back then to write an online diary - you had to have your own web space, and you had to know how to create pages and use HTML. These were barriers that prevented the average Internet user from keeping a diary.
>
> If this central journaling location could also provide a simple method for writers to post entries to their journals, without needing to know obscure codes and without needing to pay for server space, it could be a good thing. (Interviewbooks.com)

Ableson decided to create such a central journaling location with a simple method for posting entries and started Open Diary. Not only was it the first community that supplied an easy method to create your own diary it also was the first to implement the possibility of leaving notes on other people's diaries. Enabling notes allowed for interaction between writers and readers and it was "an innovation that led to the interactive nature of today's blogs and blog comments."[30] Open Diary's notes "appear at the bottom of diary entries, and can be read by anybody else reading the diary – unless they are private notes."[31] The format of the notes resemble the way we now know comments, which are listed below the blog entry. Notes on Open Diary are limited to 400 characters unless a Plus account is purchased allowing to post notes up till 1,200 characters.[32] The Plus account also provides a 'Note History' where "the

---

30   http://en.wikipedia.org/wiki/Opendiary (accessed November 22, 2007)
31   http://www.opendiary.com/help/notes/leaving_notes_on_diaries.htm (accessed November 22, 2007).
32   The ticking down of the 400 characters resembles Twitter that also ticks down, but with a nervous red effect that you are going into a danger zone. http://www.opendiary.com/leavenote.asp?authorcode=D482256&entry=10007 (accessed November 22, 2007)

system keeps track of notes as you leave them, so you can review notes you've left at any time, and see where and when you left them."[33] The Note History deals with distributed notes in the Open Diary community by centralizing notes you have left on a single page.

2.2.3 Blogger (1999)

The year 1999 was an important turning point in the history of blogs as many web publishing and blogging tools and services were released by companies such as Pitas.com, Open Diary and LiveJournal. The company Pyra Labs helped to popularize blogging even further by offering a free place to easily and quickly set up and maintain your own page with Blogger. Blogger is a web-based tool that allows blogs to be "be hosted internally by Blogger on the blogspot.com Internet domain, on a user's own domain, or externally on the user's own server (through FTP or SFTP)"[34] The release of the free and easy-to-use software platform Blogger is seen as a breakthrough moment for blogs (Leaver, 2006 & Blood 2004, p. 54).

Blogger is believed to have permanently changed the weblog culture with an important interface decision (Blood, 2004, p. 54).

> Consider Pitas (www.pitas.com), another early Weblog updater that provided users two simple form boxes: one for a URL and one for the writer's remarks. Hitting the "post" button generated a link followed by commentary. Blogger was simpler still, consisting of a single form box field into which bloggers typed whatever they wanted. I sometimes wonder whether the new bloggers knew enough HTML to construct a link. Whether they did or not, Blogger was so simple that many of them began posting linkless entries about whatever came to mind. [...]
>
> So, with the overwhelming adoption of Blogger, and without an interface that emphasized links as the central element of the form, the blog-style Weblog was born. Much controversy ensued in the original Weblog community: These are diaries, not Weblogs; Weblogs are about links. Evan Williams has said that he understood early on that Weblogs are about the format, not the content[4]. (Blood, 2004, p. 54)

Weblogs (logging links on/from the web) became blogs (blog-style weblogs) by deemphasizing the link and placing the focus on the format instead of the content. Blogger helped to change the format of the weblog and define the format of the blog

---

33 http://www.opendiary.com/upgrade.asp (accessed November 22, 2007).
34 http://en.wikipedia.org/wiki/Blogspot (accessed November 30, 2007).

by introducing another innovation, the permalink, in 2000:

> Permalinks gave each blog entry a permanent location— a distinct URL—at
> which it could be referenced. Previously, Weblog archives had been navigable
> only through browsing. Now, bloggers could reference specific Weblog entries as
> elegantly as they referenced any online source. The feature was so useful it
> became a canonical component of the standard Weblog entry [3]. (Ibidem, p.
> 54-55)

The permalink is a hyperlink fixed in time that signifies an update as a "link to a
weblog is expected to point to a perennially changing page, with 'permalinks' for any
individual entry, and notification for each change." (O'Reilly, 2005)

Blogger was a free and easy-to-use service which became incredibly popular but
in 2001 Pyra Labs, the company that created Blogger in 1999, was struck by the dot-
com crash as investors became reluctant to invest in new companies. As Blogger
gained popularity the servers could not handle the massive amount of new users
anymore and the service slowed down. In 2001 Pyra Labs asked their users to help
them to raise money for a new server by contributing to the 'Blogger Server Fund.'
This eventually allowed them to buy three new servers. In 2003 Blogger was sold to
Google because "a company of Google's size could give Blogger the resources we
needed to do things better, faster, bigger."[35] The acquisition of Blogger by Google was
the first step into the world of blogging and also meant the expansion of the Google
world:

> We are lurching toward a state in which Google and the Web are imbricated, if
> not merged, in an irresistible way. As Google multiplies its services and the
> content that sits on its servers, Web users find more reasons to remain in the
> Google world for hours at a time. (Vaidhyanathan, 2007b)

In 2005 Google would further expand its search strategy and services with the launch
of Google Blog Search (see paragraph 2.4.4).

2.2.4 Greymatter (2000)

As Rebecca Blood described, Blogger's power lay in its easy-to-use interface. Blogger
served a specific kind of user that did not need or have any HTML knowledge.
Greymatter, an often unmentioned piece of software in blog history, may have been

---

35 http://web.archive.org/web/20030603164457/blogger.com/about/blogger_google_faq.p
yra (accessed November 30, 2007).

the first to serve a more tech-savvy kind of user that wanted to customize its blog and keep control over it:

> Noah Grey: Greymatter is an opensource program for creating weblogs and journals with - as far as I know it was the first such program that you could run yourself without relying on any outside service beyond your control. It was also (again, as far as I've ever been aware) the first to offer integrated comments, searching, completely customisable templates and output, automatic image and popup handling, support for multiple users/authors, and the ability to configure nearly every aspect down to the last detail - things that are now considered standard and indispensable in weblog/journal programs. (Grey in Turnbull, 2003)

Creator Noah Grey places Greymatter in between the development of the easy-to-use Blogger and the power-and-control Movable Type:

> Greymatter broke the ground and created the "do it yourself" model and "it's all about power and control" ethic that MT and others have picked up on and taken so much farther than GM did, but I don't think the next level beyond that has been reached yet. (Grey in Turnbull, 2003)

Greymatter is still being developed but it has never become as popular as Movable Type would become.

2.2.5 Movable Type (2001)

The first public version of Movable Type was relased on October 8, 2001 by Six Apart. It addressed a different market than Blogger and

> MT can't replace Blogger; we don't want it to. There are clear advantages to having a web-based, centrally hosted application. I just felt I out-grew the features of Blogger and was a bit frustrated with the server downtimes and associated problems. (Trott in Turnbull, 2001)

Greymatter had also been developed to address the issues that arise from a centrally hosted application with hardly any possibilities to extend and customize and Movable Type continued this trend with "a system that is modular, extensible, and easy to use." Movable Type offers a "modular library code that Perl programmers can extend for their own purposes. And we plan on adding XML-RPC and SOAP APIs to the system."[36] (Trott in Turnbull, 2001) These additions would lead to the development

---

36 "An application programming interface (API) is a source code interface that an operating system or library provides to support requests for services to be made of it by computer programs." http://en.wikipedia.org/wiki/Application_programming_interface (accessed December 2, 2007).

of the trackback which was implemented in version 2.2. in 2002.[37] It was developed as "a framework for peer-to-peer communication between weblogs; it can track cross-weblog discussions, it can provide remote content repositories, it can emulate guest authoring, etc. " It was a major innovation that

> automated crossblog talk itself. Trackback allows bloggers to ping other Weblogs, placing a reciprocal link—a trackback—in the entry they have just referenced. Previously, bloggers scoured referrer logs to discover references to their sites. Trackback has made these formerly invisible connections visible, inviting instant response. Trackbacks, often interspersed among site comments, emphasize the conversational nature of the Weblog form while collating for readers all available responses to an entry. (Blood, 2004, p. 55)

Trackback renders the invisible link visible and shows the connections between blogs. The feature became widely used and popular and has been adopted since by several blog platforms. The downside of the automation of sending and receiving trackbacks is that blogs have become vulnerable to spam.

Movable Type was a very widely used and popular blog software package until the release of Movable Type 3.0 in 2004. This release included a new licensing structure which put restrictions on using Movable Type for free. Many users migrated to other blog software such as the new and upcoming WordPress. The new licensing structure of Movable Type is often regarded as "the beginning of real growth for WordPress." (Web Hosting Folio, 2006)

2.2.6 b2/WordPress (2003)

WordPress was developed by Matt Mullenweg and Mike Little as the official successor of b2/cafelog. Mullenweg and Little created the WordPress b2 fork after an announcement that b2 was no longer going to be developed:

> Matt took the opportunity to fork, that is develop his own software starting with the existing code available in b2, and bring it up to the requirements of web standards (Web Hosting Folio, 2006).

The first version of the new b2 fork WordPress was released on May 27, 2003 which included major (standard compliant) code updates, a focus on web standards, typography and a new administration interface (WordPress, 2003). The growth of

---

37   http://www.movabletype.com/blog/2002/06/version-22-released.html (accessed December 2, 2007).

WordPress started in 2004 when many of the Movable Type users migrated and the next breakthrough moment came in 2005:

> 2005 was a very exciting year for WordPress, as it saw the release of our 1.5 version which was downloaded over 900,000 times, the start of hosted service WordPress.com to expand WP's reach, the founding of Automattic by several core members of the WP team, and finally the release of version 2.0. (WordPress, About)

WordPress has known an immense growth ever since and on December 2, 2007 the latest WordPress version 2.3.1 had been downloaded over 363,083 times. WordPress is currently the most popular blog software and a recent study from Google showed that 0.8% of **all websites**[38] had been created with WordPress (CNET News.com, 2007).

2.3 Blog Software Concentration

A recent poll on the popular blog ProBlogger titled 'What Blogging Platform Do You Use' revealed that WordPress.org (36%), Blogger.com (25%) and WordPress.com (15%) were the most used platforms (ProBlogger.com, 2007). In comparison to the same poll in 2006 the amount of Movable type and "other" blog platforms users had dropped. The 2006 poll revealed that a large number of users that used other platforms than the ones listed and a fair amount of users created their own platforms or hand-coded their blogs.

This small poll reflects the general trend of the increasing amount of software concentration in blog software and Darren Rowse from ProBlogger notes that

> That these three platforms take up 76% of responses says something in and of itself with the two varieties of WordPress being used by 51% of ProBlogger readers. (Rowse, 2007)

The top three of blog platforms currently consist of WordPress, Blogger and Six Apart. WordPress includes WordPress.org (self hosted blog software) and WordPress.com (blog hosting service.) Blogger includes blogs hosted on Blogger's Blogspot domain or on the user's own server. Six Apart is the company behind the Movable Type,

---

38 Of course this astonishing number includes a huge amount of abandoned or dead blogs, spam blogs and the use of WordPress as a content management system for other purposes than blogging.

TypePad, Vox, TypeKey and SplashBlog blog platforms[39].

An analysis of the Technorati Top 100 by CMS Wire revealed that 34% of the blogs use WordPress and 16% use Movable Type – together accounting for half of the surveyed blog sites – and "the rest are either custom jobbies or a smattering of other platforms which are, relatively speaking, eating dust." (Conroy, 2008)

2.4 The Rise of Blog Engines

With the phenomenon of weblogs (1997) and blogs (1999) on the rise the first attempt to list all the blogs was made by Cameron Barrett on CamWorld in November 1998 (Riley, 2005). This original list is still available on his blog and is now published as his blogroll[40]. The rise of the engines started in 1999-2000 and begun with (semi-) manual attempts to index and organize blogs. The growing popularity of blog software runs parallel to the rise of the blog engines.

2.4.1 EatonWeb Portal (1999)

Another early initiative in indexing blogs was made by Brigitte Eaton who listed all the blogs she knew about on her blog. She created the EatonWeb portal in 1999 and the list grew from fewer than 50 to over a hundred in less than a year: (Blood, 2000 & Kahney, 2000)

> "I can't keep up with them all," said Brigitte Eaton, who maintains the list and publishes her own weblog. "I think part of the explosive growth is due to tools being developed that allow easy entry. Blogger, Pitas, Groksoup, and Editthispage, all have gotten huge numbers of participants. None of those services were around six months ago." (Kahney, 2000)

The EatonWeb Portal had just one criterion for the inclusion of a weblog:

> every weblog that is submitted gets validated by me to make sure it fits my loose

---

39   While writing this thesis Six Apart sold LiveJournal to a Russian media company named SUP in order to focus on the development of its own products. LiveJournal was originally developed by early blogger Brad Fitzpatrick who sold his company Danga Interactive to Six Apart in 2005. http://en.wikipedia.org/wiki/Live_Journal (accessed January 27, 2008)

40   The term was coined by Doc Searls on December 17, 2000 and is "The section of a weblog that lists the sites that the blogger reads on a regular basis. This is usually located on the side of a blogger's frontpage, or on a separate page linked off of the frontpage." http://www.scripting.com/2003/04/21.html (accessed December 3, 2007).

definition of a weblog: something that's organized chronologically and updated fairly regularly. i also make sure they've been around for about a month. it weeds out the one's where people give up after a few weeks.[41]

Brig evaluated all submissions by a simple criterion: that the site consist of dated entries. Webloggers debated what was and what was not a weblog, but since the EatonWeb Portal was the most complete listing of weblogs available, Brig's inclusive definition prevailed. (Blood, 2000)

This definition of EatonWeb determined whether or not a blog would be included in the most complete listing of weblogs available. EatonWeb defined the weblog by its format but added to it the idea of blogging as a practice. The portal was a product of manual labor and functioned as the first blog index and search engine. Blogs were listed under 'alphabetically,' 'categorically,' 'not in english,' 'search' and 'new.'[42] Later a weblog top 10 was published on the front page which was based on click-throughs:

i do realize that one could very easily screw with the system, but i'm hoping people won't. besides, i keep an eye on it and i do have a relatively good idea of how many hits/day extremely popular weblogs get. if someones weblog has 10X that, i'll probably just wipe the counter or the weblog itself if it happens a lot.[43]

This explanation of the vulnerability of the ranking system shows the great amount of control over the manually kept index by Brigitte Eaton. The list was kept up-to-date manually which eventually led to problems with the increasing amount of blogs and Eaton wrote on October 28, 2001 that "i've got a serious backlog of weblogs that need to be validated. if you've got too much time on your hands and want to help out, let me know."[44] EatonWeb became an established source and other lists such as Weblog Madness stopped listing blogs and added its list to EatonWeb[45]. The task of manually running the portal finally became too much work and with a failing database Brigitte Eaton announced on the 1st of June 2005 that

i'm considering selling the portal. i don't have nearly enough time to devote to it anymore. if i don't sell it, i'll probably be shutting it down in the near future. anyone who's seriously interested, please email me at brig@eatonweb.com [46]

41 http://web.archive.org/web/20011214025229/portal.eatonweb.com/about.php3 (accessed November 21, 2007)

42 http://web.archive.org/web/20001109062400/http://portal.eatonweb.com/ (accessed November 21, 2007)

43 http://web.archive.org/web/20011214025229/portal.eatonweb.com/about.php3 (accessed November 21, 2007)

44 http://web.archive.org/web/20011202073928/http://portal.eatonweb.com/ (accessed November 21, 2007)

45 http://web.archive.org/web/20020802091553/www.larkfarm.com/weblog_madness.htm (accessed November 21, 2007)

46 http://web.archive.org/web/20050610085626/http://portal.eatonweb.com/ (accessed

EatonWeb was quickly sold to Ahmed from Bloggy Network and eventually relaunched as the Blog Flux Directory. In 2007 EatonWeb was sold to Splashpress Media who relaunched it while Bloggy Network kept working on the Blog Flux Directory (Farooq, 2007.)[47]

2.4.2 Weblogs.com (1999)

"Dave Winer created Weblogs.com in 1999 as a service to index weblogs. He re-launched it as a ping server in November 2001 [...] as blogs started to grow beyond the point where it was feasible to re-index them regularly." (Arrington, 2005)

The creation of a ping server marked an important moment in blogs as it allowed blogs to notify a server of an update. The ping server acts as a centralizing server that collects updates from the distributed nodes (blogs) in the network. Weblogs.com is an open system meaning that the data about recently updated blogs can be used by anyone, including blog search engines. A WordPress blog notifies ping servers by default which means that a WordPress blog is almost automatically included in the engines.

Weblogs.com and similar services display a list of the most recently updated blogs on their front pages. The data can also be accessed through an API and used by the engines. The Weblogs list was updated hourly and showed "when which sites changed."[48] The engines contributed to the increasing focus on freshness and the 'perceived freshness fetish' in the blogosphere. (Helmond, 2007b)

November 21, 2007)

47   The history of EatonWeb is very rich and definitely deserves its own Wikipedia entry (it currently doesn't have one.) A few other noteworthy features of EatonWeb were
- The graveyard of dead blogs "the [dead?] link is a way of notifying me that a weblog no longer exists or its url has changed. it won't delete it from the database, it just gets flagged. "
http://web.archive.org/web/20011214025229/portal.eatonweb.com/about.php3
- Similar blogs based on keywords: "i wish people would stop coming up with more useful ideas. luckily, this one was relatively easy to do. i've created lists of weblogs with similar keywords. take a look at the column on the right."
http://web.archive.org/web/20020923110104/http://portal.eatonweb.com/

48   http://web.archive.org/web/20000303173908/http://www.weblogs.com/ (accessed December 3, 2007)
Interestingly enough the list was timestamped in hours Pacific. The current trend is to timestamp updates with "x minutes ago" or "x hours ago" pointing to a increasing focus on freshness on the internet where time is related differently to geography.

2.4.3 DayPop (2001), Blogdex (2001) and Technorati (2002)

The next generation of engines focused not only on indexing and freshness but also on other important features of the blog namely links.

> The Technorati 100, Blogdex, and Daypop Top 40 are all representatives of the same general type of algorithm: notification of update, extract out the links, increment the count for any matched link, with the top n number of link holders placed on the list, ordered by number of links. Though the data is treated differently–Technorati persists the number of links per page, while Daypop and Blogdex are only interested in reflecting 'fresh' links–the concept, and hence, the algorithm is the same: when a link to a specific page is encountered, add a link to the source to a list, and increment a counter. Then adjust the list accordingly. (Powers, 2005)

Daypop was originally launched in 2001 as a "a current events search engine"[49] but later it presented a top 40 that became very popular in the blog community[50]. In 2001 Cameron Marlow created Blogdex at the MIT Media Lab which was

> a system built to harness the power of personal news, amalgamating and organizing personal news content into one navigable source, moving democratic media to the masses. at current, blogdex is focused on the referential information provided by personal content, namely using the timeliness of weblogs to find important and interesting content on the web. (Internet

Marlow points to the 'timeliness of weblogs' that could serve as a meme tracker[51] by focusing on referential information related to time. The index was created around the popularity of the link in order to track "diffusion in the weblog community." (Marlow)

Technorati was created in 2002 by David Sifrey as a blog index with a focus on inlinks. It combined already existing features of Daypop and Blogdex into a new

---

49 http://web.archive.org/web/20010902230917/www.daypop.com/about.htm
     (accessed December 3, 2007)
50 http://web.archive.org/web/20011130145613/http://www.daypop.com/
     (accessed December 3, 2007)
51 Or news tracker as preferred by Alex Barnett:
     The word 'memetracking' or ('meme tracker') has been used to describe services such as Memeorandum, Megit, Tailrank and Chuquet. [...] A meme, as defined by Richard Dawkins in his 1976 book The Selfish Gene, is
     "a unit of cultural transmission, or a unit of imitation. [...] Examples of memes are tunes, ideas, catch-phrases, clothes fashions, ways of making pots or of building arches."
     [...] There is imitation going on at these these sites: 'oh, she's reporting this news, it's interesting, so I will'. These sites track the act of passing on units of news. In this sense these sites are tracking the memetic quality of blogs. However, generally speaking, these sites are not tracking the spreading of 'ideas' or 'memes'. They are tracking bits of news being passed on from one blogger or site to another. A bit of news is not a meme, nor an idea, it is a bit of news. Memeorandum, Megit and Chuquet are not 'meme' trackers. They are news trackers. (Barnett, 2006)

service that "reverses Shirky's power law to make it work in favor of the under-linked, providing greater exposure to lesser-known bloggers" (Lampa, 2004). It did not only serve as a blog index and search engine but also as a community tool with its focus on inlinks and the profiling of interesting blogs. As Technorati gained popularity Daypop, Blogdex and many other blog engines[52] ceased to exist.

2.4.4 Google Blog Search (2005)

Google entered the blog engine market relatively late on September 14, 2005 in order to expand their universal search strategy (Goldman, 2005). As the main search engine for the web it introduced Google Blog Search to expand their vertical search focusing on a specific sphere, namely the blogosphere. As the owner of Blogger and with its solid reputation and index of web search it soon became Technorati's main competitor. Not only can we see a trend of concentration in blog software but also in blog engines:

> Recent trends suggest that the search engine market is shrinking, with fewer large players guiding users' online behavior than ever before. This suggests that decisions made by just a few players in this landscape can have considerable repercussions for what material is realistically within the reach of users.
> (Hargittai, 2007)

The implications of the engines will be further be addressed in chapter four.

2.5 Three Stages in Blogs

This history of blogs, blog software and blog engines shows what Arie Altena refers to as the three stages in blogs. In the first stage the weblog is the product of an author logging links from the web with optional commentary. In the second stage the blog is an invitation to conversation with the introduction of comments and the permalink. In the third stage the blog is a form of syndication, aggregation and dispersion (Altena, 2007)[53]. The three different stages in blogging still exist but point to an

---

52 Ari Paparo's 'Big List of Blog Search Engines' from December 31, 2003 lists a lot of now extinct engines: http://www.aripaparo.com/archive/000632.html (accessed December 5, 2007)
53 The question arises whether blogs are now moving into a a fourth stage which may defined as 'Blogs as CMS' even though it refers more to a shift in the usage of the software than the

increasing relationship between the software and the engines. The next chapter will describe how this symbiotic relationship has developed. By deconstructing the software the software-engine connections can be described from a technical point of view.

practice of blogs:
http://www.annehelmond.nl/2007/04/16/deaf07-ubiscribe-collocollaboracontentquery/

3. SOFTWARE-ENGINE RELATIONS

The previous chapter has shown how the rise of blog software runs parallel to the rise of the engines. The question arises if there is an increasingly symbiotic relationship between blog software and blog engines. In order to address this question the first part of this chapter will describe the technical background of the medium. It addresses the different layers of the software, how they are entangled and how they frame the blogging experience.

The second part describes the canonical features of the software and their connections to the engines. It will show how software establishes relationships with the engines by using the underlying protocols of the blog's core features. This influences both how the blog and the blogosphere are constructed.

3.1 The Technical Background of WordPress

The blog as a medium consists of several parts that are completely interwoven but are considered as separate components, namely the server, the database, the software and the browser. WordPress software, where you can write posts and manage your blog, acts as an interface to the database on the server where most of the blog content is stored.

When blog software is discussed usually the interface is meant and not the database. The interface is meant to hide the database from the user but before you are able to install WordPress on the web server of a hosting company you need to create a database on the server. This is one of the few times that the user is confronted with the actual existence of the database. The fact that many web hosting companies offer automatic installing and upgrading of WordPress means that some users will never be confronted with the existence of a database. This process masks the database and renders it invisible.

In the current discourse blogging is never described as 'writing to the database' but it is always seen as 'writing in one's blog.' Instead of seeing blogs as self contained units where blog posts content are stored it makes them seem like a part of the larger

Web. Blogs and WordPress software are usually accessed by using a web browser, an often neglected object in the discourse on blogging.

3.1.1 The Browser

The browser is often taken for granted when talking about blogging. However, it plays an important role as blog software is accessed and run through the browser and blogs are typically displayed in the browser[54]. The blog as a medium is encapsulated within the browser. This is exemplary of the internet that consists of different types of layers that encapsulate one another and communicate through protocol:

> Developed in 1978, the OSI Reference Model is a technique for classifying all networking activity into seven layers. Each layer describes a different segment of the technology behind networked communication, as described in chapter 1.
>
> Layer 7: Application
> Layer 6: Presentation
> Layer 5: Session
> Layer 4: Transportation
> Layer 3: Network
> Layer 2: Data link
> Layer 1: Physical (Galloway, Protocol, p. 130)

The application layer includes the browser and file transfer protocol (FTP) programs. The first weblogs relied mainly on file transfer protocols such as FTP to manually upload new hand-coded web pages to the server. An important shift in blogging occurred when blog software simplified the process of updating by creating a complete automatic update environment in the browser[55]. The act of enveloping one layer, the database, into another, the browser, is a concealment of the source. This may be done in order to create continuity:

> Despite being a decentralized network composed of many different data fragments, the Internet is able to use the application layer to create a compelling, intuitive experience for the user. This is the secret of continuity. [...]

---

54 "The Flock browser is a browser which focusses on social networking web services and includes Flickr, Facebook, YouTube, del.icio.us, Technorati, Photobucket, and various popular blogging and news aggregation services." http://en.wikipedia.org/wiki/Flock_%28web_browser%29 (accessed January 5, 2008)
It would be interesting to see if different browsers shape the blogging experience differently and if they influence the daily blogging routine.

55   FTP is still used for blogging to upload manually the software and new themes, plugins and widgets. It is also used for updating the software by replacing the old files with new ones.

45

> Continuity, then, is defined as the set of techniques practiced by webmasters, that, taken as a totality, create this pleasurable, fluid experience for the user. As a whole they constitute a set of abstract protocological rules for the application layer. (Galloway, 2004, p. 64)

By incorporating the blog software into the browser a continuous blogging practice is created. This practice usually consists of reading other blogs, commenting and checking statistics on the web by using a web browser. "Browser-based blogging ensures continuity with surfing rather than, say, Microsoft Word." (Stevenson, 2008) Even though Word 2007 includes blogging capabilities its unpopularity may be explained due to the fact that it breaks the continuity of the blogging practice because the program resides outside of the native blog environment. Even though it is a familiar writing environment, it is mainly based on print and does not take the network the blog resides in into account. Users for example dislike the fact that tagging is not incorporated[56]. Paragraph 3.1.2 describes how (offline) weblog clients may be used when continuity is disrupted.

The browser communicates with the database by sending and receiving HTTP requests. The Hypertext Transfer Protocol (HTTP) is the general method to retrieve and publish web pages on the web. When writing a post you are filling in the fields of a HTML form that after publishing is sent to the database using the HTTP POST command. Forms are used for several reasons: it is one of the standard ways to receive user input from a web page. An associated script can process the form and can perform arbitrary computation. It is currently the most commonly used method for using the browser as a front end to applications (Krishnan). This means that the form is used to enter the post that is compiled in the browser into the database. The form has been the dominant form of automatic blogging since the programming language Perl and CGI[57] scripts were first used in blogging to send content to the server for processing.

The server forms an important part of the blog as it is the place where the software is installed. It is one of the main requirements for WordPress software to run. The official WordPress 2.3.1 documents state that the following is required to

---

56   Blog comments on 'Blogging from Word 2007':
http://blogs.msdn.com/joe_friend/archive/2006/05/12/595963.aspx (accessed January 14, 2008)

57   "The Common Gateway Interface (CGI) is a standard protocol for interfacing external application software with an information server, commonly a web server."
http://en.wikipedia.org/wiki/Common_Gateway_Interface (accessed January 28, 2008)

run: a web server, MySQL version 4.0 or greater and PHP version 4.2 or greater.[58] The next paragraph will deal with the requirement of the web server.

3.1.2 The Web Server

Blog software must be installed on a web server in order to be able to run. A web server is

> A computer that is responsible for accepting HTTP requests from clients, which are known as Web browsers, and serving them HTTP responses along with optional data contents, which usually are Web pages such as HTML documents and linked objects (images, etc.). (Wikipedia[59])

This means that your own computer can act as a web server if a web server such as the popular Apache server has been installed. The limitations are that your blog is only accessible through your (dynamic) IP address[60] and when the computer is turned on and online. Most Internet Service Providers will automatically assign an IP address when the modem goes online. This means that the 'address' of your blog, your numerical IP address, is subject to change. This is not the case when the blog is hosted with a hosting company or if your Internet Service Provider has assigned you a static IP address. A hosting company can offer a domain name, storage space and bandwidth space to run the blog for a fee. The blog can be installed on the web server of the hosting company and, depending on the uptime of the host, your blog will always be online at the same address.

The always online mentality seems to be an a priori condition for blogging. Blogging originated in a culture in which being online 24/7 became more and more dominant and cheaper than ever as a result of the widespread rise of broadband. A 2006 survey conducted by the Pew Internet & American Life Project shows that bloggers have a higher number of broadband connections and are online more often than their non-bloggers (Lenhart, 2006). Broadband penetration slowly started in 1999 around the same time that Blogger was starting up. There are hardly any

---

58 No mention of a browser in the official requirements at http://www.wordpress.org/about/requirements (accessed January 4, 2008)
59 http://en.wikipedia.org/wiki/Webserver (accessed December 28, 2007)
60 "An IP address (Internet Protocol address) is a unique address that certain electronic devices currently use in order to identify and communicate with each other on a computer network utilizing the Internet Protocol standard (IP)—in simpler terms, a computer address." http://en.wikipedia.org/wiki/IP_address (accessed January 12, 2008)

47

statistics on the growth of the blogosphere before 2002. This makes it hard to determine whether there is a correlation between the growth of broadband and the growth of the blogosphere. However, statistics show a tremendous increase in the penetration of broadband between 2001 and 2006 (OECD 2006) and the blogosphere also has grown enormously between 2002 and 2006 (Rainie 2005; Sifry 2007). It is very likely that the two are correlated because blog software resides within the browser and requires one to be online. Even though nowadays a broadband subscription is often cheaper than a dial-up connection it is not available everywhere and not everyone can afford it. Blogging has always been very popular among travelers and many of the first blogs were travel diaries (see paragraph 2.2.2). A great many travel areas often have Internet hotspots and wi-fi and UMTS are becoming more and more ubiquitous making it easy to go online. However, we can't take the 'always online, everywhere' for granted as a blog post by Chris Garrett titled 'Blogging On the Road … Or That Was the Plan!' points out:

> Best laid plans eh? I figured I had this mobile blogging thing sussed. With my trusty MacBook and PDA combo I am good for network, wifi, bluetooth, 3G/GPRS/UMTS. I didn't count on rural Manitoba. (Garrett, 2007)

This poses the interesting question whether it is possible to blog offline? Although blogging has almost become synonymous with being online there are two ways to blog offline. The first option is using a Weblog Client which is "software you run on your local machine (desktop) that lets you post to your blog via XML-RPC"[61]. The interface from a Weblog Client such as BlogJet or Windows Live Writer resembles the WordPress interface with a WYSIWYG-editor to format posts. You can write your post offline, completely format it, save a draft version, add a category or tags and publish it to your blog when you have online access. One of the main advantages is that you do not have to go online to write a post. This feature can prove convenient if you are in an environment where no internet connection is available or where it is still very expensive.

Weblog client software is mainly popular because of its offline writing features, easy-to-use markup features, easier YouTube embedding and saving posts for backup reasons[62]. It would be interesting to conduct further research on the use of offline blog

---

61   http://codex.wordpress.org/Weblog_Client (accessed January 12, 2008)
62   http://www.reviewsaurus.com/software-reviews/review-of-blogjet-simple-but-powerful-blogging-desktop-client/
http://www.thetechandcents.com/2007/11/blogjet-editor-in-linux-review.html

editors and how it influences or disrupts the blogging practice in terms of continuity (paragraph 3.1.1).

Another alternative is to install WordPress on your own computer instead of a external hosting server. There are various instruction manuals online that use the XAMPPLITE software to install WordPress on your own computer or a USB-stick. XAMPP is a software package that includes all the requirements for WordPress:

> XAMPP is a free, cross-platform standalone server, consisting mainly of the Apache HTTP Server, MySQL database, and interpreters for scripts written in the PHP and Perl programming languages. XAMPP's name is an acronym for X (any of four different operating systems), Apache, MySQL, PHP and Perl. (Wikipedia[63])

This method also works for installing WordPress on a USB-stick thus turning it into a portable version. There are several methods to synchronize the local files on your USB-stick and the files on your remote web server. This offline version of WordPress is very useful for both offline blogging and testing. The offline installation can be used as a test ground for trying out themes, plugins and widgets or modify them without the risk of breaking your blog due to erroneous CSS or PHP modifications (Dagza, 2007). The third requirement is the popular database MySQL version 4.0 or greater.

3.1.3 The Database and Code

3.1.3.1 The Database Form

Databases play an important role in our current age and Lev Manovich argues in *The Language of New Media* that the database logic is "a new symbolic form of the computer age [...] a new way to structure our experience of ourselves and of the world" (Manovich, 2001, p. 219). Manovich documents the dominance of the database form in new media and argues that the database form has really flourished on the Internet (p. 224). He defines the database as "a structured collection of data" (p. 218) and sees a web page as a structured collection of data because it consists of several elements that are defined by HTML.

---

http://nevon.typepad.com/nevon/2004/09/blogjet_and_ect_1.html
http://www.techcrunch.com/2006/08/13/windows-live-writer/ (accessed January 12, 2008)

63   http://en.wikipedia.org/wiki/XAMPP (accessed December 12, 2008)

The Hypertext Markup Language (HTML) is a static markup language which makes early web pages fairly simple databases that could be queried by search engines based on the HTML elements. Server sided scripting languages such as Perl and PHP launched a web filled with dynamic pages, including blogs, and rich databases that can be queried. In static web pages the database resides within the page itself as a structured collection of HTML elements. In dynamic web pages the database resides outside of the web page and content is generated after a query before sending it to a client. WordPress blogs use such an external database to store its content.

3.1.3.2 Formal Aspects and Limitations of WordPress

The second requirement for WordPress to run is a MySQL database. Before installing WordPress on your server you need to create a database on the server. A lot of hosting companies offer a "one-click" automatic installation of WordPress that will automatically create a database for the user. My hosting company supplies me with a control panel that enables me to manage my website. I can create subdomains, e-mail aliases, mailinglists, ftp access, backups and more from a simple control panel. The Powertools option includes a wide range of scripts/software to install with just a few clicks. The user only has to supply a location for the installation and accept the user agreement. This quick installation will automatically create a new database for you unless you choose to use an already existing database.

This easy automatic installation of WordPress means that some users may not even be aware of the existence of the database. This is especially the case for WordPress.com users that do not have to install WordPress software or configure a MySQL database but only have to supply a username and password. The fact is that the database is one of the most important parts of the blog because it stores most of the blog content and what is even more important the connections between the content.

User generated content including pictures and movies are often stored in the external databases such as Flickr or YouTube. These services provide different ways of integrating the externally hosted content in your blog. Flickr offers an HTML code[64]

---

64 <a href="http://www.flickr.com/photos/silvertje/2118492396/" title="Birthday Presents by silvertje, on Flickr"><img src="http://farm3.static.flickr.com/2401/2118492396_7cbb642e5d_o.jpg" width="900"

that can be pasted into the blog to embed a picture. It also offers a Blog This! button that sends the photo to the blog using the MetaWeblog API which is implemented as an XML-RPC web service.

Content is stored 'internally' in the wp-content/uploads folder on the server. Uploading content through the WordPress interface is restricted to 42 predetermined content types. Uploading an 'unknown' content (MIME[65]) type results in a "File type does not meet security guidelines" message. As it is nearly impossible to include every single MIME type administrators can set their preferences to unfiltered upload which enables them to upload anything. This also poses a potential security risk as viruses may be uploaded by other users, which is why file upload type is restricted by default.

By default both file type and file size upload are restricted in WordPress. The blogger is also restricted by the maximum amount of web space and traffic as determined by the web host. This may be a factor for bloggers choosing to use additional databases such as Flickr for storage. Blog posts in WordPress are stored in a MySQL database which is a relational database:

> Strictly speaking, a relational database is merely a collection of relations
> (frequently called tables). [...] A relation is defined as a set of tuples that all have
> the same attributes. This is usually represented by a table, which is data
> organized in rows and columns. In a relational database, all of the data stored in
> a column should be in the same domain (i.e. data type) (Wikipedia[66])

Illustration 3 shows the MySQL database, as depicted by phpMyAdmin[67], with the WordPress database tables on the left and the rows and columns within the wp_posts table on the right that contains all the information related to posts. A post consists of various fields of information such as ID, author, post date, category, ping status,

---

height="598" alt="Birthday Presents" />
65 "Multipurpose Internet Mail Extensions (MIME) is an Internet Standard that extends the format of e-mail to support:
   · text in character sets other than US-ASCII;
   · non-text attachments;
   · multi-part message bodies; and
   · header information in non-ASCII character sets.
77. [...] The content types defined by MIME standards are also of importance outside of e-mail, such as in communication protocols like HTTP for the World Wide Web. HTTP requires that data be transmitted in the context of e-mail-like messages, even though the data may not actually be e-mail." http://en.wikipedia.org/wiki/MIME (accessed January 27, 2008)
66 http://en.wikipedia.org/wiki/Relational_Database (accessed January 12, 2008)
67 "An administrator's tool of sorts, phpMyAdmin is a PHP script meant for giving users the ability to interact with their MySQL databases."
   http://codex.wordpress.org/phpMyAdmin (accessed January 27, 2008)
   As the database itself does not impose any structure on the data (see page 54), in this representation of the MySQL database structure on the data is imposed by phpMyAdmin.

51

content, etc. All these fields can be queried by PHP and new fields can be added. This allows for an enormous amount of possibilities in sorting and presenting data.
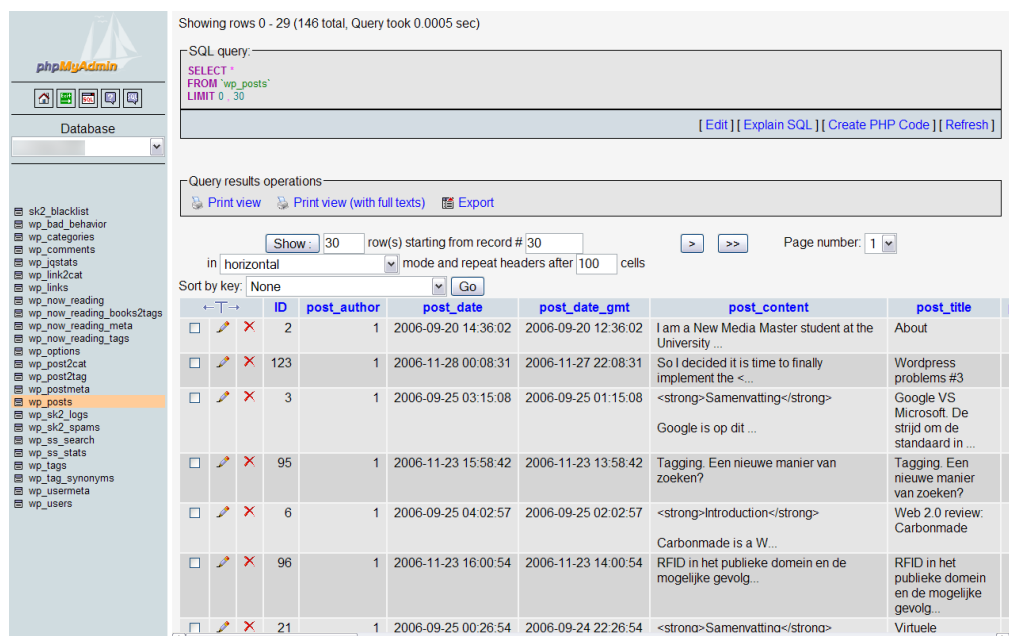

Illustration 3: MySQL blog database

WordPress communicates with the MySQL database using PHP queries. Both the WordPress software and the blog pages are written in PHP. The server-sided scripting language was developed as a set of Personal Home Page tools to solve common problems to create a web page with basic interactivity such as a guestbook in 1995 by Rasmus Lerdorf. Previously Perl was used to create dynamic homepages and Lerdorf developed PHP out of Perl to replace the restrictive use of Perl (Lerdorf).

The first dynamic homepages used Perl to write CGI scripts to create contact pages and guestbooks. The weblog software Movable Type has used Perl since 2001:

> I'm sure that the reasoning behind using Perl was two-fold: Ben Trott was a killer Perl programmer, and PHP wasn't installed on as many web servers (nor was it as popular) as it is now. (Rundle, 2006)

Matt Mullenweg, WordPress' founding developer, started to make a basic HTML web page. When he wanted to add some basic functionality such as a contact form and a guestbook he used other people's Perl scripts to modify them according to his needs. He found this modifying (of the CGI-bin and h_modding files) quite frustrating and when he discovered PHP he never moved back. Mullenweg actually used the Perl-

based Movable Type blog software before he switched to b2 (the precursor of WordPress) that is written in PHP.

Perl and PHP can both be used in blog software to create blogs. Which programming language is used depends on the software maker's personal preference. Movable Type's Ben Trott was a 'killer' Perl programmer and PHP is WordPress' Matt Mullenweg favorite language for web programming. Once he discovered PHP he never moved back to Perl (Nearthwort, 2007).

The adaptation of the easy-to-learn PHP partly explains the popularity of WordPress. WordPress relies heavily on its community that tests new versions, reports bugs, suggests different code solutions, writes plugins and writes themes. Many plugins and themes have been written by WordPress users in the PHP web scripting language that is easy to learn, read and adjust. The notion of plugins and themes will be further discussed in paragraph 3.2.5.

When we are posting a blog post we are actually communicating with the database by feeding the database fields. Nearly every field in the WordPress Write interface corresponds with fields in the database. WordPress is essentially the construction of an interface to the database. We need this interface to masque the complexity of the database.

> Users are for the most part not aware of the physical reality that structures the Internet. Tim Berners-Lee, the inventor of the worldwide web, writes that "The job of computers and networks is to get out of the way, to not be seen … The technology should be transparent, so we interact with it intuitively" (quoted in Galloway 2004: 65; Berners-Lee 1999: 159). Galloway likens this concealment of the inner workings of the Internet to other phenomena in which the apparatus of production remains hidden: the classic Hollywood film that masks continuity editing or the commodity that removes the process of production (Galloway 2004: 65). (Pietrobruno, 2007, p. 8)

Writing posts in the graphical user interface of WordPress is less complex than directly writing posts in the database. The intuitive interface masks the database and creates a continuous blogging experience within the browser.

The WordPress interface presents us different database tables in the same window; nearly everything related to the writing of a post is gathered under the 'Write Post' tab. The Write Post screen includes the title, content, tags, trackbacks, categories, author, timestamp and more. In the database some of these elements are stored in different tables. Whereas title, content, author and timestamp all belong to the wp_post table, wp_categories and wp_tags are separate tables (see illustration 4).

53
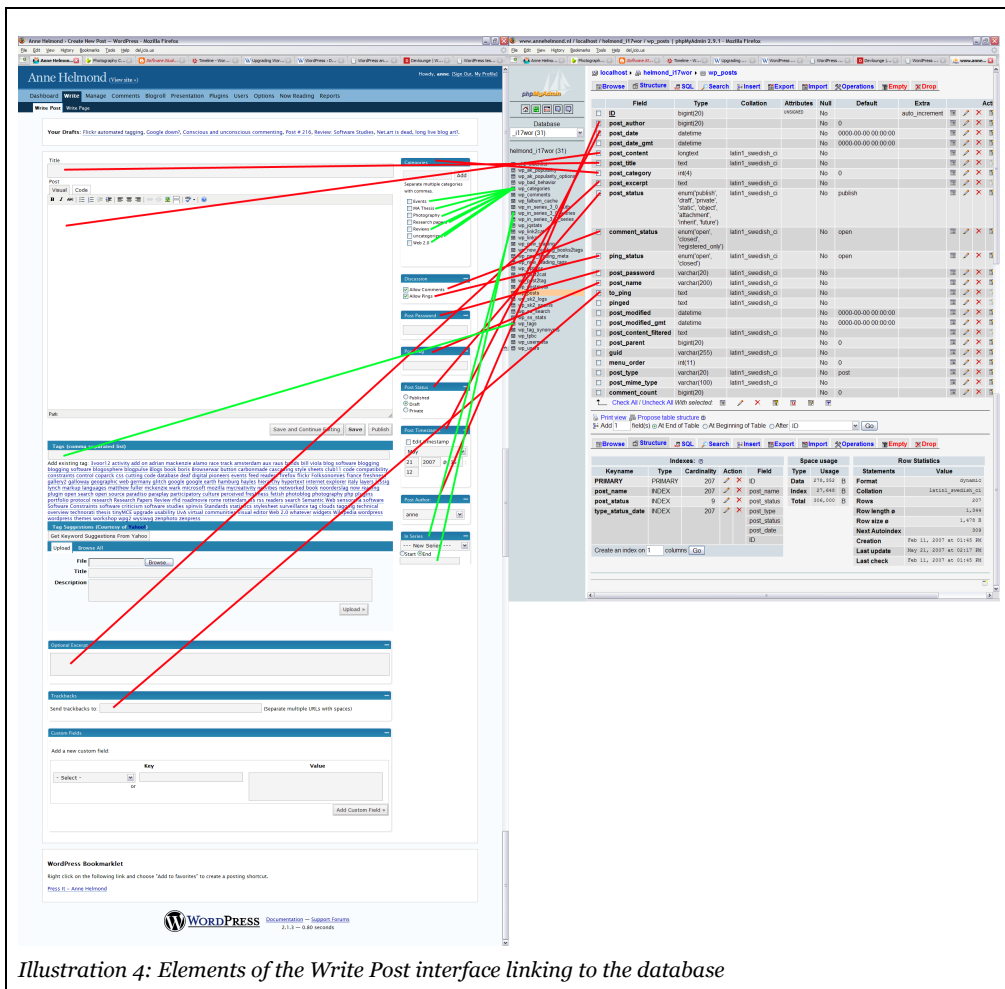
Illustration 4: Elements of the Write Post interface linking to the database

When publishing a post (or saving a draft version before publishing the post) different parts of the database are filled with information.

The database itself does not impose any structure on the information because "As a cultural form, the database represents the world as a list of items, and it refuses to order this list." (Manovich, 2001, p. 225) Structure is temporarily achieved when a user requests to read a blog post and the page is filled with the relevant fields of the database. Each page specifies which fields are requested from the database using PHP queries to extract the information. The visible structure is defined by theme and plugins used.

One of the characteristics that almost all blogs share is that blog posts are

arranged in chronological order that places the latest post on top. This is defined by a function on the main page (index.php) that tells the database to get all the posts and sort them by date in descending order. The function is defined in The Loop which is the main process of WordPress and defines the standard way blogs are seen.

The index.php of a theme contains The Loop which is the core of WordPress and refers to the main process of WordPress. The Loop contains the following code:

```php
<?php
get_header();
if (have_posts()):
    while (have_posts()):
        the_post();
        the_content();
    endwhile;
endif;
get_sidebar();
get_footer();
?>
```

It asks the database to get posts if there are any posts and if so, post them. How to post them is defined in the function get_posts($args) in post.php in an array which is a sequence of data elements.

```php
$defaults = array('numberposts' => 5, 'offset' => 0,
'category' => 0,
'orderby' => 'post_date', 'order' => 'DESC', 'include' =>
'', 'exclude' => '', 'meta_key' => '', 'meta_value' =>'',
'post_type' => 'post', 'post_status' => 'publish',
'post_parent' => 0);
```

The function tells the database to order the posts by date in descending order. This function is at the core of the functioning of WordPress and determines the ways blogs look. The reverse-chronological structure of the blog is coded into the heart of nearly every WordPress theme. Readers can sort the blog in different ways themselves such

as searching by category but because of this function the posts within the requested category will always be sorted in reverse-chronological order by date.

3.1.3.3 Alternative Uses

Two blog themes that do not conform to the regular way blogs are structured are the CLI1/CLI2 and Commodore themes written by Rod McFarland. The themes resemble a command line interface because

> if one considers the blog as a personal database, then as everyone knows databases are most efficiently queried via a command line. Well, maybe not everyone, but I find myself using the mysql command line more often than not. Once you learn the syntax you can do much more than you can through a traditional Web interface, which is essentially a wrapper around a few prepackaged SQL queries. (McFarland, 2007)
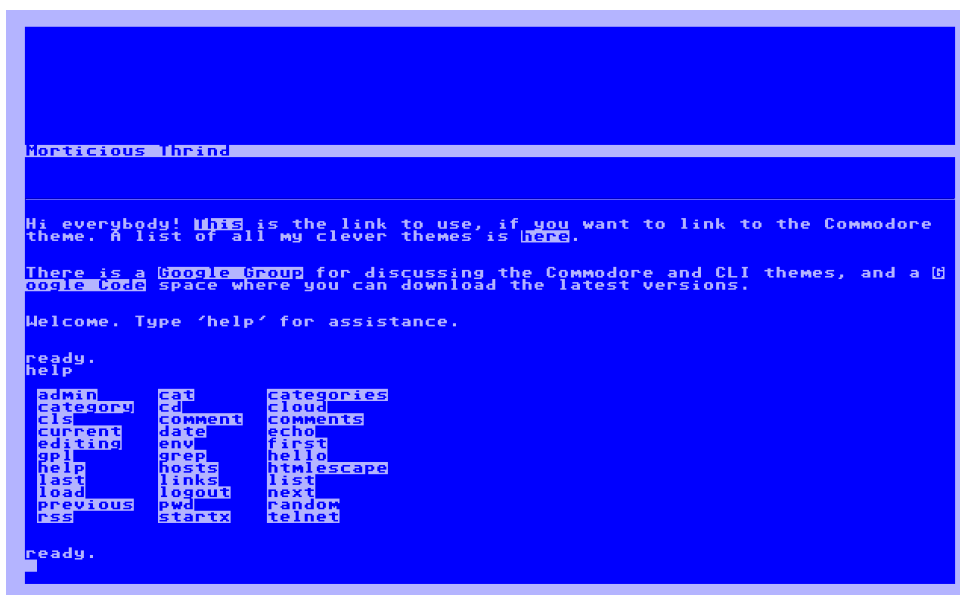
Illustration 5: The Commodore64 theme

These themes are written to display the content of the database in a way that defies the regular blog structure. McFarland 'breaks' The Loop that creates the reverse-chronological standard blog:

> Yes, I break "The Loop". Maybe it was the capitalization that made me attack it :) Architecturally, my choice to use header.php rather than index.php was arbitrary (probably I should just have used index.php). The Loop can't handle random access. There is a degenerate version of it (in lib/start-display.php) to show

whatever would have been shown on a regular page visit, but on subsequent calls to the blog via AJAX, there is no Loop. The main handler for commands is actually interpret.php (functions.php is used by the theme control panel under WP Admin). (McFarland, 2007)

You can browse the blog by using Unix commands such as "ls" to list article titles, "cd" to browse through categories to read blog entries, "cat [articlenum]" to read a post (Striegel, 2007). The theme uses JavaScript to translate the commands into WordPress commands in interpret.php to retrieve content from the database. This makes the theme very Search Engine Optimization (SEO) unfriendly:

> It's *horrible* for SE, Andrea. If you want Google hits, I'd recommend running a parallel blog with a conventional theme, like my /rod and /rodcli (hack wp-settings in your CLI version so that the theme directory is hard-coded, like wp-content/themes/cli/). CLI uses JavaScript for just about everything you can see on the page. Spiders won't see anything other than the minimal text you see on first load. (McFarland, 2007[68])

Illustration 6: The CLI 2.0 theme

Due to high demand the second release of the theme was made somewhat search engine friendly:

---

68   http://themes.wordpress.net/columns/1-column/1418/cli-10/ (accessed January 14, 2008).

> For the SE people: The theme will allow search engine indexing (CLI 1 made
> heavy use of JavaScript links). CLI 2 provides a subtle <A> link to a list of all
> articles. (McFarland, 2007[69])

Even though the theme defies the regular blog structure it is still 'more or less'
connected to the engines as the only question in the Frequently Answered Questions
shows:

> SEO?
>
> - Yeah, more or less. There's a link on the page (it's the cursor) that
>   generates a plain-link list of all the posts/pages in the blog. That should
>   be google-able, and if you've enabled 'show posts on first page' (or
>   whatever I called it) on the theme control panel, then when the spider
>   follows those links the post text will be indexed.
>
> - Or, if you're super-concerned about your PageRank, you could sniff for
>   'bot' or 'Slurp' in the user-agent and automatically generate a plain-text
>   version of the blog rather than a CLI version. I haven't found this
>   necessary though. (http://code.google.com/p/wordpress-cli/wiki/FAQ)

Themes allow for the customization of the display of the database behind WordPress.
This theme is not meant to be a 'useful' WordPress theme but it plays with the limits
of WordPress' malleability. WordPress Themes can be 'manipulated' to the extent that
they adhere to certain rules (see paragraph 3.2.7). The theme questions how we
perceive blogs as they have a very distinct look and feel that is completely different
from a unix-like command line approach:

> I liked the aesthetic, mostly, but after that I was seduced by the efficiency. And if
> a blog is really mostly about constructing one's own identity in the virtual world,
> then I really don't see why facilitating one's own memory would be a wrong thing
> to include with the same software. (McFarland, 2007)

The command line interface themes break with continuity as the visitor is taken out of
the familiar blog experience in which the database is masked.

> Totally cool and also almost totally unusable. I mean, if you don't want people to
> read your blog, ever, use that theme. (Theme Viewer, Comment on CLI 1.0,
> 2007[70])

The command line interface is usually found within an operating system or a software
applications. It is not expected to be stumbled upon while browsing the world wide

---

69 http://themes.wordpress.net/columns/1-column/1630/cli-20/ (accessed January 14,
    2008).
70 http://themes.wordpress.net/columns/1-column/1418/cli-10/ (accessed January 14,
    2008).

web.

> At first I thought that there was an error in the themes files; then I noticed.
> (Ibidem)

It is not a theme for general consumption, as most blog readers will not be familiar with a command line interface. The theme was developed partly because of a personal preference to access the WordPress database with familiar queries and the personal usage aspect of the theme is shared with other users:

> There's a real cool Unix command line theme I like, but most people probably
> won't be able to use it - still, I might do it anyway. Cause it's for ME, not YOU!
> (Ibidem)

The theme breaks with the standard blog way blogs are displayed and experience. It breaks with continuity as a tester of the theme notes that "When I finished playing around with it, I even typed "exit" because my mind was in shell mode." (Ibidem) No longer did the user feel he was visiting a blog but rather was browsing through an operating system.

3.1.3.4 The Databody

Whereas Manovich embraces the database as a new symbolic form, Mark Poster points to the drawbacks of the database in the sense that it restricts meaning:

> I content that the database imposes a new language on top of those already
> existing and that it is an impoverished, limited language, one that uses the norm
> to constitute individuals and define deviants. A database arranges information in
> rigidly defined categories, or fields (Poster, 1990, p. 95)

We can see these rigidly defined categories or fields in the WordPress MySQL database. The properties of the fields are restricted in the sense that some fields can only contain numbers, dates or letters and can have a restriction in size. For example the authors of a blog are reduced to an integer and categories are restricted in size to 55 characters.

By adding data to the database we are not only feeding our own WordPress MySQL database but also other databases such as Google or Technorati. Once a blog post is published it can be crawled[71] by search engines and other indexes. We do not

---

71   "A web crawler (also known as a web spider or web robot) is a program or automated
     script which browses the World Wide Web in a methodical, automated manner. Other less
     frequently used names for web crawlers are ants, automatic indexers, bots, and worms

know how they read our data nor how they interpret and store it. Not only do they have their own rigid defined fields but

> [...] the structure or grammar of the database *creates* relationships among pieces of information that do not exist in those relationships outside of the database. In this sense databases constitute individuals by manipulating relationships between bits of information. (Ibidem, p. 96)

Our data is very likely used to create new user profiles and improve search mechanisms. By blogging we are feeding our data body that is "the total collection of files connected to an individual" (Critical Art Ensemble, p. 145) These files are not only stored in our own database but also dispersed across different other databases by techniques such as RSS and pingback. Our database is part of a wider network of databases that we consciously (by pinging) or unconsciously (by being crawled) become part of. These other databases can create new relationships from our data and combine data from other sources. We cannot completely control where our data goes or what is being done with it and Poster warns us about this:

> We see databases not as an invasion of privacy, as a threat to a centered individual, but as the multiplication of the individual, the constitution of an additional self, one that may be acted upon to the detriment of the "real" self without that "real" self ever being aware of what is happening. (Ibidem, p. 97-98)[72]

Blogs contribute to this multiplication of the individual and the creation of an additional self. This self is both consciously and unconsciously indexed and created by the engines. The second part of this chapter will show how the blog is connected and automatically feeding the engines through its distinctive features.

(Kobayashi and Takeda, 2000)."
"This process is called web crawling or spidering. Many sites, in particular search engines, use spidering as a means of providing up-to-date data. Web crawlers are mainly used to create a copy of all the visited pages for later processing by a search engine that will index the downloaded pages to provide fast searches."
http://en.wikipedia.org/wiki/Web_crawler (accessed January 6, 2008).

72 It may be argued that we are more and more aware of databases and engines that store and serve the additional self. We are becoming more aware of the fact that an engines as Google is used to collect information on the individual in preparation of a job interview. Also 'Social-Networking Sites Catch the Eye of Employers'

3.2 Blog Features and the Engines

Blogs contain several distinguishing features allowing them to disperse and distribute themselves through syndication, aggregation and dispersion. Blog software has standardized these features and, for example, TrackBack, Pingback and RSS have become official standards while the RSS-alternative Atom was filed as an Proposed Standard with the Internet Engineering Task Force. [73]

Some of the canonical features of the blog are directly connected to the engines and others contribute to the construction of the blogosphere through the engines. This part will depict how the blog and the blogosphere are constructed through the typical features and protocols of the blog.

3.2.1 Posts and Pages

The post is a native blog format "which utilizes the medium to its best advantage by proffering frequent updates and richly hyperlinked text." (Hourihan, 2002) The regular blog chronology consists of time-stamped posts in reverse chronological order on top of which a new blog post will be placed. "Posts are part of an ongoing document that is categorised and usually related to a point in time " (WordPress.com)

Pages act differently as they reside outside of the blog chronology and are "not part of that same ongoing document but they complement it, add to it, provide additional information. They are not categorised. They may not be related to time."[74] Posts are part of the blog chronology and part of a dynamic document. In contrast, pages are not part of the blog chronology and reside outside of the dynamic document.

---

73   The Atom standard has been developed by a large group of people but the actual request for approval as an Internet Standard, the RFC[1] has been filed by Google.[1] As of 2003 Google started using Atom as the syndication format for most of its services such as Blogger, Google News and Gmail. This was the start of a syndication Some of these features are directly connected to the engines and others contribute to the construction of the blogosphere through the engines.

   1 Request for Comments (RFC) documents are a series of memoranda encompassing new research, innovations, and methodologies applicable to Internet technologies."
   http://en.wikipedia.org/wiki/Request_for_Comments (accessed 24 January, 2008)
   2 http://tools.ietf.org/html/rfc5023 (accessed 24 January, 2008)

74   http://faq.wordpress.com/2006/11/11/a-post-and-a-page/ (accessed December 2, 2007)

A recent Google Alert[75] showed me that Google also makes a distinction between posts and pages and indexes them differently. Google lists my post in the Google Blog Search index while it lists my pages in the Google Web index. WordPress and Google are reinforcing a static (page)/dynamic (post) distinction that is segregating both the blog and the web. Posts are listed in Google Web Search and Google Blog Search but pages are only listed in Google Web Search.

Pages are not considered to be part of the blog or the blogosphere by both WordPress and Google. Google Blog Search only indexes blogs by its site feed[76] and because pages "are not syndicated through the top-level feed channel like posts are (that is, the latest posts are sequentially listed as items in the feed)" (Kafkaesqui, 2007) they are not included in the Google Blog Search index. WordPress uses RSS to syndicate through the "top-level feed channel" and paragraph 3.2.4 will explain how it acts as a centralizing force.

Both Google and WordPress see pages as an additional element that is not part of the blog itself. Technorati too does not list pages in its index, only posts. It does, however, show only the latest posts from its index. Technorati has recently decided to drop posts that are over six months old from its search results:

> We're in the midst of some economization, performance fixes and retooling that have required taking some data offline. The data is not lost but our priorities are to prefer keeping recent data online. Most people don't notice :) (Kallen, 2007)

Technorati decided to display the old posts from the index no longer because of 'economical' reasons. Not only does this point to an increasing focus on freshness but also to the indexing problems that Technorati seems to be facing. This poses the question whether the engines can handle a maturing blogosphere (Helmond, 2007a). As there are over 175,000 new blogs being created every day, the engines face the challenge of indexing more and more blogs, more and more blog posts and even more links.

---

75 "Google Alerts are email updates of the latest relevant Google results (web, news, etc.) based on your choice of query or topic." http://www.google.com/alerts
    I receive an e-mail whenever my name "Anne Helmond" is mentioned in either Google News, Blogs, Web, Video or Groups.
76 A site feed is the address of the site's syndication feed, usually based on the RSS, Atom or RDF standard/protocol. Google Blog Search includes blogs that publish a feed in either RSS or Atom. http://www.google.com/help/about_blogsearch.html

3.2.2 Links

The early practices of blogging included logging links to other blogs of interest and referencing other blogs. This has inflicted blogging with a notion of sociality but

> As Rebecca Blood has argued, acknowledging sources in blogs is not necessarily informed by a deeply thought-through ethics, but rather a norm of social expectations. (Leaver, 2006)

The popular practice of linking has given birth to new formats of the link that are native to the blog, namely the permalink, the trackback and the pingback. These formats allow the blog to disperse, connect and syndicate itself. It allows for communication with other blogs and with the engines. A new attribute of the link, nofollow, originated as an initiative of the engines and was adopted by blog software makers.

The practice of linking is clearly visible in the blogroll which is a collection of links to other blogs, usually located in the sidebar of the blog. Links in blogrolls are either regarded as recommendations or as a personal list of blogs the blogger reads on a regular basis:

> This form evolved early in the development of the medium both as a type of social acknowledgement and as a navigational tool for readers to find other authors with similar interests. While the term pays homage to the practice of logrolling (the exchange of political favors and influence), a link within a blogroll indicates a general social awareness on behalf of the author. (Marlow, 2004, p.3)

The blogroll requires attention as it needs to be updated and reflected upon once in a while:

> A link made on a blogroll is made explicitly as a statement of social affiliation. By placing a link to another weblog, one assumes that the author either endorses that weblog, wishes to promote it, or claims to read it on a regular basis. Blogroll links are also updated much less frequently than the weblog itself, allowing these references to go out of date. Furthermore, once a link is made, there is a disincentive for removing it, as one feels guilty about taking traffic away from someone who was once an associate. (Marlow, 2004, p.4)

The blogroll is a characteristic of the blog that seems to be slowly disappearing with the maturing of the medium. There are several incentives to remove the blogroll from the blog. First, some bloggers prefer to link to other bloggers in blog posts instead of in the blogroll. Second, a long list of links in the sidebar disrupts the design of the blog. Third, the list needs constant updating in order to reflect a representative list. Fourth, constant requests for 'link exchanges' in the blogroll and the pressure of social acknowledgements. danah boyd decided to remove her blogroll as the maintenance

aspect of a blogroll outweighed the social aspect:

> Thus, my blogroll quickly became odd - it was a collection of the professional blogs that i read and those that i felt as though i should be reading more often. Plus, there were people who asked to be on my blogroll and then there were those who were upset because they weren't listed (even though i read their blog daily). It became weird, outdated and nowhere near representative of me.
>
> Seeing as this is the most personal of my public blogs, i decided that there is no need to list off some strange aspect of my identity in order to feel socially appropriate. (boyd, 2003)

A final reason for removing the blogroll is that the function of the blogroll as a list of blogs that the blogger reads on a regular basis has changed with the advent of site feeds. Bloggers now keep up with other blogs by subscribing to their feeds which makes this function of the blogroll less needed. The blogroll originated within the blog community as a social acknowledgement through linking.

The permalink, the trackback and the pingback also originated within the blog community and were adopted by the blog software community to ease and automate the practice of linking. In popular blog software links are placed with the simple click of a button in the WYSIWYG-editor to insert the URL instead of manually inserting the HTML code. While this means another move away from the code[77] leaving a link in the comments still requires HTML knowledge. Writing posts in WordPress is done with the help of a WYSIWYG-editor, which is enabled by default, but comment fields on blogs rarely have such editors.

This paragraph addresses the relationship between the software and the engines by looking at the new link types that originated from the practice of linking in blogs.

---

[77] On December 31, 2005 WordPress 2.0 was released which included major changes in the writing interface that is considered to be the "core of blogging" by the WordPress developers (Mullenweg, 2005: http://www.wordpress.org/development/2005/12/wp2/ (accessed December 2, 2007).). The WYSIWYG editor from TinyMCE was implemented and was turned on by default. It signifies a move away from code as bold or italic text or a link could be created with the push of a familiar (word processing software) button instead of using the right HTML code. Version 2.1 would include a tabbed editor that "allows you to switch between WYSIWYG and code editing instantly while writing a post." (WordPress 2.1 Ella, 2007: http://www.wordpress.org/development/2007/01/ella-21/ (accessed December 2, 2007).)

3.2.2.1 The Permalink

The permalink is the permanent link for a blog post. It originated from a practice that allowed fellow bloggers to point to a specific entry. Jason Kottke[78] was among one of the first bloggers to implement a permanent link on his blog:

> Finally. Did you notice the little icon next to today's date? It links to the permanent URL for this entry. Use that URL to bookmark or link to this particular entry. (Kottke, 2000)

The feature was noticed the next day by another blogger who wrote:

> by the way, Jason implements the feature I'd love to see in Blogger: persistent references. (van Oosten de Boer, 2000)

This, in turn, was picked up by Matt Haughey who

> had discussed a permalink style feature with Blogger co-founders, Evan Williams and Paul Bausch the previous weekend and Paul Bausch had pointed out that it was technically feasible to produce permanent links in Blogger, using a feature written by him that allowed the ID of a post to be placed in a Blogger template. In response to Kottke's blog, on March 6, 2000, Matt Haughey posted the technical details on his own weblog, which helped open the way to widespread adoption. (Wikipedia[79])

Not long after van Oosten de Boer's call for the implementation of 'persistent references' and the publication of the technical specification of the format by Blogger employee Matt Haughey, Blogger incorporated the feature. The implementation of the permalink into the software can be seen as the incorporation of a blogging practice that was becoming very popular and widespreading:

> It was created by everyone working there, and from listening intently to ideas from people using Blogger. (Bausch, 2003)

The permalink was quickly adopted by other bloggers and other blog software and

> In a medium whose currency is links, weblogs without permalinks were at a sudden disadvantage. Hand-coders had to invent ways to reproduce this feature if they wanted to be referenced on other blogs. (Blood, 2004, p. 54-55)

There is no standard format for the permalink and different blog software generates different permalink structures. WordPress offers three types of permalink structures:

- Default (nicknamed "Ugly") » http://example.com/?p=N
- Date and name based (nicknamed "Pretty") »

---

78   Then boyfriend of, now husband of Meg Hourihan who co-founded Blogger.
79   http://en.wikipedia.org/wiki/Permalink (accessed December 3, 2007)

65

http://example.com/year/month/day/post-name
- Numeric » http://example.com/archives/?p=N
- Custom

Users can choose their own permalink structure but the date and name based structure seems to be most common. Not only are "pretty" permalinks easier to read for humans, the structure of the permalink is also very important in relation to the engines. Search engines give importance to words in URLs and in SEO techniques for optimizing your blog for the engines it is often advised to remove date information from the permalink. Google engineer Matt Cutts simply states "I wouldn't bother with year/month/day in blog urls; I'd just use the first few words from the title of the post in the url." (Cutts in Wilson, 2005) While the goal of this thesis is not to write a manual for SEO techniques it shows the tight relationship between the software and the engines.

Permalinks have become a very important feature in blogs and serve the fundamental unit of the blog, the post:

> blogs are built on a post paradigm instead of a page paradigm that the web was
> originally built upon. I really think permalinks made that distinction clear and
> allowed the by-post model to blossom. It made every contribution to a site
> behave like a distinct unit that could be linked to, quoted from, and extracted. It
> is funny how we don't even think about it anymore. (Haughey, 2003)

The permalink treats the blog post as "a discrete object -- with its own link -- within the context *of* a page." (Hourihan, 2003) It also started treating the blog post comment as a discrete object with the adaptation of permalinks for comments.

The permalink has changed the medium of blogs and practice of blogging as it offers a stable reference point which makes cross-blog talk and trackback and pingback possible. The idea of a permanent link was not new[80] but the permalink was widely implemented by bloggers to deal with the distinct feature of the blog: the post. The permalink has blossomed by blogs and blog software and has now generally been adopted on the web[81] as a standard.

---

80 Berners-Lee, Tim. "Cool URIs don't change." *Hypertext Style*, 1998.
    http://www.w3.org/Provider/Style/URI (accessed December 7, 2007).
81 Websites such as The New York Times also offer permalinks to its articles. Its news articles
    resemble the blog post as a discrete object that can later be moved into the archive:
    "Using this link will ensure access to the article, even after it becomes part of the NYT
    archive." Popup (click Share + Permalink) on the http://www.newyorktimes.com
    (accessed December 7, 2007)

3.2.2.2 Trackback and Pingback

WordPress uses the permalink to construct the format of the trackback which is:

> a way to notify a website when you publish an entry that references it. When you send it a trackback, a link with a short excerpt of your entry will appear on the referenced website. (Adlam, 2005)

Trackback is a manual notification system because blogger A has to insert the relevant trackback URLs in a field underneath the post. Blogs can be configured to receive trackbacks automatically which means that if blogger A sends a trackback to blogger B it will automatically appear on B's blog.

WordPress has been enabled to automatically receive trackbacks by default. This has made trackbacks very vulnerable to spam as spammers will send a trackback which includes a link to their site thus boosting their ranking in the engines.

Trackback was developed to facilitate blog-to-blog communication but spammers found a way to exploit the trackback protocol in order to gain higher rankings in the engines. The increasing amount of comment, trackback and pingback spam gave birth to a new link attribute, the nofollow, which was developed by the engines and blog software community in an attempt to prevent spam. Trackback spam and the nofollow attribute and its relation to the engines will be further discussed in paragraph 4.3.1 and 4.3.2.

Pingback was developed to address the flaws of trackback. Pingback is also an automatic notification system that allows you "to notify a weblog of your entry just by posting its permalink directly in the content of your blog entry. No special trackback link necessary." (Adlam, 2005) The main difference between trackback and pingback is that a pingback is sent out automatically and it includes an automatic verification process from the pinged blog to verify the existence of the link. Pingback turned out not to be the solution as spammers found a way to exploit the protocol of pingback too.

WordPress' default settings are set to receive trackbacks and to send/receive pingbacks automatically. These default settings reinforce the self-referential and self-sustaining nature of the blogosphere.

3.2.2.3 The Social Bookmarking Link

A new feature exploiting the unit of the post and its (perma)link to the fullest is the social bookmarking link. It allows the blog to disperse itself across the network and the engines. The social bookmarking link acts as fuel for the engines as it feeds the engines with data by sharing links and posts.

The social bookmarking link can be used to submit a post to an engine. The amount of engines is growing every day and also the amount of buttons of services to submit social bookmarking links to for example Digg, Slashdot, Technorati Favorites, Spurl, Furl, Sphere, del.icio.us, Reddit, Newsvine, Stumble Upon, Facebook and Sphinn (see illustration 7).

The ShareThis button (available as a WordPress plugin) strives to be a one button solution by providing 'all' these services under a single click. It acts as a centralizing service to store all your social bookmarks in one place.

These buttons serve two main purposes: bookmarking and popularizing (achieving a higher ranking in the engines). A button to a social bookmarking site such as del.icio.us enables the blog reader to easily bookmark the post to the engine for further reference. More often these links are added in an attempt to boost traffic to the blog. Adding a blog post to an engine such as Digg may result in increasing traffic to the blog.

Engines like Digg act as a centralizing force for popular stories often contributed by blogs. Different types of web content including blog posts are submitted to Digg and may end up on the front page if they receives enough votes (Diggs.) The front page of Digg is watched by a large number of people and a link there can result in a lot of visitors which may lead to the 'Digg effect.' The phrase 'Digg effect' is derived from 'Slashdot effect' which is a "the term given to the phenomenon of a popular website linking to a smaller site, causing the smaller site to slow down or even temporarily close due to the increased traffic." (Wikipedia[82])

The social bookmarking link acts on the dispersive nature of the blog that allows itself to be included in the engines. This link is often linked to the engines by a specific form of annotation, the tag, as will be described in paragraph 3.2.5.

---

82   http://en.wikipedia.org/wiki/Slashdot_effect (accessed January 5, 2008)

Illustration 7: The social bookmarking link

3.2.3 Ping

Ping[83] plays a central role in the aggregation and dispersal of the blog. If a new blog post is published (or an old post is updated) WordPress automatically sends an XML-RPC signal, or a 'ping', to one or more ping servers. The initiation of the first ping server, Weblogs.com, marked an important moment in blog history since ping servers are now considered

> [...] the backbone of the blogosphere infrastructure and are a crucially important piece of the real-time web.

Ping servers are the backbone of the blogosphere infrastructure because they act as centralizing forces in the distributed network. They collect the most frequently published posts which may be used by the engines to update their indices[84].

> [...] Ping servers sit between search engines and blogs. Without ping servers, search engines like feedster, technorati and pubsub would not know if and when blogs and other content sources updated. They would regularly have to index these sites to find out if they've updated. With over 10,000,000 blogs, re-indexing every hour would be a massive undertaking - every few minutes, impossible. Instead, they re-index only when a source notifies the ping server that it has updated. It is a much more efficient way for the real-time web to keep "real-time" with only a few minute delay between posting and indexing.
>
> There are many ping servers today (see links below for lists), but Weblogs.com was the first and is by far the largest, as it is built into nearly all blogging software. (Arrington, 2005)

Ping servers act as an intermediary between search engines and blogs. Ping has been built into WordPress since its beginning and it currently pings one server by default, Ping-o-Matic. The Ping-o-Matic server was created by WordPress' Matt Mullenweg in order to automate manual pinging:

> At first, it was focused at people who were pinging manually and had to visit ten different sites to do so, then we rolled out the XML-RPC weblogs.com interface so publishing software could talk to Ping-O-Matic directly. We provide a super-fast single ping people can send and we mirror the data to whoever can handle it. (Evans, 2004)

Ping-o-Matic acts as a meta-server which relays your ping to other update services

---

83 Ping is frequently know as "a computer network tool used to test whether a particular host is reachable across an IP network." http://en.wikipedia.org/wiki/Ping (accessed January 5, 2008). In blogging it is used as a notification method.
84 Depending on the openness of the ping server. The Weblogs.com ping server is an openly available service but Technorati's ping server is closed and only used to update Technorati's own index.

including Weblogs.com. "In turn, Update Services process the ping and updates their proprietary indices with *your* update. Now people browsing sites such as Technorati or Sphere can find your most recent posts!"

Most ping services supply extensive documentation on how to ping their service. Technorati for example includes information on how to configure your blog software to ping Technorati directly. It also supplies the code to build the ping request into custom blog software or other services. The services rely on blogs and blog software to ping their servers and supply extensive instructions to do so.

Ping is used as an important notification system for the engines that a blog has been updated. RSS is used as a publication method to notify your blog readers that your blog has been updated.

3.2.4 Feeds

Feeds are at the core of the syndication of the blog:

> A feed is a function of special software that allows feed readers to access a site, automatically looking for new content and then posting the information about new content and updates to another site. This provides a way for users to keep up with the latest and hottest information posted on different blogging sites.
>
> There are several different kinds of feeds, read by different feed readers. Some feeds include RSS (alternately defined as "Rich Site Summary" or "Really Simple Syndication"), Atom or RDF files. (WordPress Codex[85])

Blogs are constantly updated and feeds have made it easier for readers to be notified of updates. Instead of going to a blog every single day to see if it has been updated a feed reader can be used that will automatically look for new content. Feed readers are used to manage blog subscriptions. They are a personal centralizing service that allows you to keep track of the tens or hundreds of blogs you may be reading.

Instead of visiting every single one of them to look for updates, the feed reader will collect recent updates. Feeds are not send out but retrieved and updates are handled upon request by feed readers and engines. This means that if a blog post has been edited or updated after you have retrieved the feed you will not be notified of any changes. Even though it is rare, it may happen that significant changes have been made after you retrieved the blog's feed.

---

85   http://codex.wordpress.org/WordPress_Feeds (accessed December 8, 2007)

I use Netvibes to keep up with my feed subscriptions which is a personalized start page with integrated feed reader. It allows me to organize my feeds per customized category. An extra browser plugin detects a blog's site feed automatically and adds it to Netvibes with a single click[86].

Feed readers have changed the perception of blogs as blogs are often accessed through feed readers instead of the user visiting the actual blog. Netvibes offers two display options: a 'Feed view' or 'Show website.' The feed view shows the blog post in plain text and the website view shows the blog post as it appears on the blog with the complete markup. A plain text view speeds up the process of keeping up with all blog subscriptions but it is sometimes considered a lack of respect to the blog's design.

The WordPress default site feeds only syndicates the five most recent posts[87], which reinforces the distinct unit of the post. Comments are offered in a separate feed and pages are not syndicated at all. Blogs often use partial feeds that display only a part of the total blog post in an attempt to lure readers back to the blog itself where advertisements may generate money for the blogger. Partial feeds also offer minor protection from the blog being syndicated for other purposes (see paragraph x).

A site feed acts as a centralizing force in the blog because it collects information on the most recently updated posts of the blog in one place, the feed. By default WordPress gathers the five most recent posts in its syndication feed. This feed can then be accessed by different sources such as feed readers and engines.

The default site feed generated by WordPress does not include the additionally added social bookmarking links. These have to be manually coded into the feed or by an external service that can add these links for you[88]. A new trend in feeds is to redirect your feed to an external service that can monitor your feed, supply you with feed statistics and offer feed customization. It acts as a centralizing service to manage

---

86   In 2003, feed readers were envisioned by Wired as the answer to 'manage information overload.' The integration of an 'add to Netvibes' button in the browser has made it so easy to subscribe to a feed that they seem far removed from the answer. http://www.wired.com/techbiz/it/news/2003/08/60053 (accessed December 11, 2007)

87   WordPress also offers a separate feed for comments.

88   On December 20, 2007 WordPress.com released 'Enhanced Feeds that enables WordPress.com users to finetune their feed with the following widgets: 1. Comment Count, 2. Categories, 3. Tags, 4. "Add to Del.icio.us" link, 5. "Digg it!" link, 6. "Stumble it!" link. Comment Count is turned on by default but you can switch all the features on or off as you please. http://www.bloggingpro.com/archives/2007/12/20/wordpresscom-enhanced-feeds/ (accessed December 25, 2007).
WordPress.com is the hosted service of WordPress and is often considered a test ground for the implementation of new features. This new feature acknowledges that blogs are often read in feed readers only and acknowledge their connection to the engines.

your distributed blog subscribers. It does not provide information on exactly who is subscribed to your feed but it keeps track of the number of subscribers, the type of feed reader used to read the feed and the source of the latest traffic.

The service also allows the blogger to easily add additional information to the feed. Feedburner offers a service named 'FeedFlare' that

> allows publishers to easily build "interactivity" into the content they create, making it simple for subscribers to tag, email or share their content with others. Publishers can include a variety of services including a live display of the number of comments to each post, the ability to email the author directly and show the number of blogs that link to their item. (FeedFlare Overview and FAQ[89])

In the same way as social bookmarking links can be added to the blog post to connect to the engines, FeedFlare adds these links to the blog's feed and it

> provides publishers with an easy way to tap into many of the popular social media services available today such as Digg, del.icio.us, Sphere and Facebook. They now can associate other Web services with each content item, bringing them closer to their readers and helping to expand their own reach online, drive traffic to their site, attract advertising or promote a personal cause. (Ibidem)

FeedFlare is one of the many ways in which a feed is directly connected to the engines. Adding social bookmarking links to the feed takes into account that blog posts are not always read on the blog itself but by feed readers as well. The same goes for adding advertisements to feeds which Google probably realized as well. On June 3, 2007 Feedburner was bought by Google and feeds were added as the perfect advertisement distribution platform to its portfolio.

3.2.5 Categories and Tags

Blog software standards and protocols enable the communication between the blog and the engines. Category names in the blog post URL serve as keywords for the engines and Technorati uses categories as tags. Technorati automatically translates categories into tags for its engine[90]. It also offers bloggers the correct code syntax to establish communication on the Technorati channel for supplying tags. Technorati has developed the `rel="tag"` HTML attribute that

---

89 http://www.feedburner.com/fb/a/publishers/feedflare (accessed January 26, 2008)
90 While writing this thesis Technorati changed updated their site and their usage of tags a few times. Technorati currently uses the blog's tags for their Tag Search Page: http://technorati.com/tag (accessed January 26, 2008)

indicates that the destination of that hyperlink is considered a user designated "tag" of the current page, or major portion of the current page (i.e. a blog post), e.g. by placing this link on a page,

```
<a href="http://technorati.com/tag/tech" rel="tag">tech</a>
```

the author indicates that the page has the tag "tech". (Çelik, 2005)

This attribute has been implemented by WordPress[91] pointing to a two-fold relationship between Technorati and WordPress. Before the implementation of tags in WordPress Technorati read WordPress' categories. After the addition of tags WordPress used Technorati's established `rel="tag"` micro-format to identify tags. Technorati standardized the tag attribute in order to supply a communication channel for the blog software:

> Protocol is synonymous with possibility [...] Protocol outlines the playing field for what can happen, and where. If one chooses to ignore a certain protocol, then it becomes impossible to communicate on that particular channel. No protocol, no connection. (Galloway, 2004, p. 167)

Tags can only be communicated with Technorati by using the `rel="tag"`.

Blog posts are often filed under categories for navigational and search driven purposes. Not only do categories guide a blog visitor through the blog but they also guide the search engine's crawlers. Tags are used to provide meta information about the blog post that extends beyond categories. Tags describe the content of the blog post and tags are used by the engines to index blog posts. Tagging blog posts is different from other tagging practices on the web as the object that is being tagged, the blog, is only being tagged by one person, the blogger. Websites on the social bookmarking site del.icio.us are tagged by different people, creating a folksonomy[92]. Blog posts are only tagged by the 'owner' of the object, the blogger, creating a personal taxonomy. Problems arise when these tags are used by Technorati on the Tag page as it pulls the tags out of the context of the blog:

> Using tags from one blog to another blog has problems for the same reason that Bitzi and all other publisher tags have and had problems, they are missing identity of the tagger AND a clear common object being tagged. (Vander Wal, 2007)

---

91 http://trac.wordpress.org/changeset/5568 (accessed December 25, 2007).
92 "Folksonomy is the result of personal free tagging of information and objects (anything with a URL) for one's own retrieval. The tagging is done in a social environment (usually shared and open to others). Folksonomy is created from the act of tagging by the person consuming the information."
http://www.vanderwal.net/folksonomy.html (accessed 27 January, 2008)

74

On top of that there are "many variations of tag use surfacing or tag terms being used to attract traffic for non-related content." (Vander Wal, 2007) The tags are displayed in a tagcloud to depict the 100 most popular tags. Tag clouds are also often used on individual blogs.

A tag cloud is a visualized list of tags in the form of weighted or sized tag cloud to display the content of a blog. The bigger the tag in the cloud, the more prevalent the tag is used on the blog. A tagcloud can quickly visualize the interests of the blogger and the focus of the blog (see illustration 8).

Illustration 8: Tag cloud from the Masters of Media blog. January 27, 2008.
http://mastersofmedia.hum.uva.nl/

Several WordPress plugins were developed that made tagging posts and displaying tag clouds possible. Two widely used tagging plugins were the Ultimate Tag Warrior and Jerome's Keyword plugin. These plugins were very popular and the integration of tagging into the core of WordPress was a widely heard request. WordPress finally implemented tagging in version 2.3 including a feature to import old tags.

By adding widgets and tags to its core WordPress is increasingly entangled in a web of relations with the engines. The relationship between the software and the engines is two-fold: the software embraces the engines and the engines embrace the protocols within the software. WordPress implicitly acknowledges this relationship by implementing features that connect to the engines but it also explicitly states that "WordPress, straight out of the box, comes ready to embrace search engines." The default settings in WordPress, such as providing feeds and pinging the engines, feed your blog to the engines. WordPress is the engine's friend by default as acknowledged by Google's Matt Cutts who praised WordPress for its efforts as "software that is made to do SEO pretty well." (Jaquith, 2007) In the third phase of the blogs as a form of syndication, aggregation and dispersion the blog software is set up to syndicate and disperse the blog and automatically include it in the engines.

75

WordPress does offer an opt-out feature under Options >> Privacy where the 'visibility' of the blog can be controlled. By default the blog is set to "I would like my blog to be visible to everyone, including search engines (like Google, Sphere, Technorati) and archivers" which can be changed into 'I would like to block search engines, but allow normal visitors." (WordPress Admin Panel) Not only does this setting block the engines[93] but it also blocks the ping function, which sends an update notification ping to the engines, from being executed.

The medium of the blog finds itself in an entangled relationship with the engines which works both ways. The blog software is very aware of the engines and adjusts itself accordingly by supplying software that is ready to "embrace" the search engines. The engines, on their behalf, provide tools, plugins and instructions for both the blog software and the blogger to optimize the relationship[94].

3.2.6 Plugins and Widgets

WordPress strives to keep the core of the software light by keeping the amount of code and integrated functionality small:

> The core of WordPress is designed to be lean, to maximize flexibility and minimize *code bloat*. Plugins offer custom functions and features so that each user can tailor their site to their specific needs. [...] *Plugins are tools to extend the functionality of WordPress*. (WordPress Codex[95])

Users can expand the 'minimal' functionality of WordPress by adding plugins or widgets[96]. Widgets are a special kind of plugin that enable simple drag and drop

---

93 Causes "<meta name='robots' content='noindex,nofollow' />" to be generated into the <head> </head> section (if wp_head is used) of your site's source, causing search engine spiders to ignore your site. Causes hits to robots.txt to send back: User-agent: * Disallow: / http://codex.wordpress.org/Privacy_Options_SubPanel (accessed January 5, 2008)

94 An interesting example of this two-fold relationship is the presentation Matt Cutts, head of Google's Webspam team , gave at WordCamp 2007. WordCamp is the official conference for WordPress users and developers. Cutts gave a presentation on 'Whitehat SEO tips for bloggers' aka 'How to optimize your blog for the engines with techniques that adhere their policy.' In the last slide he presented his personal 'WordPress Wish list that included "Prevent content at multiple url aliases" Just a few weeks later WordPress implemented "SEO-friendly URL redirection, Canonicalize more URLs."' a function that addressed the issue on Cutt's wish list. This case illustrates how entangled the relationship between the software and the engines is.

95 http://codex.wordpress.org/Plugins (accessed December 8, 2007)

96 "A WordPress plugin is a program, or a set of one or more functions, written in the PHP scripting language, that adds a specific set of features or services to the WordPress weblog, which can be seamlessly integrated with the weblog using access points and methods

configuration in the WordPress interface.

> A widget is a self-contained piece of code that you can move into, out of and anywhere inside the sidebar area of your blog. You can use widgets to personalise your blog and deliver information you want. (WordPress.com, Widget)

In order to use widgets the Widgets plugin from Automattic had to be installed and the WordPress theme had to be widget-ready. Plugins and widgets enable users to modify the blog according to its needs and some plugins, such as the Ultimate Tag Warrior, have become so popular that WordPress has integrated the tagging functionality into its core. The easy-to-use widgets, that after a simple configuration sometimes only require a simple drag-and-drop to be enabled, became so popular that WordPress added them to its core in version 2.2 (May 16, 2007). This illustrates the close interaction between the (open source) WordPress community and the WordPress developers.

A wide range of plugins and widgets have been developed by the blog community, blog software developers and the engines. These plugins are scattered across various sources: the Official WordPress Plugin Directory[97], the WordPress Plugin Database[98] and the individual community developers blogs. The Official WordPress Plugin Directory (with 1,274 plugins) and the WordPress Plugin Database (with 2851 plugins) act as central plugin repositories although not all developers submit their plugins to these repositories but offer them on their own blog instead.

Some plugins and widgets allow the blog to be easily connected to the engines. The Official WordPress Plugin Directory contains 52 Google plugins that vary from adding a Google Map to your blog to the automatic creation of a sitemap that makes it easier for Google to index your blog. The plugins work both ways: they allow easy implementation of Google services into the blog but they also configure your blog so that it can be used by Google easily.

Plugins can be used to custom query fields of the database in order to expand on the default functionality. They can query the blog's own database or an external database such as Flickr through an API. The integration of widgets to the core of WordPress acknowledges the active use and simple configuration of plugins to extend the software. The large amount of plugins that allow for communication with the

---

provided by the WordPress Plugin Application Program Interface (API)."
http://codex.wordpress.org/Writing_a_Plugin (accessed January 27, 2008)

97 http://wordpress.org/extend/plugins/ (accessed January 27, 2008)

98 http://wp-plugins.net/ (accessed January 27, 2008)

engines points to an increasing relationship between the blog and the engines.

Widgets have become a popular way of 'accessorizing' and personalizing the blog as

> Widgets also serve as a key method of distribution for other sites. Social music sites iLike, Last.fm, and imeem allow members to embed modules for searching for music, sharing playlists, and playing songs onto their pages. [...] Rather than visiting individual sites for different types of information or entertainment, users are increasingly designing their own online destinations -- whether blogs, profiles, or other personal pages -- and adding content from multiple sources. (Joyner, 2008)

Widgets are used to embed the scattered self into one place, the blog, creating 'the widgetized self' (Baym, 2007)

> What I need, and what I think everyone else needs too (even if they want to hang out in Second Life a lot) is my own portal www.nancybaym.me that I can just set up with a collection of widgets that bring all the sites I care about to me. Start pages on steroids. In my dream portal I can read and write to all the sites I want without having to leave my page. I can leave comments on blog posts, post to an online discussion on a forum and do everything else I want to do — and make it available to others — from my own little spot. Widgets gone wild. (Baym, 2007)

The widgetized self is not a solitary and antisocial self that withdraws from (social) sites in order to confine itself to the blog:

> It shouldn't just be about displaying me to an audience, it should be about allowing me to engage with my communities and affiliations in a way that is (at least to some degree) visible to an audience, so that they see my communities as well as me if they drop by my portal. (Baym, 2007)

While widgets are often used for display purposes only, they can also serve more engaging purposes. Widgets, less concerned with displaying external content and more with enabling social interactions, could change the relation between the blog, its audience and (external) content:

> what would happen if, say, our blogs became less portals of ourselves and more portals of our friends? (David Silver in Baym, 2007)

The popularity of widgets can be seen in the large amount of 'widget ready' WordPress themes.

3.2.7 Themes

The blog's theme determines the displayed content and layout of a blog. A WordPress theme is a collection of PHP files generating the structure and a CSS file generating the style of the blog. The main WordPress page is the index.php that includes The Loop which is the core of WordPress and refers to the main process of WordPress (see 3.1.3). The index.php file used to be only file used by WordPress to generate the blog content:

> Before the advent of Themes, WordPress generated content using a single file, `index.php`, and files to support comment display and submission. A single style sheet controlled the presentation. All other pages, including the category and archive pages, were generated by passing parameters to the `index.php` page. (WordPress Codex[99])

With the introduction of themes bloggers can:

> Physically Separate Components
> The new WordPress modular template files system provides a method to define separate physical PHP files for the different components of your WordPress site. This allows creation of unique designs and functionality for many special pages, such as category archives, monthly archives, and the individual entry pages.
>
> Quickly Change Layout and Design
> It allows users with appropriate permissions to quickly change the layout of the entire site by uploading a new theme and essentially flipping a switch in the admin panel. (Ibidem)

A theme is an overlay that "modifies the way the site is displayed, without modifying the underlying software." The theme specifies which database fields are queried and displayed on the final generated page. The accompanying stylesheet determines the actual style of the layout. "At the very minimum, a WordPress Theme consists of two files: `style.css` and `index.php`" According to WordPress developer Ryan Boren the `index.php` "is the master template for the theme, and it must exist."[100]

As previously described `index.php` includes The Loop which generates the reversed-chronological structure of the blog. In theory it is possible to create a template without The Loop (it would only display data from one post) but in practice The Loop is a canonical feature. WordPress recommends building custom themes

---

99 http://codex.wordpress.org/Using_Themes (accessed January 11, 2008)
100 http://boren.nu/archives/2004/11/10/anatomy-of-a-wordpress-theme/ (accessed January 11, 2008)

upon one of the two existing default themes because

> these have been through the presses by WordPress designers and testers, as well as bazillions of users who are more rigorous on these things than the developers. These are solid code, for the most part, and a good starting point. From there, do whatever your imagination desires. (WordPress Codex[101])

The standard blog favors a reverse chronological ordered post. The structure of the theme determines both the structure and layout of the blog and its relation to the engines.

WordPress' default themes are (search) engine friendly but customizing a theme may break some of the relations with the engines. WordPress itself offers an extensive guide on how to "maintain your WordPress site's optimal friendliness towards search engine spiders and crawlers."[102] The guide covers tips on how to code for the engines, how to write for the engines, how to optimally configure your blog for the engines and how to submit your blog to the engines. SEO for WordPress is an important part of the blogger under the relationship of software-engine relations.

---

101 http://codex.wordpress.org/Designing_Themes_for_Public_Release (accessed January 11, 2008)
102 http://codex.wordpress.org/Search_Engine_Optimization_for_WordPress (accessed January 12, 2008)

4. BLOGGERS UNDER THE INFLUENCE OF SOFTWARE-ENGINE RELATIONS

The previous chapter described how blog software facilitates relationships between blogs and between blogs and engines. It described how the engines use standardized aspects of the blog such as the site feed to establish communication channels. The engines are increasingly entangled in the medium and practice of blogs which has implications on several levels. The influence of the engines on the medium and practice of blogs asks for a critical examination of this relationship. Google, as the prevailing search engine, has been the subject of numerous critical investigations. In *The Googlization of Everything* (a book in progress[103]) Said Vaidhyanathan wishes to examine the effects of Google (and search engines in general) on the politics of culture, knowledge and information (2007a) by addressing three key questions:

> The phenomenological: What does the world look like through the lens of Google?;
>
> The cultural and communal: How is Google's ubiquity affecting the production and dissemination of knowledge?;
>
> The political: How has the corporation altered the rules and practices that govern other companies, institutions, and states? (Vaidhyanathan, 2007a)

These questions may also be applied to address the consequences of the increasing software-engine relations. The phenomenological question addresses what the blogosphere looks like according to the engines and how the blogosphere is constructed through software-engine relations. The standardization of blog features such as the permalink, trackback, pingback and site feeds has not only facilitated blog to blog communication but also the automatic indexing of blogs. This, in its turn, has enabled the engines to construct the blogosphere.

The blogosphere[104] is seen as the totality of blogs or "the overall community of

---

103 http://www.googlizationofeverything.com/2007/09/hi_welcome_to_my_book.php (accessed January 26, 2008)

104 The term was coined by Brad L. Graham as a joke in a post that made fun of the coinage of the word 'blog': "Goodbye, cyberspace! Hello, blogiverse! Blogosphere? Blogmos? (Carl Sagan: "Imagine billions and billions and billions of blogs.")"

blogs and bloggers, which is interlinked through a large number of cross-references between individual blog entries." (Bruns & Jacobs, 2006, p. 5) The blogosphere can be understood as a co-production of different actors, namely the blogger, the blog, the blog software, protocol and the blog engines. In order to conceptualize the blogosphere it must be constructed either by visual or textual means. Matthew Hurst from the Data Mining blog published several beautiful visualizations of the blogosphere based on enormous datasets[105]. The visualizations show that some parts of the network are closely knit whereas others are fairly isolated. The blog community LiveJournal, for example, is hardly connected to the rest of the blogosphere.

The engines also show us that it makes more sense to speak of numerous blogospheres instead of a single blogosphere. Previous research, conducted with the Digital Methods Initiative, showed that different engines construct different blogospheres[106]. Blogs construct the blogosphere through syndicating, pinging and linking using permalinks, trackbacks, pingbacks and blogroll. The engines use syndication feeds to index blogs and use different types of links to determine the ranking of a blog. The engines construct the blogosphere for us by their algorithms based on links and freshness. As different engines use different algorithms and keep different indices they construct different blogospheres for us. This is the result of both the cultural/communal and the political implications of software-engine relations.

The first part of this chapter (paragraph 4.1 and 4.2) wishes to address the cultural and communal question by looking at the daily blogging practice and how this practice is influenced by the software-engine relations. The second part (paragraph 4.3 and 4.4) aims to address the political question by looking at the nofollow attribute as an example of the political implications of the software-engine relations on the blogosphere.

---

http://www.bradlands.com/weblog/comments/september_10_1999/ (accessed January 5, 2008)

105 http://discovermagazine.com/2007/may/map-welcome-to-the-blogosphere (accessed January 5, 2008)

106 http://wiki2.issuecrawler.net/twiki/bin/view/Dmi/Nofollow (accessed January 5, 2008)

4.1 The Blogging Routine

This paragraph describes my personal blogging practice and the effects of the software-engine relations on this practice. The blog as a medium and bi-product of practice is very diverse and cannot be described by a single account. All bloggers have their own blogging practices and construct their own typical blogs by using a particular theme and plugins. This account is based on my own blogging experiences, after having used the major blog software WordPress over a period of seventeen months, in combination with blogging experiences from and as described by other bloggers.

4.1.1 Statistics

My daily blog routine within the network starts with checking my blog's statistics. I am especially interested in the links and referrers to my blog. First, I check my inlinks at Technorati[107] which also determines my authority within the blogosphere based on the amount of fresh inlinks over the past six months. 'The Art of Backlinking' (VanFossen, 2007) describes several techniques to obtain links to your blog because "In the world of Google PageRank™, links to your blog count."

Then I login to WordPress administration panel to check the incoming links in the dashboard (see illustration 9). The dashboard was introduced in version 1.5 as a way to keep "you up to date with the latest happenings on your blog and the most important news from around the WordPress world, keeping you connect to the latest developments, updates, news, and enhancements."[108] The dashboard acts as a centralizing feature by integrating the latest WordPress news from various sources into the administration panel of the software itself. It also shows the latest activity on the blog by providing the most recent incoming links.

---

107 http://www.technorati.com/blogs/www.annehelmond.nl?reactions (accessed December 11, 2007)
108 http://wordpress.org/development/2005/02/strayhorn/ (accessed December 11, 2007)
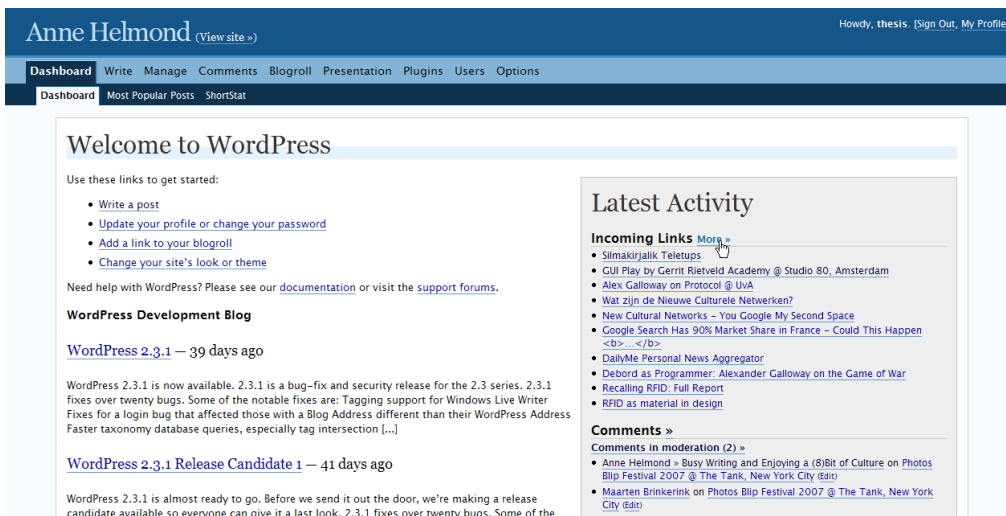
Illustration 9: WordPress 2.3.1 - Dashboard

I have configured Google Analytics Statistics for my blog providing me with detailed information on my blog visitors and traffic. It shows that the engines act as a major entry point into my blog as 59.38% of the visitors find my blog by search engines, 29.14% by referring sites and 11.48% by direct traffic (Google Analytics, December 17, 2007.) Google Analytics also shows how much my blog revolves around the unit of the post. Only 5.66% of the total amount of visitors enter my blog by the front page. The other landing pages are either blog posts, tag pages (by Technorati or my tag cloud) or static pages (archives, categories or about). My statistics confirm how much my blog relies on (search) engines which reflects a more general trend which is of influence on how certain bloggers write their content with the engines in mind (see paragraph 4.2). Google also owns Feedburner, a feed tool that I use to check my feed readers' statistics and the number of people that are subscribed to my feed.

All these statistics services are external to my blog and require an account. To be able to quickly check my blog's statistics I have installed the WordPress ShortStat plugin. It integrates stats into my blog's dashboard so that I can easily check my stats from within my blogging environment (see illustration 10). I look at these stats several times a day, not so much for the numbers, but for the keywords that have been used to end up at my blog. I am also particularly interested in the referrers showing me from which sites visitors have come. If the referrer is not a search engine I will visit the site to see where and in what context the site has linked to me.

Illustration 10: WordPress plugin WP ShortStat

I visit the ShortStat page in my dashboard several times a day by a shortcut in my browser. It is not only a distraction in between replying to e-mail or reading blogs, but it is also an intense curiosity that has slowly turned into an obsession. Blogger Justin Smith labels his own blog stats checking as an obsessive compulsive disorder:

> I run a blog too (http://flawedlogic.org) and I swear I have OCD when it comes to checking my stats. I have to check them at the VERY least once an hour, and that's if I'm extremely busy with something else. If I'm not in the middle of anything, I could check my visitor logs once or twice every 5 minutes. (Smith, 2007)

And jules feels himself a prisoner of statistics:

> Methinks a traffic stats support group would be a welcome distraction....
> I feel your pain. I'm chained to traffic logs. (jules, 2007)

The ProBlogger blog notes that we're not alone and "perfectly normal" as stat addiction is very common among bloggers. Even though Problogger states that "most bloggers 'grow out of it'" (Rowse, 2007) I am far from cured of this addiction. Blogger Moolanomy represents the bloggers that feel that their stat checking behavior has become counter-productive and is determined to "Stop Wasting Time and Focus on Productive Tasks:"

> So I will be checking my blog stats, discussion forums, and co.mments

85

(comments tracking tool) only once per day…may be twice. (Moolanomy, 2007)

Many bloggers spend a large amount of time checking their statistics but reducing the amount of time or quitting checking statistics altogether is just as hard as quitting any other addiction:

> Lets face it, bloggers love stats and rankings, even if they are based upon meaningless data. For some it is an ego thing, or just a measure of their own worth or progress to achieve whatever goals they have set for themselves, and for others it is monetary. (Beard, 2008)

I keep a list of my Technorati ranking over time, not because I am focused on constantly increasing my ranking but to check how my ranking is developing. I am aware of the fact that Technorati constantly changes its algorithm to determine the ranking and that spam blogs are included which may artificially boost my ranking. The rankings do not represent a Truth but they serve as a reflection point of my role in the blogosphere. Statistics are often used by bloggers for this 'reflectional' purpose.

> Well anyway, thanks for reading. I've got to go check Site Meter again and see if I'm still 479,027 on Technorati. I did mention my new Technorati obsession, right? (Mush, 2006)

They may turn into vanity when displayed in plain sight on the blog. Blog statistics are often displayed on blogs as a status symbol. Commonly displayed statistics are the number of feed subscribers, the number of Diggs a blog post has received, Google PageRank, Alexa rank and Blog Metrics (see illustration 11).

Illustration 11: Blog statistics

The era of statistics brings us widely available statistics which we welcome with open arms. Computers and databases have enabled the easy and automatic registration and quantification of data. Statistics on the web have been widely available since the mid 1990s. Nedstat, for example, has provided its users with (free) web statistics since 1996. However, there is a trend of increasingly available statistics for every single detailed aspect of our web behavior. Previously these statistics were often only

internally available for the services in order to gather data on their customers. Google has made consumer profiling its core business:

> One of the great attractions of Google is that it appears to offer so many powerful services for "free," that is, for no remuneration. But there is a non-monetary transaction at work between Google and its users. We get Web search, email, Blogger platforms, and YouTube videos. Google gets our habits and predilections so it can more efficiently target advertisements to us. Google's core business is consumer profiling. It keeps dossiers on all of us. Yet we have no idea how substantial or accurate these digital portraits are. (Vaidhyanathan, 2007a)

Even though we do not know what is exactly being observed we are more and more aware of the fact that every single detail of our web behavior is being monitored. We accept this fact in return for our 'own' statistical data. We let engines trace our every single step in order to see what steps other people have taken to end up on our blogs:

> Discussing the role of the consumer, Siva notes another Google illusion - that of the free service. We pay for Google with our data - our searching habits, our surfing habits - and this fuels Google's cash cow, personalized advertising. Siva calls for a renewed approach to understanding this kind of consumer surveillance, one that pushes aside the tired model of the panopticon (which Foucault analyzed in Discipline and Punish). He cites some of the ways surveillance has changed: it is private rather than state-run, and we don't know how much they know. Most of all, we're encouraged to transgress - to enjoy! as Zizek would say - rather than forced to reform as in Bentham's model. That is, on the Web we need to show our true selves. (Stevenson, 2007a)

Statistics play a major role in my blogging routine as well as keeping up with my blog subscriptions by using Netvibes (see paragraph 3.2.4.) Reading other blogs often includes the practice of bookmarking and commenting on other blog posts.

4.1.2 Commenting

I often comment on other blogs to engage in the discussion, share relevant information or post a link. There are two kinds of comments: comments left by other people on your blog and comments left by you on other blogs. Managing these two types of comments is very different. Comments left on your blog are stored in a central place, the blog or an external comment server. Comments left by you on other blogs are scattered around the blogosphere.

I receive an e-mail notification when people comment on my blog. I have enabled e-mail notification in my WordPress administration panel because my e-mail inbox is always open and my blog administration panel is not. This is my preferred

method of managing my comments but it is not ideal when dealing with tens or hundreds of comments a day. Some bloggers choose to outsource their comments to an external service such as Haloscan or Intense Debate. These services add an external commenting system to the blog which may offer extra features such as spam protection or threaded comments. Haloscan is a particularly interesting service as it displays comments "on a blog using JavaScript, which isn't seen by search engines as part of your site." (Beard, 2007) It separates the blog not only by outsourcing the comments but also by displaying them in a separate popup which is not indexed by the engines[109].

Indexing and tracking comments are relatively new in the blogosphere. Keeping track of comments left on other blogs is hard in the distributed network of the blogosphere. Not only do you have to remember on which blogs you commented but you will also have to check manually to see if there are any new comments. I use several features to deal with the distributed nature of commenting. The 'subscribe to comments' plugin allows blog commenters to be notified of follow-up comments by e-mail. It sends an e-mail that includes the subsequent comments on a blog post. This feature can only be used if the owner of the blog has installed this plugin.

The most important tool I use to bring distributed conversations together is the comment aggregator CoComments. The CoComments service acts as a centralizing force that collects all comments in one place. The additional CoComments Firefox plugin automatically sends a comment to the CoComments server if the blog is supported. The service supports all major blogging platforms including WordPress and services such as Flickr and Digg. The service offers instructions on how to configure manually unsupported blog platforms which requires some effort on the blog owner's behalf. CoComments wishes to track and index the 'commentosphere'[110]

The post is the discrete unit of the blog that constitutes the blogosphere and the comment is the discrite unit of the blog post that constitutes the commentosphere. Comments now have permalinks and RSS feeds too, making them discrete units that can be referenced or indexed. The comment functions differently from the blog post as the comments RSS feed is not used for indexing the commentosphere. CoComment works by voluntary participation to track your own comments and to be able to track conversations you are not participating in. Unlike the engines it does not offer a

---

109 These are the default settings for HaloScan.
110 http://blogs.mediapost.com/behavioral_insider/?p=151 (accessed January 24, 2008)

generally searchable index for comments. Another important feature that really distinguishes the blog comment from the blog post is the nofollow attribute which will be described in paragraph 4.3.2.

The practice of commenting is not only disrupted by the decentralized nature of comments but also by comment spam (see paragraph 4.3.1). This has led to many bloggers disabling comments on their blogs but spam is not the only reason:

> For me the decision to turn off comments wasn't so much about the spam. Spam was only a minor factor because decent filters can take care of blog spam well enough.
>
> The main issue was the opportunity cost — the amount of time and energy it takes to handle comments and what other choices I have for investing that time and energy. (Pavlina, 2005)

Handling comments, moderating and replying to them, are an important part of blog maintenance.

4.1.3 Maintaining

Blogging is not only a process of writing but also of maintaining. Setting up a blog includes choosing a theme and plugins to customize the blog to your needs and wishes. I chose to modify an existing theme to add a personal design touch and added several plugins to enhance my blog. Over time, new WordPress versions have been released that include important security updates but also additional features. WordPress updates often include changes to the code that may render plugins or themes obsolete or erroneous. The process of maintaining a blog includes updating the software and updating themes or plugins. WordPress 2.3 offered a new function that shows if a new update for a plugin has been released. Previously you would have to check the homepage manually for every plugin to see if a new version had been released. Blogging is not only writing but also constantly tweaking small pieces of code and adjusting the layout.

The sidebar is the feature of the blog that is most commonly used to insert plugins to customize the blog. The sidebar often includes links to recent or most popular posts, a blogroll, the archives and content retrieved from other databases. The sidebar usually offers the possibility to insert widgets that automatically generate the code to insert external content in your blog. Both the external services and the

engines offer widgets to connect your blog to their service.

An important part of blog maintenance is keeping the blog spam free. The blog is submitted to comment spam in the form of comments, trackbacks and pingbacks. Comment spam is one of the reasons to outsource comments as it may reduce the time spent in handling comments.

A more commonly used option is to install anti-spam plugins. WordPress 2.0 and later include the integrated Akismet anti-spam plugin[111]. This plugin requires a WordPress.com API key which in its turn requires a WordPress.com registration whether or not you host your blog there. Two other popular anti-spam plugins that do not require an API key are Spam Karma 2 and Bad Behavior. I use these plugins to handle and manage my comment spam[112]. I receive a daily digest from Spam Karma in my e-mail that provides me with my daily spam statistics. It also shows an excerpt of comments that have been marked as spam but may be actual comments. Comments are checked against a central server and marked as spam if they meet certain criteria. It may happen that a valid comment is marked as spam which requires the manual approval from the blogger which will retrieve the comment from the database and display it on the blog[113].

Blogging without any spam filters is hardly imaginable as Spam Karma has caught **29393** spam comments since I first installed it in September 2006. This is nothing compared to the 4,007,755,188 spam comments Akismet has caught since its introduction in December 2005.

Spam has changed the practice of linking as some spam protection methods may prevent bloggers from commenting or leaving links in comments.

---

111 Akismet was developed by Automattic, founded by WordPress founder Matt Mullenweg.
112 Some blogs add CAPTCHA's (Completely Automated Public Turing Test to Tell Computers and Humans Apart) or mathematical equations to their comment forms. Personally I am not a big fan of these spam protection methods as they provide an extra hurdle in commenting. Posting a comment requires filling in a form with your name, e-mail and (optionally) URL. Spam engines automatically fill in these fields and to separate a bot from a 'true' commenter another field is added to 'prove you are human.'
113 Sometimes valid comments are not even added to the moderation queue but immediately purged as spam. The commenter will be notified of this event and fortunately some commenters will take the effort to compose an e-mail to notify the author, which happened to me last week: "hi your filter told me i'm spam. read comments @ WYSIWYG. Cya :))" Nothing is more frustrating than taking the time to comment and then find out your comment as been marked as spam.

4.1.4 Linking

Early blogging practices involved logging interesting links. As the medium and practice of blogging have evolved the blog no longer solely relies on the link. As a fairly newcomer to blogging I hardly ever log links but usually write extensive blog posts. I log my links on the social bookmarking site del.icio.us. This practice is what early blogger Jorn Barger refers to as true blogging when he shares his blogging insights with Wired at the 10th anniversary of the coinage of the word weblog:

> A true weblog is a log of all the URLs you want to save or share. (So del.icio.us is actually better for blogging than blogger.com.) (Barger, 2007)

This comment shows how the practice has developed over the past ten years. When writing a blog post it is common courtesy to link to sources or other relevant blog posts. Blog software has enabled to send the other blog a notification of this link:

> Trackback is an intentional way of notifying other blogs because WordPress requires you to manually enter the blog's trackback link. It also allows you to send a notification to another blog even if you don't explicitly link to them in the post. This may be done in an attempt to include the other blog in the conversation. On top of that trackbacks may be considered "the real letters of recommendation on the web." (Helmond, 2007c)

Trackback, as a manual notification method, allows you to choose which blogs you would like to notify of your blog post. It is possible to notify blogs that have not been mentioned in the blog post in order to attempt to engage them in the conversation. Trackbacks allow blogs to interconnect by notifying each other.

After its introduction in 2002 trackback has become very popular but in 2005 Tom Coates declared trackback dead because it "has been killed by spam and by spammers." Many bloggers turned off the automatic receipt of trackbacks and/or removed a visible trackback link to prevent their blogs from being spammed.

Sending a trackback is a very conscious act of involving other blogs in the conversation. It is an explicit link notification system. It requires the blogger to be aware of the existence of a trackback URL and then copy-paste it into the required WordPress field. As a relatively new blogger I did not even know it had to be done manually. With the demise of a visible trackback link I had become used to the automatic pingback. Pingback automatically notifies all the links in your blog post. It is an implicit link notification system. With trackback you can choose if you want to notify other blogs, and if so, which. Pingback does not offer a choice, once it has been enabled , it will notify all the links in the blog post.

Pingback was developed to combat some of the flaws of trackback but as it is an automatic notification system over a manual notification system trackback seems to be disappearing. This is especially visible among 'new' bloggers such as myself who were not even aware of the difference between pingback and trackback or of the existence of trackback:

> After about 12 months of blogging, I finally realized what the trackback field was for in the WP template. I was still not comfortable understanding the difference between trackbacks and pings, your explanation above is much clearer to me than anything I found on the WP site.

> I try to include trackbacks in my posts, but I don't think they are very successful, because I usually do not find an actual "trackback link" on the blog I want to link to, and try to link to the particular posting (page), but rarely see anything from my blog show up on the blog I am trying to trackback to. (Big Fella, 2008)

With the demise of a visible trackback link new bloggers will mainly rely on pingback as it is an easy-to-use automatic system:

> Basically, I'm in support of whatever's simplest and easiest for regular bloggers to understand and use. So... pingbacks for now. I'm a big fan of making blogging about free expression and less about technical stuff. (Vigneault, 2008)

Pingback has automated the process of notifying other blogs. Link notification is no longer a conscious act (by trackback) but it has become an integral part of blogging. Linking is also a matter of politics as a link is counted as a vote in the ranking for engines. In the SEO world linking is often referred to as giving 'link love' or 'link juice.' Asking for links in the form of a link exchange may be considered impolite if done directly and may also hurt your rankings. Building links is referred to as 'link bait'

> Link bait is any content or feature within a website that somehow baits viewers to place links to it from other websites. Matt Cutts defines link bait as anything "interesting enough to catch people's attention."[3] Link bait can be an extremely powerful form of marketing as it is viral in nature.

> The quantity and quality of inbound links are two of the many metrics used by a search engine ranking algorithm to rank a website. Link bait creation falls under the task of link building, and aims to increase the quantity of high-quality, relevant links to a website. Part of successful linkbaiting is devising a mini-PR campaign around the release of a link bait article so that bloggers and social media users are made aware and can help promote the piece in tandem. Social media traffic can generate a substantial amount of links to a single web page. Sustainable link bait is rooted in quality content. (Wikipedia[114])

---

114   http://en.wikipedia.org/wiki/Link_bait (accessed December 27, 2007)

Social media and sites such as Digg and Stumble Upon are often used as engines to build a link bait campaign (Goedegebuure). The social bookmarking buttons on blogs are part of the practice of link bait.

I am not involved in the practice of link bait but I am involved in the politics of giving (or not giving) link love. With every blog post I link to, I am aware that my link may count as a vote for the engines and may show up as a pingback on the other post. WordPress automatically sends out a pingback for every link I create and if the other blog (automatically) accepts pingbacks it will be displayed in the comment section. As a result of software-engine relations links in posts and links in comments (including trackbacks and pingbacks) are treated differently (see paragraph 4.3.2 on nofollow) and are also of influence on how the blogosphere is constructed. Google has changed the practice of linking by:

> Connecting PageRank to economic systems such as AdWords and AdSense corrupted the meaning and value of links by turning them into an economic exchange. Through the turn of the millennium, hyperlinking on the web was a social, aesthetic, and expressive editorial action. When Google introduced its advertising systems at the same time as it began to dominate the economy around search on the web, it transformed a basic form of online communication, without the permission of the web's users, and without explaining that choice or offering an option to those users. (Dash, 2007)

By connecting PageRank to economic systems the practice of linking changed as it was no longer a social practice but also an economic practice. This is most visible in the practice of SEO, search engine optimization.

4.2 Abiding Software-Engine Relations

The software establishes relations with the engines by tags, ping and RSS after publishing a post. Some bloggers keep these relations in mind while writing a blog post. They are aware of the software-engine relations and write, tag and link blog posts with the engines in mind. This practice is commonly referred to as SEO, also known as search engine optimization. The software-engine relations influences how one blogs and leads to writing recommendations such as

> Write Your Content with Searchers in Mind
>
> How do you find information on the Internet? If you are writing something that you want to be "found" on the Internet, think about the words and phrases someone would use to find your information. Use them more than once as you

write, but not in every sentence. Learn how search engines scan your content,
evaluate it, and categorize it so you can help yourself get in good favor with
search engines. (WordPress Codex, Search Engine Optimization for
WordPress[115])

At the Q&A of the European Bloggers (Un)Conference Cory Doctorow from the
popular blog Boing Boing explains why he things Boing Boing[116] became the number
one blog in the blogosphere.

First of all he hopes they have a great distinctive taste in content and write well.
Second is that they keep the engines in mind while writing a blog post. By thinking in
search engine keywords they compose descriptive titles instead of clever titles. They
also keep in mind what the blog post would look like in a feed reader because the
engines and feed readers usually only show the blog post title and an excerpt.
According to Doctorow a blog post title must be able to stand on it's own, as it is what
is primarily displayed in engines and feed readers.

Blogs are often accessed and read by feed readers which leads to
recommendations as 'Writing With Post Excerpts and Feed Excerpts in Mind'

Checking out what excerpts search engines showcase from your blog will help you
understand a little more about SEO (Search Engine Optimization) techniques,
and the critical value of excerpts and well written introductory paragraphs.
(VanFossen, 2006)

The practice of SEO is used to achieve a higher ranking in the engines for either
monetary reasons and/or to increase the blog's visibility. This may leads to spam
practices that exploit the software-engine relations within the blogosphere.

---

115 http://codex.wordpress.org/Search_Engine_Optimization_for_WordPress (accessed
     December 4, 2007)
116 Boing Boing was the most popular blog according to Technorati's Popular Blog list for a
     long time. It recently dropped to a number four position but Boing Boing's Technorati
     profile still refers proudly to this position:

     It's the most popular blog in the world, as ranked by Technorati.com, and won
     the Lifetime Achievement and Best Group Blog awards at the 2006 Bloggies
     ceremony. (http://www.technorati.com/pop/blogs/) (accessed December 4,
     2007)

4.3 Exploiting Software-Engine Relations

The political dimensions of the software-engine relations should be addressed as:

> Search engines are some of the most popular destinations on the Web—
> understandably so, given the vast amounts of information available to users and
> the need for help in sifting through online content. While the results of
> significant technical achievements, search engines are also embedded in social
> processes and institutions that influence how they function and how they are
> used. (Hargittai, 2007)

The engines function as centralizing power structures within the blogosphere and their coalescing power allows us to conceptualize the blogosphere (Galloway, 2007 & Weltevrede, 2007, p. 3). Some blog software protocols can be seen as centralizing forces within the blog itself. Blog feeds for example act as both centralizing and distributing forces of the blog and are used to communicate with the engines. Together with the blog software the engines construct the blogosphere for us.

Paragraph 3.2.4 explained how the blogosphere is constructed by site feeds. Feeds and pings are important features of blogs as they allow the blog to disperse and distribute itself. They are both the strength of blogs but at the same time they show the exploits in the blogosphere. Exploits are "networks that work too well" (Galloway & Thacker, 2007, p. 6) and trackback/pingback spam and nofollow show what happens when the blogosphere works 'too well.' Spam shows what happens if the daily blogging routine collides with the software-engine politics:

> while search engine relevancy teams battle against spam in the interests of
> bringing a high-quality result to the user, their own advertising formats – paid-
> placement ads – are at the heart of a this new generation of spam. (Van
> Couvering, 2004)

The next part will describe the software-engine politics at full play in the case of spam.

4.3.1 Spam

Spam is an incredibly successful exploit of the network (Galloway & Thacker, p. 145.) Spammers exploit the flaws ("some would say features" p. 146) of the blogosphere by feeds and pinging. Resembling a virus, spam takes "advantage of technical standardization to propagate through the network" (p. 84). Spammers use the standardized protocols of site feeds, ping and pingback for their own means.

95

In the blogosphere spammers are referred to as sploggers and splogs are blogs created for spam purposes. Spammers usually exploit the network for monetary reasons:

> But whereas email spammers try to induce recipients to buy products, sploggers and other Web spammers make most of their money by getting viewers to click on ads that run adjacent to their nonsensical text. Web page owners – the spammer, in this case – get paid by the advertiser every time someone clicks on an ad. [...]
>
> Because the ad money is effectively available only to Web sites that appear in the first page or two of search results, spammers devote enormous efforts to gaming Google, Yahoo, and their ilk. Search engines rank Web sites in large part by counting the number of other sites that link to them, assigning higher placement in results to sites popular enough to be referred to by many others. To mimic this popularity, spammers create bogus networks of interconnected sites called link farms. Blogs – most of which are in essence little more than collections of links with commentary – are particularly useful elements in them. The result, Dash says, "is what you'd expect: The blogosphere is increasingly polluted by spam." (Mann, 2006)

Spammers thrive on the distinctive ability of the blog to syndicate, disperse and aggregate itself. Spam in the blogosphere consists of comments, trackbacks and pingbacks that often link to a product page, splog or sportal[117]. Spambots are used to automate the posting of spam links in comments, trackbacks and/or pingbacks.

Trackback spam exploits the trackback protocol that allows for inter-blog communication. Most blogging software is configures to automatically receive trackbacks making it vulnerable to spam. In order to combat this flaw in the protocol the pingback protocol was developed which requires a verification of the existence of the posted link. This led to the creation of splogs and sportals which can verify the existence of the link but also include ads and/or links to commercial sites. Splogs are usually created with WordPress software because it is free, easy to set up and supports all the necessary features such as trackback, pingback, feeds and ping but also plugins to display ads and import feeds.

Splogs often scrape the content of existing blogs by either importing a blog's site feed to displaying an excerpt of an existing post or by scraping the ping servers (see illustration 12).

---

117 "Web pages consisting almost entirely of pay-per-click links, all of which shunt netsurfers to legitimate commercial Web sites, collecting money along the way for the spammers." http://www.wired.com/wired/archive/14.09/splogs.html (accessed December 27, 2007).
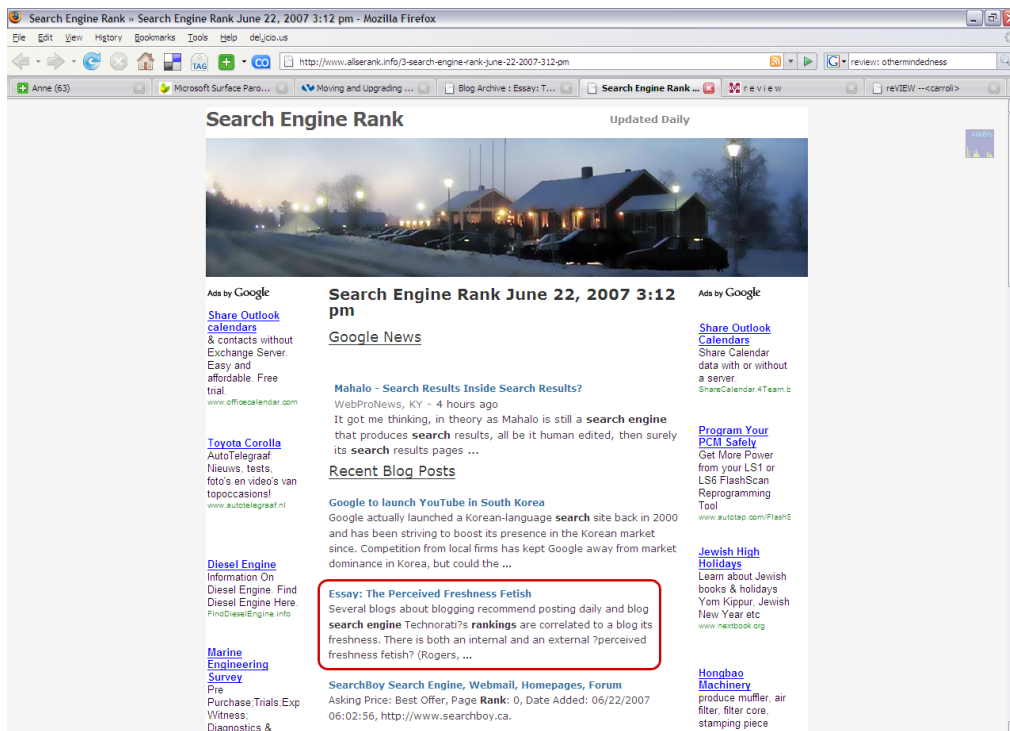
Illustration 12: A splog that scraped one of my blog posts (in red) based on keywords

As described in paragraph 3.2.3 ping servers act as centralizing forces in the blogosphere that receive and display the freshest content. Blog posts on the ping servers can be scraped for a certain keyword or tag. This leads to themed splogs, splogs that only gather posts on a particular topic. Splog comments can often be recognized because the excerpts in the trackback and pingback all display the same text around the excerpt. The phrase "wrote an interesting post today, here's an excerpt" became a widely seen phrase in the blogosphere. As the spam filters from spam plugins adjust accordingly comments including this sentence were soon banned. Spammers soon came up with a new type of blog scraping:

> Jonathan calls them "Spinning Spammers", and they are using Plugins and utilities for *synonymized scraping* of your blog. The process scrapes your blog feed and then "translates" your content using synonyms to replace recognizable words. (VanFossen, 2007)

Spammers and spam-fighters are fighting a constant battle whereby each party is constantly re-adjusting itself. Bloggers try to fight comment spam by installing anti-spam plugins. Fighting spam has become such an integrated part of the daily blogging routine that WordPress integrated the Akismet plugin into the core. The engines came

up with their own solution to prevent comment spam, the nofollow attribute.

4.3.2 Nofollow

The nofollow[118] attribute was introduced by Google as

> A non-standard HTML attribute value used to instruct search engines that a hyperlink should not influence the link target's ranking in the search engine's index. It is intended to reduce the effectiveness of certain types of search engine spam, thereby improving the quality of search engine results and preventing spamdexing from occurring in the first place. (Wikipedia[119])

> Spamdexing involves a number of methods, such as repeating unrelated phrases, to manipulate the relevancy or prominence of resources indexed by a search engine, in a manner inconsistent with the purpose of the indexing system.[1] [2] Some consider it to be a part of search engine optimization, though there are many search engine optimization methods that improve the quality and appearance of the content of web sites and serve content useful to many users.[3 (Wikipedia[120])

Three days after Google's announcement WordPress adopted the nofollow attribute and added it to links placed in the comment section[121]. Other major blog software makers including Blogger, Six Apart and LiveJournal also adopted the attribute and it became a standard in blog software.

The attribute is automatically applied to any comment in the comment section, whether it is a legit comment or spam link, trackback or pingback. Nofollow is integrated into the core of the software with no option to turn it off except for manually adjusting the code or installing a plugin that will remove the attribute. Nofollow may have been a solution to spamdexing as it prevents spam comments from entering and flooding the index but it did not reduce the amount of spam. On top of that it does not acknowledge trustworthy and legit comments. This has led to the rejection of nofollow by many bloggers who wish to acknowledge legit comments

---

118 "nofollow is a misnomer in this case since it actually tells a search engine "Don't score this link" rather than "Don't follow this link." This differs from the meaning of nofollow as used within a robots meta tag, which does tell a search engine: "Do not follow any of the hyperlinks in the body of this document." http://en.wikipedia.org/wiki/Spam_in_blogs (accessed December 28, 2007).
On top of that the engines treat the attribute differently as Google will actually follow the link but not index it and Yahoo! both follows and indexes the link. (http://www.seo-blog.com/rel-nofollow.php) (accessed January 11, 2008)

119 http://en.wikipedia.org/wiki/Nofollow (accessed December 28, 2007)

120 http://en.wikipedia.org/wiki/Spamdexing (accessed December 28, 2007)

121 http://trac.wordpress.org/changeset/2117 (accessed December 28, 2007).

and contribute to their ranking. The practice of linking is important to bloggers because "as content recommenders - as citation specialists - we create value with and benefit from the work of others." (Stevenson, 2007c)

To acknowledge the value of the practice of linking a NoNofollow movement and[122] a DoFollow community[123] were established and several dofollow plugins[124] were relased. These dofollow plugins, with names such as Link Love, GoodLinks and NoNofollow remove the nofollow attribute from links in comments. They insert the links back into the indexes of the engines.

4.4 The Blogosphere Under the Influence of Software-Engine Relations

Daily blogging practices show us the increasingly symbiotic relationship between the software and the engines and "the daily grind of spam-cleaning brings users directly into the mess of software-engine politics." (Stevenson, 2007b) As we saw with nofollow, this means that bloggers are subject to a software-engine regime. This regime determines how the blogosphere is constructed by the protocological control of the engines.

The software-engine relations show us how the blog and the blogosphere are constructed by the blog's canonical features as implemented in the software. The technical software-engine relations and the daily blogging routine collide in the case of spam and nofollow. Spam shows us the entanglement of the software with the engines by using the software to create splogs and sportals and by using exploits in blog protocols to syndicate existing blog content and post spam comments. Bloggers try to deal with these exploits by installing plugins while the blog software deals with these issues in collaboration with the engines. Nofollow points to this relationship and puts bloggers under a software-engine regime.

The blogosphere is constructed by an 'abstract cooperation'[125] between bloggers, blog software and the blog engines. In this case the abstract cooperation is "all the

---

122 http://nonofollow.net/ (accessed January 11, 2008)
123 http://www.bumpzee.com/no-nofollow/ (accessed January 11, 2008)
124 http://andybeard.eu/2007/02/ultimate-list-of-dofollow-plugins-banish-nofollow-from-comments-and-trackbacks.html (accessed January 11, 2008)
125 As used by Christoph Spehr in 'Out-Cooperating the Empire?
     Exchange between Geert Lovink and Christoph Spehr on Creative Labour and the Hybrid Work of Cooperation (July 2006)' http://www.networkcultures.org/geert/out-cooperating-the-empire-exchange-with-christoph-spehr/ (accessed January 11, 2008)

working-together we aren't necessarily aware of – from Internet Protocols to the 'added value' of hyperlinks for Google's search engine." (Stevenson, 2007b) The bloggers construct the blogosphere by the act of linking, commenting, tagging and bookmarking. The blog software cooperates by providing blog standards and protocols such as tags and site feeds that are picked up by the engines and pings that notify central servers of new content. The engines construct the blogosphere by providing us access to their 'totality' of blogs.

The blog software makes sure that the construction of the blogosphere by the engines is maintained as we can see in the case of nofollow. But what would happen if bloggers would "disable the Nofollow attribute *en masse*, invite the spammers in and watch as Google rankings become unsettled?" (Stevenson, 2007c) What if the ping servers went down to an overload of spings?[126]

---

126 Spam ping, pings sent from spam blogs.

5. CONCLUSION

Blogs both as a medium and bi-product of practice have changed over the past ten years under the regime of software-engine relations. Not only has the number of blogs increased incredibly but also their relation to the engines. Bloggers are increasingly entangled into the disarray of software-engine politics that leaves the blogger with the obsessive checking of statistics and the removal of spam comments that have by-passed the spam filters. It has also significantly changed one of the oldest blogging practices, namely that of linking. The social bookmarking links that feed the blog to the engines have given birth to the so-called Digg effect, the demise of the blogroll points to the feedreader as the new habitat of the blog and the new HTML link attribute of nofollow points to the increasingly unhealthy relationship between software and engines.

The history of blogs, blog software and the engines shows how this relationship has developed through the standardization of blog features. Blog software has helped to define the format of the blog by adding the reverse-chronological structure to its core and by developing and introducing certain blog features in collaboration with the blogging community. The permalink and the trackback were developed to serve the common practice of linking and made interlinking blogging practices visible. The pingback automated this proces of linking by automatically sending a notification to the blog that it is being linked to. This notification method, the ping, and the feed protocols have enabled the construction of a blogosphere by the engines. The blogosphere as the totality of blogs is a distributed network in which the engines act as coalescing and centralizing power structures.

The engines are increasingly entangled in the medium and practice of blogs which has implications on a phenomenological, cultural and political level. Different engines construct different blogospheres but they all adhere to the native unit of the blog, the post. The blog post clearly distinguishes the blog from its ancestor, the personal homepage which was based on the unit of the page. The post allows for frequent updates and the new permanent link to the blog post has enabled it to become the building stone of the blogosphere.

Blogs are web native objects that thrive on links, distribution, syndication and

updates. These distinctive features are built into the software and used to communicate with the engines. The blogger is both enabled and constrained by a web of relations between the blog, blogger, blog software and blog engines. Both the blog and the blogosphere are under the influence of software-engine relations. These relations work two ways as the engines use blog software protocols to index and construct the blogosphere and blog software integrates engine-defined standards.

The symbiotic relationship with the engines has led to different blogging practices including Search Engine Optimization that keep the engines in mind while constructing a blog and writing blog posts. The morally ambiguous practices of SEO and spam exploit the features of the blogosphere for financial reasons. As spam unsettles the daily blogging routine and the engines, a steady software-engine regime is enforced.

It gives us reasons to think about what would happen if bloggers would break out of this regime by disabling the default settings in the software.

REFERENCES

Adlam, Teli. "WordPress Trackback Tutorial." *WordPress SEO and Blog Marketing*. Blog Post.
  December 27, 2005. http://www.optiniche.com/blog/117/wordpress-trackback-tutorial/
  (accessed December 6, 2007).

Alton-Scheidl, Roland. "Redesigning the Concept of Bulletin Board Systems: Towards the
  Needs of an Emerging Mass Media." *Arachnet Electronic Journal on Virtual Culture* 1, no.
  4 (June 30, 1993). http://serials.infomotions.com/aejvc/aejvc-v1n04-alton-scheidl-
  redesigning.txt (accessed November 26, 2007).

Altena, Arie. "UbiScribe: Collocollaboracontentquery?" Public colloquium. April 15, 2007.
  http://www.deaf07.nl/index.php?option=com_program&act=event&task=show&id=37&.
  (accessed December 3, 2007).

Armonth. "Fix for incoming links from Technorati in WordPress." Blog Post. *SigT*, July 2, 2007.
  http://sigt.net/archivo/fix-for-incoming-links-from-technorati-in-wordpress.xhtml
  (accessed December 26, 2007).

Barger, Jorn. "Top 10 Tips for New Bloggers From Original Blogger Jorn Barger." *Wired*,
  December 15, 2007. http://www.wired.com/culture/lifestyle/news/2007/12/blog_advice
  (accessed January 7, 2008).

Bausch, Paul. "On Permalinks and Paradigms..." Blog Post Comment. *plasticbag.org*, June 11,
  2003.
  http://www.plasticbag.org/archives/2003/06/on_permalinks_and_paradigms/#comment
  -1116 (accessed December 5, 2007).

Baym, Nancy. "The Widgetized Self." Blog Post. *Online Fandom*, April 9, 2007.
  http://www.onlinefandom.com/archives/the-widgetized-self/ (accessed January 27, 2008).

Beard, Andy. "IzeaRanks & RealRank - How Many Lies Can You Tell Your Advertisers?" Blog
  Post. *Andy Beard - Niche Marketing*, January 10, 2008.
  http://andybeard.eu/2008/01/izearanks-realrank.html (accessed January 11, 2008).

Berners-Lee, Tim. "WorldWideWeb: Summary - alt.hypertext." Online Posting. *Google Groups*,
  Aug 6, 1991. http://groups.google.com/group/alt.hypertext/msg/395f282a67a1916c
  (accessed November 29, 2007).

Blood, Rebecca. "How blogging software reshapes the online community." *Commun. ACM*.
  47.12 (2004): 53-55.

Blood, Rebecca. "Weblogs: A History And Perspective." Blog Post. *Rebecca's Pocket*, September
  7, 2000. http://www.rebeccablood.net/essays/weblog_history.html (accessed November
  22, 2007).

Blood, Rebecca. *The Weblog Handbook: Practical Advice on Creating and Maintaining Your*

Blog. Da Capo Press, 2002.

Blood, Rebecca. *We've Got Blog: How Weblogs Are Changing Our Culture*. Perseus, 2003.

boyd, danah. "removing my blogroll." Blog Post. *apophenia*, November 16, 2003.
http://www.zephoria.org/thoughts/archives/2003/11/16/removing_my_blogroll.html
(accessed January 27, 2008).

boyd, danah. "A Blogger's Blog: Exploring the Definition of a Medium." *Reconstruction*, 6.4,
2006. http://reconstruction.eserver.org/064/boyd.shtml (accessed 6 September 2007).

Bristow, Helena. "Software Studies: ABOUT." *Software Studies @ UCSD*, May 18, 2007. http://
softwaretheory.net/2007/05/about-software-studies-ucsd.html (accessed January 23,
2008).

Broache, Anne, and Declan McCullagh. "Blogging's roots reach to the '70s | CNET News.com."
CNET News.com, March 20, 2007. http://www.news.com/2100-1025_3-6168685.html?
part=rss&tag=2547-1_3-0-5&subj=news (accessed November 26, 2007).

Bruns, Axel. *Gatewatching: Collaborative Online News Production*. New York: Peter Lang
Publishing, 2005.

Bruns, Axel, and Joanne Jacobs. *Uses of Blogs*. New York, USA: Peter Lang Publishing, 2006.

Çelik, Tantek. "RelTag." *Technorati Developers Wiki*, January 10, 2005.
http://developers.technorati.com/wiki/RelTag (accessed December 25, 2007).

Chandler, Daniel. "Personal Home Pages and the Construction of Identities on the Web."
August 18, 1998. http://www.aber.ac.uk/media/Documents/short/webident.html (accessed
January 5, 2008).

CNNIC. *The Survey Report on Blogs in China 2007*. China Internet Network Infomation
Center, 2007. http://www.cnnic.cn/html/Dir/2007/12/27/4954.htm (accessed January 7,
2008).

Conniff, Michael. "Just what is a blog, anyway?" *USC Annenberg. Online Journalism Review*,
September 29, 2005. http://www.ojr.org/ojr/stories/050929/ (accessed September 6,
2007).

Conroy, John. "Results: Most Popular CMS in Technorati's Top 100." Blog Post. *CMS Wire*,
January 23, 2008. http://www.cmswire.com/cms/micro-cms/results-most-popular-cms-
in-technoratis-top-100-002198.php (accessed January 24, 2008).

Cooper, Peter. "Wall Street Journal Tries to Re-Write Blogging History." Blog Post Comment.
Duncanriley.com, July 16, 2007. http://www.duncanriley.com/2007/07/15/wall-street-
journal-tries-to-re-write-blogging-history/#comment-158722 (accessed November 29,
2007).

Dagza, Ralph. "How to install WordPress on a USB stick." Blog Post. *Ralph dot FM*, March 8,
2007. http://ralph.fm/2007/03/08/how-to-install-wordpress-on-a-usb-stick/ (accessed
January 6, 2008).

Dash, Anil. "Google and Theory of Mind." Blog Post. *Anil Dash*, December 14, 2007.

http://www.dashes.com/anil/2007/12/google-and-theory-of-mind.html (accessed January 7, 2008).

Delio, Michelle. "Thousands of Blogs Fall Silent." *Wired*, June 15, 2004. http://www.wired.com/culture/lifestyle/news/2004/06/63856 (accessed November 29, 2007).

Digital Methods Initiative. "WebHome < Dmi < TWiki." *Welcome to the Digital Methods Initiative.* http://wiki2.issuecrawler.net/twiki/bin/view/Dmi/WebHome (accessed November 22, 2007).

Dodge, Don. "Search engines are the Start page for the Internet." Blog Post. *Don Dodge on The Next Big Thing*, December 13, 2007. http://microsoftstartupzone.com/blogs/the_next_big_thing/archive/2007/12/13/search-engines-are-the-start-page-for-the-internet.aspx (accessed January 23, 2008).

Farooq, Ahmed. "Bloggy Network: Blog Flux." *Tech Soapbox*, July 4, 2007. http://www.techsoapbox.com/bloggy-network-blog-flux/ (accessed December 3, 2007).

Evans, Meryl K. "Matt Mullenweg." *Digital Web Magazine*, November 24, 2004. http://www.digital-web.com/articles/matt_mullenweg/ (accessed December 3, 2007).

Fuller, Matthew. *Behind the Blip: Essays on the Culture of Software.* Brooklyn, USA: Autonomedia, 2003.

Fuller, Matthew. "Software Studies Workshop" Piet Zwart Institute. http://pzwart.wdka.hro.nl/mdr/Seminars2/softstudworkshop (accessed 11 January 2007).

Fuller, Matthew, ed. *Software Studies.* Cambridge, USA: The MIT Press (forthcoming).

Galloway, Alexander R., and Eugene Thacker. *The Exploit: A Theory of Networks.* Minneapolis: University Of Minnesota Press, 2007.

Galloway, Alexander R. *Protocol: How Control Exists after Decentralization.* Cambridge: The MIT Press, 2004.

Garrett, Chris. "Blogging On the Road … Or That Was the Plan!" Blog Post. *chrisg.com*, April 2, 2007. http://www.chrisg.com/blogging-on-the-road-or-that-was-the-plan/ (accessed January 6, 2008).

Gillmor, Dan. *We the Media: Grassroots Journalism By the People, For the People.* New Ed. Sebastopol, USA: O'Reilly Media, Inc., 2006.

Goedegebuure, Dennis. "Link Love & other SEO." *The Next Corner.* http://www.thenextcorner.com/link-love.html (accessed December 27, 2007).

Goldman, Jason. "Find out what's happening with Blog Search." Blog Post. *The Official Google Blog*, September 14, 2005. http://googleblog.blogspot.com/2005/09/find-out-whats-happening-with-blog.html (accessed December 3, 2007).

Gudrais, Elizabeth. "Googling Google." *Harvard Magazine*, November-December 2007. http://harvardmagazine.com/2007/11/googling-google.html (accessed January 23, 2008).

Haughey, Matt. "On Permalinks and Paradigms…" Blog Post Comment. *plasticbag.org*, June

11, 2003.
http://www.plasticbag.org/archives/2003/06/on_permalinks_and_paradigms/#comment-1116 (accessed December 5, 2007).

Hayles, N. Katherine. *My Mother Was a Computer: Digital Subjects and Literary Texts*. University of Chicago, 2005.

Hargittai, E. The social, political, economic, and cultural dimensions of search engines: An introduction. *Journal of Computer-Mediated Communication*, *12*(3), article 1. 2007. http://jcmc.indiana.edu/vol12/issue3/hargittai.html (accessed January 27, 2008)

Helmond, Anne. "Can Google Handle the Maturing Blogosphere?" Blog Post. *The Blog Herald*, October 28, 2007(a). http://www.blogherald.com/2007/10/28/can-google-handle-the-maturing-blogosphere/ (accessed December 6, 2007).

Helmond, Anne. "Essay: The Perceived Freshness Fetish," *Anne Helmond: New Media Master research blog*. 22 June 2007(b). http://www.annehelmond.nl/2007/06/22/essay-the-perceived-freshness-fetish/ (accessed 12 September 2007).

Helmond, Anne. "On Using Manual and/or Automatic Link Notification Systems." Blog Post. *The Blog Herald*, January 7, 2008. http://www.blogherald.com/2008/01/07/on-using-manual-andor-automatic-link-notification-systems/ (accessed January 8, 2008).

Helmond, Anne. "Software Studies Review." (Unpublished). University of Amsterdam, June 2007(c).

Hewitt, Hugh. *Blog: Understanding the Information Reformation That's Changing Your World*. New Ed. Nashville, USA: Thomas Nelson, 2006.

Hourihan, Meg. "On Permalinks and Paradigms..." Blog Post Comment. *plasticbag.org*, June 12, 2003. http://www.plasticbag.org/archives/2003/06/on_permalinks_and_paradigms/#comment-1124 (accessed December 5, 2007).

Hourihan, Meg. "What We're Doing When We Blog." Blog Post. *O'Reilly Network*, June 13, 2002. http://www.oreillynet.com/pub/a/JavaScript/2002/06/13/megnut.html (accessed November 24, 2007).

Interviewbooks.com. "Net Success Interviews - OpenDiary.com." http://www.interviewbooks.com/opendiary_com.htm (accessed November 22, 2007).

Introna, Lucas & Nissenbaum, H. "Defining the Web: The Politics of Search Engines." *Computer*. 33, no. 1 (January 2000): 54-62.

Jaquith, Mark. "Matt Cutts praises WordPress SEO benefits." Blog Post. *Mark on WordPress*, December 7, 2007. http://markjaquith.wordpress.com/2007/12/07/matt-cutts-praises-wordpress-seo-benefits/ (accessed December 11, 2007).

jules. "Blog Stats: Oranges vs. Apples vs. Plums vs. Pears." Blog Post Comment. *Mark Evans*, April 3, 2007. http://www.markevanstech.com/2007/04/03/blog-stats-oranges-vs-apples-vs-plums-vs-pears/#comment-4944 (accessed January 11, 2008).

Joyner, April. "Widgets: From Hype to Hits." *Fastcompany.com*, January 2008.

http://www.fastcompany.com/articles/2008/01/widgets.html (accessed January 17, 2008).

Kafkaesqui, "Do pages send out a RSS feed?," Online Posting. *WordPress › Support*, 5
December 2007. http://www.wordpress.org/support/topic/146439?replies=3#post-658523
(accessed 6 December 2007).

Kahney, Leander. "The Web the Way It Was." *Wired*, February 23, 2000.
http://www.wired.com/culture/lifestyle/news/2000/02/34006 (accessed December 3,
2007).

Kahney, Leander. "Dot-Com Begs for Bucks." *Wired*, January 4, 2001.
http://www.wired.com/techbiz/media/news/2001/01/40979 (accessed November 24,
2007).

Kallen, Ian. "Technorati is Toast: Drops Older Posts from Index." Blog Post Comment. *Zoli's
Blog*, November 4, 2007. http://www.zoliblog.com/2007/11/04/technorati-drops-old-
posts-from-index/#comment-6922 (accessed December 5, 2007).

Kottke, Jason. "Finally. Did you notice the." Blog Post. *kottke.org*, March 5, 2000.
http://www.kottke.org/00/03/finally-did-you-notice-the (accessed December 5, 2007).

Kramer, Staci D. "Weblogs.com Rises From the Flames." *Wired*, June 23, 2004.
http://www.wired.com/culture/lifestyle/news/2004/06/63953 (accessed November 29,
2007).

Krishnan, Ramayya. "Understanding Web Technology." http://64.233.183.104/search?
q=cache:9tzRhd4uMk4J:www.andrew.cmu.edu/user/rk2x/dssclass.dir/webtech.ppt
+Krishnan+using+the+browser+as+a+front+end+to
+applications&hl=en&ct=clnk&cd=3&client=firefox-a (accessed January 6, 2008).

Lampa, Graham. "Imagining the Blogosphere: An Introduction to the Imagined Community of
Instant Publishing." *Into The Blogosphere: Rhetoric, Community, and Culture of Weblogs*,
2004. http://blog.lib.umn.edu/blogosphere/imagining_the_blogosphere.html (accessed
December 3, 2007).

Leaver, Tama. "The Blogging of Everyday Life." *Reconstruction 6.4 (2006)*.
http://reconstruction.eserver.org/064/leaver.shtml (accessed October 25, 2007).

Lenhart, Amanda & S. Fox. *Bloggers. A portrait of the internet's new storytellers.*. Pew
Internet & American Life Project, 2006. http://www.pewinternet.org/pdfs/PIP
%20Bloggers%20Report%20July%2019%202006.pdf (accessed January 5, 2008).

Lenhart, Amanda, et. al. *Teens and Social Media*. Pew Internet & American Life Project, 2007.
http://www.pewinternet.org/pdfs/PIP_Teens_Social_Media_Final.pdf (accessed January
5, 2008).

Lerdorf, Rasmus . "Do You PHP?"
http://www.oracle.com/technology/pub/articles/php_experts/rasmus_php.html (accessed
28 February 2007).

Lovink, Geert. "Blogging, the nihilist impulse." *Eurozine*, January 2, 2007.

http://www.eurozine.com/articles/2007-01-02-lovink-en.html (accessed November 27, 2007).

Marlow. "About | overstated." http://overstated.net/about (accessed December 3, 2007).

Mann, Charles C. " 14.09: Spam + Blogs = Trouble." *Wired*, September 2006. http://www.wired.com/wired/archive/14.09/splogs.html (accessed December 27, 2007).

Manovich, Lev. *The Language of New Media*. Cambridge, USA: The MIT Press, 2001.

Marlow, Cameron. "Audience, structure and authority in the weblog." Presented at the International Communication Association Conference. May 2004. http://alumni.media.mit.edu/~cameron/cv/pubs/04-01.pdf (accessed January 27, 2008).

McFarland, Rod. "Rethinking the Blog as Database: My First Post on the Blog Herald." Blog Post Comment. *Anne Helmond* . http://www.annehelmond.nl/2007/10/16/rethinking-the-blog-as-database-my-first-post-on-the-blog-herald/ (accessed January 14, 2008).

Merholz, Peter. "Play With Your Words." Blog Post. *peterme.com*, May 17, 2002. http://www.peterme.com/archives/00000205.html (accessed November 23, 2007).

Moolanomy. "Life Hacks, Productivity, and Blogging." Blog Post. *Moolanomy*, September 29, 2007. http://www.moolanomy.com/186/life-hacks-productivity-and-blogging/ (accessed January 11, 2008).

Mush, Dullard. "An Obsession With Stats and Other New Blogger Musings." Blog Post. *Dullard Mush*, September 7, 2006. http://dullardmush.blogspot.com/2006/09/obsession-with-stats-and-other-new.html (accessed January 11, 2008).

Naraine, Ryan. "VeriSign Acquires Dave Winer's Weblogs.com." *eWeek.com*, October 6, 2005. http://www.eweek.com/article2/0,1895,1868270,00.asp (accessed November 29, 2007).

Nardi, Bonnie A., Diane J. Schiano, and Michelle Gumbrecht, *Blogging as social activity, or, would you let 900 million people read your diary?* Chicago, Illinois, USA : ACM Press, 2004.

Nardi, Bonnie A. et al., "Why we blog," *Communications of the ACM*, 47 .12 (2004), 41-46 .

Nearthwort. "Nearthwort Obtain." February 21, 2007. Podcast. http://nearthwort.com/2007/02/21/nearthwort-podcast-14-wordpress-founder-matt-mullenweg/ (accessed 24 February , 2007).

OECD. "OECD Broadband Statistics to June 2006." (2006). 14 April 2007 http://www.oecd.org/sti/ict/broadband (accessed November 23, 2007).

O'Reilly, Tim. "What Is Web 2.0. Design Patterns and Business Models for the Next Generation of Software." *O'Reilly Network*, September 30, 2005. http://www.oreillynet.com/lpt/a/6228 (accessed November 24, 2007).

Pavlina, Steve. "More on Blog Comments." Blog Post. *Steve Pavlina . com*, October 15, 2005. http://www.stevepavlina.com/blog/2005/10/more-on-blog-comments/ (accessed January 27, 2008).

Pietrobruno, Sheenagh. "The Immensity within the Minute: Forging Digital Space." Inter: A

European Cultural Studies : Conference in Sweden 11–13 June 2007. Linköping Electronic
Conference Proceedings. http://www.ep.liu.se/ecp/025/053/ecp072553.pdf (accessed
January 14, 2008).

Poster, Mark. *The Mode of Information: Poststructuralism and Social Context*. University Of
Chicago Press, 1990.

Powers, Shelley. "The Business of Algorithms." *Burningbird*, August 12, 2005.
http://burningbird.net/connecting/the-business-of-algorithms/ (accessed December 3,
2007).

Rainie, Lee. *Pew Internet: Blogosphere*. Pew Internet & American Life Project, 2005.
http://www.pewinternet.org/PPF/r/144/report_display.asp (accessed January 3, 2008).

Ratzkinzin, Joe. "Will Someone Who Actually Cares About Blogging Please Write the History Of
it?" Blog Post Comment. *CrunchNotes*, August 11, 2007. http://www.crunchnotes.com/?
p=419 (accessed November 29, 2007).

Riley, Duncan. "A short history of blogging." Blog Post. *The Blog Herald*, March 6, 2005.
http://www.blogherald.com/2005/03/06/a-short-history-of-blogging/ (accessed
December 3, 2007).

Rogers, Richard. *Information Politics on the Web*. Cambridge, USA: The MIT Press, 2004.

Rowse, Darren. "Declutter your Sidebar." Blog Post. *ProBlogger*. August 10, 2007.
http://www.problogger.net/archives/2007/08/10/declutter-your-sidebar/ (accessed
November 23, 2007).

Rowse, Darren. "Help! I'm Addicted to Checking My Blog's Stats!" Blog Post. *ProBlogger*. June
4, 2007. http://www.problogger.net/archives/2007/06/04/help-im-addicted-to-checking-
my-blogs-stats/ (accessed January 11, 2008).

Rowse, Darren. "What Blogging Platform Do You Use? [POLL RESULTS]." Blog Post.
ProBlogger, December 3, 2007. http://www.problogger.net/archives/2007/12/03/what-
blogging-platform-do-you-use-poll-results/ (accessed December 3, 2007).

Rundle, Mike. "Movable Type vs. WordPress, My Opinion." Blog Post. *Business Logs*, May 9,
2006.
http://www.businesslogs.com/blogging_software/movable_type_vs_wordpress_my_opini
on.php (accessed January 12, 2008).

Schaap, Frank. "Links, Lives, Logs: Presentation in the Dutch Blogosphere." *Into the
Blogosphere: Rhetoric, Community, and Culture of Weblogs*, 2004.
http://blog.lib.umn.edu/blogosphere/links_lives_logs.html (accessed January 7, 2008).

Seabrook, John. "Home on the Net." October 16, 1995.
http://www.levity.com/seabrook/homenet.html (accessed January 6, 2008).

Sifry, David. "Sifry's Alerts: The State of the Live Web, April 2007." Blog Post.
http://www.sifry.com/alerts/archives/000493.html (accessed September 6, 2007).

Schmidt, Jan. "Blogging Practices: An Analytical Framework." *Journal of Computer-Mediated*

Communication, 12(4), article 13. (2007).
http://jcmc.indiana.edu/vol12/issue4/schmidt.html (accessed January 23, 2008).

Schmidt, Jan, M. Wilbers and M. Paetzolt. "Use of and satisfaction with blogging software."
Research Centre „New Communication Media" Working Paper 06-04 (2006).
http://www.fonk-bamberg.de/pdf/fonkpaper0604.pdf (accessed November 23, 2007).

Smith, Justin. "Blog Archive » Blog Stats Checking Obsession." Blog Post. *My Life With IT* ,
July 12, 2007. http://itlife.escrito.info/2007/07/12/blog-stats-checking-obsession/
(accessed January 11, 2008).

Smith, Steve. "Tracking The 'Commentosphere'." Blog Post. *MediaPost's Behavioral Insider*,
April 27, 2007. http://blogs.mediapost.com/behavioral_insider/?p=151 (accessed
December 27, 2007).

Stevenson, Michael. "New Network Theory - Siva Vaidhyanathan." Blog Post. *Masters of Media*
, June 28, 2007(a). http://mastersofmedia.hum.uva.nl/2007/06/28/new-network-theory-
siva-vaidhyanathan/ (accessed January 27, 2008).

Stevenson, Michael. Personal Skype Conversation. December 28, 2007(b).

Stevenson, Michael. Personal E-mail Conversation. January 13, 2008.

Stevenson, Michael. "Pingback Spam, Popularity and Protecting Investments." Blog Post.
Masters of Media , October 9, 2007(c).
http://mastersofmedia.hum.uva.nl/2007/10/09/pingback-spam-popularity-and-
protecting-investments/ (accessed December 28, 2007).

Striegel, Jason. "Command Line Interface for WordPress." *Hackszine.com*.
http://www.hackszine.com/blog/archive/2007/04/command_line_interface_for_wor.htm
l (accessed May 20, 2007).

Truscello, Michael. "Behind the Blip: Essays on the Culture of Software (review)." *Cultural
Critique.* Number 63 (Spring 2006): 182-187.

Truscello, Michael."The Rhetorical Ecology of the Technical Effect," *Technical Communication
Quarterly*, 14.3 (2005), 345-351 .

Turnbull, Giles. "Introducing MovableType." Blog Post. *WriteTheWeb*, October 2, 2001. http://
writetheweb.com/Members/gilest/old/115/ (accessed November 30, 2007).

Turnbull, Giles. "The heart of the matter." Blog Post. *WriteTheWeb*, February 7, 2003.
http://writetheweb.com/Members/gilest/noahgrey/ (accessed November 30, 2007).

Turnbull, Giles. "The state of the blog." Blog Post. *WriteTheWeb*, February 28, 2001.
http://writetheweb.com/Members/gilest/old/107/ (accessed September 6, 2007).

Vaidhyanathan, Siva. "The Theology of Google." June 28, 2007(a). Transcription of Conference
Talk at the New Network Theory Conference at the University of Amsterdam.

Vaidhyanathan, Siva. "Where is this book going?" Blog Post. *The Googlization of Everything*,
September 25, 2007(b).
http://www.googlizationofeverything.com/2007/09/where_is_this_book_going.php

(accessed January 16, 2008).

Van Couvering, Elizabeth. "New Media? The Political Economy of Internet Search Engines." Conference of the International Association of Media & Communications Researchers (IAMCR) at Porto Alegre, Brazil, 2004.

Vander Wal, Thomas. "A Stale State of Tagging?" Blog Post. *Off the Top :: vanderwal.net*, August 30, 2007. http://www.vanderwal.net/random/entrysel.php?blog=1945 (accessed January 27, 2008).

VanFossen, Lorelle. "Playing with WordPress.com New Sidebar Widgets." Blog Post. *Lorelle on WordPress*, March 10, 2006a. http://lorelle.wordpress.com/2006/03/10/playing-with-wordpresscom-new-sidebar-widgets/ (accessed January 11, 2008).

VanFossen. "Putting Some Thought Into Blog Categories and Tags." Blog Post. *Lorelle on WordPress*, March 7, 2006b. http://lorelle.wordpress.com/2006/03/07/putting-some-thought-into-blog-categories-and-tags/ (accessed November 28, 2007).

VanFossen, Lorelle. "Spinning Spammers Steal Our Blog Content." Blog Post. *Lorelle on WordPress*, November 15, 2007. http://lorelle.wordpress.com/2007/11/15/spinning-spammers-steal-our-blog-content/ (accessed December 27, 2007).

VanFossen, Lorelle. "The Art of Backlinking." Blog Post. *The Blog Herald*, December 24, 2007. http://www.blogherald.com/2007/12/24/the-art-of-backlinking/ (accessed December 25, 2007).

VanFossen, Lorelle. "WordPress Widgets: The Next Generation of WordPress Plugins" Blog Post. *Lorelle on WordPress*, February 13, 2007. http://lorelle.wordpress.com/2007/02/13/wordpress-widgets-the-next-generation-of-wordpress-plugins/ (accessed September 6, 2007).

VanFossen, Lorelle. "Writing With Post Excerpts and Feed Excerpts in Mind." Blog Post. *Lorelle on WordPress*, July 21, 2006b. http://lorelle.wordpress.com/2006/07/21/writing-with-post-excerpts-and-feed-excerpts-in-mind/ (accessed December 27, 2007).

van Oosten de Boer, Caroline. "It's a long way into." Blog Post. *prolific.org classic*, March 6, 2000. http://prolific.org/classic/2000/03/06/its_a_long_way_into/ (accessed December 5, 2007).

W, James. "Will Someone Who Actually Cares About Blogging Please Write the History Of it?" Blog Post Comment. *CrunchNotes*, July 16, 2007. http://www.crunchnotes.com/?p=419 (accessed November 29, 2007).

Walker-Rettberg, Jill. "final version of weblog definition." Blog Post. *jill/txt*, June 28, 2003. http://jilltxt.net/archives/blog_theorising/final_version_of_weblog_definition.html (accessed November 22, 2007).

Web Hosting Folio. "The Evolution of WordPress | Blogging." Web Hosting Folio. December 10, 2006. http://www.hostingfolio.com/blogging/the-evolution-of-wordpress (accessed November 29, 2007).

Weltevrede, Esther. "Google and the failure of protocol. How centralization exists after protocol." (Unpublished thesis proposal). January 12, 2007.

Wilson, Nick. "Exclusive: Google Teaches Bloggers How To Rank." Blog Post. *Performancing.com*, December 13, 2005. http://performancing.com/node/397 (accessed December 6, 2007).

Zetter, Kim. "How Humble BBS Begat Wired World." *Wired*, June 8, 2005. http://www.wired.com/culture/lifestyle/news/2005/06/67776?currentPage=3 (accessed November 26, 2007).

APPENDIX 1: MY RSS FEED LIST

New Media

1. 901am- http://www.901am.com/
2. Andy Beard - Niche Marketing - http://andybeard.eu/
3. Anil Dash - http://www.dashes.com/anil/
4. Anne Helmond - http://www.annehelmond.nl/[127]
5. apophenia - http://www.zephoria.org/thoughts/
6. ariealt.net - http://ariealt.net/blog
7. Boing Boing - http://www.boingboing.net/
8. Checklist - http://www.erwinblom.nl/
9. Collective Thoughts - http://collective-thoughts.com/
10. Deep Jive Interests - http://www.deepjiveinterests.com/
11. Deeplinking - http://deeplinking.net/
12. First Monday Podcast - http://www.firstmondaypodcast.org/
13. gapingvoid: "cartoons drawn on the back of business cards"- http://www.gapingvoid.com/
14. Go Digital, Feel Analog - http://nonormalnews.org/blog
15. The Googlization of Everything http://www.googlizationofeverything.com/
16. Grand Text Auto - http://grandtextauto.org/
17. hypernarrative.com, a weblog by Wilbert Baan about Art, Media and Technology - http://www.hypernarrative.com/wordpress
18. jill/txt - http://jilltxt.net/
19. kenspeckle - http://kenspeckle.net/blog
20. neWMW - http://newmw.wordpress.com/
21. {nowhere.now.here} - http://katalinlovasz.net/pictures
22. maartenbrinkerink.net/blog - http://maartenbrinkerink.net/blog
23. Mashable! - http://mashable.com/
24. Masters of Media - http://mastersofmedia.hum.uva.nl/
25. Mastersofmedia.hum.uva.nl blogroll - http://mastersofmedia.g4wd.net/

---

127 Subscribing to your own feed is important for various reasons including knowing what your posts look like in a feed reader and also to verify that your feed is working.

26. Pearsonified - http://www.pearsonified.com/

27. plasticbag.org - http://www.plasticbag.org/

28. Read/WriteWeb - http://www.readwriteweb.com/

29. serial consign - design / research - http://serialconsign.com/

30. Tama Leaver dot Net - http://www.tamaleaver.net/

31. TechCrunch - http://www.techcrunch.com/

32. VPRO Kenniscentrum - http://weblogs.vpro.nl/kenniscentrum

33. Web Worker Daily - http://webworkerdaily.com/

34. What's New Media? - http://whatsnewmedia.org/

35. wordpress.justlol.net - http://wordpress.justlol.net/

36. WRT: Writer Response Theory - http://writerresponsetheory.org/wordpress

37. xkcd.com - http://xkcd.com/

Blogging

1. About:Blank - http://www.aboutblank.nl/index.php

2. BloggerTalks - http://bloggertalks.com/

3. Blogging Pro - http://www.bloggingpro.com/

4. BlogTalk Reloaded - http://2008.blogtalk.net/blog

5. Comments on: Someone must be blogging about blogs, right? - http://ask.metafilter.com/79749/Someone-must-be-blogging-about-blogs-right/

6. Hack WordPress - http://hackwordpress.com/

7. The Blog Herald - http://www.blogherald.com/

8. Lorelle on WordPress - http://lorelle.wordpress.com/

9. Performancing.com - Helping Bloggers Succeed - http://performancing.com/

10. ProBlogger Blog Tips - http://www.problogger.net/

11. Weblog Tools Collection - http://weblogtoolscollection.com/

12. WordPress Tutorials And Blogging Tips - http://www.easywordpress.com/labs

Design

1. A List Apart - http://www.alistapart.com/
2. Bartelme Design | Journal - http://www.bartelme.at/
3. Brand New - http://www.underconsideration.com/brandnew/
4. B R I G H T – tech/life/style - http://www.bright.nl/
5. Daily Slurp Gallery - http://www.dailyslurp.com/
6. Design Melt Down - http://www.designmeltdown.com/
7. Devlounge - http://www.devlounge.net/
8. fadtastic - a multi-author web design trends journal - http://fadtastic.net/
9. FFFFOUND! / EVERYONE - http://ffffound.com/
10. It's Nice That - http://www.itsnicethat.com/
11. Naar Voren » Artikelen - http://www.naarvoren.nl/
12. Naar Voren » De Zijlijn - http://www.naarvoren.nl/zijlijn/
13. Photography and Art, Mostly... - http://www.openorigins.com/
14. PSDTuts - Photoshop Tutorials and Links - http://psdtuts.com/
15. Smashing Magazine - http://www.smashingmagazine.com/
16. Devlounge - http://www.devlounge.net/

General Blogs

1. 9rules Notes: Last 35 Posts - http://9rules.com/notes/%3E
2. Lifehacker - http://lifehacker.com/
3. lifehacking.nl - http://lifehacking.nl/
4. Lolcats 'n' Funny Pictures - I Can Has Cheezburger? - http://icanhascheezburger.com/
5. maartenbrinkerink.net/blog - http://maartenbrinkerink.net/blog
6. McSweeney's Lists - http://mcsweeneys.net/links/lists/
7. Netvibes.com Blog - http://blog.netvibes.com/

The categories reflect the tab divisions in my Netvibes RSS Reader. My RSS feed list is constantly changing as new blogs are added and old ones are removed in order to

protect myself from an exponentially growing list. My criteria for removing blogs from my daily subscription list are very diverse:

- The blog has not been updated for over weeks or months.
- The blog was updated too often, several times a day.
- The focus of the blog turned out to be very commercial.
- The blog was not giving me new content. It just reproduced content I had already read elsewhere.
- The focus of the blog changed into a direction that no longer interested me.
- The blog was becoming very repetitive.
- The blog was just quoting or linkdumping and not adding any personal notes or critique.

This list reflects a total overview of the blogs I have been subscribed to from September 2006 to January 2008.

APPENDIX 2: ILLUSTRATIONS

Illustration 13: http://www.geocities.com/SunsetStrip/3500/et.html

Illustration 14: My 1998 Homepage

GEOFICTIE   250 ML ID   GOSSIP   △ 29   PHOTOGRAPHY   ANIMATIES
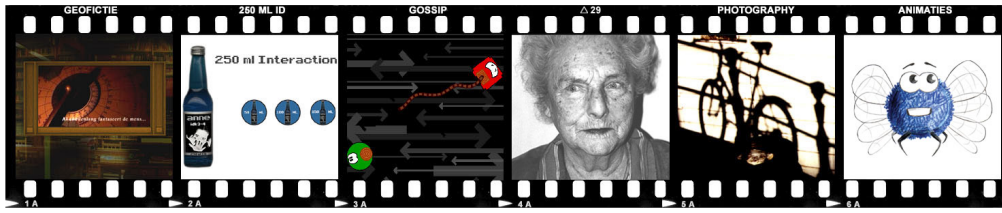
(c) 2001 anne.helmond@student-kmt.hku.nl

Illustration 15: My 2001 HKU Homepage (See DVD for detailled illustrations)