Christoph Meinel · Justus Broß
Philipp Berger · Patrick Hennig

Blogosphere and its Exploration

Springer

Blogosphere and its Exploration

Christoph Meinel • Justus Broß • Philipp Berger •
Patrick Hennig

Blogosphere and its Exploration

Springer

Christoph Meinel
Hasso Plattner Institute for Software Systems Engineering
Potsdam
Germany

Justus Broß
Hasso-Plattner Insitute
Potsdam
Brandenburg
Germany

Philipp Berger
Hasso-Plattner-Institute
Potsdam
Brandenburg
Germany

Patrick Hennig
Hasso-Plattner-Institute
Potsdam
Brandenburg
Germany

ISBN 978-3-662-44408-5 ISBN 978-3-662-44409-2 (eBook)
DOI 10.1007/978-3-662-44409-2

Library of Congress Control Number: 2015936135

Springer Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer-Verlag GmbH Berlin Heidelberg is part of Springer Science+Business Media (www.springer.com)

Preface

The Web does not just connect machines, it connects people.

Tim Berners-Lee [BL09], known as
the Inventor of the World Wide Web

With an immensecirculation of more than 200 million Weblogs worldwide, there are many good reasons for a closer consideration of the (Web)blog phenomenon. In fact, blogs provide one of the most important human-centric applications in the World Wide Web. Blogs can be accessed and read by anybody with access to the Internet without any prior registration process and almost everybody has the opportunity to create and write his or her own blog and share experiences and opinions with the rest of the world. In this way, a completely new channel has been created to support freedom of expression and foster its potential around the world. The result is a wide array of meaningful discourse on all areas of life within blogs and between blogs manifested in so-called blog posts. Given the high level of participation and volume of information, it is of great interest and benefit to identify what is going on in this vibrant world of blogs we call the blogosphere. Increasingly, it is crucial to consider the blogosphere in the context of other Web services that have emerged in the last years, such as Facebook, Twitter, Youtube, or Instagram. Alongside these other social media channels, blogs and the blogosphere continue to play an integral role in our daily digital life. While one might claim never to read a blog, when looking deeper and taking note of the origin of information on ordinary websites and other Web services, we frequently find that this information originates from the blogosphere.

This book represents an attempt to fully review the phenomenon of the blogosphere. The intention is to provide a reliable guide to understanding and analyzing the unimaginable number of very diverse blogs, each consisting of innumerable posts, which in their entirety make up the blogosphere, and go on to answer the questions of how to grasp the complexity of the blogosphere and extract useful knowledge from it. In setting out to write this book, our central aim was to increase the reader's awareness and understanding of the blogosphere phenomenon, including its structure and characteristics. This can be achieved through a better understanding of individual blogs and their particular technical characteristics, as well as a deeper knowledge of how a single blog is embedded and interconnected within the entire blogosphere. The shape and form of the blogosphere can be

described using the analogy of different *continents*. In our description the defining features and characteristics of the continents are illustrated by paradigmatic example blogs.

Following on from the structural analysis, we provide details of the available methods and describe the complex challenge of automatically retrieving information from the abundance of data contained in the blogosphere. Finally, we present our blog search platform, called BLOGINTELLIGENCE, and describe all the tools and features we have developed during the last couple of years to explore the blogosphere.

We believe that this book provides a solid reference for researchers as well as for all others interested in understanding the mechanics of the growing blogosphere. Not only does this book offer an insider's perspective on the characteristics of the blogosphere, it also gives a detailed technical understanding of the methods and technologies involved. These include crawling, analyzing, extracting, and visualizing, which are used to first process and then present the enormous amount of data from the blogosphere in a useful and informative way. Overall, we spent several years in researching what the blogosphere is ultimately made of and according to what mechanisms it functions and constantly redefines itself. This book represents the quintessence of all these years of existing research by integrating and extending the PhD dissertation of Justus Bross [Bro12] as well as the master theses of Patrick Hennig [Hen12] and Philipp Berger [Ber12], and thereby building the customized blog search platform BLOGINTELLIGENCE, all supervised by Christoph Meinel.

There are many people we would like to thank for their valuable assistance and support behind the scenes discussing, investigating, and developing. Their input made it possible to mine, analyze, and, last but not least, visualize the immense amount of data and information that can be retrieved from the blogosphere. In particular, we thank all our colleagues at the chair of *Internet Technology and Systems*, part of the vibrant and stimulating research landscape formed by all the chairs and scientists at *Hasso Plattner Institute*. We also express our gratitude to Springer Verlag, in particular to Dorothea Glaunsinger and Hermann Engesser for their trust in the success of our book project. Finally, thanks to our student assistants and colleagues Martin Boissier, Stephan Detje, Maximilian Jenders, Thomas Kellermeier, Matthias Kohnen, Julian Niedermeier, Steffen Pade, Willi Raschkowski, Keven Richley, Patrick Schilff, Adrian Sieber, and Lennard Wolf for their hard work, input, and committed support, as well as to all students who have worked with us in various roles during their bachelor's or master's studies.

Potsdam, Germany
November 2014

Christoph Meinel
Justus Broß
Philipp Berger
Patrick Hennig

Contents

List of Abbreviations

AJAX	Asynchronous JavaScript and XML
API	Application programming interface
captcha	Completely Automated Public Turing test to tell Computers and Humans Apart
CeBIT	Centrum für Büroautomation, Informationstechnologie und Telekommunikation
CEO	Chief executive officer
CERN	European Organization for Nuclear Research
CMS	Content Management Systems
CPU	Central processing unit
CSS	Cascading Style Sheets
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
HANA	High-Performance Analytic Appliance
HDFS	Hadoop Distributed File System
HITS	Hyperlink-Induced Topic Search
HITS	Hypertext-Induced Topic Selection
HPI	Hasso Plattner Institute
HTML	HyperText Markup Language
HTTP	Hypertext Transfer Protocol
ICT	Information and Communication Technology
IR	Information Retrieval
JVM	Java Virtual Machine
LDAP	Lightweight Directory Access Protocol
MDS	Multidimensional Scaling
MIT	Massachusetts Institute of Technology
MOOC	Massive open online course
OCW	OpenCourseWare
ORM	Object Relational Mapper
PDA	Personal Digital Assistant
POV	Point of View
PR	Public Relations

RAM	Random-Access Memory
RIA	Rich Internet Applications
RPC	Remote Procedure Call
RSS	Rich Site Summary
SaaS	Software as a Service
SOC	Sense of Community
Splog	spam Blog
SQL	Structured Query Language
SSO	Single-Sign-On
tf-idf	term frequency–inverse document frequency
TLS	Transport Layer Security
URL	Uniform Resource Locator
WWW	World Wide Web
XHTML	Extensible HyperText Markup Language
XML	Extensible Markup Language

Part I
Understanding the Blogosphere

Chapter 1
Introduction: The Blogosphere

Our main goal is to show new ways and means to extract reliable and valuable knowledge of the blogosphere. Following an abstract view of the blogosphere from two different angles, we dive deeper into the diverse varieties of blogs and introduce some interesting ones. Then, we continue our journey by collecting requirements for retrieving new knowledge and showing the path from content collection to data mining and knowledge visualization. After this we present a tool that actively supports the extraction of knowledge and show two mining functionalities included in the aforementioned tool. At the end, we will discuss our expectations for the future trends of the blogosphere and social media analytics in general.

1.1 Origins of Social Network Research

Social network research has its origin in the first studies of group dynamics in the 1930s by Harvard researchers [Dio00]. Over the years the investigation of information flow and influence propagation has continued and it is now possible to distinguish between two theories, i.e. Gestalt theory [Wer38] and the structural-functional anthropology [Par48], which eventually ended up as a common theoretical framework for the understanding and analysis of social groups in large and small scale communities. During his studies, Jacob Moreno [Mor41] introduced *social configuration* the first graph formalization for social groups, called a *sociogram*, where each node represents an entity and each edge a relation such as friendship, organizational interaction, or other point of contact. It was shown that, although designed for small arbitrary graphs, the mathematical graph theory also applies to social networks, although such networks are on a much larger scale than those problems studied before [Bar69].

C. Meinel et al., *Blogosphere and its Exploration*,
DOI 10.1007/978-3-662-44409-2_1

1.2 From the Web to the Social Web

The first proposal for the World Wide Web (*WWW* or simply *the Web*) was made at CERN[1] by Sir Tim Berners-Lee in 1989, and further refined while working with Robert Cailliau in 1990. The web was originally conceived and subsequently developed to meet the demand for automatic information sharing between a few scientists working in different universities and institutes around the world. The basic idea of the WWW was to merge the technologies of personal computers, computer networking and hypertext into a powerful and easy to use global information system. Since its initial take up in the early 1990s, the WWW has experienced tremendous acceptance and growth. Starting with the first web server and only a few web pages developed and operated by Sir Berners-Lee and his team at CERN in December 1990, the web has since evolved into an agglomeration of more than a billion web hosts on the Internet in 2014 (this count had not even reached 300 million in early 2011),[2] which serve an estimated number of about 33 billion Web pages[3] [dK14]. Figure 1.1 illustrates this ever expanding use of the Internet. Nowadays, the web is a platform used by hundreds of millions of people to publish and share information *online* and to reference related resources on the web through so-called hyperlinks as discussed by Meinel et al. [MS13, MS09]. By creating a vast, global and easy to use network of information, it has revolutionized the way people disseminate and exchange knowledge. The full impact of the web on human society has yet to be understood, but we already know the velocity, convenience and reach with which people can distribute or retrieve information on the Internet is unprecedented in human history.

1.3 Dimensions of Web 2.0

Various scholars [BL00, O'R05, MO07, VH07, HV09] described the transition from the early Internet to Web 2.0 as the climax of a variety of technological and underlying social developments which are summarized in Fig. 1.2.

The enormous improvements in availability, speed, reliability and network bandwidth made during the past 15 years are referred to as the first dimension *net infrastructure*. This also refers to the advances made in programming and software, in particular with respect to extensions in client-side scripting that have brought, for instance, the AJAX technology, as well as developments in server-side programming (see 1 in Fig. 1.2).

[1] European Organization for Nuclear Research, located near Geneva, Switzerland.

[2] Web Server Survey, last retrieved on August 01, 2014, Netcraft2011.

[3] Statistics by WorldWideWebSize.com, as reported in July 2014.

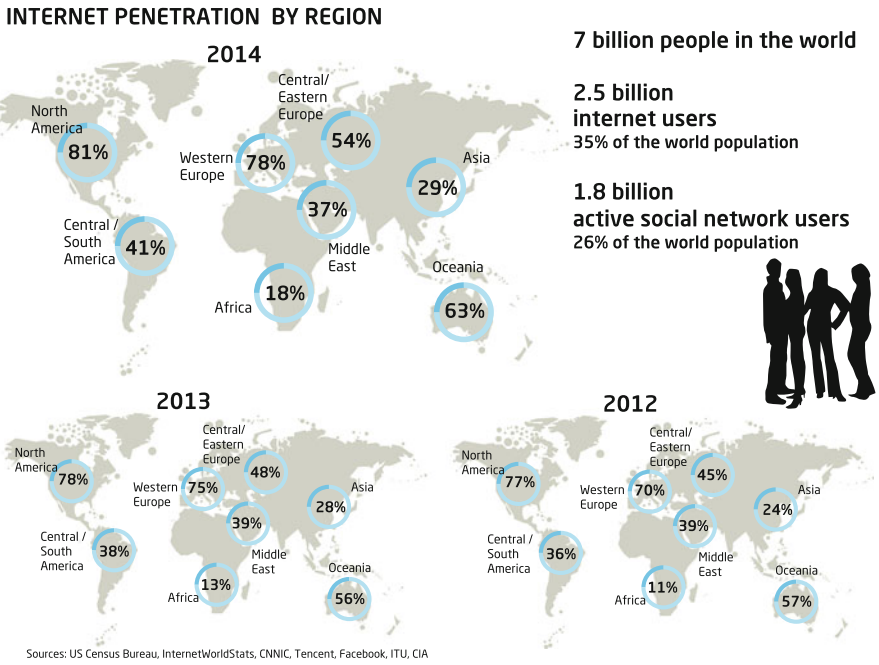

Fig. 1.1 Internet penetration for 2012, 2013, and 2014. For each region the percent of population that has internet access is shown. In 2014, 35 % of the world population are online – that is 2.6 billion internet users. Nearly 72 % of internet users are also active social network users [WeA14]

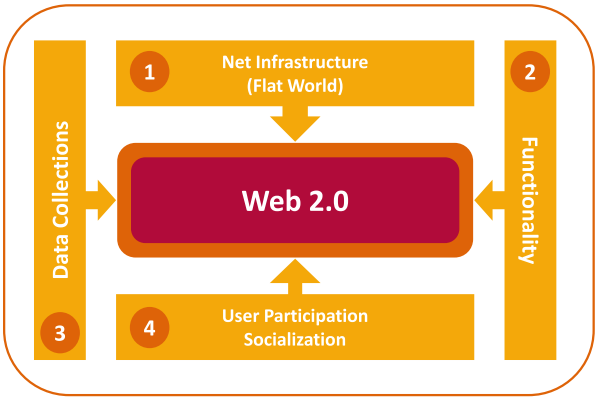

Fig. 1.2 The four main dimensions determining Web 2.0 (Adapted from Hoeren et al. [HV09])

Based on technologies such as AJAX or programming languages like Ruby, the functional dimension (see 2 in Fig. 1.2) has brought along RIAs[4] and a migration of applications from the desktop to the web. By ruling out the need for bug fixing, local installation or updating, software can now be obtained as *Software as a Service* (short SaaS) over the Internet. Nonetheless, these services imply that data resides on the web, with all the associated debates about data privacy.

The third dimension (see Fig. 1.2) refers to the comprehensive creation of data collections by computers as well as by humans that have become a part of daily routine. Computers constantly save weblog data and click paths, maintain search engine indexes and crawl sites on the web. Meanwhile, users register for online services, use tagging for self-organizational purposes and write emails as well as evaluations, comments, or articles in online diaries and blogs. These constantly growing data collections have led to a multitude of (informational) services that are built on top of these collections, including data mining [WF05], recommendations for use and action, the creation of profiles, online communities, personalization of web sites, or context-dependent advertising by search engines such as Google, Bing! and others [VM06, HV09].

Finally, the social dimension of Web 2.0 embraces different formats and services that support communication, interaction and collaboration between users (see 4 in Fig. 1.2). Weblogs, as publications by individuals or small groups, fall within this description, along with photo- or video portals such as YouTube[5] or Flickr,[6] collaborative websites like Wikipedia,[7] the collective tagging of web-resources on Delicioius[8] and Pinterest[9] or services that support the establishment of social networks like XING,[10] Facebook[11] or LinkedIn.[12] While e-mail, homepages and chatrooms have all supported communication and interaction between users, the aforementioned new formats have now introduced new mechanisms of coordination, patterning and networking that connect single usage-episodes and contributions. Consequently, the paradigm of Web 2.0 is, according to O'Reilly, the understanding of the Internet as a platform and not solely as a publication channel nor the basis of solitary applications [O'R05]. Due to the inter-connected nature of information across various websites, services and users, (partial) publics can finally evolve on the basis of Web 2.0 formats. The support of social structures along with the evolvement

[4] Short for Rich Internet Applications: web applications with characteristics comparable to those of desktop applications.

[5] http://www.youtube.com/

[6] http://www.flickr.com/

[7] http://en.wikipedia.org/

[8] http://www.delicious.com/

[9] http://www.pinterest.com/

[10] https://www.xing.com/

[11] http://www.facebook.com/

[12] http://www.linkedin.com/

of (partial) publics is also recognized within the concept of *social software*. It "refers to those online-based applications and services that facilitate information management, identity management and relationship management by providing (partial) publics of hyper-textual and social networks" [BS07, p. 32]. However, these four dimensions cannot be strictly isolated in any Web 2.0 application or scenario, in reality they interact highly, build upon, and complement each other [HV09].

1.4 Web 2.0 and Weblogs

The terminology and visions of the early debates around the *new Internet* could in fact already be discovered in publications of the early popularization phase of the Internet in the nineties. Even back then, the Internet as well as all the services and forms of communication supported by it, was already being recognized as a revolutionary phenomenon that could fundamentally change social communication. Those visions ranged from a possible *revitalization of the public* [Rhe95], a *digital revolution* [Neg95] to an *electronic agora* [Deb99]. Indetifying its emancipatory potential, Levy [Lev97] declared, with reference to the metaphor of the *virtual agora*, that such a development would "enable the public to proclaim many-voiced statements, directly and without detours via any kind of intermediator".

The central arguments in this debate were to a large extent similar to the ones regarding the discussion about *participative journalism*: The vision of a stronger involvement of the recipients in any public communication as well as the overcoming of intermediators in favor of direct and unlimited communication. The utopia of Brecht from the thirties to change the prevalent *distributional system* into a *communicational system* in view of the developing television technology, has at last become true [Bre67, pp. 127 ff.]. The Internet has now reached most areas within entertainment, research, business, science, and beyond, characterized by the Web's transition from a medium where people exclusively consume information to a medium where people both consume but also contribute content in a variety of forms. Ramakrishnan and Tomkins, for instance, noted only 7 years ago that 10 GB of *user-generated or user driven content* was created in the WWW on a daily basis in 2007 [RT07]. If we take a look at statistics for 2014, the numbers are not mentioned as daily measurement anymore. Nowadays we are talking about what is happening within an *Internet minute* and the numbers are even more spectacular [Soc14].

In other words, the Web heavily benefits from user contributions and user-generated content (UGC) [HV09]. This development is closely associated with the frequently predicted collapse of traditional journalism, which is already becoming increasingly obsolete now that everybody has the capability of becoming their own reporter or commentator.

Particularly popular among Web 2.0 formats, and widely rumored to have the potential to provide direct and unlimited communication, are weblogs – commonly known as *blogs*. A blog has a journal-like structure containing several articles,

called *posts*, ordered by their entry date. Each post consists of a title, a publication date, and its main content. The author of a blog, called a *blogger*, has two possible schemes to sort his post collection known as *categories* or *tags*. Categories introduce a hierarchical sorting schema that enables the author to group posts together. In contrast, tags are important keywords attached to a post that highlight aspects of the posts and improve the detectability of a post.

Weblogging systems are specialized, but easy-to-use, Content Management Systems (CMS) with a strong focus on updatable content, social interaction, and interoperability with other web authoring systems. The technical solutions agreed upon among developers of weblogging systems are a fine example of how new, innovative conventions and best practices can be developed on top of existing standards set by the World Wide Web Consortium and the community. Applications like these that offer a simplified mode of participation in today's Internet in contrast to earlier traditional web applications, are now described as *Web 2.0 applications* while the concurrently developing *Participation Internet* has been referred to up to now as *Web 2.0* [O'R06]. The cumulative *social* character of the Internet contrasts sharply with traditional mass media communication channels such as the printing-press, television or radio, since these only offer a unidirectional form of communication. The Internet offers all its users real interaction, communication and discussion. This is also why blogs are referred to as one of the most frequently used *social media tools* [CH07].

Since the end of the 1990s weblogs have evolved to become an essential component of today's cyber culture [HSBW04]. In 2008, the worldwide number of blogs totalled around 130 million [Smi08, Tec09a] increasing to even more than 260 million blogs in 2014. Compared to around 60 million blogs in 2006, this highlights their increasing importance in today's Internet society on a global scale as illustrated in the following by Fig. 1.3.

Meanwhile, the point of origin of weblogs is indefinable since their potential areas of application are numerous. Beginning with personal diaries, reaching over to knowledge and activity management platforms in private or business contexts alike, and finally to enabling content-related and journalistic web offerings [Kir07] the range of content is illustrated by Fig. 1.4. Single weblogs are embedded into a complex superstructure: An independent and segmented public that dynami-

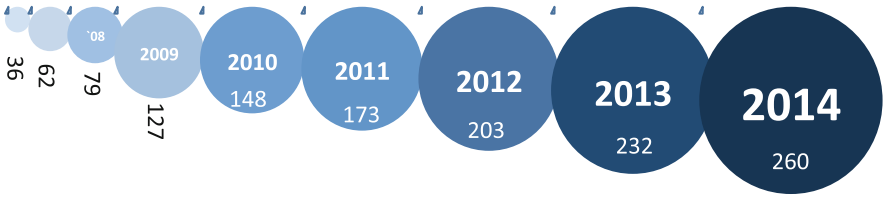

Fig. 1.3 Blog writing – usage trends 2006–2014. Blogging shows no signs of slowing its growing prominence in popular culture and society. The number of blogs increased from 35,77 (2006) million to 260,47 million (2014) [Smi08]

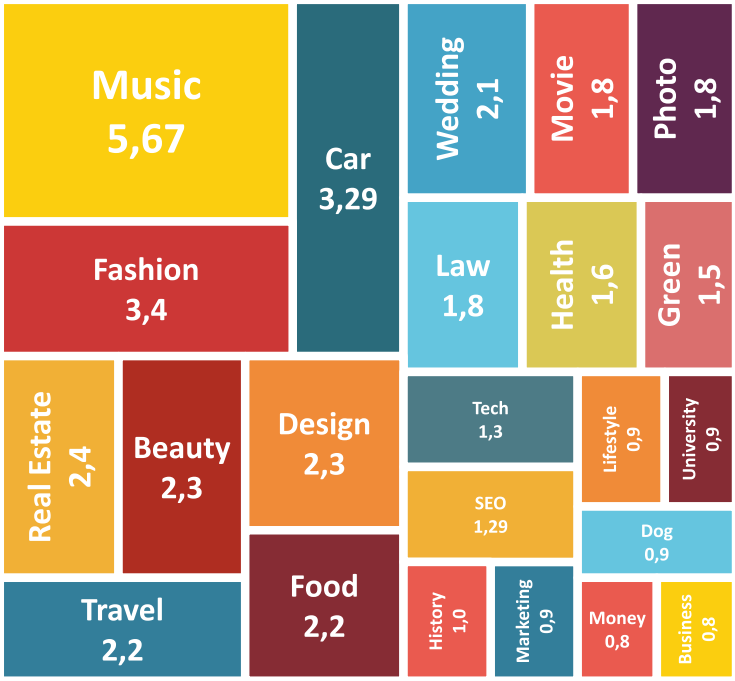

Fig. 1.4 Topics blogged about in 2013. The numbers refer to the number of blogs for each topic in millions. The potential areas of application for weblogs are numerous [Gai13]

cally evolves and functions according to its own rules and with ever-changing protagonists, altogether forming a closely interlinked network also known as the *blogosphere*. Its global interconnectedness, and the corresponding aggregation of individual knowledge, creates a gigantic and constantly changing archive of open source intelligence [Sch06b]. For this reason weblogs are often grouped into the family of *social software* [BQNM10].

1.5 Publishing Revolution or Utopian Fallacy

The debate on photo- and video portals, wikis or social networks and particularly weblogs is currently characterized by a level of excitement that exhibits obvious parallels to discussions about the impact of the Internet in the early phase of its popularization. The association of new formats under the headword *Web 2.0* [Joh05, O'R05] reflects a leap forward to a new version of the Internet, since the latest media formats allow for an entirely new quality of user-participation, integration and networking. Some have even called it a "lurking media revolt" and the "most extensive cultural transformation ever on this planet" that is leading to "fundamentally different democratic structures" [MÖ5].

However, around the same time as the burst of the dot-com bubble and the turn of the millennium, a more measured judgment in communication science gained acceptance. Sullivan et al. [Sul02] refer to weblogs as "a publishing revolution more profound than anything since the printing press", other communication scientists evaluate weblogs more carefully as "a new structural change of the public" [VNR07, BB06]. From the communication scientist's perspective, Altmeppen, for instance, comments that "online journalism can only be identified within the subsidiaries of traditional media in the Internet" [BB06, p. 132], while Schönhagen, for instance, states that all identified modes of social communication also exist *outside* the Internet and expresses the view that computer-based communication will not generate a profound change in societal communication in general. This leads to the prediction that will almost certainly not substitute traditional mass-media [Sch04]. Similarly, the vision referring to a revival of direct forms of public address has not become reality [Mei98, p. 16]: "The prophecy that with the Internet a global agora will emerge, in which the parts of speaker and listener are clearly assigned, has apparently not come true. Instead we experience the digital data network evermore as a broadcast-medium with a backward channel." The predictions of fundamental changes on the basis of new technologies' potential stem from a perception that sees social change as the inevitable result of technological change. Science and Technology Studies (short *STS*) have however, shown that technical issues are always a result of social actions as well as societal forces and influences, particularly with regards to the development and adoption of new technologies. Dennis McQuail correctly puts this in a nutshell: "Technology only proposes, while society disposes" [Nev00]. Throughout the process of technology-adoption numerous variations regarding usage-expectations and actual utilization can often be observed. While many of the new characteristics of a novel technology are initially ignored, accustomed practices gain in importance over the course of the institutionalization phase, which are than carried over to the new technology [MS05]. This is why astounding divergences between initial usage expectation and the later actual utilization can, in retrospect, often be observed. The unpredictability in this regard is particularly present in the domain of the Information- and Communications- Technology (short *ICT*) [LL06]. "The computer as a universal tool machine and the Internet as an open network of interconnected computers are not 'determined' and thus allow for numerous forms of usage scenarios. They are at the time 'recombinant' and therefore enable the modification, innovation and recombination of single modules of technological systems" [Sch06b, p. 41].

1.6 Weblogs vs Facebook, Twitter & Co.

Weblogs are never isolated. They exist in a mixed environment of real-life social interactions, other social networks, news portals and traditional webpages. Bloggers use a variety of sources to substantiate their allegations and support their arguments. Blogs are also increasingly used as a marketing channel for traditional pages and

new services. Besides marketing and referencing, bloggers also use other channels as distribution frameworks for their own posts. Tradtional bloggers now often make the jump to other types of social networks because these networks tend to offer more tailor-made features and even the opportunity to reach a bigger audience.

One example of this is the upcoming social network or social bookmarking service Pinterest[13]: Former fashion and design bloggers have nearly moved to Pinterest because it best fits their needs for visual information sharing and bookmarking [Mil12].

The interaction process on Pinterest fosters the resharing and recombination of content between these users. Nevertheless, the majority still use their blogs to initially publish content and also link it on their other social networks. In this way the blogger still has full access to the content and is able to directly observe the usage statistics of his posts.

The same applies to other social networks such as LinkedIn, Twitter or Facebook, although these networks also require the creation of small content bites to keep the interest of other users, so alongside their normal posts, bloggers create small Facebook or Twitter posts reporting their current activities.

The main advantage of other social networks is the huge user base, the high level of activity of users and the automatic generation of individual news feeds which are generated specifically for each user. This enables users to also see related posts that friends have liked/pinned/retweeted and through this the speed of information flow increases dramatically.

As shown in Fig. 1.5, the user base of Facebook is greater than the total number of blogs. However, it has to be noted that only 10 % of all users of a social community are contributors according to the 90-9-1 rule [BH11]. Although Facebook has over 1.3 billion users, only 130 million of these are active content creators. In the blogosphere, the number of 260 million blogs is actually the number of creators (at least one-time creators). Thus, it can be argued that the blogsphere's community is similar in size to the active Facebook community, although as a decentralized network the blogosphere cannot be measured from a single point of the Internet.

Social networks filter content, for example Facebook controls the content users can see in their news feeds based on their interests and on the internal ranking algorithm factors of Facebook itself. In contrast to social networks, blogs cannot be censored or controlled by a central authority except for search engines or huge blog hosting platforms and the content of a blog is only filtered by the blog author

[13] www.pinterest.com

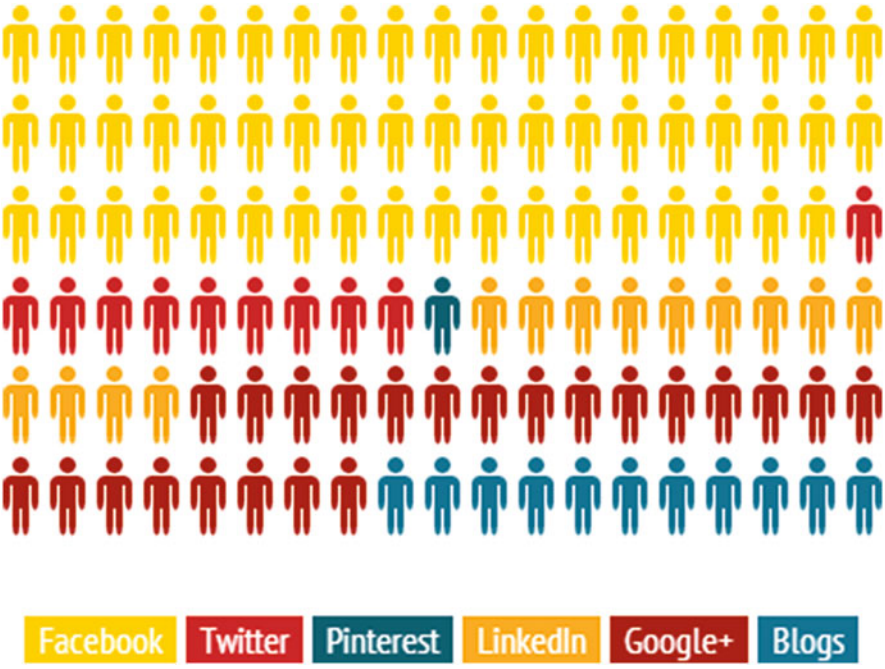

Fig. 1.5 **Users** of Facebook, Twitter, Pinterest, LinkedIn, Google+ in comparison to blogs. Each stickman symbolizes approximately 20 million users [Smi08]

himself. Furthermore, the actual durability of information published on blogs is extremely high. Even decades after the creation of a post it is still detectable via search engines or blog archives.

Chapter 2
Micro-perspective

This chapter provides a comprehensive overview about weblogs as standalone entities. In particular, we discuss characteristic features of blogging software in general, its most important technical built-in features, as well as fundamental hosting issues. Also, this chapter presents a comprehensive review of prior weblog research by constructing an extensive weblog typology. As the title suggests, this chapter treats weblogs as single entities which accounts for the title of this chapter: the Micro-Perspective.

2.1 Weblogs: The Smallest Entities of the Blogosphere

The words weblog and blog derive from the terms *web* and *log* and a blog is no more than a specific type of website or page. Blogs originated as online diaries with entries, also known as *posts*, usually written in reverse chronological order with the most recent entry displayed first. Nowadays, there are countless weblogs around, which use a wide variety of presentation styles and cover a vast range of topics. Single blog posts combine textual parts with images and other multimedia data, and can be directly addressed and referenced via an URL (Uniform Resource Locator) in the World Wide Web. Readers of blog posts can publish their personal opinion about the topic covered in a highly interactive manner by commenting on a post. These comments can however be subject to moderation by the author of a blog.

2.2 Blogging Software and Platforms

While the first blogs around were simple websites that were regularly updated with new posts (or comments), the end of the 1990s saw the emergence of open-source and free-to-use *blog hosting services*. These service providers subsequently offered

C. Meinel et al., *Blogosphere and its Exploration*,
DOI 10.1007/978-3-662-44409-2_2

a user-friendly and ready made blog service that enabled users with any level of computer skills to generate and publish content accessible to all Internet users. From this point on, anybody capable of using a simple text-editor program could thus actively take part in the unrestricted exchange of opinions over the web [ML08]. Nowadays, weblogging systems are more specialized, but still easy-to-use CMS (Content Management Systems) with a strong focus on updatable content, social interaction, and interoperability with other Web authoring systems. The technical solutions agreed upon among developers of weblogging systems are fine examples of how new, innovative conventions and best practices can be developed and superimposed on existing standards set by the World Wide Web Consortium[1] and the community.

Deciding which product to ultimately use is a challenge in itself that requires a careful evaluation prior to installation. It starts with the assessment of which requirements the software should eventually meet: The central issues for these considerations are the system requirements, on which the software of choice will run. The most widely accepted blogging software systems available either require a web server with Perl, PHP (or Ruby on Rails respectively) or Java with a Servlet-Container. For data storage, a relational database or data system is deployed. Some blog software incorporates functionality that requires additional software, such as GD or Image-Magick libraries for graphics processing and other products even allow functionality-extensions through the implementation of plug-ins. Following on from basic system requirements, the next consideration is blogging functionality which is to a large extent dependent on the designated field of application of the blog implementation. For example, does the user intend to install more than just one blog entity? In a corporate context it is particularly important to consider the assignment of permissions prior to any installation. For instance, does the chosen software differentiate between an administrator, editorial staff, authors or regular users? What about the commenting function in the blog? Can it be administered or controlled that comments can be written anonymously or only after registration in the blog? In addition to a built-in search functionality, a WY-SIWYO-editor, as well as the option to backup drafts of articles in the process of writing, is helpful. A personal design is also very important for a lot of blog-enthusiasts. For this reason, the user should carefully assess whether the blog software allows for alterations to the graphical design via the exchange of reversible templates and skins or CSS formatting, for instance. Bloggers may even be looking for a system back-end in their own language that could also allow the administration of multilingual posts. HTML as an output format is not noteworthy, but this should hold equally true for XHTML or standards-compliant HTML. Syndication formats that allow for the automatic distribution of content, such as ATOM or RSS should be supported by default. Email-Notification, for example whenever a new comment on your post is waiting in the moderation queue, is often unavailable. You might also be interested in advanced editing functionality of media formats such as pictures, movies or music that goes beyond

[1] https://www.w3.org