

# 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 immense circulation 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 Kellmermeier, 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

Blogosphere and its Exploration

Meinel, C.; Broß, J.; Berger, P.; Hennig, P.

2015, XIV, 268 p. 81 illus., 75 illus. in color., Hardcover

ISBN: 978-3-662-44408-5