

Semantic Web and Beyond
Computing for Human Experience

Harry Halpin

Social Semantics

The Search for Meaning on the Web

 Springer

Social Semantics

SEMANTIC WEB AND BEYOND

Computing for Human Experience

Series Editors:

Ramesh Jain
University of California, Irvine
<http://ngs.ics.uci.edu/>

Amit P. Sheth
Wright State University
<http://knoesis.wright.edu/amit/>

As computing becomes ubiquitous and pervasive, computing is increasingly becoming an extension of human, modifying or enhancing human experience. Today's car reacts to human perception of danger with a series of computers participating in how to handle the vehicle for human command and environmental conditions. Proliferating sensors help with observations, decision making as well as sensory modifications. The emergent semantic web will lead to machine understanding of data and help exploit heterogeneous, multi-source digital media. Emerging applications in situation monitoring and entertainment applications are resulting in development of experiential environments.

SEMANTIC WEB AND BEYOND Computing for Human Experience addresses the following goals:

- brings together forward looking research and technology that will shape our world more intimately than ever before as computing becomes an extension of human experience;
- covers all aspects of computing that is very closely tied to human perception, understanding and experience;
- brings together computing that deal with semantics, perception and experience;
- serves as the platform for exchange of both practical technologies and far reaching research.

Harry Halpin

Social Semantics

The Search for Meaning on the Web

Foreword by Henry S. Thompson

 Springer

Harry Halpin
Computer Science and W3C/CSAIL
Laboratory
Massachusetts Institute of Technology
Cambridge, MA, USA

ISSN 1559-7474

ISBN 978-1-4614-1884-9

ISBN 978-1-4614-1885-6 (eBook)

DOI 10.1007/978-1-4614-1885-6

Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2012937296

© Springer Science+Business Media New York 2013

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

*I dedicate this thesis to my father and mother,
Harry Halpin Sr. and Rebecca Halpin. One
must always remember that our parents knew
us before we even knew ourselves.*

Foreword

The World Wide Web demands our attention, not only in terms of its role as a major contributor to the increasing centrality of technology in society, but also in purely scientific terms.

To say that the Web is ubiquitous, at least in the so-called developed world, is commonplace to the point of vacuity. But ubiquity alone doesn't require scientific exploration. After all, paved roads are ubiquitous (and socially important) too, but they don't engender a lot of scientific interest. What makes the scientific study of the Web attractive is the unique nature of the thing that binds the Web together, that actually makes it a Web: the Uniform Resource Identifier, or URI. What makes this study urgently *necessary* is the huge economic and social pressure behind innovation on the Web, and the resulting stresses on its foundational technologies.

If we don't have clear and explanatory theories of these technologies, we are at real risk of breaking the Web by accident. Different sectors of Web usage are pushing the technology in different directions, and without care there is no guarantee that the result will remain coherent, not only intellectually, but literally. Interoperability is the *sine qua non* for a distributed architecture – if we lose that, we lose everything.

All of this makes the work reported here, and the unique combination of philosophical and experimental perspectives that it adopts, of very special value. As things stand today, theory and practice differ about the nature of URIs. By bringing multiple novel sources of insight to bear on the problem, this work offers real hope of progress towards bridging that gap, and giving us a sound basis for the future growth of the Web.

Edinburgh, UK

Henry S. Thompson

Preface

There may seem to be no more abstract and theoretical pursuit than the study of *meaning* itself. There's even no a priori reason why individual 'minds' should be capable of understanding how meaning somehow exists in a world that is - at least according to the more mature science of physics - ultimately atomic in nature. Yet far from feeling alone in a world devoid of meaning, I take it for granted that we strive in a rich social world that is overflowing with undiscovered facets. The Web is the universal method of intertwining and sharing these facets with each other. Representations are the texture of our life on the Web. Even if the task is impossible, the calling is worthwhile. Across the Mediterranean, I find the courage of Egypt contagious as I follow their digital photos and text in my Web browser.

A friend once said that the world is not composed of atoms, it is composed of stories. So this book can be considered the trace of my particular story. The story I am telling in this book is not exactly the story I had hoped to tell as a graduate student. This book is to a large extent a reworked and highly edited version of my thesis, and as such suffers from the problems that any thesis has, namely that the studies it comprises were done as small shots in the dark in order to reveal some aspect of a much more sophisticated question. There is much to be done, like formalizing a computational theory of sense and evaluating such a theory. At the time of writing these studies I did not have access to the Web-scale data-sets or processing power needed to formulate a testable theory of social semantics, and as I sit here in Yahoo! Research, I cannot but be amazed by the fact that I have an entire copy of the Web accessible from my desktop. Looking back, the idea of social semantics came upon me in a visit to Amsterdam when I first arrived in Europe: meaning is not something we possess alone, but something we create together. To this day, I still cannot think of a better way to phrase the hypothesis of social semantics.

Barcelona, Spain

Harry Halpin

Acknowledgements

The majority of this book was written as my thesis *Sense and Reference on the Web* at the University of Edinburgh, so this book would not have been possible without the support of my community of friends and colleagues across the globe and in Edinburgh. In particular, this thesis would not have been possible had it not been for the unwavering support of my advisor, Henry S. Thompson, who encouraged me to pursue considering Web architecture a first-rate citizen of inquiry, a brave act few advisors would have been willing to do. I would also like to thank Andy Clark for philosophical inspiration and Victor Lavrenko for his invaluable help on the empirical evaluation. Conversations and support from other colleagues at Edinburgh have been important. However, even more support has come from the global community of Web hackers and Semantic Web researchers. I have been particularly privileged to have had numerous discussions with Pat Hayes and Tim Berners-Lee on these subjects, and I hope I have accurately given an exegesis of their debate. My time at Duke, where I was fortunate enough to study under Fredric Jameson and Michael Hardt, has had a decisive if subterranean influence on this book. Various friends and co-authors deserve my gratitude. In particular, I would like to single out Rob Didham, Kavita Thomas, Dan Connolly, Brandon Jourdan, Jochen Leidner, Maciej Zurawski, Priya Reddy, Malamo Korbetis, Claire Grover, Richard Tobin, Peter Buneman, Phil Wadler, Valentin Robu, Jonathan Oppenheimer, Michael Wheeler, Laura Gomez, Dirk Bollen, Hana Shepard, Dan Brickley, Orit Halpern, Paolo Bouquet, Nicholas and Rita Tishuk, Ras Al-Majnuun, Alexandre Monnin, Arturo Escobar, and everyone in Bilston, the Forest Cafe, and Carrboro. Particular acknowledgments must be given to Brian and Mooness Davarian, both of whom masterfully proofread the entire book. Others shall not be named to protect the innocent. Lastly, I have found intellectually invaluable my time at the Santa Fe Institute, the Oxford Internet Institute, the Island seminar with Brian Cantwell Smith, and the Interface Seminar at Duke University – and more recently, my time with Peter Mika at Yahoo! Research Barcelona. Special gratitude must go to the late Karen Spärck Jones, who called me out of the blue and encouraged this strangely philosophical approach to information retrieval and semantics when I was first beginning.

Contents

1	Introduction	1
1.1	Scope	3
1.2	Summary	4
1.3	Notational Conventions	7
2	Architecture of the World Wide Web	9
2.1	The History of the Web	10
2.1.1	The Man-Machine Symbiosis Project	10
2.1.2	The Internet	12
2.1.3	The Modern World Wide Web	14
2.2	The Terminology of the Web	16
2.2.1	Protocols	18
2.2.2	Information Encoding and Content	21
2.2.3	Uniform Resource Identifiers	26
2.2.4	Resources	29
2.2.5	Digitality	35
2.2.6	Representations	37
2.3	The Principles of Web Architecture	42
2.3.1	Principle of Universality	42
2.3.2	Principle of Linking	45
2.3.3	Principle of Self-description	47
2.3.4	The Open World Principle	48
2.3.5	Principle of Least Power	49
2.4	Conclusions	50
3	The Semantic Web	51
3.1	A Brief History of Knowledge Representation	52
3.2	The Resource Description Framework (RDF)	56
3.2.1	RDF and the Principle of Universality	57
3.2.2	RDF and the Principle of Linking	58
3.2.3	RDF and the Principle of Self-description	60

3.2.4	RDF and the Open World Principle	61
3.2.5	RDF and the Principle of Least Power	63
3.3	Information and Non-information Resources	65
3.4	An Ontology of Web Architecture	68
3.4.1	Resources and URIs	69
3.4.2	Information Resources	71
3.4.3	Web Resources and Web Representations	73
3.4.4	Media Types	75
3.4.5	Hypertext Web Transactions	77
3.4.6	Modeling the Semantic Web and Linked Data	79
3.5	The Semantic Web: Good Old Fashioned AI Redux?	81
4	Theories of Semantics on the Web	85
4.1	The Identity Crisis	85
4.2	Sense and Reference	89
4.3	The Logicist Position and the Descriptivist Theory of Reference ...	91
4.3.1	Logical Atomism	94
4.3.2	Tarski's Formal Semantics	96
4.3.3	Logical Descriptions Unbound on the Web	98
4.4	The Direct Reference Position and the Causal Theory of Reference	101
4.4.1	Kripke's Causal Theory of Proper Names	101
4.4.2	Putnam's Theory of Natural Kinds	102
4.4.3	Direct Reference on the Web	103
4.5	Sense and Reference on the Web	105
5	The Semantics of Tagging	107
5.1	Making Sense of Tagging	107
5.1.1	Related Work	110
5.1.2	The Tripartite Structure of Tagging	111
5.2	Detecting Power Laws in Tags	112
5.2.1	Power Law Distributions: Definition	112
5.2.2	Empirical Results for Power Law Regression for Individual Sites	113
5.2.3	Empirical Results for Power-Law Regression Using Relative Frequencies	115
5.2.4	The Dynamics of Tag Distributions	116
5.3	The Effect of Suggestions on Tagging	120
5.3.1	Models of Collaborative Tag Behavior	121
5.3.2	Experimental Design	124
5.3.3	Results	125
5.3.4	Influence of Tag Suggestions on the Tag Distribution	128
5.4	Constructing Tag Correlation Graphs	132
5.4.1	Methodology	133
5.4.2	Constructing the Tag Correlation (Folksonomy) Graphs	134

- 5.5 Identifying Tag Vocabularies in Folksonomies Using Community Detection Algorithms 136
 - 5.5.1 Using Community Detection Algorithms to Partition Tag Graphs 137
 - 5.5.2 Edge Filtering Step 138
 - 5.5.3 Normalized vs. Non-normalized Edge Weights 138
 - 5.5.4 The Graph Partitioning Algorithm 139
 - 5.5.5 Experimental Results 140
- 5.6 Comparing Tags to Search Keywords 143
 - 5.6.1 Data Set and Methodology Employed 144
 - 5.6.2 Discussion of the Results from the Query Log Data and Comparison 145
- 5.7 Conclusions 147
- 6 The Semantics of Search 149**
 - 6.1 Introduction 149
 - 6.2 Is There Anything Worth Finding on the Semantic Web? 152
 - 6.2.1 Inspecting the Semantic Web 152
 - 6.2.2 Selecting Queries for Evaluation 155
 - 6.2.3 Relevance Judgments 156
 - 6.3 Information Retrieval for Web Search 161
 - 6.3.1 Vector Space Models 162
 - 6.3.2 Language Models 164
 - 6.4 System Description 166
 - 6.5 Feedback Evaluation 167
 - 6.5.1 Hypertext to Semantic Web Feedback 168
 - 6.5.2 Semantic Web to Hypertext Feedback 171
 - 6.6 Pseudo-Feedback 174
 - 6.7 Inference 175
 - 6.8 Deployed Systems 177
 - 6.8.1 Results 177
 - 6.8.2 Discussion 179
 - 6.9 Future Work on Relevance Feedback 181
 - 6.9.1 Experimental Conclusions 182
 - 6.9.2 The Equivalence of Sense and Content 183
 - 6.10 The Representational Nexus 185
- 7 Social Semantics 187**
 - 7.1 Introducing Social Semantics 187
 - 7.1.1 The Hidden History of Search Engines 189
 - 7.1.2 Against Private Language 191
 - 7.2 Representations as Dynamic Ontology 193
 - 7.3 A Statistical Semantic Web 197
 - 7.4 Towards Collective Intelligence 200

References 205

Index 217

Chapter 1

Introduction

You have abandoned the old domain, the old concepts. Here you are in a new domain, for which new concepts will give you the knowledge. The sign that a real change in locus and problematic has occurred, and that a new adventure is beginning, the adventure of science in development.

Louis Althusser (1963)

This book is an inquiry into representation. Given the almost impossibly wide scope of possible kinds of questions pertaining to representations, we will deploy an analysis that is simultaneously both historical and scientific by restricting our inquiry to an investigation of representations on the World Wide Web. Yet regardless of our careful scoping, we will nonetheless be blindly driven into the realm of *semantics*, the hard question of how meaning is assigned to representation – a question that is as hard, it seems, as that of the more popular hard problem of consciousness (Chalmers 1995). The nature of representation is no longer fashionable to even pursue in philosophy or even in artificial intelligence; it is a problem whose immensity overwhelms us. As a subject matter the apparent phenomenon of *reference*, the suspiciously mysterious – and so perhaps even non-existent! – connection between a representation and that which it represents, verges upon the totality of our social relationship with the world. From Plato’s Theory of Forms to the evolution of representation in artificial life (Halpin 2006), science is littered with theories of the semantics, all of which equally purport to solve this thorny matter in one way or another. One would be forgiven in thinking, given the lack of clear success of any theory so far, that perhaps the question is unscientific or simply intractable in nature, yet that compels us with only a more irresistible attraction.

At first glance, representation and semantics seem strangely old-fashioned, particularly given the current enthusiasm for embodiment in cognitive science, which in its more extreme versions leads to claims of “intelligence without representation” (Brooks 1991). Yet this fetish for embodiment may be strangely disciplinary

and – although radical on the surface – actually ends up being a reactionary fad when viewed within the context of a larger landscape outside academic philosophy and cognitive science. In particular, computer science – with the exception of the peculiarly anthropomorphic line of research of artificial intelligence (AI) – does not seem to care about embodiment. In his *One Hundred Billion Lines of C++*, computer scientist-turned-philosopher Brian Cantwell Smith notes that in artificial intelligence, debates over representation tend to frame the debate as if it were between “classical” logic-based symbolic reasoners and some “connectionist” and “embodied” alternative ranging from neural networks to epigenetic robotics (1997). Smith then goes on to aptly state that the kinds of computational systems discussed in artificial intelligence and philosophy tend to ignore the vast majority of existing systems, for “it is impossible to make an exact estimate, but there are probably something on the order of 10^{11} – one hundred billion lines of C++, in the world. And we are barely started. In sum: symbolic AI systems constitute approximately 0.01% of written software” (1997). What Smith fails to mention is that the same small fraction likely holds true of “non-symbolic AI” computational systems such as robots, artificial life, and old-fashioned connectionist networks (an exception may soon be made for the machine-learning that runs phenomena such as advertising and search on the Web). As raw statistics of deployed systems by themselves hold little intellectual weight, no doubt a philosopher could argue that the vast majority of computational systems may have no impact on our understanding of representation and intelligence. In other words, what the vast majority of the planet is doing with computation and representation – which is increasingly focused on the World Wide Web – is simply intellectually uninteresting. In this book we argue otherwise.

Although one can easily deny that anything resembling digital representations exists ‘inside the brain,’ it is much harder to argue that there are no digital representations on the Web. As one clicks from web-page to web-page, it seems that the Web is nothing but a vast network of digital representations. The thesis of this book is that the wide class of computational systems outside of those traditionally considered by artificial intelligence or philosophy presents what Cantwell Smith calls a “middle distance” where questions of representation (and perhaps even intelligence) come to the forefront in a peculiarly obvious manner and are likely more tractable than they are for humans, given the relative complexity of computers and humans (Smith 1996). At the present moment, with all the totalizing attraction of a black hole, computational systems the world over are becoming part and parcel of the World Wide Web, described by Tim Berners-Lee – the person widely acclaimed to be the ‘inventor’ of the Web – as “a universal information space” (1992). We further argue that not only may the Web reveal general insights about the nature of representation, but its unique historical status as the first *actual* universal information space may prompt an entire re-thinking of semantics. When asked to consider this hypothesis, Michael Wheeler – a philosopher who is well-known for his Heideggerian defense of embodiment – surmises that “the power of the Web as a technological innovation is now beyond doubt” but “what is less well appreciated is the potential power of the Web to have a conceptual impact on cognitive science” and so the Web may provide a new “fourth way” in addition to the

“three kinds of cognitive science or artificial intelligence: classical, connectionist, and (something like) embodied-embedded” (2008). While countless papers have been produced on the technical aspects of the Web, very little has been done explicitly on the Web *qua* Web as a subject matter of interest to philosophy. This does not mean there has not been interest, although the interest has come in particular more from the side of those engineers working on developing the Web rather than those already entrenched in philosophy, linguistics, and artificial intelligence (Halpin et al. 2006; Bouquet et al. 2007 2008). In this spirit, what we will undertake in this thesis as a whole is to apply many well-known philosophical theories of reference and representation to the phenomenon of the Web, and see which theory survives – and finally, if the Web points a way to a *new* theory of semantics, which we surmise may be a social semantics.

1.1 Scope

The World Wide Web is without a doubt one of the most significant computational phenomena to date. Yet there are some questions that cannot be answered without a *theoretical* understanding of the Web. Although the Web is impressive as a practical success story, there has been little in the way of developing a theoretical framework to understand what – if anything – is different about the Web from the standpoint of long-standing questions of representation and semantics in philosophy. While this situation may have been tolerable so far, serving as no real barrier to the further growth of the Web, with the development of the Semantic Web, a next generation of the Web “in which information is given well-defined meaning, better enabling computers and people to work in cooperation,” these philosophical questions come to the forefront, and only a practical solution to them can help the Semantic Web repeat the success of the hypertext Web (Berners-Lee et al. 2001). At this moment, there is little doubt that the Semantic Web faces gloomy prospects – and perhaps for good reason. On first inspection, the Semantic Web appears to be a close cousin to another intellectual project, known politely as ‘classical artificial intelligence’ (also known as ‘Good-Old Fashioned AI’) an ambitious project whose progress has been relatively glacial and whose assumptions have been found to be cognitively questionable (Clark 1997). The initial bet of the Semantic Web was that somehow the *Web* part of the Semantic Web would somehow overcome whatever problems the Semantic Web inherited from classical artificial intelligence, in particular, its reliance on logic and inference as the basis of meaning (Halpin 2004).

This thesis is explicitly limited in scope, concentrating only on the terminology necessary to phrase a single, if broad, question: How can we determine the meaning of a Uniform Resource Identifier (a URI, such as <http://www.example.org>) on the Web? Although the thesis is interdisciplinary, as it involves elements as diverse as the philosophy of language and machine-learning, these elements are only harnessed insofar as they are necessary to phrase our central thesis and present a possible solution. Due to constraining ourselves to the scope of the Web and

the topic of representation, this thesis is not an attempt to develop a philosophy of computation (Smith 2002), or a philosophy of information (Floridi 2004), or even a comprehensive “philosophy of the Web” (Halpin 2008b). These are much larger projects than can fit within the scope of a single book or even a single individual’s life-long calling. However, in combination with more fully-formed work in philosophy, we hope that at least this book provides a starting point for future work in these areas. So we use notions from philosophy selectively, and then define the terms in lieu of our goal of articulating the principles of Web architecture and the Semantic Web, rather than attempting to articulate or define the terms of a systematic philosophy or with reference to the many arguments over these terms in analytic philosophy. Many of the terms in this thesis could be explored much further, but by virtue of our scoping are not explored, as to constrain the book to a reasonable size. Unlike a philosophical work, in this book counter-arguments and arguments are generally not given for terminological definitions, but instead references are given to the key works that explicate these terms further.

This thesis does not inspect every single possible answer to the question of *What is the meaning of a URI?*, but only three distinct positions. An inspection of every possible theory of meaning and reference is beyond the scope of the thesis, as is an inspection of the tremendous secondary literature that has accrued over the years. Instead, we will focus only on theories of meaning and representation that have been brought up explicitly in the various arguments over this question by the primary architects of the Web and the Semantic Web. Our proposed solution of social semantics rests on a theory of meaning, a neo-Wittgensteinian theory, that is one of the most infamously dense and infuriatingly obscure theories of meaning.

Finally, while the experimental component of this book has done its best to be realistic, it is in no way complete. Pains have been taken to ensure that experiments, unlike much work in the Semantic Web, at least uses real data and users, and are properly evaluated over a range of algorithms and parameters. Our work on tagging systems takes its data from a real system, *del.icio.us*, as well. While various parts of the experiments could no doubt be optimized and scaled up still further, these experiments should be sufficient to motivate our movement towards social semantics, although a full formalization of such a theory and testing of it would require access to the data of a large-scale search engine such as Google, which for the time being is outside of scope. For future work, we would like to pursue the formalization and large-scale testing of social semantics.

1.2 Summary

The thesis of this book must be stated in a twofold fashion: first to analyze the problem, and then to propose a solution. To analyze the problem of representation on the Web, one must ask the question: **What is the meaning of a URI?** First, we will clarify the problem that the Web is a kind of new language that can be defined by its engineering conformance to the principles of Web architecture, but

nonetheless inherits problems regarding sense and reference from the philosophy of natural language. So there is no easy way out of the hard question of representation. Our proposed answer is then that only a theory of representation and semantics that takes into account the socially grounded use of a multiplicity of representations is sufficient to provide the meaning of a representation on the Web, from which the meaning of a peculiar URI can be derived. In essence, we turn the question on its head; instead of saying that a URI can have its meaning only by virtue of what representations can be accessed from it, we instead say that the network of representations and their use provides the meaning of a URI. Thanks must be given to co-authors for letting me expand upon our earlier findings and re-use our earlier words. The term ‘we’ is deployed in order to acknowledge their contributions. Note that all previously published versions of work in this book have been edited, amended, and otherwise expanded.

In order to orient the reader to the Web, we give an extended introduction to its history and its architecture in Chap. 2, while introducing the philosophical terminology in concert with examples from the Web. In Chap. 3 we propose that the Semantic Web, as embodied by the Resource Description Framework (RDF), is a kind of URI-based knowledge representation language for data integration and illustrate it by providing the elements of Web architecture in terms of a formal Semantic Web ontology. The ontology in particular is joint work with Valentina Presutti, whose latest version is published as *The identity of resources on the Web: An ontology for Web architecture* in the journal *Applied Ontology* Halpin and Presutti (2011). These works have in earlier forms been published as *An Ontology of Resources: Solving the Identity Crisis* (Halpin and Presutti 2009) with Valentina Presutti and my early essay *The Semantic Web: The Origins of Artificial Intelligence Redux* (Halpin 2004).

In Chap. 4 we illustrate the crisis of the Semantic Web: There is no answer to the aforementioned question of how to assign meaning to a URI. There are at least two distinct positions to this question on the Semantic Web, each corresponding to a distinct philosophical theory of semantics. The first response is the **logicist position**, which states that *the meaning of a URI is determined by whatever model(s) satisfy the formal semantics of the Semantic Web* (Hayes 2004). This answer is identified with both the formal semantics of the Semantic Web itself and the traditional Russellian theory of names and its descriptivist descendants (Russell 1905). While this answer may be sufficient for automated inference engines, this answer is insufficient for humans, as it often crucially under-determines what kind of things the URI identifies. As the prevailing position in early Semantic Web research, this position has borne little fruit. Another response is the **direct reference position** for the Web, which states that *the meaning of a URI is whatever was intended by the owner*. This answer is identified with the intuitive understanding of many of the original Web architects like Berners-Lee and a special case of Putnam’s ‘natural kind’ theory of meaning. This position is also a near relative to Kripke’s famous response to Russell (Kripke 1972; Putnam 1975). Further positions that have been marginal to the debate on the Web, such as that of semiotics, are not explored.

A much shorter version of this work has been previously published as *Sense and Reference on the Web* in the journal *Minds and Machines* (Halpin 2011).

Then we dive from the heights of theory to the depths of experimental work. In Chap. 5 we begin the exploration of an alternative form of discovering the meaning of a representation, namely that of ‘bottom-up’ collaborative tagging systems, where users simply ‘tag’ a resource with a term they find useful or descriptive and so define the ‘sense’ of a URI as a set of terms. We commit a number of experiments to determine if these tags converge over time and over a diversity of resources. Then in Chap. 6 we extend this exploration to search engines, considering the ‘bag-of-words’ produced by a document to be equivalent to a set of tags, and therefore the sense of the URI. In particular, we explore this using documents from both the Semantic Web and the hypertext Web, and use relevance models to combine them. The study of tagging was previously published as *The Complex Dynamics of Collaborative Tagging* in *ACM Transactions on the Web* co-authored with Valentin Robu and Hana Shepard (Halpin et al. 2007; Robu et al. 2009), while the user study was co-authored with Dirk Bollen as *An Experimental Analysis of Suggestions in Collaborative Tagging* (Bollen and Halpin 2009). A few elements of the study of search engines and relevance feedback was previously published as *Relevance Feedback between Web Search and the Semantic Web* with Victor Lavrenko, who co-wrote some of the text and the equations (Halpin and Lavrenko 2011b) with a longer version published as a journal article entitled *Relevance feedback between hypertext and Semantic Web search: Frameworks and Evaluation* in the *Journal of Web Semantics* (Halpin and Lavrenko 2011a).

We finally turn to formulate a third position in Chap. 7, *social semantics*, which states that since the Web is a form of language, and as language exists as a public mechanism among multiple agents, then the meaning of a URI is determined by the socially-grounded use of networks of representations on the Web by ordinary users. As vague as this position seems at first glance, we argue this analysis of meaning and representation is the best fit to how natural language works, and it supersedes and even subsumes the two other positions. Furthermore, it goes beyond a certain quietism about natural language attributed to Wittgenstein as well as a certain belief in the occult powers of some ‘mental’ lexicon. Ideas in this version were previously published with Andy Clark and Michael Wheeler as *Towards a Philosophy of the Web: Representation, Enaction, Collective Intelligence* (Halpin et al. 2010). The entire Ph.D. thesis was submitted and approved to University of Edinburgh, with Yorick Wilks being the external examiner, as *Sense and Reference on the Web* (Halpin 2009b), with the precis being published with Henry S. Thompson as *Social Meaning on the Web: From Wittgenstein to Search Engines* in *IEEE Intelligent Systems* (Halpin and Thompson 2009).

As Wittgenstein would say, one must remember that every “language game” comes with a “form of life” (Wittgenstein 1953), and the Web is a new form of life that goes beyond the philosophy of natural language, and leads us straight into a new philosophy of dynamic machinic and human assemblages, a philosophy-to-come of collective intelligence.

1.3 Notational Conventions

In order to aid the reader, this book employs a number of notational conventions. In particular, we only use “double” quotes to quote a particular author or other work. When a new word is introduced and used in an unusual manner to be clarified later, we use ‘single’ quotes. The use of ‘single’ quotes is also used when a word is supposed to be understood as the word *qua* word, a mention of the word, rather than a use of the word. When a term is defined, the word is first labeled using ***bold and italic*** fonts, and either immediately followed or preceded by the definition given in *italics*. Mathematical or formal terms are *italicized*, as is the use of *emphasis* in any sentence. Finally, the names of books and other works are often italicized. In general, technical terms like HyperText Transport Protocol (HTTP) are often abbreviated by their capitalized initials. The World Wide Web is usually referred to as the Web. One of the largest problems is that this whole area historically has had a rather ad-hoc use of terms, so we hope this fairly rigorous notational convention helps separate the use, mention, definition, and direct quotations of terms.