The Epistemology of Intelligent Semantic Web Systems

Synthesis Lectures on the Semantic Web: Theory and Technology

Editor
Ying Ding, ***Indiana University***
Paul Groth, ***Elsevier Labs***

Synthesis Lectures on the Semantic Web: Theory and Application is edited by Ying Ding of Indiana University and Paul Groth of Elsevier Labs. Whether you call it the Semantic Web, Linked Data, or Web 3.0, a new generation of Web technologies is offering major advances in the evolution of the World Wide Web. As the first generation of this technology transitions out of the laboratory, new research is exploring how the growing Web of Data will change our world. While topics such as ontology-building and logics remain vital, new areas such as the use of semantics in Web search, the linking and use of open data on the Web, and future applications that will be supported by these technologies are becoming important research areas in their own right. Whether they be scientists, engineers or practitioners, Web users increasingly need to understand not just the new technologies of the Semantic Web, but to understand the principles by which those technologies work, and the best practices for assembling systems that integrate the different languages, resources, and functionalities that will be important in keeping the Web the rapidly expanding, and constantly changing, information space that has changed our lives.
Topics to be included:

- Semantic Web Principles from linked-data to ontology design
- Key Semantic Web technologies and algorithms
- Semantic Search and language technologies
- The Emerging "Web of Data" and its use in industry, government and university applications
- Trust, Social networking and collaboration technologies for the Semantic Web
- The economics of Semantic Web application adoption and use
- Publishing and Science on the Semantic Web
- Semantic Web in health care and life sciences

The Epistemology of Intelligent Semantic Web Systems
Mathieu d'Aquin and Enrico Motta
2016

Entity Resolution in the Web of Data
Vassilis Christophides, Vasilis Efthymiou, and Kostas Stefanidis
2015

Library Linked Data in the Cloud: OCLC's Experiments with New Models of Resource Description
Carol Jean Godby, Shenghui Wang, and Jeffrey K. Mixter
2015

Semantic Mining of Social Networks
Jie Tang and Juanzi Li
2015

Social Semantic Web Mining
Tope Omitola, Sebastián A. Ríos, and John G. Breslin
2015

Semantic Breakthrough in Drug Discovery
Bin Chen, Huijun Wang, Ying Ding, and David Wild
2014

Semantics in Mobile Sensing
Zhixian Yan and Dipanjan Chakraborty
2014

Provenance: An Introduction to PROV
Luc Moreau and Paul Groth
2013

Resource-Oriented Architecture Patterns for Webs of Data
Brian Sletten
2013

Aaron Swartz's A Programmable Web: An Unfinished Work
Aaron Swartz
2013

Incentive-Centric Semantic Web Application Engineering
Elena Simperl, Roberta Cuel, and Martin Stein
2013

The Epistemology of Intelligent Semantic Web Systems

Mathieu d'Aquin and Enrico Motta

www.morganclaypool.com

ISBN: 9781627051613 paperback
ISBN: 9781627050005 ebook

DOI 10.2200/S00708ED1V01Y201603WBE014

A Publication in the Morgan & Claypool Publishers series
SYNTHESIS LECTURES ON THE SEMANTIC WEB: THEORY AND TECHNOLOGY

Lecture #14
Series Editors: Ying Ding, *Indiana University*
Paul Groth, *Elsevier Labs*
Founding Editor Emeritus: James Hendler, *Rensselaer Polytechnic Institute*
Series ISSN
Print 2160-4711 Electronic 2160-472X

The Epistemology of Intelligent Semantic Web Systems

Mathieu d'Aquin and Enrico Motta
Knowledge Media Institute, The Open University

SYNTHESIS LECTURES ON THE SEMANTIC WEB: THEORY AND TECHNOLOGY #14

MORGAN & CLAYPOOL PUBLISHERS

ABSTRACT

The Semantic Web is a young discipline, even if only in comparison to other areas of computer science. Nonetheless, it already exhibits an interesting history and evolution. This book is a reflection on this evolution, aiming to take a snapshot of where we are at this specific point in time, and also showing what might be the focus of future research.

This book provides both a conceptual and practical view of this evolution, especially targeted at readers who are starting research in this area and as support material for their supervisors. From a conceptual point of view, it highlights and discusses key questions that have animated the research community: what does it mean to be a Semantic Web system and how is it different from other types of systems, such as knowledge systems or web-based information systems? From a more practical point of view, the core of the book introduces a simple conceptual framework which characterizes Intelligent Semantic Web Systems. We describe this framework, the components it includes, and give pointers to some of the approaches and technologies that might be used to implement them. We also look in detail at concrete systems falling under the category of Intelligent Semantic Web Systems, according to the proposed framework, allowing us to compare them, analyze their strengths and weaknesses, and identify the key fundamental challenges still open for researchers to tackle.

KEYWORDS

semantic web, linked data, intelligent systems, knowledge engineering, knowledge-based systems, ontologies

Contents

Preface

In this book, we take both a conceptual and a practical view of the Semantic Web. The Semantic Web is a young discipline, even if compared only to other areas of computer science. Nonetheless, it already exhibits an interesting history and evolution. From a very general vision which, as presented in the foundational article by Berners Lee et al. (2001), integrated different techniques and approaches (agents, databases, knowledge representation, web technologies, etc.), the focus quickly moved to establishing and standardizing a core set of technologies for the representation, distribution, and access to knowledge and data on the Web—e.g., RDF, OWL, SPARQL, etc. Once these core technologies were established, the research community was then able to focus on developing applications and building a large-scale web of data, in accordance with the Linked Data principles.

Having gone in a rather accelerated manner through this cycle, from vision to impact, this is the time to see where this is all going. This is the conceptual aspect of this book: The *Epistemology* of Intelligent Semantic Web Systems. The debate that started as soon as the idea of the Semantic Web was put forward by Tim Berners Lee and colleagues, mostly from the Artificial Intelligence area, has indeed never stopped: What does it mean to be a Semantic Web system? How is it different from other types of systems, such as knowledge systems or web-based information systems? Our goal is not necessarily to provide definitive answers to these questions, but to highlight and discuss in a concise manner the key elements that need to be considered and understood when engaging with this new class of systems.

First, we look at the Semantic Web as part of the Web, and characterize it as a conceptual construct, part of a stack of structures and networks, rather than a stack of technologies (Chapter 1). The idea here is that, by understanding what makes the Semantic Web something of a higher level of abstraction than the Web itself, we have a better view of the implications in terms of the challenges to be addressed, and of the opportunities for new applications and systems to be developed. This view of the Semantic Web naturally leads to thinking of it as a distributed, networked system where knowledge is shared and managed globally. It also leads us to think of it as a type of knowledge system with specific characteristics to do with being distributed, networked, and thus open and without centralized control.

Second, we consider conceptual models established in the field of Knowledge-based Systems and explore how they can be adapted to the open, decentralized nature of the Semantic Web (Chapter 2). In doing so, we establish a simple conceptual framework to characterize Intelligent Semantic Web Systems in terms of the functions and components they include. The goal here is not only to better understand what makes an Intelligent Semantic Web System, but also to establish a reference model useful both to recognize and compare Intelligent Semantic Web

Systems, and also to guide their design. This framework can be seen as defining the core of the book, joining up the conceptual discussion on the nature of the Semantic Web with the more practical view on what it means to develop an Intelligent Semantic Web System. In contrast with other characterizations, here we do not focus on actual technologies but on the various conceptual elements that need to be present when designing Intelligent Semantic Web Systems.

The practical aspect of this book is therefore in this focus on Intelligent Semantic Web *Systems*, and in the deliberately rapid jump from debating the nature of the Semantic Web, to looking at what it means concretely. Hence, once we have established the general, conceptual framework, we look in detail at concrete systems that, according to this framework, fall into the category of Intelligent Semantic Web Systems (Chapter 3). The goal here is that, by looking at concrete examples, we can see how the general model captured by the Intelligent Semantic Web System framework is instantiated in practice. In addition, the model also allows us to compare different systems, thus analyzing their relative strengths and weaknesses and their conformance to our framework.

Having clarified the nature of Intelligent Semantic Web Systems, we are then left with the question: "What next?" Hence, in the final chapter, we look at the key open research challenges that come out of our analysis. To some extent, here we come to two seemingly contradictory conclusions: On the one hand, the area of Intelligent Semantic Web Systems has evolved into a mature field with research directly leading to major opportunities and impacts already clearly visible in many academic and commercial systems. On the other hand, the area is still in its infancy with many fundamental challenges still open for researchers to tackle.

This contradiction is one of the reasons why this book is especially suitable for those (e.g., students) who are starting research in this area and as support material for their supervisors. Much of the material in this book comes from our own experience of more than a decade working as researchers and application developers in the Semantic Web area and from the many discussions on these issues we have had with other members of the research community (see for example d'Aquin et al., 2008a). Much of the thinking here also comes from the experience of establishing in 2003 the first international summer school dedicated to Semantic Web Research (SSSW). SSSW is still today the key educational event in the Semantic Web area and has evolved dramatically since its first edition, reflecting the evolution of the area but also, to some extent, contributing to shaping it. This book is a reflection on this evolution, with the aim to take a snapshot of where we are at this specific point in time, and also to show what the future will be like, or at least should be like!

Mathieu d'Aquin and Enrico Motta
April 2016

Acknowledgments

The content of this book is based on research started by our group more than a decade ago. It derives, directly or indirectly, from the work of the past and present members of this group, namely: Alessandro Adamou, Sofia Angeletou, Elizabeth Cano Basave, Carlo Allocca, Claudio Baldassarre, Emanuele Bastianelli, Enrico Daga, Maurizio di Matteo, Martin Dzbor, Salman Elahi, Miriam Fernandez, Jorge Gracia, Laurian Gridinoc, Davide Guidi, Tom Heath, Yuangui Lei, Ning Li, Shuangyan Liu, Vanessa Lopez, Andriy Nikolov, Francesco Osborne, Michele Pasin, Silvio Peroni, Dnyanesh Rajpathak, Marta Sabou, Angelo Antonio Salatino, Lucia Specia, Keerthi Thomas, Ilaria Tiddi, Victoria Uren, Maria Vargas-Vera, Paul Warren, Fouad Zablith and many others who have been, in one way or another, associated with us. We are immensely grateful to have had the opportunity to work with such brilliant researchers and to debate with them the state, direction, and impact of the semantic web, linked data, and Intelligent Semantic Web Systems. We are also grateful to all our colleagues, who came for a visit or we met at conferences, seminars, and other events, and with whom we discussed and wrote about the challenges and opportunities of the semantic web. Finally, we thank our families and friends for being there and mostly bearing with us.

Mathieu d'Aquin and Enrico Motta
April 2016

CHAPTER 1

Characterizing the Semantic Web

The Semantic Web, and by extension semantic web technologies, is very young in comparison to other computing disciplines, such as databases and artificial intelligence—and indeed even the Web is very young in comparison with these disciplines.[1] As a result, as is usually the case with new phenomena, it will probably take time to develop a comprehensive understanding of the Semantic Web and distinguish its fundamental aspects from the purely coincidental ones. Hence, it is no surprise that defining the Semantic Web has been, in the past dozen years of active research in the area, both a difficult thing to do, and an evolving exercise. For example, many initial solutions in this area put a lot of emphasis on the "semantics" aspect, giving an interpretation to the word that relates it to the long standing area of artificial intelligence and (logic-based) knowledge representations [Genesereth and Nilsson, 1987]. Taking this view, the Semantic Web has been characterized as a web in which information is interpreted and reasoned upon by software agents acting on our behalf [Berners-Lee et al., 2001]. The key here is the association of web documents with formal semantics, expressed by means of logical models encoded in web *ontologies* [Staab and Studer, 2009]. The ability to achieve logical inferences on such models was seen as the true hallmark of a *Semantic* web, with the *web* aspect of considering the distribution of knowledge and information in a global, collaborative network being left in the background.

This logicist stance has of course led to very valuable work, producing standards for expressing formal semantics in ontologies [Horrocks et al., 2003], and also inference systems able to scale to vast amounts of semantically characterized data (e.g., see Sirin et al., 2007). As a result of adopting this perspective, many of the early Semantic Web applications were actually rather similar to earlier knowledge-based systems, with an additional emphasis on web interfaces and knowledge sharing and reuse [d'Aquin et al., 2008a, van Harmelen et al., 2009]. Specifically, this similarity was primarily a result of focusing more on the "semantic" rather than the "web" aspect, and emphasizing therefore the ability to reason with formal representations of knowledge, rather than the "web-like" ability of operating in an open and distributed world [d'Aquin et al., 2008a].

Other characterizations of the Semantic Web have instead focused on the technological elements. When considering the Semantic Web as a platform for the global exchange of machine

[1] The origin of the semantic web is generally associated with the article Berners-Lee et al. [2001], and the start of the Web can also be linked to an article (a proposal) written by Tim Berners-Lee in 1989 (see `http://www.w3.org/History/1989/proposal.html`), while database systems were already an established field of study in the 1960s and the first artificial intelligence conference took place in 1956 (see Russell and Norvig, 2003).

readable information, and given that the Web has only been made possible through the adherence to standard technologies, it is natural to characterize the semantic web also through a stack of standard technologies [Horrocks et al., 2005]. In such a view, a semantic web system is one which uses RDF, URIs, and HTTP to model and share information, and which reuses standard vocabularies (i.e., ontologies) to structure data in such a way that they can be reused by applications, using web technologies as a medium for their distribution. By and large, this is the view at the core of the linked data movement [Bizer et al., 2009], which emphasizes the open publication of data using standard web technologies and the use of the architecture of the Web (URIs and links, see below) for the representation of data.

Our view is that both these perspectives are unsatisfactory (for rather different reasons) and we will therefore try to elaborate what we believe is the essence of this new technology that is the semantic web: how "being a web" makes it fundamentally different from traditional knowledge-based systems and how the technologies employed are merely a reflection, or an instantiation, of more fundamental properties. Of course, this is not purely done as an academic exercise (even if it is an interesting one). By doing so, we hope to clarify what are the fundamental characteristics of intelligent semantic web systems, regardless of the languages and technologies they use, and how they can be understood as part of a new discipline at the boundaries of different fields, including software engineering, artificial intelligence, web development, data processing, human-computer interaction, and others.

1.1 IT IS A STACK AFTER ALL!

To achieve a more fundamental understanding of the semantic web, we need to consider first what is the Web. Naturally, it is tempting to describe the Web by highlighting the underlying technology as the core defining element of the concept: "the Web is whatever uses HTTP over a network." Here however, the more conceptual view might be as easily expressible: "The Web is a graph of documents connected by hyperlinks."

This might appear too simple to do justice to the importance and impact of the Web, but what is really interesting is the implications of such a basic conceptual definition. Documents on the Web have web addresses (URIs) and connecting them is made by simple reference to these web addresses. This means that the network formed by the Web is a purely conceptual network that transcends the physical network and the software infrastructure on top of which it is sitting. To say it plainly, the power of the Web is that it does not matter where, by whom, and using which tool a document was created and published on the Web, for it to be part of the network and connected to the rest of the global space of web documents. This is achieved because the Web abstracts from these considerations and assumes they are dealt with appropriately by means of lower level technologies.

Considering this, it is quite natural to envision the semantic web similarly as a stack, where higher level concepts abstract from the lower level, more concrete realizations. This has been considered in particular through the well known "semantic web technology stack" (see Figure 1.1)

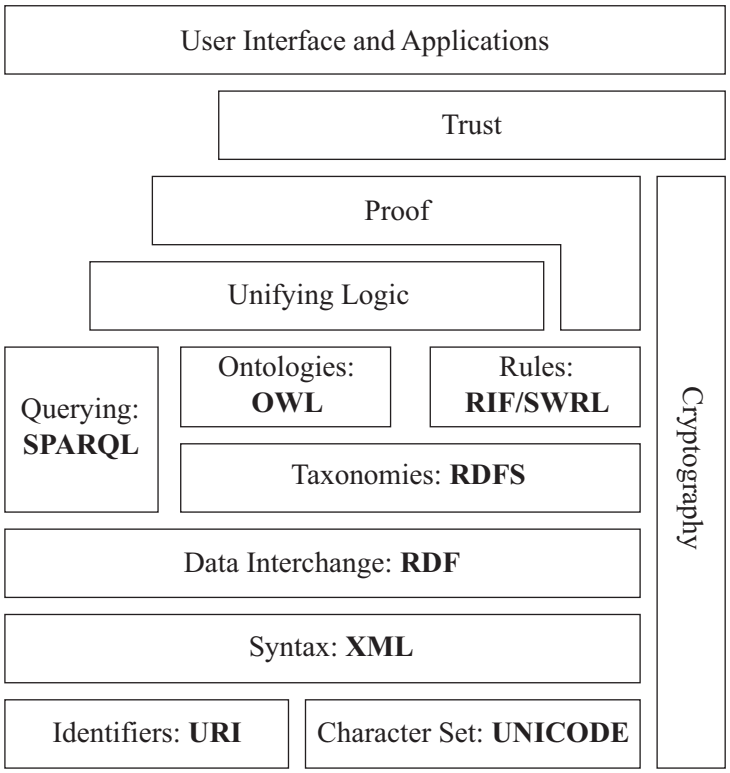

Figure 1.1: One of the instantiations of the Semantic Web Technology Stack (from http://en.wikipedia.org/wiki/Semantic_Web_Stack).

first mentioned in Berners-Lee [2003] and further refined by numerous authors (e.g., Horrocks et al., 2005). There is however something fundamentally unsatisfactory about using this stack to explain and define what is the semantic web, and in particular trying to characterize what kind of intelligent systems can be developed as semantic web applications. The reason is that this is simply a stack of "accidental" technologies, describing how existing technologies (URIs, RDF, etc.) rely on each other, and how future standards might eventually rely on these elementary technologies. The stack does not explain the concepts that define the essence of the semantic web, but (possible) relationships between components, which might be used to implement it. For example, the mention of URIs in this stack is a reflection of the need for the Semantic Web to include a mechanism for expressing global, universal identifiers for data objects and abstract concepts. URIs might be the best (and possibly only) candidate for this, but it is only one instantiation of this concept. Similarly, RDF in this stack is an instantiation of the notion of a distributed, graph-based data model, while RDFS is simply the schema model for such a data model.

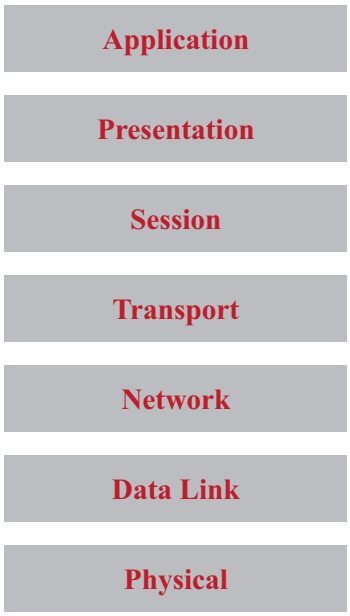

Figure 1.2: The OSI model—networking layers.

Nonetheless the idea of characterizing the Semantic Web as a stack of ever more abstract notions is appealing. Indeed, we can take inspiration from the Open Systems Interconnection (OSI) stack used to categorize networking technologies (Zimmermann [1980]; see Figure 1.2). Here, the idea is that each layer manipulates objects and notions at a higher level of abstraction than the one below and simply assumes that the layer underneath is able to fulfill its role in the process of transferring information. The physical layer is about cables, signals, and hardware, the data link layer about reliably transferring data from one machine to another through a direct connection, the network layer about transferring data from one point of the network to another, the transport layer about delivering packets of information between nodes of a network, the session layer about maintaining connections, the presentation layer (mostly) about the representation of information in a machine-independent way and the application layer about whatever might be done with the ability to transfer information from one (virtual) place to another. Each of these features might be implemented through a number of different technologies, achieving the same purpose in different ways. For example, the physical layer might be achieved through Wifi or Ethernet or, at transport level, one might use TCP or UDP. Conversely, each layer also encapsulates elements of the higher level layers, without needing to process or understand them. Indeed, the physical layer for example only assumes bits of data to be transferred, and transforms these bits into signals. Similarly, the transport layer (e.g., TCP) does not need to interpret the content

of the information to be transferred from the layers above (e.g., emails, web pages, etc.), and only needs to care about implementing a protocol to get any arbitrary packet of information from one node to the other. In the OSI stack, the Web belongs to the top layer, called (somewhat misleadingly in this case) the application layer, which means that the Web abstracts from the physical and implementation details that require information to be located and transferred between different places, machines, and systems.

The Semantic Web has been described as "an extension to the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation" [Berners-Lee et al., 2001]. Similarly, linked data can be seen both as relying on the basic infrastructure of the Web for the purpose of data sharing, and also as a pragmatic reduction of the original Semantic Web vision. While the vision of the Semantic Web was essentially about giving meaning to information on the Web, linked data is essentially about a pragmatic way to represent and deliver such information as a true part of the Web (using URIs and web links as the basis for a global graph data model, see Bizer et al., 2009). Based on this simple line of reasoning (i.e., the Semantic Web as a layer defined over the Web), we can then consider the Semantic Web as part of a stack, which defines a set of ever more abstract notions. Specifically, we start with the basic mechanisms for the encapsulation and transmission of information in documents (the Web), we move then to the materialization of this information as processable and interconnectable data entities (linked data), then to the interpretation of this information through the application of semantics (the semantic web)—see Figure 1.3.

This characterization can be directly related to the one separating the "symbol level" from the "knowledge level" in knowledge representation and knowledge-based systems [Newell, 1982] and indeed it can be seen as the adaptation of the classic analysis from Newell to the new scenario emerging with the semantic web. The symbol level (here web technologies and linked data) is system oriented and mostly concerned with the mechanisms used to represent and manipulate information. The knowledge level (here the semantic web) is concerned with the more abstract understanding of this information, especially the conclusion that an agent can draw from it, independently of the way it is represented. The lower layers (the symbol level) of this new semantic web stack are quite clearly defined and have been analysed already in detail in the literature. Hence, in this book, we essentially focus on the higher levels of the stack. Specifically, we are interested in analyzing the class of intelligent systems that can be designed by taking advantage of the distributed knowledge available through the semantic web, which in turn relies on the distributed, structured information layer provided by linked data.

In the next sections, we analyze these different layers in some detail. However, in contrast with earlier analyses (e.g., Antoniou and van Harmelen, 2008, Heath and Bizer, 2011, Hitzler et al., 2011) that focused on specific technical aspects of the Semantic Web (such as the common information and knowledge representation formats/standards/practices), we strive to keep the discussion independent from a purely technological perspective (even though we will refer to

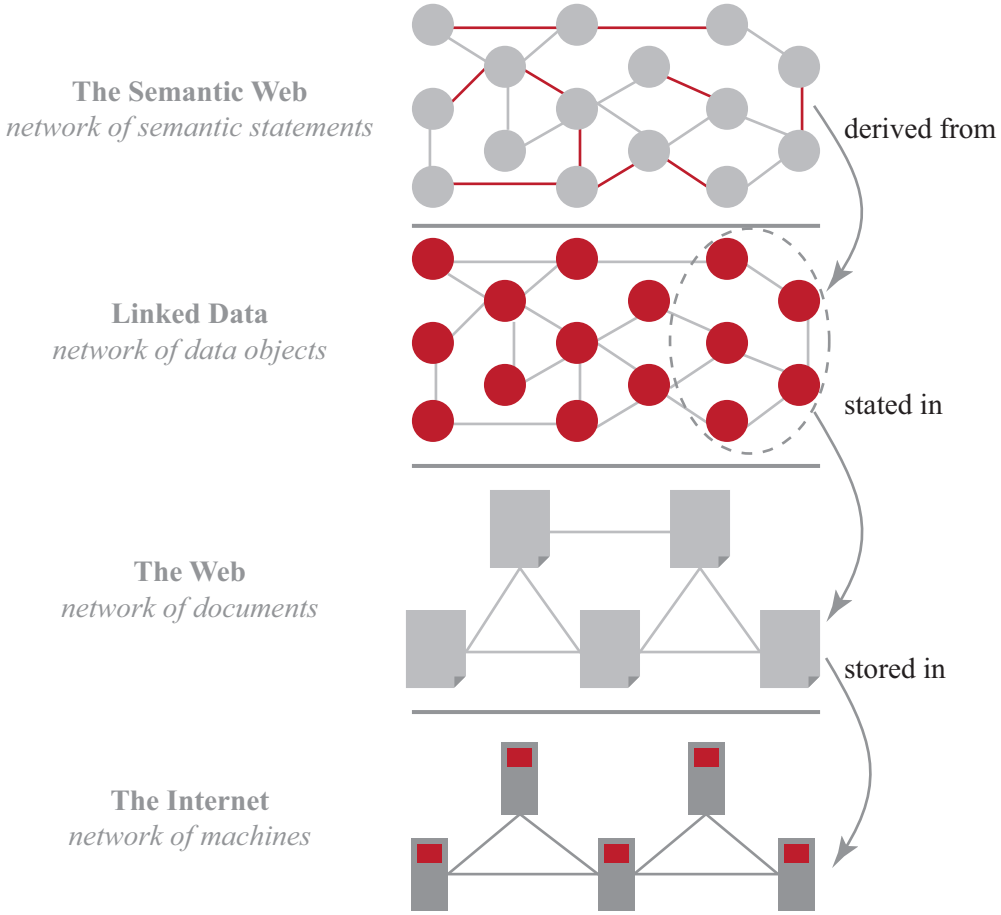

Figure 1.3: Layers of the Semantic Web.

current implementations) and focus instead on the abstract functionalities that the different layers are meant to realize.

1.2 THE WEB AS A LAYER ON TOP OF THE INTERNET

As discussed above, there may be a lot of different aspects that can be argued as being central to the definition of the semantic web. This "faceted" view on a technological platform actually already starts with the Web. Indeed, one might for example consider the Web from the point of view of networking technologies. It has already been mentioned that web technologies appear in the upper, application part of the OSI network layer stack. What this means is that *whatever the Web does* is encapsulated down into actual networking technologies for data transfer, machine-to-machine

communication, etc. The layer just below the Web in this case is the internet (materialized through the IP protocol) which is itself an abstraction from the more concrete and technologically constrained lower layers (the internet literally being a network of networks, which might be very heterogeneous from the point of view of the bottom layers).

From another point of view, the Web can be described as a hypertext system [Conklin, 1987], i.e., a system made of documents that link into each other through hyperlinks. The Web, from this point of view, can be seen as a particular instantiation of a hypertext system, having specific properties (e.g., no centralized control and consistency requirements) and relying on specific technologies and protocols, most notably HTTP [Fielding et al., 1999].

What is interesting however is that the Web can be seen as both things at the same time: the realization of a hypertext system that sits on top of global networking technologies. Hence, while following a link on the Web is concretely translated into (or encapsulated into) sending a request to a specific server, in a specific network on the internet, this mechanism is abstracted into the notion of accessing a document that is conceptually connected to another document in a way that is not affected by the technological realization of this conceptual link. This creates a global network of documents which connect across networks, machine, systems, organizations, and locations.

1.3 LINKED DATA AS A LAYER ON TOP OF THE WEB

The question of the relationship between linked data and the semantic web is a difficult one to answer, as the area has clearly spun off from the semantic web research community and many researchers may consider linked data as the concrete realization of a semantic web [Bizer et al., 2009]. Here we propose however a novel characterization of linked data, as a lower-level layer of the semantic web, which provides the basis for knowledge to be distributed, networked, and shared based on the principles that the linked data community has put forward. We first however discuss what makes linked data principles so interesting and successful, and we characterize them as "using the architecture of the Web for data linking."

Indeed, besides the technological considerations of representation languages, query mechanisms, and storage facilities, the basic idea of linked data is to rely on the same principles that have allowed the web of documents to become such a globally significant system, for the purpose of enabling the sharing of information entities rather than documents. Analogously to the Web, the fundamental element here is the identification of these information objects using web addresses (URIs) and their connection through hyperlinks. Taking a simple example (see Figure 1.4), such a linked data entity can be used to represent a person (Mathieu), which would then be assigned a URI to identify him (e.g., `http://data.open.ac.uk/person/0e5d4257051894026ea74b7ed55557e7`). Another information entity can be a particular publication, such as d'Aquin et al. [2005], which is also associated with a URI (`http://data.open.ac.uk/oro/43798`), or the higher-education institution where Mathieu is working, The Open University (`http://education.data.gov.uk/id/school/133849`). Crucially, as in the case of the Web, these

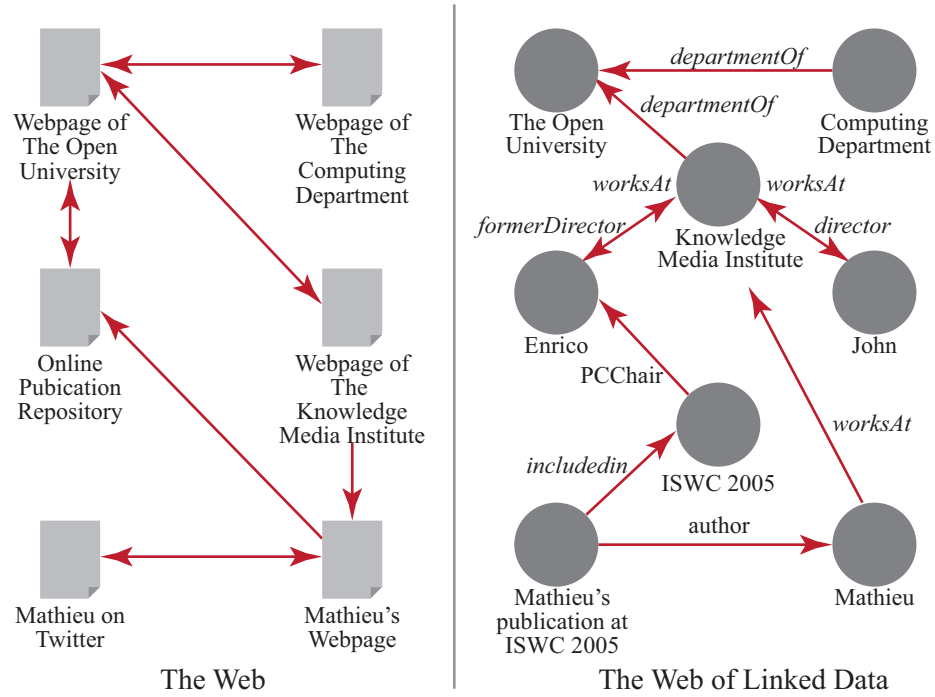

Figure 1.4: Linked data as the application of web principles to information objects.

different identifiers are not produced and maintained centrally but in a distributed fashion, by different data owners. Links can then be built between such information objects, the linked data entities, in ways similar to the hyperlinks of the web of documents. However, a crucial difference here is that such data entities are also labeled by the type of relationship that relates them. Thus, one can declare that Mathieu is one of the *co-authors* of the publication and an *employee* of the Open University. In other words, the basic mechanisms of the Web—universal identifiers (URIs), a simple protocol for obtaining resource representations from these identifiers (HTTP) and representation formats for resources that can embed links to other identifiers—are used both to publish and to model the data in a globally accessible graph.

These basic principles might seem simple, and naturally much debate has taken place with respect to their technological realization. They are however truly the essence of what makes linked data such a powerful and successful paradigm: the essential property of being independent and abstracting from the artificial boundaries and accidental connotations created by technology. In the case of the Web, one does not need any special access to declare a web link between her web documents and those of others. Here, information about Mathieu and his publications sits in a different information system (`data.open.ac.uk`) from what concerns The Open University as

an organization (data.gov.uk). Still, all these information objects exist and can be used within a single common, collective data graph where these differences are abstracted and removed [Heath and Bizer, 2011].

This is the reason why we consider linked data as a layer on top of the Web, in the same way as the Web is a layer on top of the internet, and internet protocols form a layer on top of lower level network protocols. Data modeling and sharing is encapsulated within the basic web mechanisms designed to publish and connect documents (i.e., information containers). Information about different linked entities are "put online" as web documents and through mechanisms that concern the functioning of web servers. However, at the conceptual level of linked data, these considerations become irrelevant: these are just the lower level mechanisms that make it possible to create such a network of interconnected information objects.

Much effort has been spent on extending and adapting the basic mechanisms to access documents on the Web, so that they work consistently with data on the Web of Data. RDF [Lassila and Swick, 1999] is an obvious example, building on web standards such as XML and using URIs and links as the basis for a graph-based data modeling and representation language. An even better illustration of the need for new ways to access this higher level dimension of the Web is provided by the intensive discussions around the mechanisms through which web servers should deliver linked data, how software agents should request such linked data, and how they should interpret the response [Ayers and Völkel, 2008, Heath and Bizer, 2011].

Another significant challenge in the development of linked data, also related to the need for new technological developments, is about supporting access to the data by human users, in particular with respect to providing ways to browse and present the information content. However, many "Linked Data Browsers" suffer from relying on a rather naive approach to reproduce (rather than to abstract from) the mechanisms of the Web [Karger and Schraefel, 2006]. Hence, we believe that there is still a need to better understand how to interact with information embedded in such a large, heterogeneous, and open data graph, beyond the metaphors used for somehow simpler interactions with documents.

We will come back to these issues in later chapters, through providing a more conceptual, "knowledge level" view of the way intelligent semantic web systems interact with information on the semantic web.

1.4 THE SEMANTIC WEB AS A LAYER ON TOP OF LINKED DATA

The stack model we are considering here is a way to answer the question of the relationship between linked data and the semantic web, i.e., linked data provide an intermediary step, or layer, between the web and the semantic web. As discussed earlier, the linked data layer considers elements related to the distribution of information, the integration of information, and its publication for sharing online, as encapsulated within the architecture of the Web. The semantic

web layer instead considers notions of a higher degree: how such information is interpreted and processed within a network of knowledge entities, rather than as raw data.

Such notions and considerations have been the focus of the research community on the Semantic Web since the early days and quickly materialised into specific technologies. In particular, ontologies [Staab and Studer, 2009] were chosen as the main paradigm for the representation of knowledge on the semantic web, with languages such as OWL [McGuinness and van Harmelen, 2004] devised to enable the use of formal knowledge representation formalisms (namely description logics; Baader, 2003) to encode distributed, sharable knowledge on the Web. While the aspects of data representation were not a strong focus in this view of the semantic web, the idea that we are promoting here, that the Semantic Web is a layer above linked data, is still clearly illustrated by the way such technologies are being used nowadays: the ontologies are used as a semantic structure above the data, relating them to formal notions which make it possible, to a certain extent, to manipulate them at a more abstract level of meaning than the one defined by the raw linked data entities. Following up from our example in Figure 1.4, each data object in the linked data graph might be associated (typed, labeled) with abstract concepts from web ontologies (e.g., Mathieu is a Person, The Open University is an Organization/Educational Institution/University). Similarly, the relationships used to connect these data objects can be defined as part of ontologies, themselves shared and published using the same web mechanisms as the more concrete linked data (e.g., the relation "worksAt" is defined, at a certain URI, as a relationship between a person and an organization). The role of ontologies here is therefore seen as both providing a structure for the data (for agents to know what to expect from these data) and a shareable, logical expression of the intended meaning of the concepts which data objects instantiate.

Generalizing from this, we can see that the relationship between the semantic web layer and the linked data layer provides a much steeper abstraction step than the one between linked data and the web. Indeed, it is the same abstraction step which has classically been characterized as moving from information to knowledge, where the latter is viewed as information interpreted and integrated in such a way that it may lead to the production of more knowledge, through inference [Aamodt and Nygard, 1995]. This notion of inference is actually central to the notion of semantics as understood, originally, in knowledge representation and later in the initial views on the Semantic Web that led to the focus on ontologies. In some ways, it is what makes the Semantic Web a layer above linked data, as the central idea of the Semantic Web is to make explicit, through more or less complex inferences, the knowledge which is encapsulated within the data sources integrated and made available through linked data.

Now, even if ontologies (and therefore description logics; Baader, 2003) have been a clear focus of semantic web research in the last decade, this idea that the Semantic Web represents the knowledge level (in the sense of Newell, 1982) of the (symbol-level) linked data is not only materialised through ontological and logical inferences. More generally, we take the view consistent with the motto "A little semantics goes a long way" [Hendler, 2007], that the Semantic Web is as

much materialised by simple mechanisms such as interpretable links between linked data entities, basic taxonomies, lightweight but shared schemas or analytical processes, as it is from complex, formal and logical mechanisms, such as the ones found in description logics. In other words, in the semantic web environment, simple inferences at a very large scale can be more valuable than complex inferences in a closed system [d'Aquin et al., 2008a].

Having established this principle does not however help us to solve our main challenge: how do we build intelligent systems that benefit from the semantic web? Being a level of abstraction above linked data means that the issues regarding the way in which human users and software agents interact with the Semantic Web become even more prominent here. The web paradigm of interacting with information through browsing documents becomes irrelevant in an environment made of a large, global network of inferable information. In the next chapter, we turn to the large body of work done in the area of knowledge-based systems, and especially to the ways in which knowledge-based systems have been abstracted from the base technologies used to implement them, in order to characterize them through more abstract notions, such as knowledge and symbol levels, as well as the TELL and ASK protocol [Lakemeyer and Levesque, 2000]. The straightforward approach here would be to think of the development of intelligent systems on the semantic web as creating knowledge-based systems where the Semantic Web is the knowledge base. As we will see, this view is actually an interesting starting point, as long as we take into account the challenges arising from the open nature of the Semantic Web, in contrast with the relatively closed environment of knowledge-based systems.