Contents

Chapter 1

Semantic Web

Pascal Hitzler

Kno.e.sis Center, Wright State University, Dayton, OH, USA

Krzysztof Janowicz

University of California, Santa Barbara, CA, USA

Introduction

The Semantic Web, as an interdisciplinary research field, arose out of the desire to enhance the World Wide Web in such a way that interoperability and integration of multi-authored, multi-thematic and multi-perspective information and services could be realized seamlessly and on-the-fly [8, 14]. Thus, the Semantic Web is concerned with developing methods for information creation, annotation, retrieval, reuse, and integration. It draws from many disciplines, including knowledge representation and reasoning in artificial intelligence, databases, machine learning, natural language processing, software engineering, information visualization, and knowledge management [40, 41].

Semantic (*Web*) *Technologies* are under substantial investigation in many disciplines where information reuse and integration on the Web promises significant added value, e.g., in the life sciences, in geographic information science, digital humanities research, for data infrastructures and Web services, as well as in Web publishing. At the same time, Semantic Technologies are also being picked up to enhance solutions in application areas which are not primarily targeting the World Wide Web but have to access similar challenges, such as enterprise information integration, intelligence data analysis, and expert systems.

The Semantic Web spans from foundational disciplines to application areas. In terms of size and impact of its scientific community, amount of available project funding, and industrial impact, it has emerged as a major field within Computer Science during the last decade.

1.1 Underlying Principles

As the Semantic Web is driven by an application perspective, namely to improve information retrieval beyond simple keyword search and foster semantic interoperability on the Web, there is a wide variety of proposed and employed methods and principles which differ substantially and are under continuous discussion concerning their suitability for achieving the Semantic Web vision.

Nevertheless, some approaches have become well-established and widely accepted. Probably the most commonly accepted principle is the idea of endowing information on the Web with *metadata* which shall convey the meaning (or *semantics*) of the information by further describing (or *annotating*) it, and thus enable information retrieval, reuse, and integration. Shared schemas that formally specify the conceptualizations underlying a certain application area or information community are called *ontologies*. In the light of this approach, the development of suitable languages for representing ontologies has been, and still is, a key issue for Semantic Web research.

1.1.1 Languages for Representing Knowledge on the Web

Formal languages that support the representing of information semantics are commonly called *ontology languages*, and among the many proposals for such languages which can be found in the literature, those which have become recommended standards by the World Wide Web Consortium[1] (W3C) are of primary importance; namely the *Resource Description Framework RDF* [31], the *Web Ontology Language OWL* [22, 32], and the *Rule Interchange Format RIF* [10, 11]. Syntactically, they are based on XML [12], although their underlying data models are significantly different. During the last years other serializations have gained popularity, for example Notation3 [7] and Turtle [35] for representing RDF graphs. Historically, all ontology languages can all be traced back to traditions in the field of knowledge representation and reasoning [23]. We briefly describe each of these languages in turn.

The **Resource Description Framework (RDF)** is essentially a language for representing *triples* of the form `subject predicate object`, where each of the entries is a uniform resource identifier (URI) for Web resources. Predicates are usually also called *properties*. An `object` of one triple may be the `subject` of another triple and so forth. An example for such triples would be

`ex:HoratioNelson ex:diedIn ex:BattleOfTrafalgar` and
`ex:BattleOfTrafalgar ex:during ex:NapoleonicWars`

where `ex:` is a namespace identifier so that `ex:HoratioNelson` expands to a proper URI like `http://www.example.org/HoratioNelson`. The RDF specification furthermore includes pre-defined vocabulary, the meaning of which is defined by means of a *formal semantics*, described further below. The most important examples for this are `rdf:type`, `rdfs:subClassOf` and `rdfs:subPropertyOf`, where `rdf`, respectively `rdfs`, are used as namespace identifiers for the pre-defined namespaces of RDF and RDF Schema, respectively. RDF Schema is, in fact, part of the RDF specification, but has been given a separate name to indicate that it has additional expressive features, i.e., a richer pre-defined vocabulary. The informal meaning of this vocabulary examples are as follows. A triple such as

`ex:BattleOfTrafalgar rdf:type ex:NavalBattle`

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

indicates that the individual event `ex:BattleOfTrafalgar` is an instance of the class `ex:NavalBattle` and is interpreted as set membership. A triple such as

```
ex:NavalBattle rdfs:subClassOf ex:Battle
```

indicates that every instance of the class `ex:NavalBattle` is also an instance of the class `ex:Battle`. A triple such as

```
ex:diedIn rdfs:subPropertyOf ex:participatedIn
```

indicates that, whenever some a is in a `ex:diedIn`-relationship to some b (e.g., because a triple a `ex:diedIn` b has been specified), then a and b are also in `ex:participatedIn`-relationship.

An RDF Document now represents, essentially, a finite set of such triples. It is common practice to think of such a set as representing a finite labeled graph, where each triple gives rise to two vertices (subject and object) and an edge (the predicate). The labels of such subject-predicate-object triples are URIs. Note, however, that it is possible that a URI occurring as a predicate (i.e., an edge) also occurs as a subject or object (i.e., as a vertex) in a different triple. This means that edges can at the same time be vertices, thus breaking what would commonly be considered a graph structure. The W3C recommended standard SPARQL[2] [37] serves as a language for querying RDF triples that are often stored in a so-called *triplestore*. It is currently undergoing a revision [18] which is expected to make SPARQL also useful for the other ontology languages introduced below.

The **Web Ontology Language (OWL)** has been developed for representing knowledge which is more complex than what can be represented in RDF. In particular, OWL provides pre-defined vocabulary to describe complex relationships between classes. For example, in OWL we can specify that some class is the *union* (logically speaking, the disjunction) or the *intersection* (logically speaking, the conjunction) of two (or more) other classes. Class union and intersection are two examples of *class constructors* which OWL provides. Others also involve properties or even URIs which stand for *individuals* (which should be thought of as instances of classes), such that complex relationships like *Cape Trafalgar is a headland that witnessed at least one battle*, or *Every naval battle has (among other things) two or more involved parties, at least one location, a time span during which it took place, and a reason for engaging in it* can be expressed. Conceptually speaking, OWL is essentially a fragment of first-order predicate logic with equality and can be identified as a so-called *description logic* [6].

The OWL standard defines a number of variants which differ in sophistication and anticipated use cases. OWL EL, OWL QL, and OWL RL correspond to relatively small description logics for which reasoning algorithms (see Section 1.1.2) are relatively easy to implement. OWL DL encompasses the former three and corresponds to a very expressive description logic called $\mathcal{SROIQ}(D)$. OWL Full is the most general variant and encompasses both OWL DL and RDF—but it is not a description logic, and usually attempts are made to remain within smaller fragments when modeling ontologies.

The **Rule Interchange Format (RIF)** follows a different paradigm than OWL by addressing knowledge modeling by means of rules as used in logic programming, Datalog, and other rule-based paradigms. Its main purpose is to facilitate the interchange of rules between applications, but it also constitutes a family of knowledge representation languages in its own right. In particular, RIF-Core corresponds to Datalog, while RIF-BLD (from *Basic Logic Dialect* corresponds to Horn logic (essentially, definite logic programs). Rule paradigms essentially cater for the expression of if-then type relationships such as *If*

[2] "SPARQL" is a recursive acronym which stands for "SPARQL Protocol and RDF Query Language."

two parties agree on a cease-fire, then a combat operation started by one of these parties constitutes a violation of the agreement. Rules in the sense of RIF-Core or RIF-BLD also correspond to certain fragments of first-order predicate logic [24].

1.1.2 Formal Semantics

Ontology languages can be understood as *vocabulary specification languages* in the sense that they can be used to define relationships of vocabulary terms related to a domain of interest. For example, they could be used to define a set of vocabulary terms describing spatial classes and relationships, such as *Headland*, *Coastal Landform*, *Erosion*, or *contributesTo*. Furthermore, terms can be put into relation to each other, e.g., in form of a taxonomy: *Headland* is a subclass of *Coastal Landform*, which in turn is a subclass of *Landform*. More complicated relationships are also expressible, such as "every land form has at least one event that contributed to its creation," which, for instance, relates the terms *Erosion* and *contributesTo* to *Headland*, without saying that there is only one erosion per headland or that erosions are the only kind of events that are in a *contributesTo* relation to headlands. Note how we exploit the fact that erosions are subclasses of events and headlands are landforms. The exact kinds of relationships which are expressible depend on the ontology language chosen for the representation. Such vocabulary definitions are often referred to as *terminological* or *schema knowledge*. In the context of OWL, a set of such definitions is often referred to as a *TBox*.

Ontology languages furthermore allow for the expression of *assertional* knowledge or *facts* (often called *ABox* statements in the context of OWL). This refers to statements involving instances of classes, such as "Cape Trafalgar is a headland" or "Nelson is a Person."

Each of the W3C ontology languages presented in Section 1.1.1 comes endowed with a so-called *formal semantics* which is described in model-theoretic terms borrowed from first-order predicate logic [19, 34]. This formal semantics describes in mathematical terms how to draw valid *logical conclusions* from an ontology. For example, if *Headland* is specified as a subclass of *Coastal Landform*, and the ontology furthermore specifies that "every landform was shaped by at least one event," then one such logical consequence would be that "every headland was shaped by at least one event." While this example may seem obvious, in more complex cases it is often difficult to decide intuitively whether a statement is a logical consequence of an ontology. In these cases, the formal semantics serves as a *formal specification* for the valid logical consequences, and it is in this sense in which formal semantics provides a "meaning" to ontologies.

Logical consequences can be understood as knowledge which is *implicit* in an ontology. The formal semantics states that this implicit knowledge constitutes things which are *necessarily true*, given the ontology. As such, formal semantics enables *interoperability*, since it describes consequences which can be drawn from combining previously separated pieces of knowledge. Formal semantics can also be used in various ways when creating ontologies. For example, if an ontology engineer observes that her ontology has some logical consequences which seem to be undesirable from an application perspective, then this is an indication that parts of the ontology may have to be corrected or revised.

1.1.3 Key Issues in Realizing the Semantic Web Vision

A standard template to using ontology languages would thus be as follows. Entities are represented on the Web using URIs, e.g., *Battle of Trafalgar* could be represented by a URI such as `http://example.org/BattleOfTrafalgar` and the Napoleonic Wars by `http://example.org/NapoleonicWars`. Likewise, vocabulary elements are represented by URIs, such as `http://example.org/during` or `http://example.org/participatedIn`. In a background ontology, terminological information could then express, e.g., that "during" is a transitive relation. Now consider, e.g., the situation where metadata in a historical knowledge base is provided along with the human-readable content, which states that the Battle of Trafalgar took place during the Napoleonic Wars, while another, more generic Web portal lists facts about historic periods, including the fact that the Napoleonic Wars took place during the age of the Industrial Revolution. By collecting all this information, e.g., using a Semantic Web crawler, this information could be combined, and due to the transitivity of "during" we would be able to also obtain the implicit knowledge, through logical deduction, that the Battle of Trafalgar took place during the Industrial Revolution. Even more, by adding semantics to the *participatesIn* relation discussed before, we could infer that Nelson lived during the Industrial Revolution.

The example just given indicates how Semantic Web technologies can in principle be used for information sharing, integration, and reuse across knowledge bases. However, in order to cast this idea into practical approaches, several obstacles need to be overcome. We list some of the most important ones.

- Different Web sites and knowledge bases may use different URIs for identifying entities on the Semantic Web. For information integration, we need to be able to identify these so-called *co-references*.

- Different Web sites may use different background ontologies. For information integration, these different ontologies need to be understood in relation to each other and formally related using suitable formats, preferably by using ontology languages or derivatives thereof. This issue is known as *ontology alignment*.

- Information on the Web may be *semantically heterogeneous*. For example, on the instance level different sources may specify other start and end dates for historical periods and, thus, may not agree whether the Napoleonic Wars also fall into the period known as Age of Enlightenment. On the schema level, the meaning assigned to terms such as *Headland*, *Erosion*, or *NavalBattle* may differ to a degree where they may become incompatible. Such issues are addressed by *semantic mediation and translation* research.

- Who is going to endow information on the Web with matadata? Can we automate or partly automate this process using data mining or machine learning techniques? What are good interfaces and tools for developers and domain experts to support the creation of metadata?

- Algorithms for deducing logical consequences from ontologies are usually very dependent on input which constitutes ontologies of very high modeling quality. The creation of ontologies of such high quality can currently not be automated, and their creation usually requires experts in both ontology modeling and in the application domain under consideration. Consequently, their creation and maintenance can be very expensive. The creation of methods, workflows, and tools to support ontology modeling (and any other part of the ontology lifecycle) is, thus, of utmost importance for the development of the field and its applications.

- How can we deal with evolving information, e.g., when data changes or is revised, or if vocabulary terms change their meaning over time? How can we deal with uncertainties or noise in metadata, which seem to be unavoidable in many realistic settings?

1.1.4 Linked Data

Linking Data[3] [9] is a community effort which started in 2007 and is supported by the W3C. Its declared goal is to bootstrap a global graph of interlinked data called the *Web of Data* by converting existing data sets into the format of ontology languages, and interlinking them with meaningful relationships. Linked Data has been a key driver for research and development in Semantic Web since its inception, and has in particular created substantial interest in the industry and governments around the worlds. This Web of Data is currently growing rapidly. As of 2011, it had an estimated size of about 32 billion RDF triples, with contributions coming increasingly from industry and from public sector bodies such as governments and libraries.

In its current form, Linked Data contains mainly assertional knowledge expressed in RDF, and there are only very few light-weight guiding principles for its creation. Methods and tools for making use of Linked Data, and the question how to advance from Linked Data towards the more general Semantic Web vision, are among the most prominent current research questions. Similarly, ontologies that have been developed to annotate and interlink these datasets are often light-weight to a degree where they fail to restrict the interpretation of terms towards their intended meaning [25].

1.2 Impact on Practice

In order to assess the impact of the Semantic Web field in practice, it is important to realize that Semantic Web Technologies can also be applied in contexts other than the World Wide Web. They likewise apply to (closed) intranets, but also to many other settings where information creation, retrieval, reuse, and integration are of importance. Typical examples includes enterprise information integration and knowledge management systems. In many other cases, e.g., search engines, ontologies and Semantic Web Technologies are used in the backend, and, while of major importance, are not visible to the end users.

And indeed, early adoption of Semantic Web Technologies in non-Web areas can be traced back to at least the first years of the 21st century, when the first spin-off companies left the research realm and applied their methods to enable industrial applications. In the next few years, industry interest kept growing, but slowly, and at the same time other research disciplines which rely on information management and integration adopted Semantic Web Technologies for their purposes. A leading role in this latter development was taken by the life sciences, in particular related to health care and bioinformatics.

Usage of Semantic Web Technologies on the *Web*, however, hardly happened in these years, apart from use-case studies and prototypes. One prominent exception was *Semantic MediaWiki*[4] [30], a semantic enhancement of the MediaWiki software which underlies Wikipedia.[5] It was presented in 2006 and very quickly found substantial uptake with thousands of known installations world-wide. Indeed, efforts are currently under way to customize the extension for the use in Wikipedia.[6]

Since 2007, and substantially driven by advent of Linked Data, industrial uptake is rising significantly, both on and off the Web. At the time of this writing,[7] Semantic Web

[3] http://linkeddata.org/
[4] http://semantic-mediawiki.org/
[5] http://www.wikipedia.org/
[6] http://meta.wikimedia.org/wiki/New_Wikidata
[7] March 2012

Technologies have started to deeply penetrate information technologies and can be expected to become part of the mainstream arsenal for application developers in the near future. This adoption of Semantic Web technologies does not necessarily mean the use of the exact standards described in Section 1.1, but rather the adherence to the general principles of using machine-readable and understandable metadata for information sharing, retrieval, and integration.

Adoption in practice of Semantic Web Technologies is, to-date, often restricted to the use of simple metadata, and in particular the systematic use of formal semantics is currently only possible in particular application areas and settings, since stronger solutions for some of the key issues listed in Section 1.1.3 need to be developed first (see also Section 1.3).

In the following, we list some of the most visible recent achievements of Semantic Web technologies with respect to practical applications.

- The need for meta-modeling has recently been picked up by some of the most prominent companies on the Web. Facebook's Open Graph protocol[8] is a prominent example, as is schema.org,[9] a joint effort by Bing, Google, and Yahoo! for improving Web search. RDF versions are endorsed by these companies, and the schema.org ontology is officially available in OWL/RDF.[10]

- In February 2011, the Watson system by IBM [15] made international headlines for beating the best humans in the quiz show *Jeopardy!*.[11] The performance is already being considered as a milestone in the development of Artificial Intelligence techniques related to general question answering. Semantic Web technologies have been one of the key ingredients in Watson's design.[12]

- A significant number of very prominent websites are powered by Semantic Web technologies, including the New York Times,[13] Thomson Reuters,[14] BBC [36], and Google's Freebase, as well as sites using technology such as Yahoo! SearchMonkey [33] or Drupal 7[15] [13].

- The Speech Interpretation and Recognition Interface *Siri* launched by Apple as an intelligent personal assistant for the new generation of IPhone smartphones in 2011 heavily draws from work on ontologies, knowledge representation, and reasoning [17].

- Oracle Database 11g supports OWL.[16]

- Recently, The Wall Street Journal ran an article[17] announcing that Google is in the progress of enhancing its Web search using Semantic Web technologies.

- The *International Classification of Diseases*, ICD, is the international standard manual for classifying diseases, endorsded by the World Health Organization and in use world-wide. It is currently revised, and ICD-11 (the 11th revision) is scheduled to appear in 2015. The revision is driven by a collaborative platform using OWL as underlying technology [42].

[8] http://ogp.me/
[9] http://schema.org/
[10] http://schema.org/docs/schemaorg.owl
[11] http://www.jeopardy.com/
[12] http://semanticweb.com/game-show-circuit-was-just-a-first-step-for-ibms-watson-and-deep-qa_b20431
[13] http://www.nytimes.com/, see also http://data.nytimes.com/
[14] http://www.opencalais.com/
[15] http://semantic-drupal.com/
[16] http://www.oracle.com/technetwork/database/options/semantic-tech/index.html
[17] http://online.wsj.com/article/SB10001424052702304459804577281842851136290.html

- Many governments and large companies now publish a plethora of governmental and other information as Linked Data, with England and the U.S. being early adopters.[18]

1.3 Research Issues

Currently, the Semantic Web field is mainly driven by developments regarding Linked Data (see Section 1.1.4). The amount of information available as Linked Data has shown exponential growth since its inception, and there are no indications that this will slow down soon. The Linked Data cloud indeed serves as an interlinked data source which is available in form of readily processable syntax (namely, RDF), and as such has the potential for wide-spread usage in data-intensive applications.

However, while it is certainly very helpful that Linked Data is available in RDF, there are very few general principles which would guide its creation.[19] Consequently, Linked Data still suffers from many heterogeneity challenges which the Semantic Web field has initially set out to overcome [25, 27]. In reference to the issues identified in Section 1.1.3, we note that co-reference identification is still a central and unsolved issue, even for Linked Data [43]. Likewise, efficiently addressing semantic heterogeneity in Linked Data, e.g., by means of ontology alignment, has only recently started to be studied [26], in particular for the difficult quest of providing question answering systems based on Linked Data [29]. The development of special techniques for dealing with evolving Linked Data is also in very early stages.

Thus, there currently is a significant disconnect between the factually deployed Linked Data on the Web, and the already established research results regarding the use of Semantic Web Technologies and ontology languages. It appears to be the foremost current research question, how to enable these strongly and formally semantic approaches for use on and with Linked Data. And this quest is inherently tied with the current lack of useful schema knowledge in Linked Data [25].

The question, how Linked Data can indeed be evolved into a Semantic Web in the stronger sense of the term, is indeed very controversially discussed. And while it is certainly also a question about good underlying conceptual principles, it is foremost also a question of practicability, of finding suitable next steps for development which can find significant adoption in practice.

To address the relative lack of schema knowledge in Linked Data, a foremost requirement is the availability of strong tool support which makes Semantic Web Technologies more easily accessible for practitioners. This includes the automated, semi-automated or manual generation of metadata and ontologies, powerful ontology alignment systems, lifecycle support for metadata and ontologies such as versioning approaches and revision processes, and tools which enable an easier reuse of metadata and ontologies in applications for which they were not originally developed. In all these aspects, the research community has provided significant research advances which nevertheless remain to be further strengthened before they can find wide-spread adoption.

At the same time, a significant body of best practices guidelines need to be developed, which includes ontology creation and lifecycle aspects, but also methods and processes for

[18]See http://www.data.gov/semantic and http://data.gov.uk/linked-data/.

[19]This is, in principle, a good thing, and, in fact, is in line with the "bottom-up" nature of the World Wide Web itself.

making use of Semantic Web technologies. Of growing importance in this respect is the development of useful ontology design patterns for ontology modeling [16].

Eventually, in order to meet the goals of the Semantic Web vision, strongly semantic approaches—including ontology reasoning–will have to be embraced and brought to bear on realistic Web data. In order for this to happen, researchers need to find clear answers on how to establish semantic interoperability without giving up on semantic heterogeneity [28], to scalability issues of reasoning algorithms, and to the question how to deal with metadata which arises from the strongly collaborative Web without central control, i.e., metadata which substantially varies in modeling quality [21, 25].

1.4 Summary

The Semantic Web is an interdisciplinary research field which aims at augmenting the existing World Wide Web with machine-readable and understandable metadata such that information on the Web becomes available for processing in intelligent systems. In particular, it shall establish solutions for seamless information creation, retrieval, reuse, and integration on the Web.

The key underlying technology is the use of so-called ontology languages for representing metadata information. There exist several such languages endorsed by the World Wide Web Consortium. They support formal semantics which enables automated reasoning using deductive logical methods.

Semantic Web is recently seeing a lot of adoption by the industry, and this also includes adoption for purposes other than for information on the World Wide Web. One of the driving recent developments is the publishing of significant amounts of data in ontology language formats on the World Wide Web—this information is referred to as Linked Data.

Despite its success, many core issues still require further in-depth, and partially foundational, research, such as the systematic use of deep semantics on the Web by means of automated reasoning techniques.

Defining Terms

Annotation refers to the attaching of metadata information to entities on the Web, such as web pages, text, text snippets, words or terms, images, tables, etc. These annotations are usually not visible in normal Web browsers, but can be retrieved from the source code of the web page for further processing. E.g., a picture could bear the annotation "Barack Obama," indicating that it is a picture of Barack Obama.

A **co-reference** occurs when two different identifiers are used for one real-world entity. The identification of co-references in ontological data on the Semantic Web is a challenging research problem. The term is borrowed from linguistics.

Datalog is a rule-based knowledge representation language based on function-free Horn logic. It is used in deductive databases.

Description logics are a family of closely related knowledge representation languages. Traditionally, they are strongly based on first-order predicate logic from which they inherit their open-world semantics. Description logics are usually decidable.

Horn logic is a fragement of first-order predicate logic which can be expressed in the form of rules. It is the basis for the logic programming paradigm, and thus also for the logic programming language Prolog.

Linked Data refers to data on the World Wide Web which is represented using RDF, following a few simple principles. These datasets are strongly interelinked, thus forming a network of knowledge, often refered to as the *Web of Data*. If the data is available under an open license, it is called Linked *Open* Data, but often this distinction is not made in the literature. Linked Data is considered a major milestone in the development of the Semantic Web.

Logic programming is a knowledge representation and programming paradigm usually based on Horn logic with some modifications and extensions. In particular it uses a closed-world paradigm. Prolog is the most widely known logic programming language.

Metadata is data which provides information for other data, often through the process of annotation. Metadata can be of many different forms, in the simplest case it consists only of keywords (often called *tags*), but in a Semantic Web context metadata is usually expressed using some ontology language.

Ontology is a term borrowed from philosophy. In modern Computer Science an ontology is a knowledge base which is expressed by describing relationships between concepts in a given domain of interest. The relationships are described using knowledge representation languages—then called ontology languages—, which are usually derived from first-order predicate logic or from the logic programming paradigm.

Ontology alignment refers to the establishing of formal relationships between the concepts and other terms in two different ontologies. This can be done manually, but there is also a significant area of research which develops automated ontology alignment systems.

Ontology design patterns are schemas for sets of concepts and relations occurring frequently in ontology modeling. They are typically expressed in some ontology language.

Semantics, more precisely *formal* semantics, usually refers to the notion of model-theoretic semantics in first-order predicate logic, adapted for an ontology language. The main purpose of such a formal semantics is to define a notion of *logical consequence*. In this sense, a formal semantics is an implicit specification for all the logical consequences which can be derived from an ontology or a knowledge base.

URI is an abbreviation for *Uniform Resource Identifier*. URIs are used to represent resources on the World Wide Web, e.g., websites. It is not required that an entity identified by an URI can in fact be located or accessed on the World Wide Web.

Web of Data—see Linked Data.

World Wide Web Consortium W3C is the main international organization which develops standards for the World Wide Web.

XML stands for *Extensible Markup Language*. It is a text-based format for the representation of hierarchically structured data. XML is ubiquitous on the World Wide Web, and often used for the interchange of data on the World Wide Web.

Bibliography

[1] Dean Allemang and Jim Hendler. *Semantic Web for the Working Ontologist*. Morgan Kaufmann, 2008.

[2] Grigoris Antoniou, Marko Grobelnik, Elena Paslaru Bontas Simperl, Bijan Parsia, Dimitris Plexousakis, Pieter De Leenheer, and Jeff Z. Pan, editors. *The Semanic Web:*

Research and Applications – 8th Extended Semantic Web Conference, ESWC 2011, Heraklion, Crete, Greece, May 29 – June 2, 2011, Proceedings, Part II, volume 6644 of *Lecture Notes in Computer Science*. Springer, 2011.

[3] Grigoris Antoniou, Marko Grobelnik, Elena Paslaru Bontas Simperl, Bijan Parsia, Dimitris Plexousakis, Pieter De Leenheer, and Jeff Z. Pan, editors. *The Semantic Web: Research and Applications – 8th Extended Semantic Web Conference, ESWC 2011, Heraklion, Crete, Greece, May 29 – June 2, 2011, Proceedings, Part I*, volume 6643 of *Lecture Notes in Computer Science*. Springer, 2011.

[4] Lora Aroyo, Chris Welty, Harith Alani, Jamie Taylor, Abraham Bernstein, Lalana Kagal, Natasha Fridman Noy, and Eva Blomqvist, editors. *The Semantic Web – ISWC 2011 – 10th International Semantic Web Conference, Bonn, Germany, October 23–27, 2011, Proceedings, Part I*, volume 7031 of *Lecture Notes in Computer Science*. Springer, 2011.

[5] Lora Aroyo, Chris Welty, Harith Alani, Jamie Taylor, Abraham Bernstein, Lalana Kagal, Natasha Fridman Noy, and Eva Blomqvist, editors. *The Semantic Web – ISWC 2011 – 10th International Semantic Web Conference, Bonn, Germany, October 23–27, 2011, Proceedings, Part II*, volume 7032 of *Lecture Notes in Computer Science*. Springer, 2011.

[6] Franz Baader, Diego Calvanese, Deborah McGuinness, Daniele Nardi, and Peter Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, 2007.

[7] Tim Berners-Lee and Dan Connolly. *Notation3 (N3): A readable RDF syntax*. W3C Team Submission 28 March 2011, 2011. Available from http://www.w3.org/TeamSubmission/n3/.

[8] Tim Berners-Lee, James Hendler, and Ora Lassila. The Semantic Web. *Scientific American*, 284(5):34–43, May 2001.

[9] Christian Bizer, Tom Heath, and Tim Berners-Lee. Linked Data – The Story So Far. *International Journal on Semantic Web and Information Systems*, 5(3):1–22, 2009.

[10] Harold Boley, Gary Hallmark, Michael Kifer, Adrian Paschke, Axel Polleres, and Dave Reynolds, editors. *RIF Core Dialect*. W3C Recommendation 22 June 2010, 2010. Available from http://www.w3.org/TR/rif-core/.

[11] Harold Boley and Michael Kifer, editors. *RIF Basic Logic Dialect*. W3C Recommendation 22 June 2010, 2010. Available from http://www.w3.org/TR/rif-bld/.

[12] Tim Bray, Jean Paoli, C.M. Sperberg-McQueen, Eve Maler, and François Yergeau, editors. *Extensible Markup Language (XML) 1.0 (Fifth Edition)*. W3C Recommendation, 26 November 2008. Available at http://www.w3.org/TR/xml.

[13] Stephane Corlosquet, Renaud Delbru, Tim Clark, Axel Polleres, and Stefan Decker. Produce and consume Linked Data with Drupal! In Abraham Bernstein, David R. Karger, Tom Heath, Lee Feigenbaum, Diana Maynard, Enrico Motta, and Krishnaprasad Thirunarayan, editors, *The Semantic Web – ISWC 2009, 8th International Semantic Web Conference, ISWC 2009, Chantilly, VA, USA, October 25–29, 2009. Proceedings*, volume 5823 of *Lecture Notes in Computer Science*, pages 763–778. Springer, 2009.

[14] Lee Feigenbaum, Ivan Herman, Tonya Hongsermeier, Eric Neumann, and Susie Stephens. The Semantic Web in action. *Scientific American*, pages 90–97, November 2007.

[15] David A. Ferrucci, Eric W. Brown, Jennifer Chu-Carroll, James Fan, David Gondek, Aditya Kalyanpur, Adam Lally, J. William Murdock, Eric Nyberg, John M. Prager, Nico Schlaefer, and Christopher A. Welty. Building Watson: An overview of the DeepQA project. *AI Magazine*, 31(3):59–79, 2010.

[16] Aldo Gangemi. Ontology design patterns for semantic web content. In Yolanda Gil, Enrico Motta, V. Richard Benjamins, and Mark A. Musen, editors, *The Semantic Web – ISWC 2005, 4th International Semantic Web Conference, ISWC 2005, Galway, Ireland, November 6-10, 2005, Proceedings*, volume 3729 of *Lecture Notes in Computer Science*, pages 262–276. Springer, 2005.

[17] Tom Gruber. Siri: A virtual personal assistant. Keynote presentation at Semantic Technologies Conference, June 2009. http://tomgruber.org/writing/semtech09.htm.

[18] Steve Harris and Andy Seaborne. *SPARQL 1.1 Query Language*. W3C Working Draft, 12 May 2011. Available from http://www.w3.org/TR/sparql11-query/.

[19] Patrick Hayes, editor. *RDF Semantics*. W3C Recommendation, 10 February 2004. Available at http://www.w3.org/TR/rdf-mt/.

[20] John Hebeler, Matthew Fisher, Ryan Blace, Andrew Perez-Lopez, and Mike Dean. *Semantic Web Programming*. Wiley, 2009.

[21] Pascal Hitzler. Towards reasoning pragmatics. In Krzysztof Janowicz, Martin Raubal, and Sergei Levashkin, editors, *GeoSpatial Semantics, Third International Conference, GeoS 2009, Mexico City, Mexico, December 3–4, 2009. Proceedings*, Lecture Notes in Computer Science, pages 9–25. Springer, 2009.

[22] Pascal Hitzler, Markus Krötzsch, Bijan Parsia, Peter F. Patel-Schneider, and Sebastian Rudolph, editors. *OWL 2 Web Ontology Language: Primer*. W3C Recommendation 27 October 2009, 2009. Available from http://www.w3.org/TR/owl2-primer/.

[23] Pascal Hitzler, Markus Krötzsch, and Sebastian Rudolph. *Foundations of Semantic Web Technologies*. Chapman & Hall/CRC, 2009.

[24] Pascal Hitzler and Anthony K. Seda. *Mathematical Aspects of Logic Programming Semantics*. CRC Press, 2010.

[25] Pascal Hitzler and Frank van Harmelen. A reasonable semantic web. *Semantic Web*, 1(1–2):39–44, 2010.

[26] Prateek Jain, Pascal Hitzler, Amit P. Sheth, Kunal Verma, and Peter Z. Yeh. Ontology alignment for linked open data. In P. Patel-Schneider, Y. Pan, P. Hitzler, P. Mika, L. Zhang, J. Pan, I. Horrocks, and B. Glimm, editors, *Proceedings of the 9th International Semantic Web Conference, ISWC 2010, Shanghai, China, November 7-11, 2010*, volume 6496, pages 402–417. Springer, Heidelberg, 2010.

[27] Prateek Jain, Pascal Hitzler, Peter Z. Yeh, Kunal Verma, and Amit P. Sheth. Linked Data is Merely More Data. In Dan Brickley, Vinay K. Chaudhri, Harry Halpin, and Deborah McGuinness, editors, *Linked Data Meets Artificial Intelligence*, pages 82–86. AAAI Press, Menlo Park, CA, 2010.

[28] Krzysztof Janowicz. The role of space and time for knowledge organization on the Semantic Web. *Semantic Web*, 1(1-2):25–32, 2010.

[29] Amit Krishna Joshi, Prateek Jain, Pascal Hitzler, Peter Z. Yeh, Kunal Verma, Amit P. Sheth, and Mariana Damova. Alignment-based querying of Linked Open Data. Technical report, Kno.e.sis Center, Wright State University, Dayton, OH, 2012. Available from http://www.pascal-hitzler.de/.

[30] Markus Krötzsch, Denny Vrandecic, Max Völkel, Heiko Haller, and Rudi Studer. Semantic Wikipedia. *Journal of Web Semantics*, 5(4):251–261, 2007.

[31] Frank Manola and Eric Miller, editors. *Resource Description Framework (RDF). Primer.* W3C Recommendation, 10 February 2004. Available at http://www.w3.org/TR/rdf-primer/.

[32] Deborah L. McGuinness and Frank van Harmelen, editors. *OWL Web Ontology Language Overview.* W3C Recommendation, 10 February 2004. Available at http://www.w3.org/TR/owl-features/.

[33] Peter Mika. Case Study: Improving Web Search Using Metadata. W3C Semantic Web Use Cases and Case Studies, November 2008. Available from http://www.w3.org/2001/sw/sweo/public/UseCases/yahoo/.

[34] Boris Motik, Peter F. Patel-Schneider, and Bernardo Cuenca Grau, editors. *OWL 2 Web Ontology Language: Direct Semantics.* W3C Recommendation, 27 October 2009. Available at http://www.w3.org/TR/owl2-direct-semantics/.

[35] Eric Prud'hommeaux and Gavin Carothers, editors. *Turtle—Terse RDF Triple Language.* W3C Working Draft 09 August 2011, 2011. Available from http://www.w3.org/TR/turtle/.

[36] Yves Raimond. Case Study: Use of Semantic Web Technologies on the BBC Web Sites. W3C Semantic Web Use Cases and Case Studies, January 2010. Available from http://www.w3.org/2001/sw/sweo/public/UseCases/BBC/.

[37] Andy Seaborne and Eric Prud'hommeaux. *SPARQL query language for RDF.* W3C Recommendation, 15 January 2008. Available from http://www.w3.org/TR/rdf-sparql-query/.

[38] Sadagopan Srinivasan, Krithi Ramamritham, Arun Kumar, M. P. Ravindra, Elisa Bertino, and Ravi Kumar, editors. *Proceedings of the 20th International Conference on World Wide Web, WWW 2011, Hyderabad, India, March 28 – April 1, 2011.* ACM, 2011.

[39] Sadagopan Srinivasan, Krithi Ramamritham, Arun Kumar, M. P. Ravindra, Elisa Bertino, and Ravi Kumar, editors. *Proceedings of the 20th International Conference on World Wide Web, WWW 2011, Hyderabad, India, March 28 – April 1, 2011 (Companion Volume).* ACM, 2011.

[40] Steffen Staab and Rudi Studer, editors. *Handbook on Ontologies.* International Handbooks on Information Systems. Springer, 2 edition, 2009.

[41] Rudi Studer. The Semantic Web: Suppliers and customers. In Isabel F. Cruz, Stefan Decker, Dean Allemang, Chris Preist, Daniel Schwabe, Peter Mika, Michael Uschold, and Lora Aroyo, editors, *The Semantic Web – ISWC 2006, 5th International Semantic Web Conference, ISWC 2006, Athens, GA, USA, November 5-9, 2006, Proceedings*, volume 4273 of *Lecture Notes in Computer Science*, pages 995–996. Springer, 2006.

[42] Tania Tudorache, Sean M. Falconer, Csongor Nyulas, Natalya Fridman Noy, and Mark A. Musen. Will semantic web technologies work for the development of ICD-11? In Peter F. Patel-Schneider, Yue Pan, Pascal Hitzler, Peter Mika, Lei Zhang, Jeff Z. Pan, Ian Horrocks, and Birte Glimm, editors, *The Semantic Web – ISWC 2010 – 9th International Semantic Web Conference, ISWC 2010, Shanghai, China, November 7–11, 2010, Revised Selected Papers, Part II*, volume 6497 of *Lecture Notes in Computer Science*, pages 257–272. Springer, 2010.

[43] Giovanni Tummarello and Renaud Delbru. Linked Data is Merely More Data. In Dan Brickley, Vinay K. Chaudhri, Harry Halpin, and Deborah McGuinness, editors, *Linked Data Meets Artificial Intelligence*. AAAI Press, Menlo Park, CA, 2010.

Further Information

Semantic Web as a field of research and applications is still under rapid development. Textbooks on the topic thus tend to be outdated rather quickly, unless they focus on fundamental issues which already find a broad consensus in the community. We recommend the following three books as thorough introductions: Foundations of Semantic Web Technologies [23] by Hitzler, Krötzsch, and Rudolph; Semantic Web for the Working Ontologist [1] by Allemang and Hendler; and Semantic Web Programming [20] by Hebeler, Fisher, Blace, Perez-Lopez and Dean. The first-mentioned book [23] focuses on foundations, in particular on a thorough treatment of the standardized ontology languages, including their formal semantics. The other two books [1, 20] are written from a more practical perspective—they convey more hands-on knowledge, but are less comprehensive in the formal foundations. All three books are widely acclaimed as excellent introductions.

The edited Handbook on Ontologies [40] by Staab and Studer is a popular resource which introduces many aspects of Semantic Web technologies. The chapters are written by well-known experts in the field. The book is necessarily much less systematic and thorough than a textbook, but its coverage is much wider and includes many topics which are of central importance for the Semantic Web, but which have not been developed far enough yet to be included in a standard textbook.

Primary resources for state-of-the-art developments in the field are the major journals and conferences in the area. Many journals in Computer Science and adjacent fields in fact publish papers related to Semantic Web, but there are also some prominent ones which are dedicated exclusively to the field. These include the Journal of Web Semantics,[20] the Semantic Web journal,[21] and the International Journal On Semantic Web and Information Systems.[22] Major conferences in the area are the International Semantic Web Conference[23] (see e.g. [4, 5]) and the Extended Semantic Web Conference[24] (see e.g. [2, 3]), which attract primarily researchers, and the Semantic Technology Conference[25] which targets industry. Many other major conferences also publish research papers in Semantic Web, e.g., the comprehensive World Wide Web Conference[26] (see e.g. [38, 39]).

[20] http://www.websemanticsjournal.org/
[21] http://www.semantic-web-journal.net/
[22] http://www.ijswis.org/
[23] http://swsa.semanticweb.org/
[24] http://eswc-conferences.org/
[25] See http://semanticweb.com/
[26] wwwconference.org

Acknowledgement. Pascal Hitzler acknowledges support by the National Science Foundation under award 1017225 *III: Small: TROn – Tractable Reasoning with Ontologies.* Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.