Integrated Design and Process Technology, **IDPT-2006**
Printed in the United States of America, June, 2006

A Semantic Web Status Model

A.J. Gerber, A. Barnard, A.J. van der Merwe
School of Computing
Unisa
Pretoria, South Africa
gerbeaj, barnaa, vdmeraj@unisa.ac.za

ABSTRACT: *Semantic Web* application areas are experiencing intensified interest due to the rapid growth in the use of the Web, together with the innovation and renovation of information content technologies. The Semantic Web is regarded as an integrator across different content and information applications and systems, and provide mechanisms for the realisation of Enterprise Information Systems. The rapidity of the growth experienced provides the impetus for researchers to focus on the creation and dissemination of innovative Semantic Web technologies, where the envisaged 'Semantic Web' is long overdue. Often the terms *'Semantics', 'metadata', 'ontologies'* and *'Semantic Web'* are used inconsistently. In particular, these terms are used as everyday terminology by researchers and practitioners, spanning a vast landscape of different fields, technologies, concepts and application areas. Furthermore, there is confusion with regards to the current status of the enabling technologies envisioned to realise the Semantic Web. In this paper we chart the Semantic Web landscape and provide a brief summary of related terms and enabling technologies. We also use the architectural model proposed by Tim Berners-Lee as basis to present a status model that reflects current and emerging technologies.

I. INTRODUCTION

The Semantic Web is regarded as a means to integrate different content and information applications and systems. The technologies specified to realise the Semantic Web provide several mechanisms that might find application areas within the domain of Enterprise Information Systems. *'Semantics', 'metadata', 'ontologies'* and *'Semantic Web'* are terms used regularly in the vernacular of IT researchers and are also frequently employed by IT practitioners active in these different applied research fields. These terms span various domains, technologies and concepts owing its inception to different histories. Commonality is however presented by management, usage and interpretation of data, as well as the extraction and manipulation of metadata (data about data).

It is plausible to speculate that the indiscriminate use of terminology surrounding the Semantic Web is confusing for any interested reader and it might require effort to enable a person to become familiar with the meanings and implications of these terms and their related technologies.

In this paper we chart the Semantic Web technology landscape and discuss the appropriate terms together with its history and relation to other important concepts in section II. In particular, we investigate the underlying architecture that will facilitate the founder vision of the Semantic Web as proposed by Berners-Lee [Berners-Lee et al., 2001; Berners-Lee, 2003]. As a result of this investigation we suggest a Semantic Web status model (adapted from Berners-Lee's proposed model) reflecting the current status.

In section II we provide the reader with a valuable starting point to assimilate Semantic Web terminology and associated concepts. This is followed by a status model that reflects the refinements imposed by the status quo of current technology in section III. We are of the opinion that our adaptation may provide valuable insight into the limitations of the technologies currently supporting this layered model. We conclude the paper with some comments on future research in section IV.

II. THE SEMANTIC WEB

The Semantic Web is an information space used by *machines* rather than *humans*. Instead of processing and manipulating Web information, a user would have a personal *agent* on his/her computer that would solve problems related to information overload, acquisition and discrepancy resolution [Decker et al., 2000a]. Once an agent has executed the first level of information management, a user would access or manipulate the results.

In 2001 Tim Berners-Lee [Berners-Lee et al., 2001] introduced a vision of the new Web referred to as the *Semantic Web* (presented in Figure 1 with added 'Layer' captions for reference purposes). Several Semantic Web authors refer to and even adopt this model to be the Semantic Web architecture of choice [Hendler, 2001; Fensel, 2002b; Patel-Schneider and Fensel, 2002; Thuraisingham, 2003; Antoniou and von Harmelen, 2004]. In 2003 Berners-Lee proposed a subsequent architecture [Berners-Lee, 2003] (presented in Figure 2 with corresponding added 'Layer' captions as used in Figure 1), but since this architecture has not been adopted to the same extent by Semantic Web authors, we include both the original as well as the adapted model in this discussion. In both these models of the Semantic Web, a higher level layer language use the syntax and semantics of its lower level layer.

In sections II-A through II-G we consider the different layers, as presented in Figures 1 and 2, together with its associated technologies. In the discussion of each layer we provide a brief description of the residing technologies as well as references to significant work on the relevant topic. In sections II-H and II-I we give a brief discussion of digital signatures and encryption, which serves as identification authentication as well as security mechanisms for layers three to six.

A. Layer 1

Layer 1 in both models comprises Unicode and URI (Uniform Resource Identifier) technologies. The function of these technologies is to provide a unique identification mechanism for upper language technologies.

A.1 Unicode:

Unicode aims to uniquely identify the characters in all the written languages by assigning a unique number to each character. The Unicode Standard (*Universal Character Set, Unicode/ISO10646*) specified by the Unicode Consortium is the universal character encoding standard used for representation of text for computer processing. This standard supports three encoding mechanisms, UTF-8, UTF-16 and UTF-32, allowing the same data to be encoded in a byte, word or double word format.

The emergence of the Unicode standard and the availability of supporting tools are amongst the most significant recent global software technology trends. Unicode replaces the use of legacy character sets and it allows data and text to be exchanged internationally between different systems.

Useful References - [Bettels and Bishop, 1993; Unicode, 2004]

A.2 URI:

A URI (Uniform Resource Identifier), defined as a compact string of characters that can be extended, is used to identify an abstract or physical resource. A *resource* is defined as an entity that has identity. The general URI specification of the IETF (Internet Engineering Task Force) is known as RFC2396. URLs (or Uniform Resource Locators) are a subset of URI that specifically identify resources by using their network 'location' rather than identifying the resource by name or by other attributes.

The Semantic Web would be impossible without global identification and hence the use of URIs. The future expansion of URIs into IRIs will ensure that a resource can be identified across language and character encoding boundaries and any discussion about 'meaning' has to uniquely identify the objects or resources of the discussion.

Useful References - [Berners-Lee et al., 1998; IETF, 2004; W3C, 2003, 2004e]

B. Layer 2

Layer 2 comprises of Namespaces, XML (Extensible Markup Language) as well as XML Schema technologies (Figure 1). In Figure 2, XML Schema was omitted, but it is clear from the discussion by Berners-Lee [2003] that it is included under the 'XML' caption as a technology of layer two. The function of these technologies is to provide a self-describing syntax for the upper layer language technologies. We acknowledge the existence of DTD (Document Type Declaration) as XML originally used DTD as a validation mechanism, however, DTD was replaced by XML Schema and a discussion of DTD is thus excluded.

B.1 Namespaces:

Namespaces (NS) provide a simple method for qualifying element and attribute names used in XML documents. Namespaces are identified by URI references. The W3C Namespace Recommendation [Bray et al., 1999] defines an XML namespace as a collection of names, identified by a URI reference [RFC2396], which are used in XML documents as element types and attribute names.

Useful References - [Bray et al., 1999, 2004; Srivastava, 2004; XML.com, 2005]

B.2 XML:

XML (Extensible Markup Language) specifies a standard for the exchange of data over networks, notably the Web. XML is considered to be both a *metalanguage* and a *markup language*. XML as metalanguage allows for the specification of the content of documents according to a predefined and specific structure. All documents conforming to this specification will have the same structure or represent data items in the specified structure. In addition, XML as markup language allows for the insertion of markup tags into text to define the logical structure of a document, or to add information regarding information contained in a document (metadata).

An XML document is a *text* document which in itself does not have any functionality. It is used only to describe data, information or metadata. Thus, XML is a means for defining common grammars to enable data exchange. It does not specify semantics, all parties *must agree* on the data model and document structure for XML data exchange to be successful. If an XML grammar is accepted as a standard for data exchange, any XML parser can parse the XML data and access the content if it is a valid XML document. It is however difficult to re-engineer the data model from any given XML document if the document type specification is not available.

Useful References - [Bray et al., 1999, 2004; Decker et al., 2000a; McKinnon and McKinnon, 2003; W3C, 2005; XML.com, 2005]

B.3 XML Schema:

An *XML schema* is an XML document defining the content and structure of one or more derived XML documents. Generally, a *schema* is a model for describing the structure and content of data. XML Schema is a content modelling language as well as an application of *XML* that applies only to XML-related languages and documents. In particular, an XML Schema describes a model for a whole class of XML documents. The model describes the possible arrangement of elements, their attributes and text that would be present in a schema-valid document.

Useful References - [Carey, 2003; Decker et al., 2000a; McKinnon and McKinnon, 2003; Mertz, 2001; Palmer, 2001; Srivastava, 2004; W3C, 2001a,b,c, 2004f, 2005]

C. Layer 3 / Layers 3a and 3b

RDF (Resource Descriptive Framework) and RDF Schema technologies reside on Layer 3 (refer to Figures 1 and 2). The function of these technologies is to provide a metadata description mechanism for the upper language technologies. With the positioning of RDF Schema, Layer 3b, above RDF M&S (Model and Syntax), Layer 3a, in Figure 2, Berners-Lee [2003] emphasises the importance of a vocabulary description mechanism on top of the RDF data model as part of the Semantic Web language stack.

C.1 RDF:

The purpose of RDF (Resource Description Framework) is to declare metadata that is machine-processable. RDF provides a mechanism to declare statements that describe resources by means of a basic data model. A *statement* describes an entity (resource) in terms of *properties*, which have *values*. Furthermore, an RDF statement is a subject, predicate, object triple [Decker et al., 2000b; Dong et al., 2004; W3C, 2004b]. The *subject* is the resource of the statement. The *predicate* is the property or characteristic of the subject specified by the statement (examples include creator, creation-date, or language), and the value of the property is the *object*.

In terms of the Semantic Web, the basic object-attribute-value data model is the only semantics prescribed in the RDF specifications. RDF has no other data-modelling commitments and specifies no reserved terms for further data modelling or no other mechanisms for declaring property names. For semantic interoperability RDF has significant advantages over XML primarily because of the data model used.

Useful References - [Beckett, 2004a,b; Brickley and Guha, 2004; Broekstra et al., 2002; DCMI, 2005; Decker et al., 2000a,b; Dong et al., 2004; Grant and Beckett, 2004; Hayes, 2004; Palmer, 2001; W3C, 1999, 2004b,c]

C.2 RDF Schema:

RDF Schema specifies extensions to RDF that are used to define the common vocabularies in RDF metadata statements. RDF itself provides the data model and does not prescribe any application-specific classes and properties. This is accomplished by RDF Schema. RDF Schema provides a predefined, basic type system for RDF models, thus extending RDF by assigning an externally specified semantics to specific resources. RDF Schema expressions are valid RDF expressions, and therefore RDF Schema is a semantic extension of RDF [Broekstra et al., 2002; Hayes, 2004]. Software that can interpret RDF can also be used to interpret an RDF Schema implementation; although it will not attach the intended meaning to the built-in schema definitions.

The RDF vocabulary description strategy contained in RDF Schema acknowledges that there are many techniques that enable *description of meaning* of classes and properties. To extend the description of meaning, ontology languages (such as DAML+OIL, OIL and OWL), inference rule languages and other formalisms are used.

Useful References - [Beckett, 2004a,b; Berners-Lee et al., 2001; Brickley and Guha, 2004; Broekstra et al., 2002, 2001; Decker et al., 2000a,b; Dong et al., 2004; Grant and Beckett, 2004; Hayes, 2004; Palmer, 2001; W3C, 1999, 2004b,c]

D. Layer 4 / Layers 4a and 4b

In Figure 1 'Ontology vocabulary' is depicted on Layer 4, whilst this layer is separated as 'Ontology', Layer 4a, and 'Rules', Layer 4b, in Figure 2. In Figure 2, Berners-Lee [2003] acknowledges that an ontology is a *knowledge representation language* capturing the syntax (ontology) as well as the semantics (rules) of a specific domain [McGuinness et al., 2002; McGuinness and van Harmelen, 2004]. Currently, OWL is the W3C technology representing an 'Ontology vocabulary' or 'Ontology' associated with this layer, whilst W3C research efforts aim to establish the technologies required for the implementation of the 'Rules' to be contained in this layer [Horrocks et al., 2005].

It is noted that the terminology on this layer differs from the three preceding layers, because the *functionality* rather than the *technology* is mentioned.

D.1 Ontology Vocabulary / Ontology:

An ontology specifies a machine readable vocabulary in computer systems technology descriptions. Ontologies assist in creating a common understanding for communication between people and computer applications. Generally it is defined as a shared, formal, explicit specification or conceptualisation of a particular domain [Broekstra et al., 2001; Decker et al., 2000a; Heflin and Hendler, 2001].

It is envisioned that ontologies will play a crucial role in knowledge processing, sharing, and reuse between Web applications. On the Semantic Web, ontologies may be used in applications required to search across, or merge information from diverse communities.

In early 2001 the W3C initiated a Web-Ontology Working Group (WebOnt) in order to consolidate existing Web ontology efforts (notably OIL and DAML+OIL [DAML, 2005; On-To-Knowledge, 2005]) into a Web Ontology Language. OWL extends RDF Schema in order to express complex relationships between different classes specified in RDF Schema, as well as to enhance the specification of constraints applicable to classes and properties. OWL specifies three sub-languages of increasing expressiveness [Bechhofer et al., 2004; McGuinness and van Harmelen, 2004; Smith et al., 2004]. These language are OWL Lite, OWL DL and OWL Full. Ontology designers should select the most appropriate version.

The OWL language provides a specific subset in the form of OWL DL to support existing DL (Description Logics) and to provide a language subset that possesses the computational properties required for reasoning systems [Bechhofer et al., 2004; McGuinness and van Harmelen, 2004; Smith et al., 2004]. DL is a set of knowledge representation formalisms with semantic characterisation based on standard first-order logics. DL offers a formal foundation for frame-based systems [Fensel, 2002b; McGuinness et al., 2002], where meaning is provided by interpretations that define the formal semantics of the logic [Grau, 2004].

Useful References - [Antoniou and von Harmelen, 2004; Bechhofer et al., 2004; Broekstra et al., 2002, 2001; Bussler et al., 2002; Butler, 2002; Carroll and Roo, 2004; DAML, 2005; Decker et al., 2000b; Fensel, 2002a,b; NCI, 2005; Gruber, 1993; Hayes, 2004; Heflin, 2004; Heflin and Hendler, 2001; McGuinness et al., 2002; McGuinness and van Harmelen, 2004; Niles and Pease, 2001; On-To-Knowledge, 2005; Palmer, 2001; Smith et al., 2004; Staab et al., 2000; Staab and Maedche, 2000; Van Harmelen et al., 2001; W3C, 2004a].

E. Layer 5

The Semantic Web models (Figures 1 and 2) depicts 'Logic' or 'Logic framework' residing *above* the ontology layer.

E.1 Logic / Logic framework:

Logic is regarded as the foundation of knowledge representation languages, and it is required to provide the highly expressive language constructs in which knowledge can be captured in a transparent way. A logic framework provides a well-established formal semantics which assigns unambiguous meaning to logical statements. Without a logic framework, inferencing on the Semantic Web will not be possible.

Useful References - [Antoniou and von Harmelen, 2004; Baader et al., 2003; Bechhofer et al., 2004; Berners-Lee, 1998a; Decker et al., 2000a; Fensel, 2000, 2002b; Grau, 2004; McGuinness et al., 2002; Palmer, 2001; Patel-Schneider and Fensel, 2002].

F. Layer 6

In the Semantic Web models of Berners-Lee (Figures 1 and 2), 'Proof' resides on Layer 6.

F.1 Proof:

'Proof' as concept exists within the theorem proving domain, for instance as applied in artificial intelligence [anon., 2005]. To support Semantic Web proof scenarios, *proof languages* were developed. A proof language determines the validity of a specific statement. An instance thereof generally consists of a list of inference items used to derive the information in question, as well as the associated trust information of each item [Appel and Felten, 1999; Palmer, 2001].

A Semantic Web will probably not require proof generation and in general proof validation will be adequate. The search for and generation of a proof for an arbitrary question, is typically an intractable process for many real world problems, and the Semantic Web does not require this to be solved. For perceived Semantic Web applications construction of a proof is performed according to constrained rules, and only the validation thereof is required from other parties. For example, when a user is granted access to a web site, an accompanying document explains to the web server why they should be granted access. Such proof for example, could be a chain of assertions and reasoning rules with pointers to all supporting material [Berners-Lee, 1998b].

Useful References - [anon., 2005; Appel and Felten, 1999; Palmer, 2001; Patel-Schneider and Fensel, 2002]

G. Layer 7

'Trust' resides on Layer 7 in Figures 1 and 2.

G.1 Trust:

Semantic Web interaction requires different collaborators to communicate, implying that they have to determine how to trust one another, as well as how to establish the trust levels of acquired information. When dealing with user interactions on the Web, McKnight et al. [2000] define the term *trust* as the belief that another entity is benevolent, competent, honest, or predictable in a given situation. Trust also includes the participants' willingness to depend on one another in a specific interaction. Furthermore, user trust of Semantic Web information is determined by the *source* of the information, in particular its authenticity and trustworthiness [Thuraisingham, 2003].

Within the Semantic Web the concepts trust and proof are dependent on the interaction context. However, an all-encompassing definition of context is problematic. An appropriate meaning of context is therefore explicated by means of the following example. A user on the Semantic

Web receives data from a friend regarding the best music performances. The data can be trusted as it originates from a *known* (implying verified) friend, whose musical interests are familiar. It is thus possible to *use* the data because the user either shares or disagrees with the musical tastes of the friend. Within the domain of the Semantic Web, *context* therefore assists applications or users regarding the trustworthiness and usefulness of data [Bhargava et al., 2004].

Useful References - [anon., 2005; Appel and Felten, 1999; McKnight et al., 2000; Palmer, 2001; Patel-Schneider and Fensel, 2002; Park, 2003]

H. Digital Signatures

In the Semantic Web models of Berners-Lee 'Digital Signature' is associated with layers three to six (ref Figures 1 and 2). The Digital Signature Standard (DSS) is a cryptographic standard or a particular application of public key cryptography promulgated by NIST (National Institute of Standards and Technology) [Center, 2005]. A digital signature is an electronic signature that can be used to authenticate identity. Digital signatures are easily transportable, cannot be imitated, and can be automatically time-stamped. A digital signature can be used with any kind of message, whether it is encrypted or not.

XMLDSig (XML Digital Signatures), also called XML Signatures, is an IETF/W3C joint standard that specifies how to digitally sign and verify a signature of a XML data object [Bartel et al., 2002]. XMLDSig enables digital signatures on arbitrary digital content (XML or non-XML) [Lee et al., 2005]. XML Signatures are digital signatures designed for use in XML transactions [Simon et al., 2001].

For the Semantic Web a digital signature is a mechanism used to unambiguously verify an identity such as the author of a document [Palmer, 2001]. The implementation of digital signatures on the Semantic Web could result in a system which can express and reason about relationships across the whole range of public-key based security and trust systems. It is foreseen that XMLDSig will be used in many phases in Semantic knowledge management systems, such as the authenticity verification for retrieved/updated knowledge and involved intermediaries, among others.

Useful References - [Berners-Lee, 1998b; Biddle, 1996; Center, 2005; Hess et al., 2004; McKnight et al., 2000; Palmer, 2001; Simon et al., 2001; Lee et al., 2005; Horrocks et al., 2005; Park, 2003; Bartel et al., 2002]

I. Encryption

The original model of Berners-Lee does not depict 'Encryption'. It was however added in the adapted model of Figure 2 where it is associated with layers three to six, along with 'Signature'.

Encryption is an effective way to achieve data security. XMLEnc (XML Encryption) is a W3C standard that specifies how to encrypt/decrypt an XML-formatted data object [Imamura et al., 2002]. XMLEnc supports end-to-end (as opposed to point-to-point) encryption of an XML object, which can be the whole or a part of an XML document. The document can be transmitted in XML or non-XML syntax.

On the Semantic Web it is foreseen that encryption could be used in knowledge storage, internal/external knowledge transfer as well as authentication [Lee et al., 2005].

Useful References - [Berners-Lee, 2003; Horrocks et al., 2005; Lee et al., 2005; Park, 2003; Imamura et al., 2002]

III. Status of Semantic Web Technologies

In this section we discuss the status of the technologies covered in section II. In particular we adapt Figure 2 to reflect the current status of the technologies, which is graphically depicted in Figure 3. We discuss the adaptations to this model in sections III-A to III-E.

A. 'OWL' replaces 'Ontology' on Layers 4 and 4a

As mentioned in section II-D, OWL is the W3C technology representing 'Ontology' on Layer 4a in Figure 2 in particular. This provides motivation for one of the changes reflected in Figure 3, where we replace 'Ontology' with OWL in order to be consistent with the other *W3C technologies* represented in the four bottom layers[1].

B. 'Digital Signatures' moved to Layer 1

As previously discussed in section II-H, a digital signature is a mechanism used to unambiguously verify an identity [Palmer, 2001] such as the author of a document. The concept, digital signature, and the use thereof is prevalent in e-commerce and trust negotiations application domains [Hess et al., 2004; McKnight et al., 2000].

However, within the context of the Semantic Web the notions of digital signatures, encryption, as well as it's associated roles are still vague [Patel-Schneider and Fensel, 2002; Lee et al., 2005]. Since the function of Layer 1 technologies are to provide a unique identification mechanism for the upper layer technologies, it provides motivation for 'digital signatures' to be incorporated as a technology into Layer 1. In addition, XML as an immediate upper layer technology, provides a mechanism by means of *XML signatures* to use digital signatures in XML transactions [Simon et al., 2001]. This provides further motivation for digital signatures to be moved to Layer 1.

C. Classification of the bottom four layers as 'Established Technologies'

The technologies of layers 1 through 4a were adopted as either W3C Specifications or W3C Recommendations, implying that these technologies are *established* technologies.

[1] We acknowledge Parnas' *functional modularisation of software* [Parnas, 1972], also dubbed *the separation of concerns* theory by Dijkstra [Dijkstra, 1974], which emphasises functionality rather than technology in a structured view. However, for this paper we follow the approach of Berners-Lee [Berners-Lee, 2003].

This provides the motivation for classifying the bottom four layers as established technologies:

- The Unicode Standard was adopted by industry leaders and is required by standards such as XML, Java, JavaScript, LDAP and CORBA 3.0. Unicode is specified as part of any emerging W3C standard. In particular, any application, parser or browser that adhere to the W3C standards therefore has to adhere to the Unicode Standard.
- Regarding URI, RFC2396 currently allows for a subset of ASCII only, comprising about 60 characters. In contrast IRIs (Internationalised Resource Identifiers) are being developed by the Internationalisation Activity of the W3C. IRIs are specified as a sequence of characters from the Universal Character Set (Unicode/ISO10646) which implies that an IRI can include characters from any of the written languages in the world.
- The current W3C Recommendation specifies the use of URLs to identify a *namespace* because URLs indicate domain names that should be unique throughout the Internet. However, it is envisioned that IRIs will eventually replace URLs as namespace identifiers. Namespaces are required because of naming conflicts that arise when different authors create grammars for metadata, using for instance, elements with the same name. XML namespaces avoid naming conflicts when using and re-using multiple vocabularies because an element is qualified against a namespace, thus making it a *unique* element. XML Schema is used to create a vocabulary for an XML instance, as a result XML Schema uses namespaces extensively.
- The first *XML* specification was accepted by the W3C on Feb 10, 1998. The third edition was accepted as a W3C recommendation on February 4, 2004 [Bray et al., 2004]. Numerous XML-based languages and applications were developed by organisations that share high volumes of information.
- Regarding XML Schema, the XML 1.0 Recommendation was endorsed by the W3C in 1998. In response to the shortcomings of DTD, the W3C mandated the XML Schema Working Group to develop an XML Schema Language. On May 2, 2001, the W3C endorsed the language specifications viz, *XML Schema Part 1: Structures* and *XML Schema Part 2: Datatypes*.
- RDF was released as a W3C Recommendation in February 2004 [W3C, 2004d]. In doing so, W3C endeavoured to establish a more practical technology platform as opposed to only a research project environment. This technology platform enables flexible access to structured Web data. The RDF Specification set of the W3C provides an extensive overview of RDF. RDF is used to define metadata vocabularies. One of the most quoted examples is the Dublin Core Metadata Initiative DCMI [2005].
- RDF Schema is a W3C Recommendation [W3C, 2004d] that provides a technology that will assist in the description of *meaning* using underlying basic data structures by allowing the specification of domain vocabularies. RDF Schema does not specify a vocabulary of application-specific classes and their associated properties, but it describes the mechanisms necessary to specify such a vocabulary.
- In February 2004 the OWL specification was endorsed as a W3C Recommendation. The current W3C OWL document set consists of six documents. OWL is intended as a Web Ontology language, implying that it is specifically designed to be compatible with the architecture of the World Wide Web and Semantic Web [McGuinness and van Harmelen, 2004]. OWL has already permeated the ontology engineering community and a number of existing OWL ontologies are available on the Web [DAML, 2005; NCI, 2005].

D. *Classification of the top layers as 'Emerging Functionalities'*

We distinguish between *established technologies* and *emerging functionalities*. The *emerging* functionalities reside in the top layers, including layer 4b that contains 'Rules'. We categorise established technologies as those adopted as either W3C Recommendations or Specifications in layers 1 through 4a. In contrast, emerging functionalities are the functionalities that are required to realise the Semantic Web but that are presently predominantly research efforts.

Reasoning and inferencing are two of the driving principles of the Semantic Web. New data is derived from existing data by means of *inferencing*. *Reasoning* deduct meaning and make decisions based on the acquired data. An expressive, logical language that supports reasoning is required in order to implement or use the Semantic Web. Presently, this is one of the active research focus areas of the Semantic Web domain [Fensel, 2000; Grau, 2004; Patel-Schneider and Fensel, 2002]. RDF together with its data model, as well as OWL, support the notion of machine-processable information on the Web [Decker et al., 2000b]. However, neither RDF nor OWL (residing on the data layer) has the necessary expressive power to enable the inferencing and reasoning required for the Semantic Web.

Trust and proof within the Semantic Web application domain are emergent research concepts [Patel-Schneider and Fensel, 2002]. It is, however, a crucial aspect of the eventual Semantic Web realisation due to inherent contradictions and duplications in ontology definitions [Palmer, 2001]. Applications on the Semantic Web presently depend upon *context* to manage trust and proof. It is an accepted requirement that proof checking mechanisms enhanced by digital signatures will be integrated into the eventual Semantic Web interaction and collaboration activities.

E. *Classification of the bottom four layers as the data layer of the Semantic Web*

Ontologies that build upon the lower layers depicted in Figure 2 and that capture the meaning and metadata of information are the realisation of the information space required by the Semantic Web, i.e. the existing technologies enable an information space usable by machines. We propose that the established technologies discussed in sections II-A to II-D but excluding 'Rules' (depicted in layers one to four a in Figure 3), constitute the *data layer* of the Semantic Web. It is thus not unreasonable to state that *some* maturity can already be associated with the data layer of the Semantic Web.

However, an application is required to manipulate or use a data layer. Thus, in the same vein, a Semantic Web application is required to manipulate any ontology. In the vision of Berners-Lee [Berners-Lee et al., 2001], Semantic Web data (ontologies) are used by *agents*. However, presently several unresolved issues cloud the realisation of this vision. This is supported by Patel-Schneider and Fensel who state that *"...there are layering decisions to be made whenever a new representation formalism has to be compatible with an existing representation formalism. This will occur at other points of the semantic web tower, has occurred in the past, and will undoubtedly occur in the future."* [Patel-Schneider and Fensel, 2002]. This serves as a further motivation for the term 'emerging' associated with the top three layers of the model in Figure 3.

IV. Conclusion

At present, the information overload experienced on the Web necessitates the urgent introduction of an automated information management functionality. The realisation of which constitutes the envisioned Semantic Web. The Semantic Web with its associated technologies are permeating various fields. It is therefore necessary for readers to acquaint themselves with the Semantic Web and its technologies and concepts. This may well be a daunting task as these technologies are diverse and span interrelated fields. In this positioning paper we thus provide the novice reader with among other things, a starting point to assimilate Semantic Web terminology and associated concepts. Our adapted architectural model reflects various refinements imposed by the status quo of current technology, and imparts some insight into the limitations of the technologies currently supporting this layered architecture.

-

References

ANON. **Proof General**. 2005. URL `http://proofgeneral.inf.ed.ac.uk/`.

ANTONIOU G. AND VON HARMELEN F. *A Semantic Web Primer*. The MIT Press, 2004.

APPEL A.W. AND FELTEN E.W. **Proof-carrying authentication**. In *CCS '99: Proceedings of the 6th ACM conference on Computer and communications security*, pages 52–62. ACM Press, New York, NY, USA, 1999. ISBN 1-58113-148-8.

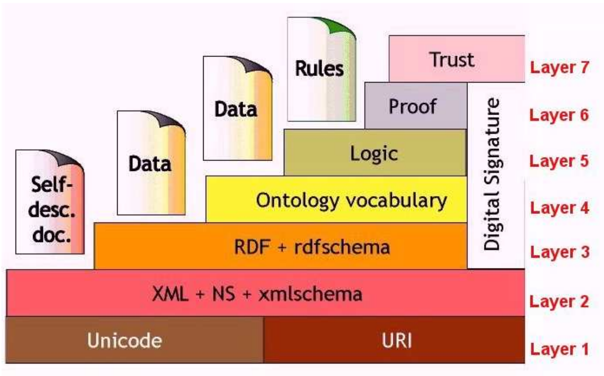

Fig. 1. The Semantic Web Architecture (Tim Berners-Lee et al. [2001])

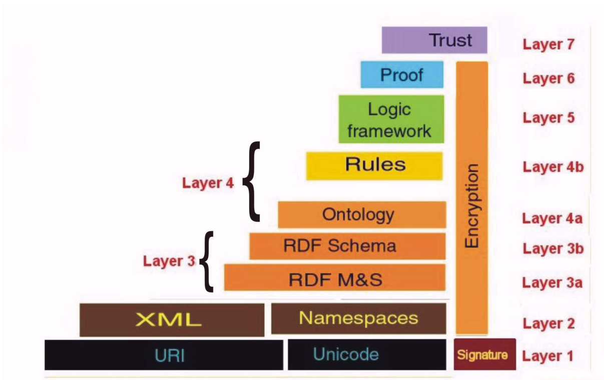

Fig. 2. The Adapted Semantic Web Architecture (Tim Berners-Lee [2003])

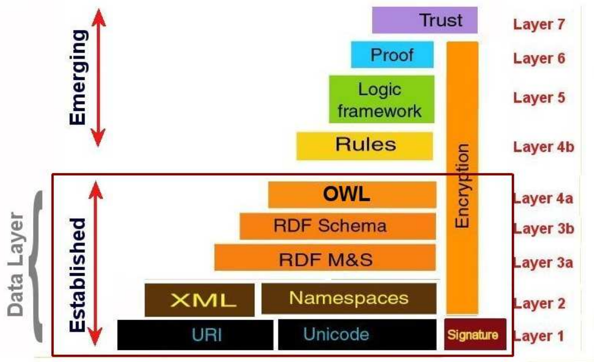

Fig. 3. The modified Semantic Web architecture

BAADER F., CALVANESE D., MCGUINNESS D.L., NARDI D., AND PATEL-SCHNEIDER P.F. *The Description Logic Handbook, Theory, implementations and application*. Cambridge University Press, 2003.

BARTEL M., BOYER J., FOX B., LAMACCHIA B., AND SIMON E. **XML-Signature Syntax and Processing**. W3C Website, 2002. URL `http://www.w3.org/TR/xmldsig-core/`.

BECHHOFER S., VAN HARMELEN F., ET AL. **OWL Web Ontology Language Reference**. W3C Website, 2004. URL `http://www.w3.org/TR/2004/REC-owl-ref-20040210/`.

BECKETT D. **Dave Beckett's Resource Description Framework (RDF) Resource Guide**. Web Document, 2004a. URL `http://www.ilrt.bris.ac.uk/discovery/rdf/resources/`.

BECKETT D. **RDF/XML Syntax Specification (Revised)**. W3C Website, 2004b. URL `http://www.w3.org/TR/2004/REC-rdf-syntax-grammar-20040210/`.

BERNERS-LEE T. **The Semantic Web as a language of logic**. W3C Website, 1998a. URL `http://www.w3.org/DesignIssues/Logic.html#Crawf90`.

BERNERS-LEE T. **The Semantic Web Road Map**. W3C Website, 1998b. URL `http://www.w3.org/DesignIssues/Semantic.html`.

BERNERS-LEE T. **Standards, Semantics and Survival**. Technical Report, Software and Information Industry Association (SIIA), 2003.

BERNERS-LEE T., FIELDING R., IRVINE U.C., AND MASINTER L. **IETF RFC2396 - Uniform Resource Identifiers (URI): Generic Syntax**. Web Document, 1998. URL `http://www.ietf.org/rfc/rfc2396.txt`.

BERNERS-LEE T., HENDLER J., AND LASSILA O. **The Semantic Web**. *The Scientific American*, May 17, 2001, 2001.

BETTELS J. AND BISHOP F. **Unicode: A Universal Character Code**. *Digital Technical Journal*, 5 no. 3(3), 1993.

BHARGAVA B., LILIEN L., ET AL. **The Pudding of Trust**. *IEEE INTELLIGENT SYSTEMS*, pages 74–88, 2004.

BIDDLE C.B. **Legislating Market Winners: Digital Signature Laws and the Electronic Commerce Marketplace**. *World Wide Web Journal (W3J)*, 1, 1996.

BRAY T., HOLLANDER D., AND LAYMAN A. **Namespaces in XML Recommendation - REC-xml-names-19990114**. W3C Website, 1999. URL `http://www.w3.org/TR/REC-xml-names/`.

BRAY T., PAOLI J., SPERBERG-MCQUEEN C.M., MALER E., AND YERGEAU F. **Extensible Markup Language (XML) 1.0 (Third Edition)**. W3C Website, 2004. URL `http://www.w3.org/TR/REC-xml/`.

BRICKLEY D. AND GUHA R. **RDF Vocabulary Description Language 1.0: RDF Schema**. W3C Website, 2004. URL `http://www.w3.org/TR/2004/REC-rdf-schema-20040210/`.

BROEKSTRA J., KAMPMAN A., AND VAN HARMELEN F. **Sesame: A Generic Architecture for Storing and Querying RDF and RDF Schema**. In *Proceedings of The Semantic Web - ISWC 2002: First International Semantic Web Conference, Sardinia, Italy*, volume Volume 2342 / 2002, page 54. 2002.

BROEKSTRA J., KLEIN M., ET AL. **Enabling Knowledge Representation on the Web by Extending RDF Schema**. In *Proceedings of the 10th International World Wide Web Conference (WWW10), Hong Kong*, volume ACM 1-58113-348-0/01/0005, page 467. 2001.

BUSSLER C., FENSEL D., AND MAEDCHE A. **A conceptual architecture for semantic web enabled web services**. *ACM SIGMOD, SPECIAL ISSUE: Special section on semantic web and data management*, 31, issue 4:24 – 29, 2002.

BUTLER M.H. **Barriers to real world adoption of semantic web technologies**. *HP Labs Technical Report*, HPL-2002-333, 2002.

CAREY P. *New Perspectives on XML - Introductory*. ISBN: 0-619-10187-3. Thomson Course Technology, 2003.

CARROLL J.J. AND ROO J.D. **OWL Web Ontology Language Test Cases**. W3C Website, 2004. URL `http://www.w3.org/TR/2004/REC-owl-test-20040210/`.

CENTER E.P.I. **Digital Signatures**. EPIC Website, 2005. URL `http://www.epic.org/crypto/dss/`.

DAML. **DARPA Agent Markup Language**. 2005. URL `http://www.daml.org/`.

DCMI. **Dublin Core Metadata Initiative (DCMI)**. 2005. URL `http://dublincore.org/`.

DECKER S., MELNIK S., ET AL. **The Semantic Web: The Roles of XML and RDF**. *IEEE Internet Computing*, 4:63–74, 2000a.

DECKER S., MITRA P., AND MELNIK S. **Framework for the Semantic Web: An RDF Tutorial**. *IEEE Internet Computing*, pages 68–73, 2000b.

DIJKSTRA E.W. **On the role of scientific thought**. Edsger W. Dijkstra Archive website, 1974. URL `http://www.cs.utexas.edu/users/EWD/transcriptions/EWD04xx/EWD447.html`.

DONG J.S., LEE C.H., LEE H.B., LI Y.F., AND WANG H. **Semantic web foundations: A combined approach to checking web ontologies**. In *Proceedings of the 13th International Conference on World Wide Web*, ISBN:1-58113-844-X, pages 714 – 722. ACM Press New York, NY, USA, 2004.

FENSEL D. **The Semantic Web and its languages**. *IEEE Intelligent Systems*, pages 67–73, 2000.

FENSEL D. **Language Standardization for the Semantic Web: The Long Way from OIL to OWL**. In *Distributed Communities on the Web: 4th International Workshop, DCW 2002, Sydney, Australia*, volume 2468 / 2002, pages 215–227. 2002a.

FENSEL D. **Ontology-Based Knowledge Management**. *IEEE Computer*, 35 no 11:56–59, 2002b.

GRANT J. AND BECKETT D. **RDF Test Cases**. W3C Website, 2004. URL `http://www.w3.org/TR/2004/REC-rdf-testcases-20040210/`.

GRAU B.C. **A Possible Simplification of the Semantic Web Architecture**. In *WWW '04: Proceedings of the 13th international conference on World Wide Web*, ISBN:1-

58113-844-X, pages 704–713. ACM Press, New York, NY, USA, 2004. ISBN 1-58113-844-X.

Gruber T.R. **Toward Principles for the Design of Ontologies Used for Knowledge Sharing**. In *International Workshop on Formal Ontology, Padova, Italy.* 1993.

Hayes P. **RDF Semantics**. W3C Website, 2004. URL `http://www.w3.org/TR/2004/REC-rdf-mt-20040210/`.

Heflin J. **OWL Web Ontology Language, Use Cases and Requirements**. W3C Website, 2004. URL `http://www.w3.org/TR/2004/REC-webont-req-20040210/`.

Heflin J. and Hendler J. **A Portrait of the Semantic Web in Action**. *IEEE Intelligent Systems*, 16:54–59, 2001.

Hendler J. **Agents and the Semantic Web**. *IEEE Intelligent Systems*, 16:30–37, 2001.

Hess A., Holt J., Jacobson J., and Seamons K.E. **Content-triggered trust negotiation**. *ACM Transactions on Information and System Security (TISSEC)*, Volume 7 Issue 3, 2004.

Horrocks I., Parsia B., Patel-Schneider P., and Hendler J. **Semantic Web Architecture: Stack or Two Towers?** *Lecture Notes in Computer Science: Principles and Practice of Semantic Web Reasoning: Third International Workshop*, 3703 / 2005, 2005. ISSN 0302-9743. doi: \bibinfo{doi}{10.1007/11552222}.

IETF. **Internet Engineering Task Force**. Web Documen, 2004. URL `http://www.ietf.org/`.

Imamura T., Dillaway B., and Simon E. **XML Encryption Syntax and Processing**. W3C Website, 2002. URL `http://www.w3.org/TR/xmlenc-core/`.

Lee J., Upadhyaya S.J., Rao H.R., and Sharman R. **The semantic e-business vision: Secure knowledge management and the semantic web**. *Communications of the ACM*, 48, 2005.

McGuinness D.L., Fikes R., Hendler J., and Stein L.A. **DAML+OIL: An Ontology Language for the Semantic Web**. *IEEE Intelligent Systems*, pages 72–80, 2002.

McGuinness D.L. and van Harmelen F. **OWL Web Ontology Language Overview**. W3C Website, 2004. URL `http://www.w3.org/TR/2004/REC-owl-features-20040210/`.

McKinnon A. and McKinnon L. *XML*. ISBN: 0-619-03514-5. Web Warrior Series, Thomson Course Technology, 2003.

McKnight D.H., Choudhury V., and Kacmar C. **Trust in e-commerce vendors: a two-stage model**. In *Proceedings of the twenty first international conference on Information systems (ICIS '00), Brisbane, Queensland, Australia*, ISBN:ICIS2000-X, pages 532–536. Association for Information Systems, Atlanta, GA, USA, 2000.

Mertz D. **XML Matters: Comparing W3C XML Schemas and Document Type Definitions (DTDs)**. IBM Developer Works, 2001. URL `http://www-106.ibm.com/developerworks/library/x-matters7.html`.

NCI. **US National Cancer Institute's Center for Bioinformatics (NCI) Thesaurus**. 2005. URL `http://www.itpapers.com/abstract.aspx?scid=442&docid=120560`.

Niles I. and Pease A. **Towards a standard upper ontology**. In *Proceedings of the international conference on Formal Ontology in Information Systems*, volume 2001, pages 2–9. 2001.

On-To-Knowledge. **On-to-Knowledge Website**. 2005. URL `http://www.ontoknowledge.org/`.

Palmer S.B. **The Semantic Web: An Introduction**. W3C Website, 2001. URL `http://infomesh.net/2001/swintro/`.

Park J.S. **Towards Secure Collaboration on the Semantic Web**. *ACM SIGCAS Computers and Society*, 33:1, 2003. doi: \bibinfo{doi}{http://0-doi.acm.org.oasis.unisa.ac.za:80/10.1145/1008773.1008774}.

Parnas D. **On the Criteria To Be Used in Decomposing Systems into Modules**. *Communications of the ACM*, 15:1053 – 1058, 1972.

Patel-Schneider P.F. and Fensel D. **Layering the Semantic Web: Problems and Directions**. In *Proceedings of The Semantic Web - ISWC 2002: First International Semantic Web Conference, Sardinia, Italy*, volume 2342 / 2002, page 16. Springer-Verlag GmbH, 2002.

Simon E., Madsen P., and Adams C. **An Introduction to XML Digital Signatures**. Website: XML.com, 2001. URL `http://www.xml.com/pub/a/2001/08/08/xmldsig.html`.

Smith M.K., Welty C., and McGuinness D.L. **OWL Web Ontology Language Guide**. W3C Website, 2004. URL `http://www.w3.org/TR/2004/REC-owl-guide-20040210/`.

Srivastava R. **XML Schema: Understanding Namespaces**. Technical Report, Oracle, 2004. URL `http://www.oracle.com/technology/pub/articles/srivastava_namespaces.html`.

Staab S., Erdmann M., Maedche A., and Decker S. **An Extensible Approach for Modeling Ontologies in RDF(S)**. In *ECDL 2000Workshop on the Semantic Web*. 2000.

Staab S. and Maedche A. **Ontology Engineering beyond the Modeling of Concepts and Relations**. In *Proceedings of the ECAI2000 Workshop*. 2000.

Thuraisingham B. **Security Issues for the Semantic Web**. In *Proceedings of the 27th Annual International Computer Software and Applications Conference*, page 632. IEEE, 2003.

Unicode. **The Unicode Consortium**. Unicode Website, 2004. URL `http://www.unicode.org/`.

van Harmelen F., Patel-Schneider P.F., and Horrocks I. **Annotated DAML+OIL (March 2001) Ontology Markup**. DAML Website, 2001. URL `http://`

www.daml.org/2001/03/daml+oil-walkthru.

W3C. **Resource Description Framework (RDF) Model and Syntax Specification**. W3C Website, 1999. URL http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/.

W3C. **XML Schema Part 0: Primer - W3C Recommendation**. W3C Website, 2001a. URL http://www.w3.org/TR/xmlschema-0/.

W3C. **XML Schema Part 1: Structures - W3C Recommendation**. W3C Website, 2001b. URL http://www.w3.org/TR/2001/REC-xmlschema-1-20010502/.

W3C. **XML Schema Part 2: Datatypes - W3C Recommendation**. W3C Website, 2001c. URL http://www.w3.org/TR/2001/REC-xmlschema-2-20010502/.

W3C. **World Wide Web Consortium (W3C) - Uniform Resource Identifier (URI) Activity Statement**. W3C Website, 2003. URL http://www.w3.org/International/O-URL-and-ident.html.

W3C. **Information on OWL**. Website, 2004a. URL http://www.w3c.oeg/2001/sw/WebOnt/.

W3C. **RDF Primer W3C Recommendation**. W3C Website, 2004b. URL http://www.w3.org/TR/rdf-primer/.

W3C. **Resource Description Framework (RDF): Concepts and Abstract Syntax**. W3C Website, 2004c. URL http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/.

W3C. **World Wide Web Consortium Issues RDF and OWL Recommendations**. W3C Website Press Release, 2004d. URL http://www.w3.org/2004/01/sws-pressrelease.

W3C. **World Wide Web Consortium (W3C) - Internationalized Resource Identifiers (IRIs)**. W3C Website, 2004e. URL http://www.w3c.org/Addressing/Activity.

W3C. **XML Querying**. Website, 2004f. URL http://www.w3c.org/XML/Query.html.

W3C. **W3C Extensible Markup Language (XML)**. Website, 2005. URL http://www.w3.org/XML/.

XML.COM. **O'Reilly XML.com**. Website, 2005. URL http://www.xml.com/.