



Semantic Web: XML and RDF Roles

FADI AL KALANI

Shaqra University, KSA.

(Received: August 16, 2015; Accepted: November 28, 2015)

ABSTRACT

This research paper presents an overview of the main concepts of the semantic web, and is aiming to present some ideas of developing the concept of semantic web using XML and RDF. In particular, the research reveals the positive impact of using RDF instead of XML through the concept of "Semantic Web", with the knowledge that many of the researches demonstrated the preference of RDF for the many reasons including simplicity, abstract syntax, and providing a data model. In particular, our main goal is to define the appropriate elements to develop a semantic web using both XML and RDF. Our approach is achieved by developing a simulated web search engine to describe and emphasize the positive role of using RDF rather than XML in web search.

Key words: XML, RDF, W3C.

INTRODUCTION

Semantic web is the future development of World Wide Web (WWW). The information stored by WWW is intended for human use. The contents in the WebPages are left to the user to read and understand also to make connections between information stored in the WebPages. Semantic web allows the user computer to draw the required connections between information in the web pages, analyze it and combine the relevant information for the user on a global scale.

This paper is organized as follows: In section 2 we give an overview of semantic web definition and basic concepts, In section 3 the structure of semantic web is explained briefly, section 4 represent the role of ontologies in the

architecture of the Semantic Web. In sections 5 and 6 briefly summarize some of the available standards for semantic interoperability, namely XML and RDF. In section 7 exploit XML for semantic interoperability and RDF showing some features that differs each. In section 8 some challenges and proposed solution is discussed and finally, section 9 summarizes and concludes and future works this paper.

Semantic web

Understanding and interpreting natural language texts are an extensive computational task which requires artificial intelligence technology. Computers have no reliable way to process the semantics. Computers as we know it are not capable of reading and understanding between lines, Interpreting irony or lyrical expressions may neither an easy nor accurate by machines. These

problems and other goals are the real motivation of such technology “semantic web” which is using the technique of storing extra data within the WebPages; meta-data. Meta-data is not to be displayed on the webpage but rather can be used by the web browser to make connections to other WebPages.

Information linked up in such a way to be process-able by machines in a mesh network defines Semantic Web where it develops languages for articulating information in a machine process-able form. [2] The idea states that the information will no longer only be intended for human readers, but also for processing by machines, enabling intelligent information services, personalized Web-sites, and semantically empowered search-engines. [1]

Today’s search engines are not allowing the search to be precise and efficient which is required by the users that grows day after another referred to the structure and size of current Web that is not allowing to make search more precise and efficient. On the other hand web contains now a huge number of documents and this number has a strong tendency to double each one or two years. The structure of documents and Web itself, probably, can be changed in “a better – machine process-able way” [2].

Creating a unified universal medium for data to be shared, exchanged, combined and processed by machines as well as by people is one of the main goals of semantic web. The Semantic Web is intended to efficiently connect personal information management, enterprise application integration, and the global sharing of commercial, scientific and cultural data.[7]

The two known kinds of search techniques that exist are Full Text Search (FTS) that processes natural languages based on queries to retrieve information like Google. The second technique is the Unambiguous Search (US) which is based on data whose semantic is already defined in the system. The Semantic Web requires interoperability on the semantic level as well as semantic interoperability requires standards not only for the syntactic form of documents, but also for

the semantic content. Proposals aiming at semantic interoperability are the results of recent the World Wide Web Consortium (W3C) standardization efforts, particularly Extensible Markup Language (XML)/XML Schema and Resource Description Framework (RDF)/RDF Schema (RDFS).[1] The specification as recommended by (W3C) is the Resource Description Framework (RDF).

W3C standard document format for writing and exchanging information on the Web emphasis that XML is mostly concerned with syntax that does not make sense without semantics, and many recent activities aim at adding more semantic capabilities to XML [4]. However, RDF is mostly concerned about semantics which is not very useful in a computer system without syntax, and many recent activities aim at providing a syntactic grounding for RDF [1].

The language that is used to display the graphics and text (data in the form of audio, video, text and image documents) in a WebPages HTML. The semantic web by introducing XML (extensible markup language), RDF (resource description framework), RDFS (RDF Schema) and OWL (web ontology language) is used to address the issue of contents semantic to describe web contents that enable automated information access.[1]

Semantic Web Architecture

The SW architecture is built on two main concepts (syntaxes) the URI and the Unicode, the former represents the data while the later support the internal text style standards. The syntaxes are called **Resource Description Framework** (RDF) which is a general metadata format used to represent information about internet resources and expand the web to be more than human-readable to be more about machine process-able information. [2]RDF Schema vocabulary descriptions are written in RDF. The extra descriptive power of RDF Schema is carried in a collection of RDF resources.

Building Semantic Web [4]

The Web is a worldwide medium for sharing data and knowledge. The requirements of such model are:

1 *Universal expressive power*: a data format must have enough expressive power to express any

form of data.

2 *Support for Syntactic Interoperability*: states how easy it is to read the data and get a representation that can be exploited by applications.

3 *Support for Semantic Interoperability*: mean the difficulty of understanding the data. Semantic interoperability is the difficulty of understanding the data and means to define mappings between unknown terms and known terms in the data while syntactic interoperability; which talks about parsing the data [4].

Ontology

Semantic Web is extending syntactic interoperability to semantic interoperability by providing a source of shared, precisely defined terms through one key which is ontology. [8]

Ontologies can play a crucial role in enabling the processing and sharing of knowledge between programs on the Web. Ontologies are generally defined as a “representation of a shared[4].

Ontology typically consists of a hierarchical description of important concepts in a domain, along with descriptions of the properties of each concept. The degree of formality employed in capturing these descriptions can be quite variable, ranging from natural language to logical formalisms, but increased formality and regularity obviously facilitates machine understanding. Ontologies are defined independently from the actual data and reflect a common understanding of the semantics of the domain of discourse.[8,2]

```

<class-def>
  <class name="plant"/>
  <subclass-of>
    <NOT><class name="animal"/></NOT>
  </subclass-of>
</class-def>
<class-def>
  <class name="tree"/>
  <subclass-of>
    <class name="plant"/>
  </subclass-of>
</class-def>
<class-def>
  <class name="branch"/>
  <slot-constraint>
    <slot name="is-part-of"/>
    <has-value>
      <class name="tree"/>
    </has-value>
  </slot-constraint>
</class-def>

```

Fig. 1: XML data mode based tree

Ontology is an explicit specification of a representational vocabulary for a domain; definitions of classes, relations, functions, constraints and other objects. Pragmatically, a common ontology defines the vocabulary with which queries and assertions are exchanged among software entities. [2]. Ontologies are not limited to conservative definitions, which in the traditional logic sense only introduce terminology and do not add any knowledge about the world. To specify a conceptualization we need to state axioms that put constraints on the possible interpretations for the defined terms.[2]

Ontology is divided into three categories i.e., Natural Language Ontology (NLO), Domain Ontology (DO) and Ontology Instance (OI). NLO is the relationship between generated lexical tokens of statements based on natural language, DO is

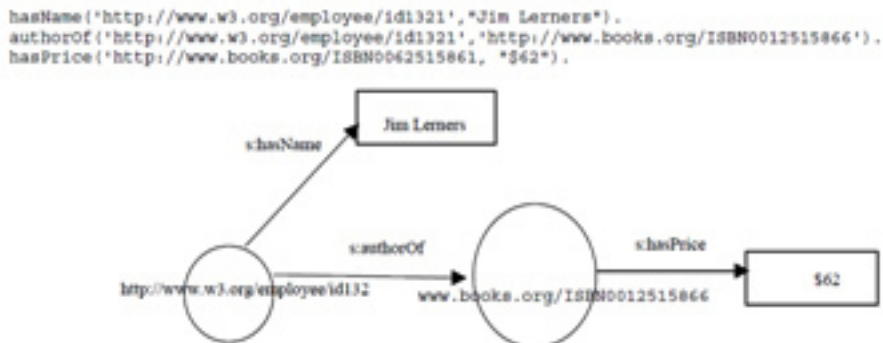


Fig. 2: Graph representation of RDF[4]

the knowledge of a particular domain and OI is the automatically generated web page behaves like an object. Ontology development process with respect to implementation point of view depends on some currently available ontology supported languages XML, RDF and OWL. [12]

Ontology languages

DAML+OIL is an ontology language specifically designed for use on the Web to describe the structure of a domain; it exploits existing Web standards (XML and RDF), also it adds familiar ontological primitives of object oriented and frame based systems, and the formal rigor of a very expressive description logic. [8]

XML

XML is the source for a rapidly growing number of software development activities. XML is intended as a markup-language for arbitrary document *structure*, as opposed to HTML, which is a markup language for a specific kind of hypertext documents. An XML document consists of a properly nested set of open and close tags, where each tag can have a number of attribute-value pairs.

Essential to XML is that the expressions of the tags and their allowed combinations is not fixed, but can be defined per application of XML. The basic data-model of XML is a labeled tree, where each tag corresponds to a labeled node in the data-model, and each nested sub-tag is a child in the tree as the figure1 below shows.

Any XML document whose nested tags form a balanced tree is a well-formed XML document. Furthermore it is possible to enforce constraints on which tags should be used, and which nesting of these tags is allowed. In XML 1.0 this is done in a Document Type Definition (DTD) [4]. When it comes to the use of XML; it fulfills two of the prescribed requirements (universal expressive power and the syntactic interoperability requirement since an XML parser can parse any XML data, and is usually a reusable component) However When it comes to semantic interoperability, XML has drawbacks where it just describes grammars but not semantic of specific domain.

XML will be the technique of choice for representing kinds of documents in product catalogs, digital libraries, scientific data repositories, and across the Web. And so it will be a major catalyst in constructing the "Semantic Web". Though, a document in XML format does not necessarily make a document's semantics explicit and more responsive for effective information searching. Searching XML documents can be made more effectively and efficiently when they are organized in a more explicit way. Level of concentration on trees of topics that reflect the user's or user community's interest profile is needed. Different areas of improvements are needed to provide an easy-to-use yet powerful and efficient search language that combines concepts from current pattern-matching languages (e.g., XPath, XQuery, etc), extracting more semantics from existing document collections by constructing structural and ontological skeletons (e.g., in the form of DTDs or XML schemas) that describe the data at a higher semantic level and can also facilitate new forms of indexing for efficiency, and classifying existing documents according to a given thematic or personalized, hierarchical ontology to make searching more effective and efficient [9].

RDF

RDF (Resource Description Framework) is a URL based syntax data representation which provides a secure and reliable mechanism for metadata exchange between web applications. RDF processes Meta data by making abstract data models based on three object types .i.e., Resource, Property and Statement. Resource is an expression, Property is an attribute describing resource and Statement is a resource having some properties and values[12]. The object-attribute-value triple (basic construction and providing basic role of RDF): an object O has an attribute A with value V. Such a triple corresponds to the relation that is commonly written as A(O,V), for example: has Price('http://www.books.org/ISBN00125 15866', "\$62").

RDF allows objects and values (1st and 3rd elements of the basic RDF triples) to be mixed: any object can play the role of a value. This notation provides natural semantic units because all objects are independent entities. No special terms for

any additional data model (similarly to XML; RDF data model provides no mechanisms for declaring the property names that are to be used) Besides the proposed Object-Attribute-Value-semantics (which is itself only informally described in the RDF standard), the RDF data model provides no mechanisms for declaring the property names that are to be used also can be represented by graph; figure2 shows a graph of some statements.[4,2]

The recent W3C recommendation to provide a standard meta-data for source description on the web is RDF [Lassila&Swick, 1999]. The difference between data and meta-data is not easy and straight forward: RDF is also capable of representing data same as meta-data.[4]

Two essential RDF Schema constructions are *subClassOf* and *subPropertyOf*. RDF objects may be instances of one or more classes; using the *type* property. The *subClassOf* property form RDF Schema allows the specification of the hierarchical organization of such classes [4]. XML is one of the serialization formats that is to be used by RDF. W3C introduced Notation 3 (or N3) as a non-XML serialization of RDF models designed to be easier to write by hand, and in some cases easier to follow. Because it is based on a tabular notation, it makes the underlying triples encoded in the documents more easily recognizable compared to the XML serialization. N3[5].

RDF vs. XML [1]

RDF is more useful than XML because it provides independent syntax serialization and abbreviation for data modeling, syntax reification and semantic based features like domain independency, vocabulary and privileges in defining terminologies used in schema language but still RDF modeling mechanism is insufficient in expressing various logical statements [12].

XML is concerned with the serialization while RDF is concerned with informational content. Solutions for the interoperability problem is easier to solve with RDF than with pure XML, since XML is syntax and its data model is tree and RDF is a data model based on a graph that uses URIs with different syntax, including an XML syntax.

Different RDF parsers are available, and are application independent, so requirement 2 which is support for syntactic is fulfilled. Regarding *Semantic Interoperability*, RDF has substantial advantages over XML: semantic units are given naturally through its object-attribute structure. In this structure all objects serve as independent entities. A domain model, defining objects and relationships of a domain of interest, can be represented naturally in RDF, so translation steps, as required when using XML, are not necessary. To find mappings between two RDF descriptions, techniques from KnowledgeRepresentation are directly applicable. *Of course this does not solve the general interoperability problem that is, finding semantic-preserving mappings between objects. However, the usage of RDF for data interchange raises the level of potential reuse much beyond parser reuse, which is all that one would obtain from using plain XML.*

Furthermore, since RDF describes a layer independent of XML, the RDF model (and software using the RDFmodel) can still be used, even if the current XML syntax is changed or disappears.[4]

Modeling semantics on the web is achieved via two main approaches: declarative and procedural semantics. In the former the meaning of an expression E is given by a mapping to another, or by stating the conclusions or properties that follow from E. The meaning of expression E can be understood without reference to any specific computational procedure. While the procedural semantics, the meaning of an expression E is given by referring to the behavior that some real or virtual procedure (or program, or machine) will exhibit on E. only way to obtain the meaning of an expression is by executing the procedure on E, and observe its behavior. The difference between a declarative and a procedural semantics loosely coincides with the difference between the XML and RDF approaches to semantics of Web-pages. As we've argued, an XML expression has no inherent semantics, and its semantics is only determined by the actions that one or more programs undertake on the XMLexpression (e.g. is tag-nesting interpreted as part-of, or subtype-of, or something else again?). An RDF expression on the other hand has a specific declarative semantics (e.g. the intended meaning

of “subClassOf”), and this is specified independent of any processor for RDF expressions (or stated otherwise: any RDF processor must conform to this intended semantics.)[4].

The main advantage of RDF over the basic XML is its simplicity. Unlike the order of elements in XML, the order of RDF properties does not matter. In addition RDF offers a very appealing and flexible solution to any web designer. RDF has an abstract syntax that reflects a simple graph-based data model, and formal semantics with a rigorously defined notion of entailment providing a basis for well-founded deductions in RDF data. XML and RDF are the current standards for establishing semantic interoperability on the Web, but XML addresses only document structure. Finally RDF better facilitates interoperation because it provides a data model that can be extended to address sophisticated ontology representation techniques [s.Deckerets 2001][4].

The Resource Description Format (RDF) [6, 9] is used to represent information modeled as a “graph”: a set of individual objects, along with a set of connections among those objects. In that role, RDF is one of the pillars of the so-called Linked Data Web (nee Semantic Web). RDF/XML is used to serialize information represented using graphs, how RDF graphs can be read and written by using the Jena software package, and how distributed graphs can be queried using the SQL query language. [3]

Semantic Web Problems and Solutions [2] Limitations

Ontology makes the abstract model of a particular domain based on set of data and structures but lacks in defining the boundaries of model. Also Size of Ontology varies with respect to the number of classes and instances; if the number of instances increased to large extent then it becomes very hard to manage manually and currently there is no as such mechanism exists to manage automatically. The Manual Ontology generation process sometime becomes very complex and time consuming especially while dealing with the large amount of data and to support the process of semantic enrichment reengineering for the building of web consisting of metadata

depends on the proliferation of ontologies and relational metadata and requires high production of metadata at high speed and low cost, which is currently also not available.

Problems

Not all parts of the SW are developed yet. May be the most important parts like agent, trust and retrieval techniques are developed very poor that strongly restricted the real usability of the SW. Technological base do not allow to use SW in good way. Technological base means the current level of hardware and software that cannot support all SW features. (e.g. “We believe that performance in terms of speed is not as important in this case as performance in terms of what is retrieved.” from OWLIR report) Software as well as hardware does not allow to easy develop a SW application and documents that brakes the mass SW’s growing.

People have no resolution to change the Web and to spend a lot of money. After all IT crisis people are very careful with all new technologies which needs a lot of money. It is mostly a psychological problem, but maybe it is the most important problem in the case of making the SW really a global scale system. Many people do not believe in Semantic Web. Poor theoretical foundation and in the absence of good news and evaluation results make people and especially computer experts more and more pessimistic.

Solutions

The core of the SW is already developed (Description languages: RDF, DAML+OIL, ontology and inference, the basic concepts of agents and proofs are taken from AI)

There is a clear plan for future. It seems the group of developers has a clear plan how to make web semantic. It gives the hope that all other problem is only technical. Industry makes computer faster and faster every minute. Thousands of unemployed but highly qualified programmers are ready to develop a new software agent for SW.

Wide using the XML created a good foundation for Semantic Web together with already developed software tools for creation SW application and documents (even poor) allow

reducing the cost of the SW building. That is a good reason to put up money right now. Active position of Scientific Community made people believe in SW. The problem of the SW is it is not a technology but philosophy of the future web existing. You trust the technology, but you believe the philosophy. Technology gives the answer, philosophy teaches to find the answer.

In Fact, many of these advantages are theoretically described and does not have a clear measure with the concept Semantic web , furthermore those advantages were built on with the assumption of the features of the both languages XML and RDF ,in addition, the impact of using RDF rather than XML is not measurable specially through the concept of semantic web. Actually, there have been different previous works dealing with similar problems.

The RDF data-model is sound and makes approaches from AI and Knowledge Engineering for establishing semantic interoperability directly applicable. Furthermore, it is universally extensible. We have shown the feasibility of our proposal by applying it to a particular ontological modeling language, and we have argued that a similar strategy should apply to any knowledge modeling language.

The Web community is currently regarding XML as the most important step towards semantic integration. We have argued why this cannot be true in the long run, and why RDF is a much better platform for this. The AI community is currently very much interested in applying many of its techniques to the Web. We have shown a generic method for Web-enabling arbitrary Knowledge Representation languages. This is an important step towards the realization of the dream of the Semantic Web.

The development of the semantic Web, and of Web ontology languages, presents many challenges. As we have seen, no DL system yet provides reasoning support for the full DAML+OIL language.

Developing a “practical” satisfy ability/ subsumption algorithm (i.e., one that is amenable to highly optimized implementation) for the whole language would present a major step forward in DL (and semantic web) research.

Moreover, even if such an algorithm can be developed, it is not clear if even highly optimized implementations of sound and complete algorithms will be able to provide adequate performance for typical web applications.[8]

REFERENCES

1. Fadi Al-Kalani, Mamoun G.Awad, Nabeel Bani Hani *Semantic Web: Improving Web Search Using RDF Instead of XML GJCST, USA, 10(15)(Ver. 1.0): (2010).*
2. Jun Cai, Vladimir Eske, Xueqiang Wang, *Semantic Web & Ontologies;*
<http://www.mpi-inf.mpg.de/departments/d5/teaching/ss03/xml-seminar/talks/CaiEskeWang.pdf>
3. <http://people.ku.edu/~grobe/SIGUCCS-semantic-web-intro/fp0518-grobe.pdf>
4. Stefan Decker, Frank van Harmelen, Jeen Broekstra, Michael Erdmann, Dieter Fense, Ian Horrocks , Michel Klein, Sergey Melnik, *The Semantic Web - on the respective Roles of XML and RDF, (2000).*
5. [http://en.wikipedia.org/wiki/Resource](http://en.wikipedia.org/wiki/Resource_Description_Framework)
6. Stefan Decker, Frank van Harmelen, Jeen Broekstra, Michael Erdmann, Dieter Fense, Ian Horrocks , Michel Klein, Sergey Melnik, *The Semantic Web: the roles of XML and RDF, IEEE, (2000).*
7. Dan Tufi_, Radu Ion, Elena Irimia, Verginica Barbu Mititelu, Alexandru Ceau_u, Dan, ^tefnescu, Luigi Bozianu, Ctlin Mihail, *Resources, Tools and Algorithms for the Semantic Web, (2006).*
8. Ian Horrocks, *DAML+OIL: Description Logic for the Semantic Web, IEEE Data Engineering Bulletin, (2002).*
9. Norbert Fuhr, Gerhard Weikum, *Classification and Intelligent Search on Information in XML, (2002)*

10. Peter F. Patel-Schneider and Jérôme Siméon, Building the Semantic Web on XML, *The First International Semantic Web Conference (ISWC2002)*.
11. http://classweb.gmu.edu/kersch/infos770/Semantic_Web_16_2/Semantic%20Web.pdf
12. Zeeshan Ahmed, Detlef Gerhard, *Web to Semantic Web & Role of Ontology*, 2010.
13. <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>