

Towards Global Semantic Web Xml On Scope As A Case Study

Eng. Haytham. T. Al-Feel

International Academy for Engineering & Media Science

Prof. M. A. Koutb

Faculty of Electronic Eng. Monouf

Dr. Hoda Suoror

Faculty of Electronic Eng. Monouf

***Abstract:** There are many problems associated with the World Wide Web: getting lost in the hyperspace according to the large amount of search results; difficulties of web administration due to the huge number of web pages available on the web today; and the web content is still accessible only to humans. Machines cannot participate in taking a decision. The solution to these problems is the semantic web. This is a new approach for building a new architecture for the web by developing languages for expressing and presenting information in both human readable and machine process able form.*

This paper presents an overview for the Extensible Markup Language (XML) which is the basic language for the semantic web. Also this paper works in depth with different languages associated to XML like XSL and CSS, highlighting simplicity and appearance for each plus its suitability to be used with XML. Illustration of the proposal structure is demonstrated through actual data of the 'Truth program' on Dream T.V.

1. INTRODUCTION

Since the last decade, World Wide Web (WWW) is occupying our lives. Everything can be done through the internet. WWW is used as a tool for collecting and sharing information and in other aspects of our lives such as in learning, in commerce, by government, for health et al. The increasing number of web pages and the dramatically increasing number of internet users however cause many problems to arise, such as the difficulties facing users to find and access the required material [1]. Moreover the administration of the web becomes very difficult. Most of the content of web pages is created by writing HTML codes. Instead of describing documents and linking them in HTML language which is not machine understandable, a new 'semantic web' approach will be used. The semantic web consists of many layers like XML and the Resource Description Framework (RDF) which will be used to describe different entities [2] such as people, places and data available on the web [3] in XML syntax.

The objective of this paper is to present the semantic web along with its functions and components. In addition, the paper will further investigate the basic component of the semantic web focusing on XML, its schema and style-sheet languages such as XSL and CSS, highlighting the main advantages and disadvantages for each and its suitability for being used with XML.

This paper is organized as follows: Section Two introduces related works; Section Three explains the meaning of the semantic web, its benefits and its architecture; Section Four explains the Extensible Markup Language (XML) as a core language for the semantic web; In Section Five, languages associated to the XML language such as DTD, XML, Schema, XSL and CSS will be explained; Section Six shows a comparison between XSL and CSS, highlighting the main advantages and disadvantages for each and its suitability for being used with XML. Conclusion & future work are discussed in Section Seven.

2. RELATED WORKS

Many researchers studied the semantic web from different points of view. Tim Berners Lee was the first to point to the idea of the semantic web in relation to RDF and XML layers [4][5]. Helfin and Handler's work is on the ontology based knowledge representation language which added features to RDF and XML [6]. Stefan Decker and others worked with XML and DTD [7], Ludovic Menet and colleagues tried to find a solution to manage complex data using XML Schema [8] while Jackie Kopecky and others focusing their work in the semantic annotation and XML Schema [9].

The full vision of the semantic web is still years away [10] and needs time to work with its full capability. To achieve the complete vision of the semantic web, we need to look again at the semantic web architecture and to work separately with each layer alone to test its strength and weakness and its relation to other layers and related technologies. This is what this paper tries to do by focusing only on XML - the base layer of the semantic web - and its related languages such as DTD, XML Schema, XSL & CSS plus how they can be integrated with each other.

3. THE SEMANTIC WEB

The semantic web develops languages for expressing information in a machine processable form [11], it provides a common framework that allows data to be shared and reused across application and community boundaries [12]. The semantic web is considered as an extension to the current web. It enables computers and people to cooperate; [13] and helps computers to process, find, read, understand and use different types of data over the World Wide Web to accomplish useful goals for users. It is like a mesh of information linked up to be accessible and processable with ease by machines on a global scale [14]. This will structure the web pages meaningfully, creating an environment where software agents can roam from page to page and carry out sophisticated tasks [15].

The Semantic web consists of seven main layers as shown in Figure 1. The first layer in the semantic web architecture is Unicode and URI, they provides a baseline for representing characters and a unique way for identifying objects in the semantic web. XML placed in layer two in the semantic web architecture, but considered as the basic layer of the semantic web. XML is considered as a set of rules for defining semantic tags that identity and break a document into parts. The XML layer will be discussed deeply in this paper. The third layer consists of the Resource Description Framework and its schema (RDF & RDF Schema). RDF is a data model for resources and relations between them. It provides simple semantics for this data model. This data model can be represented in XML syntax, while the RDF Schema is the vocabulary for describing properties and classes of RDF resources.

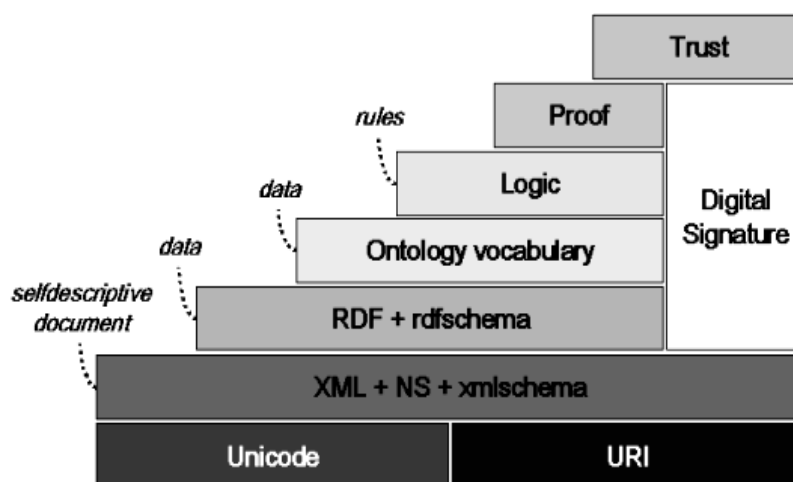


Figure 1: Semantic web layers

The Web Ontology Language (OWL) is the fourth layer in the semantic web architecture. It adds more vocabulary for describing properties and adds classes and relations between different resources. One of the important layers in the semantic is the Logic because information will be extracted from the web according to this logic. An example for logic is smart question answering. There is a relationship between Trust and Proof layers in the semantic web. There are logical operations done by information processors to make assure credibility of the information given. Once the proof has been carried out, the trust source can be sent to other agents to consume time and efforts. Because RDF syntax depends on XML syntax, the XML layer and its associated languages will be discussed further in the coming paragraphs.

4. EXTENSIBLE MARKUP LANGUAGE (XML)

XML is the language responsible for representing various types of data and documents on the web. It is a set of rules that break and identify the different parts of a document [16]. XML describes the structure of a document not the formatting, It describes what is in the document, no what the document looks like. XML allows the programmer create his own tags as needed. As shown in Figure 2, XML can be exported into different applications like a web browser such as the Internet Explorer that displays the document to the user, Databases like Microsoft SQL server can store XML data in a new record; sound documents can use XML documents as an input for a new record in sound documents, MIDI programs, printed documents and other forms.

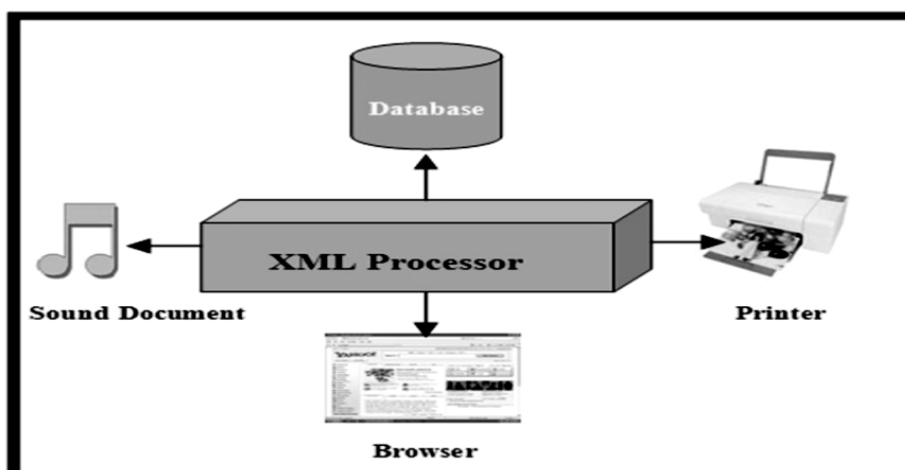


Figure 2: Different applications of XML

An XML documents has three components as seen in Figure 3: an optional prolog; a root element which is required, and material following the root, which is optional. The root element is the body of the XML document and has elements and attributes [17].

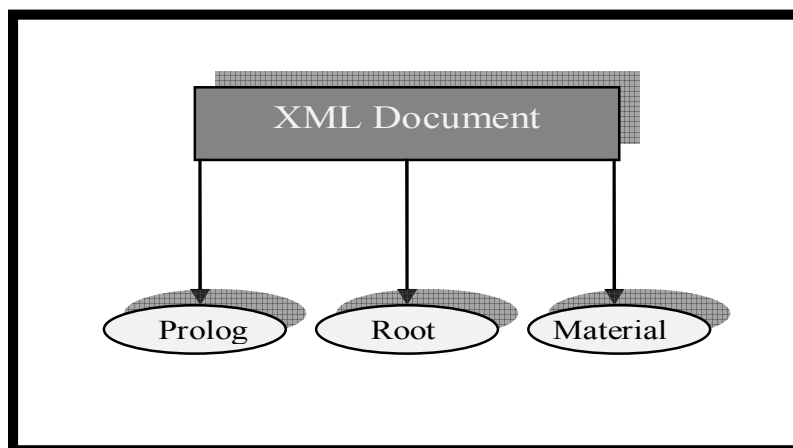


Figure 3: Components of an XML document

The prolog has XML declaration, for example `<?xml version= "1.0"?>` which defines the version used in this document. It also contains other attributes which are optional such as `encoding = "ISO-8859-1"` In this case, the document confirms that XML document uses ISO-8859-1 which is a Latin/west European character set. Also this line could include a 'standalone' attribute which means XML can stand alone without the need for other files. Stand alone takes yes or no and its value by default is no. This may be followed by a series of DTDs.

XML is a case sensitive language. Tag in XML is called an element. All elements must have a closing tag and also all XML documents must be properly nested. Attributes in XML provide additional information and must always be quoted. Element names must start by a letter or underscore.

5. ADVANCED XML CONCEPTS

XML is a family of technologies which offers useful services to accomplish important tasks. For example Cascade Style Sheet (CSS) and Extensible Style Sheet Language (XSL) are style sheets to describe the XML documents. All these concepts and others which are shown in Figure 4 will be discussed in the coming paragraphs.

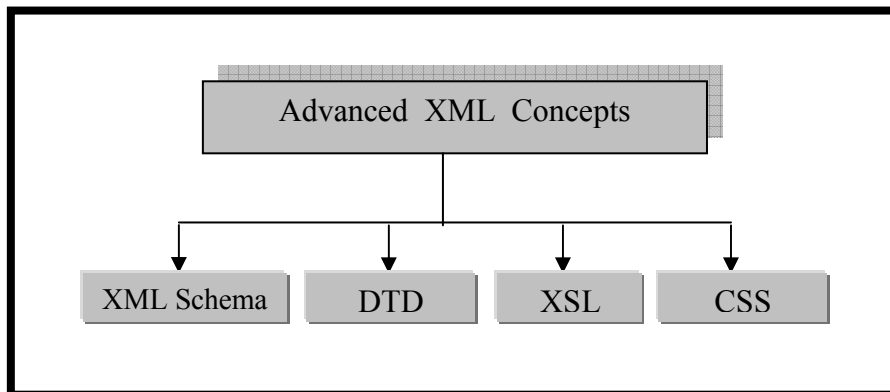


Figure 4: XML advanced concepts

5.1 DTD

A ‘Document Type Definition’ describes the overall syntax and structure for the XML document. It specifies a set of rules to which a document must conform to be valid [18]. It has two types of entries: element definition and attributes definition. An element definition `<!ELEMENT>` declares a tag and its sub elements while attribute definition `<!ATTLIST>` declares an attribute for that element [19]. String types for elements and attributes include `#PCDATA` and `CDATA` [20] while `#REQUIRED`, `#FIXED` and `#IMPLIED` are used with attributes. An attribute with `#FIXED` means the value will not change, while `#REQUIRED` means that the option is mandatory and `#IMPLIED` means that it is optional.

5.2 XML SCHEMA

XML schema has the same purpose as DTDs, but provides several improvements such as a rich set of data types that can be used with different elements [21]. Also XML schema is written according to XML specifications and syntax and this is an advantage because DTD uses another syntax which takes time from developers to understand. XML Schema increases the productivity of XML documents authoring and model customization towards specific needs [22].

5.3 CASCADING STYLE SHEET (CSS)

Cascading Style Sheet is initially developed for HTML to specify properties like font size and color. It is made up of style rules that tell a browser how to present a document [23]. When linked to XML document, it defines the properties of elements in that document [24] and separates a document structure from its presentation [25]. An XML document can be formatted in many different ways: the same document can have different presentations just by changing the Style

Sheet. Figure 5 provides the main data for the XML document which is written in a text editor such as the notepad and shown by a browser such as Internet Explorer 7.

By changing the cascade style sheet each time as in figures 6, 7 & 8 for the same XML file, the appearance for the same file through the browser will change.

```
<?xml version="1.0"?>
<?xml-stylesheet type="text/css" href="p1.css"?>
<PROGRAM>
  <TITLE>The truth</TITLE>
  <CHANEL>Dream2</CHANEL>
  <INTRODUCER>Wael elebrashy</INTRODUCER>
  <LENGTH>45 minutes</LENGTH>
  <PRODUCER>Dream T.V</PRODUCER>
  <CAMERA_MAN>AMR.SALAM</CAMERA_MAN>
  <DIRECTOR>M.yonis</DIRECTOR>
  <YEAR>2007</YEAR>
</PROGRAM>
```

Figure 5: The XML document describes "The Truth" program In Dream T.V. attached to the CSS file

```
Program {display table}
```

Figure 6: The first CSS associated to the XML document

```
PROGRAM { margin: 0.7cm 0.2cm 0.9cm 1.7cm}
PROGRAM, TITLE, CHANEL,
INTRODUCER,LENGTH,PRODUCER,CAMERA_MAN,DIRECTOR,Y
EAR { display: block }
ACT { margin-bottom: 5ex }
PROGRAM {color:green}
```

Figure 7: The second CSS attached to the XML document

By changing the different features in the code, we will have different presentations results. This means that there is an unlimited number of cascading style sheets that can be attached to the XML document. CSS has shortage and limitations when associated to XML documents. CSS can't work with XML by its full capability because it is designed especially to the Hyper Text Markup Language (HTML) and according to its standards.

```

PROGRAM { margin: 0.7cm 0.2cm 0.9cm 1.7cm}
PROGRAM, TITLE, CHANEL,
INTRODUCER,LENGTH,PRODUCER,CAMERA_MAN,
DIRECTOR,YEAR { display: block }
ACT { margin-bottom: 5ex }
PROGRAM{color:green}
PROGRAM { border-style: solid }
TITLE{ border-style: solid }
TITLE{color:red}
CHANEL{ border-style: solid }
CHANEL{color:yellow}
INTRODUCER{ border-style: solid }
INTRODUCER{color:blue}

```

Figure 8: The third CSS attached to the XML document

5.4 EXTENSIBLE STYLE-SHEET LANGUAGE (XSL)

XSL consists of two parts: XSLT and XSL-FO. XSLT is a language for transforming XML documents. The most common use of XSLT today is for transforming an XML document into either another type of XML document, HTML document or other types of text format such as C++/Java source files [26]. Also XSLT can be used as a server-side scripting solution in which XML is extracted from flat files or databases dynamically transformed by the web server as requested by the client browser.

This solution is necessary when the underlying data changes frequently [27]. XSL-FO is an XML vocabulary for specifying formatting. An Extensible Stylesheet Language consists of a number of template rules which is matched with nodes in the source to form a part of the result tree which makes the style sheet to be applicable for many documents that have similar source tree [28].

A style sheet is represented by an `xsl:stylesheet` element, another element can be put instead of this element which is called `xsl:transform`, but the end element should be as the start element. In an XSL transformation as shown in Figure 9, both XML document and XSL are processed according to the instruction in the XSL style-sheet which generates a new output. This means an XSLT transforms the XML document into an XML result document which then is represented by XSL-FO which creates a complete description of how pages should look [29].

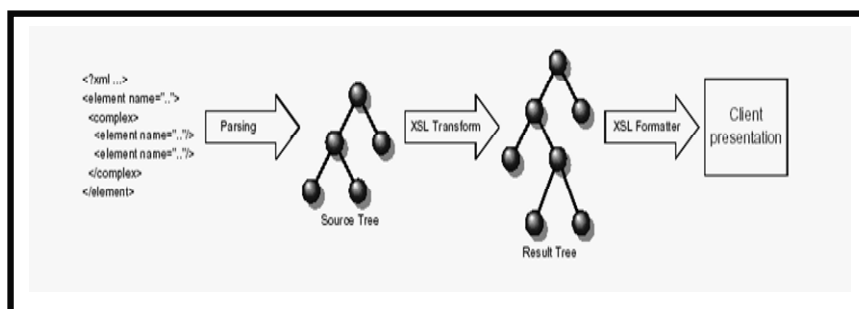


Figure 9: The transformation and formatting processes of XML document attached to XSL

An XML document can be formatted in many different ways with XSL such as CSS. Figure 10 provides the main data for the XML document. By adding the code provided in Figure 11, the document will appear through the browser. By changing the style sheet file to that in Figure 12 for the same XML file the resulting output will change. This means that the same XML file can take multiple layouts.

```
<?xml version="1.0"?>
<?xml-stylesheet type="application/xml" href="p1.xml"?>
<?xml-stylesheet type="text/xsl" href="p1.xml"?>
<PROGRAM>
  <TITLE>The truth</TITLE>
  <CHANEL>Dream2</CHANEL>
  <INTRODUCER>Wael elebrashy</INTRODUCER>
  <LENGTH>45 minutes</LENGTH>
  <PRODUCER>Dream T.V</PRODUCER>
  <CAMERA_MAN>AMR.SALAM</CAMERA_MAN>
  <DIRECTOR>M.yonis</DIRECTOR>
  <YEAR>2007</YEAR>
</PROGRAM>
```

Figure 10: The XML document describe "The Truth" program In Dream T.V attached to XSL file

```
<?xml version="1.0"?>
<xsl:transform version="1.0"
xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
  <xsl:template match="PROGRAM">
    <H1>welcome</H1>
  </xsl:template>
</xsl:transform>
```

Figure 11: The first XSL attached to the XML document

```

<?xml version="1.0"?>
<xsl:transform version="1.0"
xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
<xsl:template match="PROGRAM">
<HTML>
<HEAD>
<TITLE>
</TITLE>
</HEAD>
<BODY>
<font size="10" color="green"><xsl:value-of select="."/></font>
</BODY>
</HTML>
</xsl:template>
</xsl:transform>

```

Figure 12: The second XSL attached to the XML document

6. COMPARISON BETWEEN XSL & CSS

This study shows that XSL can control the appearance of the contents, attributes, and the ordering of different elements which is not found in CSS. XSL is more complicated than CSS but is more powerful. CSS is supported by different browsers like Internet Explorer 7, Netscape, Opera and others which give the CSS more popularity. One of the differences between XSL and CSS is in technique, CSS simply takes the source and applies some styles to it while XSL takes the source, applies some processing and creates a result tree [30]. One of the weak points which appeared when practicing with CSS is the appearance of attributes. Attributes in XML document don't appear when CSS is used. Also XML elements appear in the same order as the XML document. On the other hand, there are still some bugs with XSL with browsers caused by applicable different parts of XSL standards.

7. CONCLUSION & FUTURE WORK

The semantic web architecture consists of different layers such as RDF & OWL. These layers depend on XML as a language and its syntax. Thus, studying different parts of the XML language and associated languages is a step towards the development of the semantic web. Our study shows that XML faces some difficulties such as the applicable standards on different browsers; a large number of recent recommendations; and the need for other languages which may be written by a different syntax. Also It shows that XSL is more powerful than CSS and suitable to work with the semantic web than CSS. Our future work will be the studying of the RDF layer and the integration of the XML & RDF layers.

References

- [1] <http://www.esus2004.org/> . Last visited 5/6/2007
- [2] Glaser, H., Millard, I., Jaffri, A., Lewy, T. and Dowling, B. On Coreference and The Semantic Web. In: 7th International Semantic Web Conference, Karlsruhe, Germany,2008.
- [3] André, P., schraefel, m., Van Kleek, M. and Karger, D. Bringing the Semantic Web home: a research agenda for local, personalized SWUI. In: CHI 2008 SWUI Workshop.
- [4] Tim Berners Lee. Why RDF Model is Different from the XML Model,1999. W3 Consortium .
<http://www.w3c.org/DesignIssues/RDF-XML.htm> . Last visited 7/6/2007
- [5] Tim Berners Lee. Weaving the Web. Hyper, San Francisco, 1999.
- [6] Helfin and Hendler. Semantic Interoperability on the Web. In Proceedings of Extreme Markup Languages 2000. Graphic Communication Association ,2000,pp.111-120
- [7] Stefan Decker, Frank Van Harmelen, Jeen Broekstra, Michael Erdmann, Dieter Fensel, Ian Horrocks, Michel Klein and Sergey Melnik. The Semantic Web on the Respective Roles of XML & RDF IEEE Internet Computing, September/October 2000 (Vol. 4, No. 5) 2000.pp,63-74.
- [8] Ludovic Menet , Myriam Lamolle , Amar Zerdazi. Managing Master Data with XML Schema and UML. International Workshop on Advanced Information Systems for Enterprises, April 2008,p.53-59.
- [9] Jacek Kopecký , Tomas Vitvar , Carine Bournez , Joel Farrell. SAWSDL: Semantic Annotations for WSDL and XML Schema . IEEE Internet Computing, , November 2007.,p.60-67
- [10] Jorge Cardoso. The Semantic Web Vision: Where Are We?. IEEE Intelligent Systems, September 2007, pp.84-88.
- [11] <http://www.w3.org/designIssues/semantic.html> . Last visited 12/6/2007
- [12] <http://www.w3.org/2001/sw/> . Last visited 15/3/2007
- [13] <http://www.w3c.org> . Last visited 18/6/2008
- [14] <http://infomesh.net/2001/swintro/> . Last visited 4/4/2007
- [15] <http://www.scientificamerican.com/article.cfm?articleID=00048144-10D2-IC70-84A9809EC588EF21&catID=2> . Last visited 22/3/2007
- [16] Elliott Rusty Harold, XML 1.1 Bible 3rd Edition, Wiley publishing, 2004.
- [17] Bhavant Thuraisi Nghan, XML Databases and the Semantic web, CRC press LLC ,2002.
- [18] Peter Murray-Rust and Henry S. Rzepa, Scientific Publications in XML- Towards a Global Knowledge Base. Data Science Journal, Vol 1,2002,pp84-98.
- [19] Peter Thiemann , "A Typed Representation for HTML & XML Documents in Haskell", 2001.
- [20] Dongwon Lee & Wesley W Chu, "Constraints-Preserving Transformation from XML Document Type Definition to Relational Schema", 2000.

- [21] Michel Klein, Dieter Fensell, Frank van Harmelen and Ian Horrocks, The Relation between Ontologies and Schema Languages :Translating OIL-specifications in XML-Schema. the Workshop on Applications of Ontologies and Problem-solving Methods, 14th European Conference on Artificial Intelligence (ECAI 2000), Berlin, Germany.
- [22] Digo Fernando Veiga & Luismar Marques Porto, "XML Schema Representation as a Way to Interchange and Customize Genomic and Metabolic Models", Brazilian Archives of Biology and Technology, vol.48 ,June 2005
- [23] <http://htmlhelp.com/reference/css/q>. Last visited 17/3/2007
- [24] <http://www.webreference.com/authoring/style/shets/> . Last visited 5/5/2007
- [25] <http://meyerweb.com/eric/books/css-tdg/> . Last visited 20/3/2007
- [26] Aaron Skonnord and Martin Grudgin, Essential XML Quick Reference, A programmer's Reference to XML, Xpath, XSLT, XML Schema, SOAP and more, Addison – Wesley, 2001.
- [27] Sal Mangano ,XSLT Cookbook , O'Reilly Dec 2002.
- [28] Extensible Stylesheet Language (XSL) Version 1.0 World Wide Web Consortium Working Draft 18-Aug-98, <http://www.w3.org/TR/WD-xsl> . Last visited 19/6/2008
- [29] Jani Utriainen, "XML Parsers and XSL Processors", 2000.
- [30] Mika Kirjavainen , "XML Browsers, How XML Is Displayed in a Browser",2000.