# The suitable Schema for the Semantic Web

## A Comparative Study between DTD & Xml Schema

**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:*** *Current web contains billions of documents and has many administrative problems and limitations. The semantic web will be the extension for the current web. It is still the main topic of many researches. Many researches suppose the XML schema as the language used to make restriction to XML documents. This paper represents the Document Type Definition; compare it with XML schema to see which one is suitable for the semantic web.*

### 1. INTRODUCTION

The World Wide Web was originally created as a repository to store information. HTML added new features to the Web by linking different documents and facilitate the creation of web pages by better presentations. The Improvement of technology helps in increasing the number and quality of images, movies and other media elements. According to this, the number of web pages increased and also the number of search engines and databases have multiplied. All these factors caused a problem of how we use all this information in a correct manner. How do we find the information we need quickly without wasting time and effort. The semantic web will be the solution for all these problems. It will be an extension to the current web which helps human and machine to work together. In order to bring machine more into action, data must be put on the web in a form that machines can understand [1]. This is done by developing new languages for expressing information in a new way. XML which stands for extensible markup language is the base language to do this. It is a set of rules that breaks a document into parts and identify the different parts of the document. It describes structure not formatting [2]. There are different languages associated with the XML document their duty is to do the formatting for it, like Cascade Style Sheet (CSS) and Extensible Style Sheet Language (XSL).

There is an important concept associated with the XML document which is called validation. Validation means that XML document is written according to XML standards and associated to a schema file. Data integration requires relating multiple

schemas which is called schema mapping in order to allow scientists to retrieve data of different schemas in an integrated way [3]. The schema file provides the syntax for describing the logical structure to a document [4]. Validation is very important especially when exchanging data with others because it verifies that sending information is the same as the receiver expects and vise versa [5]. There are many schemas and formalisms have been developed to add semantics to documents [6] such as DTD, XML Schema, XDR (XML–Data Reduced) [7], RELAX NG [8], and SOX (Schema for Object Oriented XML) [9], but in this paper, DTD and XML Schema will be discussed because both describe permissible XML documents [10] and are developed by the World Wide Web Consortium (W3C) which means that they will be globally standardized.

## 2. DTD

DTD stands for Document Type Definition. It is designed first to work with SGML and has been widely used as the schema language for XML documents [11]. It is the first language used with XML as a schema and is written in a formal syntax that explain precisely which elements and entities may appear in the documents and what their content will be [5] by another way a DTD is used to ensure that XML documents conform to a common grammar [12]. Figure 1 & 2 show how can XML and DTD documents are written and attached to each other. Data in the XML document is a real data for a T.V program called the truth in Dream Channel.

```
<?xml version="1.0"?>
<!DOCTYPE PROGRAM SYSTEM “program2.dtd”>
<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 1: The XML Document Describe "The Truth" Program In Dream T.V.**

```
<!ELEMENT PROGRAM (TITLE, CHANEL+,
INTRODUCER,LENGTH,PRODUCER,CAMERA_MAN?,
DIRECTOR,YEAR)>
<!ELEMENT TITLE (#PCDATA)>
<!ELEMENT CHANEL ANY>
<!ELEMENT INTRODUCER (#PCDATA)>
<!ELEMENT LENGTH (#PCDATA)>
<!ELEMENT PRODUCER (#PCDATA)>
<!ELEMENT CAMERA_MAN (#PCDATA)>
<!ELEMENT DIRECTOR (#PCDATA)>
<!ELEMENT YEAR (#PCDATA)>
```

**Figure 2: The DTD Document Associated To The XML Document**

## 3. XML SCHEMA

An XML schema defines exactly the structure and the allowed elements in an XML file that is valid according to this schema [13]. It enables constrain to XML documents to a specific vocabulary and a specific hierarchical structure. It also allows validation of instances as shown in Figure 3 to ensure the accuracy of field values and the document structure at the time of creation. So it seems as a contract between designers, editors and programmers, to which all of them have to agree on [13].

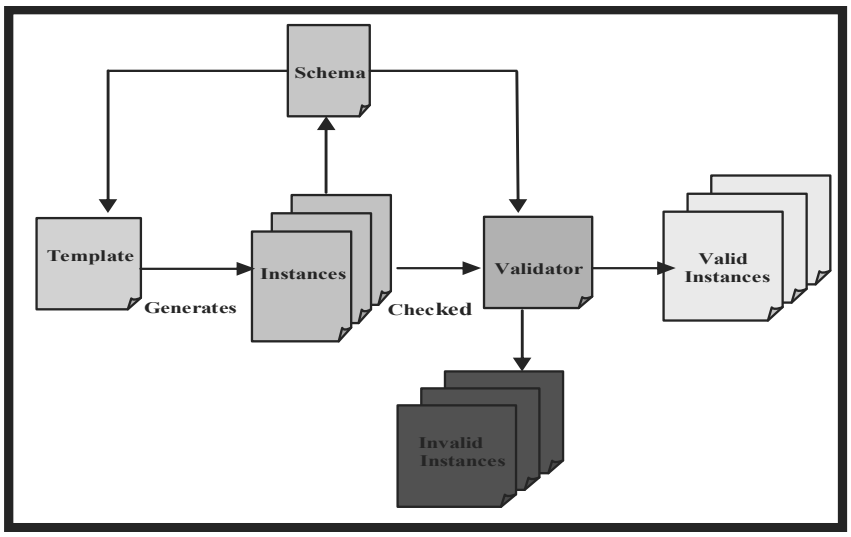


**Figure 3: The Validation Of Instances**

Figure 4 shows the XML schema document which is written to the XML document of the truth program in Figure 1.

```
<?xml version="1.0"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema">
<xsd:element name="PROGRAM" type="ProgramType"/>
 <xsd:complexType name="ProgramType">
  <xsd:sequence>
   <xsd:element name="TITLE" type="xsd:string"/>
   <xsd:element name="CHANEL" type="xsd:string"/>
   <xsd:element name="INTRODUCER" type="xsd:string"/>
   <xsd:element name="LENGTH" type="xsd:integer"/>
   <xsd:element name="PRODUCER" type="xsd:string"/>
   <xsd:element name="CAMERA_MAN" type="xsd:string"/>
   <xsd:element name="DIRECTOR" type="xsd:string"/>
   <xsd:element name="YEAR" type="xsd:string"/>
  </xsd:sequence>
 </xsd:complexType>
</xsd:schema>
```

**Figure 4: The XML Schema Document Associated To The XML Document**

## 4. THE COMPARISON:

- XML schema is written according to XML syntax which will facilitate to the programmer who knows XML to write the schema easily and also the schema will be easily understood by the XML application.
- Sub element order is one of the important aspects in XML in which element can not appear before another element if the schema specified the order. Both XML schema and DTD support this feature.
- Number of occurrences of an element in an XML document is one of the features supported by XML schema through max Occurs and minOccurs. On the other hand DTD support this feature but partially, it is not accurate like XML schema. DTD can control number of occurrences through operators like (*) which means the element will appear zero or more, while (?) means that the element will appear zero or one and (+) means that the element will appear one or more time.

- Data type is fully supported in XML schema. XML schema supports different data types like integer, float and others. For example the programmer can specify number of digits in the salary of an employee but this could not be done by DTD. DTD supports simple data type like text through (# PCDATA).

• Both XML schema and DTD support the default value for an attribute. Figure 5 shows the default value in DTD.

```
<! ELEMENT author (# PCDATA) >

<! ATTLIST author Language CDATA "Arabic" >
```

**Figure 5: The Default Value In DTD**

- Also both schema languages support the required feature for an attribute.
- The unique attribute is a feature in XML schema but it is found in DTD but by another way which is the ID to give a unique value to an attribute.
- XML schema support namespaces while DTD doesn't support.
- XML schema support inheritance like an object oriented and this is done through two options, they are extend and restrict to the base type while DTD doesn't support.

Summary for the comparison listed above, is shown in table 1.

**Table 1: Summary For The Comparison Between DTD & XML Schema**

| Features | DTD | XML schema |
|---|---|---|
| XML syntax | **X** | √ |
| Sub element order | √ | √ |
| Number of occurrences | partial | √ |
| Attribute default value | √ | √ |
| Data type | partial | √ |
| Required feature | √ | √ |
| Unique attribute | √ | √ |
| Namespaces | **X** | √ |
| Inheritance | **X** | √ |

## 5. CONCLUSION:

It seems from our study and comparison that XML schema will be suitable for working with a new technology like the semantic web. DTD will be suitable with documents and website already designed according to DTD because changing schema will be waste of time and efforts. XML schema is written in XML syntax, which will make life easy for programmers to deal with, and also it has many features like inheritance and namespaces which will help in growing of this technology.

## References


**[1]** Lucas Zamboulis. XML Schema Matching & XML Data Migration & Integration, 2003.

**[2]** Elliotte Rusty Harold. XML 1.1 Bible, 3rd Edition, 2004 Wiley Publishing.

**[3]** Dongkwang Kim, Karpjoo Jeong, Hyoseop Shin, Suntae Hwang. An XML Schema-based Semantic Data Integration , IEEE 2006.

**[4]** Aaron Skonnard and Martin Gudgin. Essential XML Quick Reference, A programmer's Reference to XML, XPath, XSLT, XML schema, SOAP and more, published by Addison-Wesley 2001.

**[5]** Elliote Rusty Harold, W. Scott Means. XML in Nutshell, 2nd edition, published by OReilly 2002.

**[6]** Inseok Song & Petra Saskia Bayerl. Models and formalisms for semantic markup, 2003.

**[7]** Frankston, C., & Thompson, H. XML-data reduced, URL http://www.ltg.ed.ac.uk/˜ht/XMLData-Reduced.htm, 1998.Last visited 16/1/2008

**[8]** Clark, J., & Murata, M. RELAX-NG specification, URL:http://www.oasis-open.org/committees/relax-ng/spec-20011203.html, 2001.

**[9]** Davidson, A., Fuchs, M., Hedin, M., Jain, Koistinen, J., Lloyd, C, Maloney & Schwarzhof, K. Schema for object-oriented XML 2.0, URL: http://www.w3.org/TR/NOTE-SOX/, 1999. Last visited 25/9/2008

**[10]** Makoto Murata, Dongwon Lee, Kohsuke Kawaguchi and Murali Mani. Taxonomy of XML schema languages using formal language theory,2001. ACM Transactions on Internet Technology (TOIT), Volume 5, Issue 4, 2005: pp.600-704.

**[11]** Shiyong Lu, Yezhou Sun, Mostafa Atay, Farshad Fotouhi. On the consistency of XML DTDs, Data & Knowledge Engineering 2005.

**[12]** Jeff Helfin, James Hendler, Semantic interoperability on the web, URL: http:\ www.cse.lehigh.edu\~helflin\pubs\ Last visited 12/1/2008

**[13]** Gerald Reif, Mehdi Jazayeri & Harold Gall. Towards Semantic Web Engineering: WEESA-Mapping XML Schema to Ontologies.