

XML STRUCTURE DEFINITION AND DATA EXTRACTION USING XML TECHNOLOGIES

Mrs.S.Kayathri

Assistant Professor/MCA, M.Kumarasamy College of Engineering, Tamilnadu, India.

Abstract

All the SGML content is not supporting in web browsers, to avoid this problem SGML content is convert in to HTML. The biggest barrier is reuse, interchange and automation. Several computers are having the in unsupported formats. That is the case exchanging data between two systems is a longest process for web developers. Huge amount of data must be converted, and unsupported format data is often lost. XML is the highly efficient derived from SGML. The main goal of XML is to specifying SGML elements are supported in web browsers then it is easy to publish SGML content on the web.

Keywords: XML, DTD, XSD, XSLT, SGML, HTML

1. Introduction

The XML is mainly used to store and retrieve the data in a text based markup language. It is specified in a structure format by using meaningful tags. XML allows a own descriptive tags for our applications. XML is a compatible with hardware and software cross platform markup language. Therefore XML can be used to transfer structured data between heterogeneous systems.

XML is used as a common data interchange format in a number of applications. In the case of a business to business e-commerce model, XML is used for exchanging data between various trading partners. There are several languages derived from XML such as MathML, which is a markup language used for representation of mathematical formula and scientific data. Similarly, Wireless Markup Language is used for web applications that can be accessed by using a cellular phone, is also derived from XML.

2. XML Documents

The XML documents consist of elements and mark up tags. It contains the large amount of data.

```
<Book>
  <First Name>Kennedy</First Name>
  <Last Name>John</Last Name>
  <Title>Welcome</Title>
  <body>Welcome to everyone</body>
</Book>
```

3. XML Tree Structure

The XML tree structure starting from the root node its end with siblings.

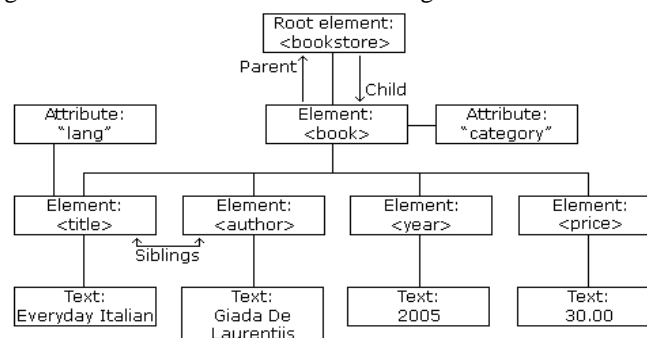


Figure 1: XML Tree

4. XML Validation

Validation is the process of validating and verifying the tags. The Xml document is match with the elements and attributes of DTD. There are two ways of validating the document.

- Well-formed XML document
- Valid XML document

5. XML Prolog

The prolog is an optional. It refers the information about the root tag. It contains the information about character encoding, document structure, and style sheet.

```
<?xml version="1.0" encoding="UTF-8"?>
```

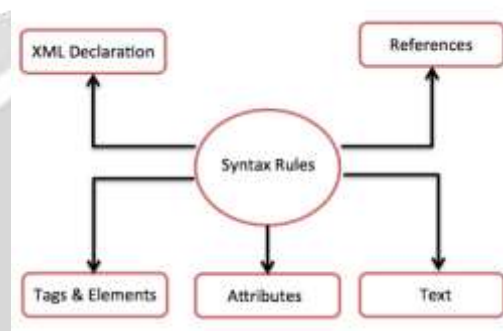


Figure 2: XML Prolog

The prolog is an optional. It refers the information about the root tag. It contains the information about character encoding, document structure, and style sheet.

6. XML DTD

The aim of a DTD is specify the structure of XML document. It is display the list of legal elements. There are two types of DTD.

1. Internal DTD

2. External DTD

6.1 Internal DTD

```
<?xml version="1.0">
```

```
<!DOCTYPE DOCUMENT
```

```
[
```

```
<!ELEMENT STOCK (CUSTOMER)*>
```

```
<!ELEMENT CUSTOMER
```

```
(Cname,Caddress,Cdate,Corder)*>
```

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

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

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

```
<!ELEMENT Corder (items)*>
```

```
<!ELEMENT item (Pproduct,Pprice)>
```

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

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

```

]>
<stock>
  <customer>
    <Cname>vaishali</Cname>
    <Caddress>1/54</Caddress>
    <Cdate>Dec14,2018</Cdate>
  <Corder>
    <item>
      <Pproduct>Sarees</Pproduct>
      <Pprice>1000</PPrice>
    </item>
    <item>
      <Pproduct>Sudithar</Pproduct>
      <Pprice>2000</PPrice>
    </item>
  </Corder>
</customer>
</stock>

```

6.2 External DTD

Step1: Create DTD for Student.dtd

```

<?xml version="1.0">
<!DOCTYPE document
[
<!ELEMENT STOCK(CUSTOMER)*>
<!ELEMENT CUSTOMER
(CName,CAddress,CDate,COrder)*>
<!ELEMENT CName (#PCDATA)>
<!ELEMENT CAddress (#PCDATA)>

<!ELEMENT Cdate (#PCDATA)>
<!ELEMENT COrder (#PCDATA)>
<!ELEMENT item Pproduct,Pprice>
<!ELEMENT Pproduct (#PCDATA)>
<!ELEMENT Pprice (#PCDATA)>
]>

```

Step 2: Create the XML Document

```

<?xml version="1.0">
<DOCTYPE STOCK SYSTEM "student.dtd">
<stock>
  <customer>
    <Cname>vaishali</Cname>
    <Caddress>1/54</Caddress>
    <Cdate>Dec14,2018</Cdate>
  <Corder>
    <item>

```

```

<Pproduct>Sarees</Pproduct>
<Pprice>1000</PPrice>
</item>
<item>

<Pproduct>Sudithar</Pproduct>
<Pprice>2000</PPrice>

</item>
</Corder>
</customer>
</stock>

```

7. XML Schema

XML schema is an alternative method of Document Type Definition. Schema is used to specify the overall structure of the document.

XML schema is represented in two ways.

1. Simple
2. Complex

7.1 XML Schema Definition

1. The simple element contains the only one text. It cannot have other elements and attributes.

```
<xs:element name="aaa" type="bbb"/>
```

2. The complex type specifies the multiple elements.

```

<xs:complexType>
<xs:sequence>
<xs:element name="Customer Name" type="xs:string"/>
<xs:element name="Customer address" type="xs:string"/>
<xs:element name="Customer Place" type="xs:string"/>
<xs:element name="Customer Delivery" type="xs:string"/>
</xs:sequence>
</xs:complexType>

```

8. XQuery

Xquery is the query based language. It is mainly used to store and retrieve the data. It is like a SQL database. It is supporting XPath expressions.

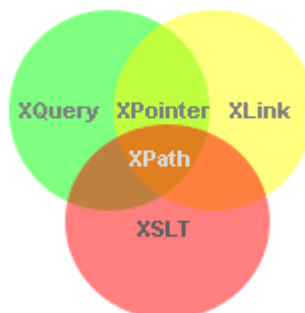


Figure 3: XML Technologies

FLWOR method is used for extract the query
For - selects the sequence of node
Let - binds the variable
Where - filters the node
Order by - sorts the node
Return - what to return

The Xquery predicate the limit to extract the data from XML document. The below query is representing the book price less than 30.

```
doc("books.xml")/bookstore/book[price<30]
```

The query extraction of the XML document is

```
<book category="Children">
  <title lang="en">Angry Birds</title>
    <author name>John</author name>
  <year>2018</year>
  <amount>25</amount>
</book>
```

9. XLink

In the HTML using the hyper reference for referiing another page,like that XLink is used to browse the link elements.

This is a simple example for XLink in XML documents.

```
<? xml version="1.0" >
```

```
<homepage xmlns:xlink="http://www.education.org/2018/xlink">
  <homepage xlink:type="simple" xlink:href="https://www.education.com">Visit website</homepage>
  <homepage xlink:type="simple" xlink:href="http://www.education.org">Visit the Page</homepage>
</homepage>
```

10. XML Related Technologies

10.1 SVG

SVG is Scalable Vector Graphics. It is specified in vector based graphics in XML. It is cannot zoomed and resized. The files can be animated. Using this technology images can be created, edited, searched, scripted and compressed. These particular images are scalable. Using SVG printing a good quality image. It is contain the some predefined shapes. The shapes are Rectangle, Circle, Ellipse, Line, Polyline, and Polygon.

```
<svg width="500" height="500">
  <rect width="300" height="600" style="fill:rgb(255,255,255);
stroke-width:3; stroke:rgb(255,255,255)" />
</svg>
```

10.2 RDF

RDF refers Resource Description Framework. It is mainly used to describe the resources on the web. It is easily understood by the computers. Now a day's RDF used in semantic web. Web information contains the exact meaning, processed by computers and integrates by computers.RDF use the Web identifiers to identify resources. It has the property and property values.

- **Resource** "https://www.education.com/rdf"
- A **Property** is a name of the resource. Each property contains the property value.

11. References

- [1] Suci, "1. Deutsch, M. Fernandez, D. Florescu, A. Levy, D. XML-QL: A Query Language for XML", <http://www.w3.org/TR/NOTE-xml-ql>.
- [2] Microsoft Corporation, XML Schema, <http://www.microsoft.com/xml/schema/reference/star.asp>.
- [3] Oracle Corporation, "XML Support in Oracle 8 and beyond", Technical white paper, <http://www.oracle.com/xml/documents>.
- [4] The Query Languages Workshop (QL'98), <http://www.w3.org/TandS/QL/QL98/>, December 1998.
- [5] Introductions to XML syntax include: Ray, Erik. "Markup and Core Concepts", Chapter 2 of *Learning XML: (Guide to) Creating Self-Describing Data* (January 2001).
- [6] <http://www.oreilly.com/catalog/learnxml/chapter/ch02.html> Walsh, Norman. "A Technical Introduction to XML", *xml.com* (October 1998). Available at <http://www.xml.com/pub/a/98/10/guide0.html>
- [7] XML Schema is published in three parts: *Part 0: Primer* (May 2001). Available at <http://www.w3.org/TR/xmlschema-0/> *Part 1: Structures* (May 2001). Available at <http://www.w3.org/TR/xmlschema-1/> *Part 2: Structures* (May 2001). Available at <http://www.w3.org/TR/xmlschema-2/>
- [8] XHTML: The Extensible HyperText Markup Language, W3C Recommendation, January 2000. <http://www.w3.org/TR/xhtml1>.
- [9] Extensible Markup Language (XML), W3C Recommendation, February 1998. <http://www.w3.org/TR/REC-xml>.
- [10] XML Schema Part 0: Primer, W3C Working Draft, April 2000. <http://www.w3.org/TR/xmlschema-0/>.
- [11] https://www.w3schools.com/xml/xml_what.asp
- [12] https://www.tutorialspoint.com/xml/index.html/wiki/Green_computing
- [1] Murugesan, San. "Harnessing green IT: Principles and practices." *Information Technology professional* 10.1 (2008): 24-33.
- [2] Mishra, Sushree. "GREEN COMPUTING." *Science Horizon* (2013): 21.
- [3] Donnellan, Brian, Charles Sheridan, and Edward Curry. "A capability maturity framework for sustainable information and communication technology." *IT professional* 13.1 (2011): 33-40.
- [4] www.cdproject.net
- [5] Google green computing report
- [6] <https://www.vpit.ualberta.ca/green/index.php?ref=home>
- [7] tmembersmittec.blogspot.in/2013/12/recent-implementation-of-green-computing.html
- [8] <http://greenworld2010.blogspot.in/2010/05/approaches-to-green-computing.html>