

Effective Web Data Presentation and Extraction Using XML Technologies

S. Kayathri^{1*}, S. Ramya² and S. Meena³

¹Assistant Professor, Department of Master of Computer Applications, M. Kumarasamy College of Engineering, Karur, Tamil Nadu. India. Email: kayathris.mca@mkce.ac.in

²Assistant Professor, Department of Master of Computer Applications, M. Kumarasamy College of Engineering, Karur, Tamil Nadu. India. Email: ramyas.mca@mkce.ac.in

³Assistant Professor, Department of Master of Computer Applications, M. Kumarasamy College of Engineering, Karur, Tamil Nadu. India. Email: meenas.mca@mkce.ac.in

*Corresponding Author

Abstract: The SGML is not supporting in web browsers, to avoid this problem SGML content is converted into HTML. The biggest barrier is reuse, interchange and automation. Several computers are having the in unsupported formats. That is the case exchanging data between the two systems is the longest process for web developers. Most of the time a very wide amount of data must be lost due to unsupported file format. The 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: DTD, HTML, SGM, XML, XSD, XSLT.

I. INTRODUCTION

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

XML is used as a encode data exchange between a client and server-side applications. In the case of a business to a business e-commerce model, it is used for encoding the data between various trading partners. There are more than a few languages derivative from XML such as MathML, which is a markup language used for the representation of mathematical formula and scientific data. Correspondingly, Wireless Markup Language is used for web applications that can be accessed by using a mobile phone, is also derived from XML.

II. XML DOCUMENTS

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

```
<Book-details>
  <name>Kennedy</name>
  <initial>J</initial>
  <title>Welcome</title>
  <body>Welcome to everyone</body>
</Book-details>
```

III. XML TREE STRUCTURE

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

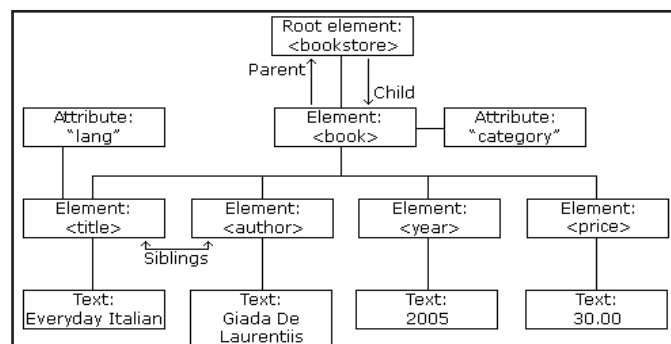


Fig. 1: XML Tree

IV. XML VALIDATION

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

- Well-formed
- Valid

V. XML PROLOG

The prolog is optional. It refers to the information about the root tag. It contains information about character encoding, document

structure, and style sheet.

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

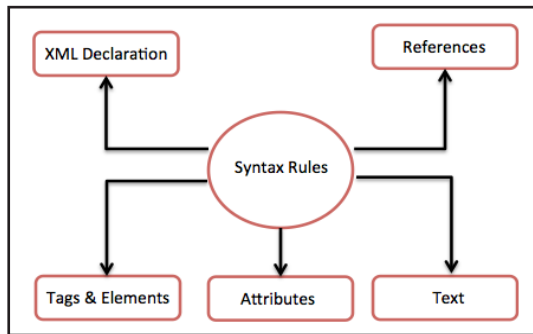


Fig. 2: XML Structure

The prolog is not a compulsory. It refers to the information about the root tag. It contains information about character encoding, document structure, and style sheet.

VI. XML DTD

The XML DTD specifies the overall structure of the document. It displays the list of legal elements.

XML types are:

- Internal
- External

A. Internal DTD

```

<!DOCTYPE DOCUMENT
[
<!ELEMENTSTOCK(CONSUMER)*>
<!ELEMENT CONSUMER
(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>
<consumer>
<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>
</consumer>
</stock>

```

B. External DTD

Step 1: Create DTD for Student.DTD

```

<?xml version="1.0">
<!DOCTYPE document
[
<!ELEMENT STOCK(CONSUMER)*>
<!ELEMENT CONSUMER
(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 XML DTD

```

<?xml version="1.0">
<DOCTYPE STOCK SYSTEM "student.dtd">
<stock>
<consumer>
<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>
</consumer>
</stock>

```

VII. XML SCHEMA

XML schema represents the overall construction structure of the XML document. It is one more way to define the XML structure. Schema is represented in two ways.

- Simple
- Complex

XML Schema Definition

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

```
<xs:element name="XML Elements" type="datatype"/>
```

2. The complex type specifies the multiple elements.

```
<xs:complexType>
```

```
<xs:sequence>
```

```
<xs:element name="CName" type="xs:string"/>
```

```
<xs:element name="CAdd" type="xs:string"/>
```

```
<xs:element name="CPlace" type="xs:string"/>
```

```
<xs:element name="CDeI" type="xs:string"/>
```

```
</xs:sequence>
```

```
</xs:complexType>
```

VIII. XQUERY

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

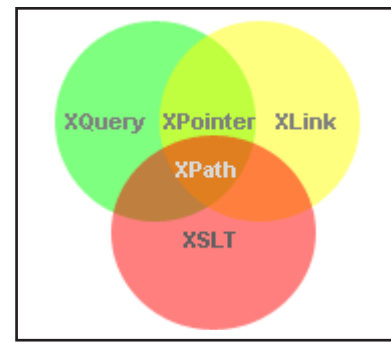


Fig. 3: XML Technologies

FLWOR method is used to 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 finds the size of extracting data from the document. The following XQuery is used to find the booking amount of less than 100.

```
doc("bookdetails.xml")/bookstore/book[price<100]
```

The query extraction of the XML document is

```
<book category="Child">
```

```
<title lang="en">Angry Birds</title>
```

```
<author name>John</author name>
```

```
< published year>2018</ published year>
```

```
<price>25</price>
```

```
</book>
```

IX. XLINK

In the HTML using the hyper reference for referring 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.edu">Welcome
```

```
</homepage>
```

```
<homepage xlink:type="simple" href="http://www.education.
org">Home Page</homepage>
</homepage>
```

X. XML TECHNOLOGIES

A. SVG

SVG is a Scalable Vector Graphics. It is specified in vector-based graphics in XML. It cannot be 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 contains some predefined shapes. The shapes are Rectangle, Circle, Ellipse, Line, Polyline, and Polygon.

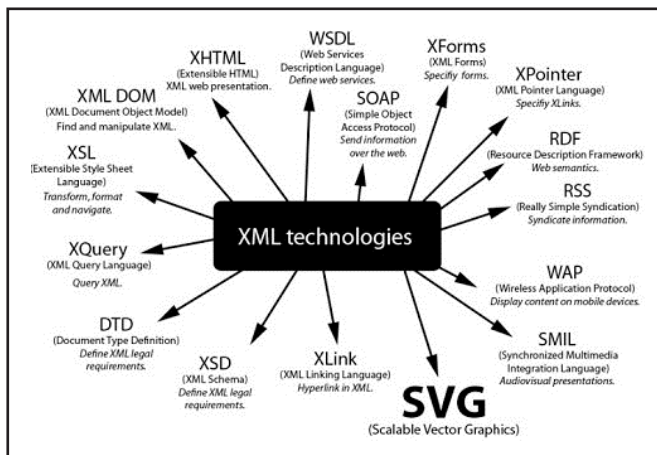


Fig. 4: XML Related Technologies

B. RDF

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

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

REFERENCES

- [1] https://www.w3schools.com/xml/xml_whatism.asp
- [2] <https://www.tutorialspoint.com/xml/index.html>
- [3] https://www.w3schools.com/xml/xml_xlink.asp
- [4] <https://whatism.techtarget.com/definition/XML-Extensible-Markup-Language>