Entering into XML

Chapter 02

Topics

- Understanding XML
- HTML vs. XML
- SGML vs. XML

What is XML?

- XML is a subset of SGML
- XML is used to describe the structure and meaning of document
- XML allows you to create own vocabularies (tag)
- XML is extensible; you can create other language using XML

XML in brief

- XML is optimized for World Wide Web
- XML can work with HTML for data display
- XML is smaller and simpler than SGML
- XML includes a hyperlinking mechanism XLL
- XML includes a styling mechanism XSL

HTML vs. XML

- Compare the two markups: HTML on the left and XML on the right.

HTML	XML
<html> <head><title>books in store</title></head> <body> <table> <caption>price list</caption> <tr><th>name</th><th>price</th> </tr> <tr><td>XML</td><td>700.00</td> </tr> <tr><td>HTML</td><td>350.00</td> </tr> </table> </body> </html>	<books> <book> <name>XML</name> <price>700.00</price> </book> <book> <name>HTML</name> <price>350.00</price> </book> </books>
he HTML above tells the browser how to format the contents	The XML above describe information (data) but does not indicate what to do with it or how to display
HTML is about displaying	XML is about describing

Why XML is needed?

→ HTML can display data but cannot structure and describe data.

→ We need language that is formal but not limited.

- So comes XML, XML is extensible and can be used to exchange data among applications
- XML is not replacement of HTML but actually complementary to HTML
- XML describes data to our application and HTML format and displays data

SGML vs. XML

- XML is a descendant SGML
- But XML is simpler and smaller than SGML
- Obviously, there are some similarities. But there are extra things in XML
- XML derived from SGML and optimized for World Wide Web

→ How XML relates to SGML?

- XML is generalized like SGML; you can define your own tag sets
- XML is self-describing like SGML
- XML documents can be validated against the rules defined in DTD like SGML

→ What are extra things included in XML?

- XML is simpler and smaller than SGML
- XML includes an hyperlinking scheme , which is described as a separate language called Extensible Linking Language (XLL)
- Goals of XML
- Use of XML
- XML standards

- XML includes a specification for style language called Extensible Stylesheet Language (XSL). XSL can transform an XML document into another format.

How XML is used

- XML is self-describing
- Each XML is created based on some rules defined in DTD
- So XML can be used to describe the structure meaning our data.
- XML can be used to work with data. A few example:

→ XML can be used as data interchange format. Since XML uses text-format standards based, data in one application's data can be converted to XML which can easily be used by other applications.

→ XML can be used for Web Data. Data content for web page can stored in XML and HTML can be used to format and display that data.

→ XML can be used as common data store that can be used in various different ways or by different applications

How XML is processed?

- XML is processed using parsers.
- Parsers breaks down the XML elements and generally converts it in a form which can be manipulated by an application.

What is parsing?

- Parsing means breaking down a document into its constituent parts and understanding the structure of the document and the relationships of the parts.

Goals of XML

- The goals of XML as defined by the creators of XML specifications:

→ XML shall be straightforwardly useable over the internet

→ XML shall support a wide variety of applications

→ XML shall be compatible with SGML

→ It shall be easy to write programs which processes XML

→ The number of optional feature is to be kept minimum, ideally zero

→ XML documents should be human-readable and reasonably clear

→ The design of XML should be formal and concise

→ XML documents shall be easy to create

→ Terseness in XML markups is of minimal importance

XML - recommendation and standards

- Current specification: XML 1.0 specification is currently approved as a recommendation by the W3C.
- MathML (Mathematical Markup Language): MathML is derived from XML. It is used to create mathematical formulae and scientific content on the web
- RDF (Resource Description Language): RDF is used for describing metadata
- XLL (XML Linking Language): It is a mechanism for linking XML documents
- SMIL (Synchronized Multimedia Integration Language): It is intended to use TV-Like multimedia on the web
- XSL (Extensible Stylesheet Language): It is a text-formatting language and is used to convert XML documents into various display format.