Web Basics - XML

Lesson 1: Introduction to XML

Lesson Objectives

- In this lesson, you will learn about:
  - Evolution of XML
  - Role of XML in Web Applications
  - Different members of XML family
  - Introduction to Namespace

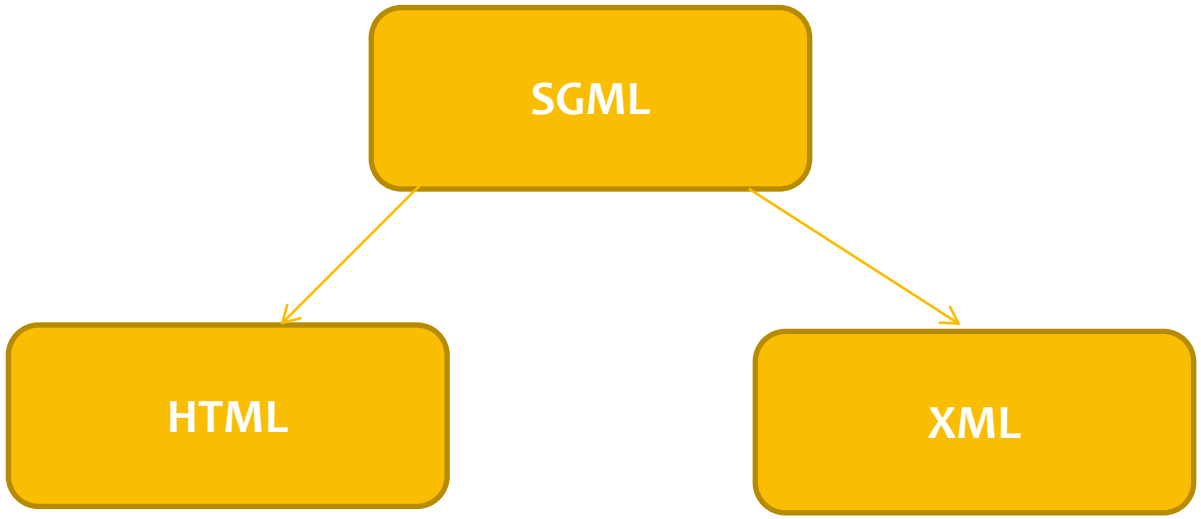
The Basics of Markup Language

- What do we mean by “Markup Language”?
  - The term “markup” is used to identify anything put within a document which either adds or provides special meaning
    (for example, italicized text)
  - A markup language is the set of rules
  - It also provides a description of document layout and logical structure

SGML

- SGML stands for Standard Generalized Markup Language
  - SGML was conceptualized in 1974 and adopted as international standard in 1986
  - SGML was born out of the basic need to make the data storage-independent
  - SGML also does not have any specific document structure, and usage of tag set is not limited
  - It does not constrain the potential of creating new document standards

Evolution of XML



Why Not Go Back To SGML?

- SGML is an incredibly rich meta language
- SGML is completely configurable
  - For example:
    - You can change the symbols for tagging from angle brackets (<tag>) to curly braces ({tag})
    - You can change the tag name lengths from 8 characters to 88 characters
- There is no style mechanism in SGML
- It is generic, just for generating customized language

HTML (Hypertext Markup Language)

- HTML which is an application of SGML, contains predefined set of tags and it is based on SGML manual
- HTML is a markup languages for web pages
- Similar to SGML
  - most tags describe meaning, not formatting
  - uses angled bracket convention (<tag></tag>)
  - based on a simple, widely compatible text format
- Different from SGML
  - HTML incorporates only one (standard document representation)

Introduction to XML

- What is needed a light-weight form of SGML which can provide well defined syntax for representing and processing document content over the web
- The answer is:
  - XML, the eXtensible Markup Language, is described as a means of structuring data
  - XML provides rules for placing text and other media into structures and allows you to manage and manipulate the results
  - XML standard is a subset of the SGML, developed in 1996 by the SGML working group

XML Design Goals

- The usage of XML was aimed at:-
  - Should be usable over the Internet
  - Should support a wide variety of applications
  - Should be compatible with SGML and XML documents should be easy to create
- Also XML can be used for
  - Data Exchange
  - Store and Retrieve Data

XML Today

- The primary functional purpose of XML is to transfer structured text and data among systems in multiple organizations
- XML, unlike HTML, does not have a fixed format
  - There are no pre-defined tags; you create your own
- Like HTML, XML uses tags. Tags are always enclosed within angled-brackets (< >)
  - XML tags define the meta information and are distributed throughout the document
- XML 1.0 is the most widely used version

XML versus HTML

- <table>
- <tr>
  <td>Apples</td>
  <td>Bananas</td>
  </tr>
- </table>

<table> tag in HTML is predefined & used for creating tabular display

<table> tag in XML could mean anything e.g its a coffee table which is a furniture

- <table>
  <name>African Coffee Table</name>
  <width>80</width>
  <length>120</length>
  </table>

XML and the Web

- XML deals with what the data is about and how to specify the data structure
  - XML represents data formats on web for the following:
    - Books
    - Financial transactions (EDI)
    - Technical manuals
    - Chemical formulae
    - Medical records
    - Museum catalog records
    - Chess games
    - Encyclopedia entries

A Family of Standards

- XML is a group of technologies
- It consists of the following specifications:
  - Extensible Style Language (XSL)
  - XML Linking Language (including Xpath, Xlink, and Xpointer)
  - XML Namespaces

Extensible Style Language (XSL)

- Cascading Style Sheets (CSS) makes it possible for the same HTML content to be easily formatted in multiple ways
- Extensible Style Language (XSL) works with XML data in a way similar to that CSS works with HTML
  - The rules created with the style language – the style sheet – should define how the content will be displayed
  - Formatting should not appear in the content itself

XML Namespaces

- XML Namespaces provide a way of assigning unique names to document constructs so that the software can operate correctly and avoid collisions
- XML Namespaces allow context to be given to the element names
  - This allows them to remain unique and thus process able

Summary

- In this lesson, you have learnt that:
  - SGML is the Standard Generalized Markup Language
  - HTML is the Hypertext Markup Language and XML is the Extensible Markup Language (meta-markup language)
  - XML is not a replacement for HTML
  - HTML tags do not say anything about the structure of the information
  - HTML lacks in link management, is not reusable, is not Object Oriented, and so on
  - Looking at the future of electronic commerce, HTML has limitations
  - XML is a project of w3c and its implementation is in the developing stage
  - XML uses features of SGML but it is easy compare to it
  - Markup Language created using XML are called XML vocabularies or XML applications

Review Question

- Question 1:Which of the following is/are true about SGML?
  - Option 1: makes Data storage independent
  - Option 2: usage of tag set is unlimited
  - Option 3: both the above
- Question 2: ___ allows to apply style to XML
- Question 3: XML namespace provides ___