Lesson 1: Introduction to XML

Web Basics - XML

Capgemini

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

Evolution of XML-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

The Basics of Markup Language:

There exist three types of markup:

- Stylistic: It determines how a document is presented (for example, the HTML tags <I> for italics, <B> for bold, and <U> for underline).

- Structural: It determines how the document is to be structured (for example, the HTML tags <P> for paragraph, <SPAN> for creating ad hoc styles in a document, and <DIV> for grouping structures that are aligned in the same way.

- Semantic: It tells about the content of the data (for example, the HTML tags <TITLE> for page title, <HEAD> for page header information, and <SCRIPT>to indicate a JavaScript in a page.)

In XML, the only type of markup that we are concerned with is structural.

Evolution of XML-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

SGML:

The **Standard Generalized Markup Language (SGML)**, from which XML is derived, has been around in various forms for quite some time now.

SGML may be best described by *what it is not*, as follows:
- SGML does not promote a specific document structure.
- There is not a limited tag set that must be used.
- SGML will not constrain the potential of creating new document standards.

SGML provides the common framework necessary to describe documents and to create new measures of conformance.
- Almost all languages that have been created to manipulate documents can trace at least a portion of their roots back to SGML.
- SGML itself is used by many large organizations, such as the United States Department of Defense, to handle complex electronic document exchanges. Avoiding the presentation features common to other document formats like PDF or even MS Word, SGML concentrates on the structure of the information.
- It does not promote one specific structure. However, it allows for the customized containment of data.

Evolution of XML-Evolution of XML
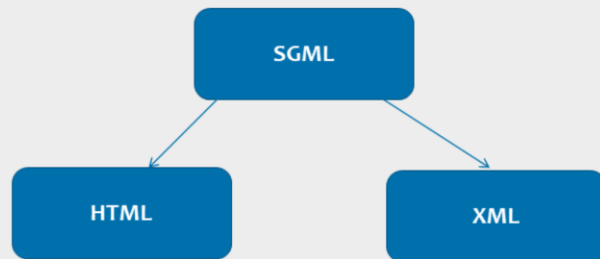
Evolution of XML-Why Not Go Back To SGML?

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

Why Not Go Back to SGML?

If HTML is limiting and SGML provides a means of describing document structures, why not simply go back to SGML?

SGML is a language for creating languages. With SGML, you can create tag sets.

Example 1: The Air Transport Association used SGML to create a tag set for aircraft maintenance documentation.

Example 2: The Society of Automotive Engineers used SGML to create a tag set for automotive service manuals.

SGML is an incredibly rich Meta language. It is completely configurable.

Example 1: You can change the symbols for tagging from angle brackets (<tag>) to curly braces ({tag}).

Example 2: You can change the tag name lengths from 8 characters to 88 characters.   All this flexibility takes up computing power on the part of tools that interact with SGML. However, for the web, you need **lightweight tools** and processes. As a subset of SGML, in many ways XML serves as a lightweight, web-compatible version of SGML.

Also SGML has some more limitations

No mainstream browser support

No support for styles

Not much support for datatype

Security limitations

Evolution of XML-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)

HTML:

The **H**yper **T**ext **M**arkup **L**anguage (HTML), which is an Application from SGML, contains predefined set of tags. HTML is designed for web publishing intended to describe structure of text.

However, HTML is **inflexible** in that it cannot allow domain-specific tag sets to be created and used without formally introducing them into the HTML DTDs.

Limitations of HTML

Browser specific commands for e.g Netscape-specific tags (e.g., <blink>),Microsoft IE tags (e.g., <marquee>)

Limited no of tags.

Some formatting commands not separating content and presentation e.g CENTER

HTML authors code to the browser's standards, not the W3C standards, therefore

Pages look different in different browsers & HTML validation is difficult

HTML talks of how and not what

Introduction of CSS to separate the look of the document to some extent

Introduction of XHTML to ensure standardization

Evolution of XML-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

Introduction to SGML and HTML:

XML, the e**X**tensible **M**arkup **L**anguage, is best 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.

The XML standard is a subset of the **Standard Generalized Markup Language**

(SGML), and was developed in 1996 by the **SGML Editorial Review Board** under the auspices of **World Wide Web Consortium (W3C)**.

XML is designed for that express purpose. XML provides a mechanism for the interchange of structured information on the web. This sort of data is required to transform the web from a publishing media to an application processing environment

Evolution of XML-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 Design Goals:
        The design goals for XML were proposed by the World Wide Web Consortium (W3C), and published in January 1998.
        The focus of XML is:
To carry the power of SGML through its ability to create user defined information about data, its ability to adapt to specific user needs, and the ability to maintain document changes.
To carry HTML's ease of use by the ability to link, its simplicity in the use of user defined tags, and its portability among different platforms.
        Some more design goals are as follows:
It shall be easy to write programs which process XML documents.
The number of optional features in XML should be kept to the absolute minimum, ideally zero.
XML documents should be human-legible and reasonably clear.
The XML design should be prepared quickly.
The design of XML shall be formal and concise.

Evolution of XML-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

About XML:
XML versus HTML:
        Both are based on SGML – the International Standard for structured information.

In HTML:	In XML:
<p>P200 Laptop	<product>
 Friendly Computer Shop	<model>P200 Laptop</model>
 $1438	<dealer>Friendly Computer Shop</dealer>
	<price>$1438</price>
	</product>

        Both XML and HTML may appear the same in your browser. However, the XML data is smart data. HTML tells how the data should look, but XML tells you what it means.

        With XML, your browser knows there is a product, and it knows the model, dealer, and price. From a group of these, it can show you the cheapest product or closest dealer without going back to the server. Unlike HTML, with XML you create your own tags, so they describe exactly what you need to know. As a result, your client-side applications can access data sources anywhere on the Web, and in any format. New "middle-tier" servers sit between the data sources and the client, translating everything into your own task-specific XML.

Evolution of XML-XML versus HTML

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

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

The Role of XML-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

The Role of XML:
XML will be most interesting to people and organizations who have:
Information resources that do not fit into the HTML mold, and
Resources that they want to make available over the web
XML was created to structure, store, and transport information. It is just plain text. Software that can handle plain text can also handle XML. However, XML-aware applications can specially handle the XML tags. The functional meaning of the tags depends on the nature of the application.
XML is as important for the web, as HTML was to the foundation of the web.
XML is everywhere. It is the most common tool for data transmissions between all sorts of applications, and is popular in the area of storing and describing information.
Some examples:
Books
Financial transactions (EDI)
Technical manuals
Chemical formulae
Medical records
Museum catalog records
Chess games
Encyclopedia entries

Different members of XML family-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

Introducing XML and its Relatives:
XML is a group of technologies.
It consists of the following specifications:

eXtensible Style Language: XSL works with the XML data in a way similar to the manner in which CSS works with HTML.

XML Linking: XML linking and addressing mechanisms are specified in three W3C Working Draft documents:

XML Path Language (Xpath): The primary purpose of Xpath is to do the actual addressing of parts rather than the whole XML document.

XML Linking Language (Xlink): It uses XML syntax to create structures to describe both "simple unidirectional links" of today's HTML as well as more "sophisticated multidirectional links". The important part of Xlink is that is defines the relationship between two or more data objects (or portion of objects) as opposed to a whole document.

XML Pointer Language (Xpointer): Xpointer builds on Xpath to support addressing into the internal structures of XML documents. Thus you can use the XML markup to link to specific parts of another document without supplying an ID reference.

XML Namespaces

Different members of XML family-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

Introduction to Namespace-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

XML Namespaces:

XML Namespaces are a way of assigning "unique names" to document constructs so that software can operate correctly and avoid collisions.

Namespaces allow context to be given to element names, which allow them to remain unique and thus processable.

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

Summary

Summary:
Why XML?

Many of the most influential companies in the software industry are promoting XML as the next step in the evolution of the web. How can they be so confident about something so new?

We can all safely bet on XML because the central ideas in this new technology are in fact very old and have been proven correct across several decades and thousands of projects.

Database Publishing:

One particularly popular application of XML will surely be the publishing of databases to the web.

Consider for instance a product database, used by the internal ordering system of a toy manufacturer. The manufacturer might want the database to be available on the web so that potential clients know the toys that are available and their price.

Rather than having someone in the web design department to mark up the data again, they can build a connection between their web server and their database using the features typically built into web servers that allows those sorts of data pipes. The designer can then make the product list beautiful using a style sheet. Pictures of the toys can be supplied by the database. In essence, the web site will be merely a view on the data in the database. As toys get added and removed from the database, they will appear and disappear from the view on the website.

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 _____