e-PG Pathshala

Subject : Computer Science

Paper: Web Technology

Module: Markup Languages

Module No: CS/WT/4

Quadrant 1 – e-text

Learning Objectives

The objective of this module is to know about the different Markup languages. This module also provides an introduction about HTML. It discusses about the basic tags such as Hyperlinks, Formatting, Headings, Paragraphs and Tables in HTML.

Introduction

Web site creation is important to create visually appealing websites. Web design is the creation and visual design of documents displayed on the World Wide Web. There are more than a handful of different scripting languages and markup languages for attractive web site creation. Web programming languages like HTML, XML and XHTML provide the tools to build and design a web site. These web programming languages or otherwise called as Markup languages are basically used for website creation and to create dynamic and interactive websites.

Before we discuss about the markup languages, let us see about the basic terminologies used in the context of Web site creation.

- **Web server** is a system on the Internet containing one or more web site.

- **Web site** is termed as a collection of one or more web pages.

- A **Web pages** is a single disk file with a single file name.

- **Home pages** are the first page in the website.

To know about how the Web works, the web information are stored in the form of Web pages. Now let us understand, the web pages are available in HTML format. The web pages are stored in the computers called Web servers in the Web server file system. The computer reading the pages is called web clients with specific web browser. The most commonly used web browsers

are Internet Explorer, Mozilla Firefox and Google Chrome etc. The web server waits for the request from the web clients over the Internet. The well known web servers are Internet Information Server (IIS) or Apache.

Web standards

The Web standards are not defined or setup by the browser companies or Microsoft, but by the World Wide Web Consortium (W3C). The specifications form the Web standards as HTML, CSS, XML, XHTML etc.

W3C - World Wide Web Consortium

The World Wide Web Consortium (W3C) is an international community to develop Web standards. It is led by the Web inventor Tim Berners-Lee and CEO Jeffrey Jaffe. W3C's mission is to lead the Web to its full potential.

W3C's long term goals for the Web are:

- o *Universal Access*: To make the Web accessible to all by promoting technologies.
- o *Semantic Web* : To develop a software environment that permits each user to make the best use of the resources available on the Web.
- o *Web of Trust* : To guide the Web's development with careful consideration for the novel legal, commercial, and social issues raised by this technology.

The History of Markup

In the early 1970s, GML called the Generalized Markup Language was developed where every tag would be defined like the following

":h1.The Content is placed here"

Since 1980s, there evolved Standard Generalized Markup Language called SGML and HTML. SGML was originally created by IBM in 1986. It is actually a meta language, meaning it is used to create other languages. Hence SGML forms the basis for the development of Markup languages like HTML, XHTML and XML.

Currently the Markup language that is widely used is eXtensible Markup Language or in short XML. It is not intended to replace HTML. Later, there evolved XHTML which does by providing better data description.
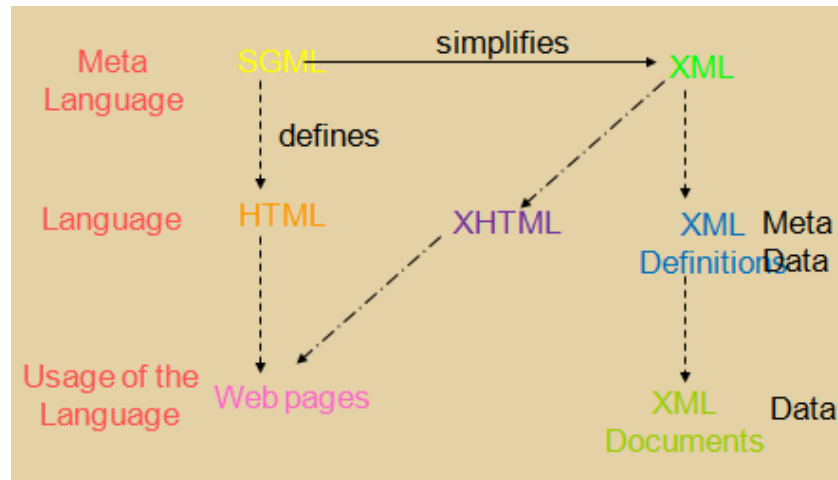
Figure 4.1 SGML, HTML and XML

The above Figure 4.1 describes about how the Markup languages evolved over time and the purpose of these languages.

SGML is a Meta language for the development of other Markup languages. Based on SGML, HTML invented by Tim Berners-Lee was developed that allows hyperlinks to display multimedia information as web pages to the web users. XML is another Markup language whose purpose is to describe information and transfer information over the Internet. XHTML is another language which combines the features of both HTML and XML modified to conform to XML standards.

HTML

The expansion of HTML is HTML HyperText Markup Language. HTML is the set of "markup" symbols or codes inserted in a file intended for display on a World Wide Web browser. The markup tells the Web browser how to display a Web page's text, images, sound and video files for the user. It is not a programming language i.e., it cannot be used to describe computations. In HTML, the individual markup codes are referred to as elements (but many people also refer to them as tags).

History of HTML

HTML is an evolving standard as new technology/tools are added even now.

- HTML 1 (Berners-Lee, 1989): very basic, limited integration of multimedia

- HTML 2.0 (IETF, 1994): tried to standardize these & other features

- HTML 3.2 (W3C, 1996): attempted to unify into a single standard but didn't address newer technologies like Java applets & streaming video

- HTML 4.0 (W3C, 1997): current standard (but moving towards XHTML) attempted to map out future directions for HTML, not just react to vendors

- XHTML 1.0 (W3C, 2000): HTML 4.01 modified to conform to XML standards

- XHTML 1.1 (W3C, 2001): "Modularization" of XHTML 1.0

- HTML 5 (Web Hypertext Application Technology Working Group, W3C, 2006): New version of HTML4, XHTML 1.0, and DOM 2 (still a work in progress), no longer based on SGML, but "backward compatible" with parsing of older versions of HTML.

- HTML 5 is referred to as a "living language".

XML

XML is the abbreviation of eXtensible Markup Language. It provides a standard way to represent information so as to allow information to be stored and interchanged among any Internet-connected devices. It is not a markup language but it is a meta-markup language that specifies rules for creating markup languages. Browsers use XML parsers to isolate and extract the information from XML documents.

XHTML

The eXtensible HyperText Markup Language or XHTML is a reformulation of HTML 4 in XML 1.0. It consists of all HTML 4.0.1 predefined components combined with XML standards. This language has been developed as a way of making XML documents that look and act like HTML documents. Using XHTML helps to strengthen the structure and syntax of the markup.

WML-Wireless Markup Language

Formerly called HDML (Handheld Devices Markup Languages) allows the text portions of web pages to be displayed on cell phones or PDAs via wireless media. It is part of the Wireless Application Protocol (WAP).

Cascading Style Sheets (CSS)

It provides a powerful and flexible way to control the details of web documents. HTML is more concerned about the content, CSS is used to impose a particular style on the document. It is named as cascading style sheets because they can be defined at three different levels to specify the style of a document called Inline, document level and external.

DHTML

It is used to describe a set of animated web documents that built from HTML, style sheets and scripts. There are three main parts of DHTML called,

- — Positioning

- — Style modifications

- — Event handing

It relies on the browser for the display and manipulation of the web pages.

Client-side and Server-side Technologies

The Table 4.1 lists the different web technologies to develop web programs or scripts at client-side and server-side.

Client-Side	Server-Side
HTML, XML	CGI/Perl
Cascading Style Sheets (CSS)	PHP
Scripting languages	ColdFusion
- JavaScript, VBScript	Scripting Languages
Java Applets	- Server-side JavaScript
ActiveX controls	- ASP, JSP, Java Servlets
Plug-ins and Helpers application	ISAPI/NSAPI programs

Table 4.1 Client-side and Server-side Technologies

Basics of HTML

Hypertext Markup Language (HTML) is a programming tool that uses hypertext to establish dynamic links to other documents. It is known as the Web's programming language and provides a general structure for creating web pages. All web pages are actually HTML files. HTML documents are simply text documents that uses simple ASCII text files to create HTML documents with HTML file extensions or suffix as .html or .htm. HTML documents can be created with Notepad in Windows and TestEdit in MAC OS. HTML editors can also be used.

The HTML documents contains the content or information of the webpage as well as special instructions called tags. Tags provide instructions on how to display text or graphics and control user inputs. Tags are generally enclosed in angle brackets: < >. Typically, there is a starting and ending tag around text. Embedded tags typically provides instruction for the structure, and appearance of the content

Quote: HTML Editors are called "WYSIWYG" meaning "*What You See Is What You Get!* "

Some of the examples of HTML Editors are Dreamweaver, FrontPage, GoLIve etc.

HTML Document Structure

A HTML document contains the information in the form of text data (content of the page). The HTML document is divided into two major parts as HEAD and BODY. The head part of the document contains the information about the page, e.g. the title and the body part contains the actual content of the page.

The basic document starts with <HTML> and ends with </HTML>

The HEAD part of the document contains information about the document namely,

- Title of the page which appears at the top of the browser window).

- Meta tags which is used to describe the information about the content that is in the document (used by Search engines).

- The script code written in JavaScript and Style sheets generally appears in the document Head.

The BODY part of the document contains the actual content of the document. This is the part that will be displayed in the browser window.

A sample HTML document looks like this,

```
<HTML>
<HEAD>
        <TITLE> My web page </TITLE>
</HEAD>
<BODY>
Content of the document
</BODY>
</HTML>
```

HTML Tags

All HTML tags are made up of a tag name and sometimes they are followed by an optional list of attributes which all appear between angle brackets < >. The content will not be displayed by the browser within the brackets unless the HTML is correctly written and the browser interprets the tags as part of the content. Attributes are properties that extend or refine the tag's functions.

Basic Syntax

Most of the HTML tags but not all have a start tag and an end tag like the following

<H1>Hello, world!</H1>

There are a few HTML tags that are standalone tags which do not use an end tag and are used for representing standalone elements on the page. Some of those tags are given below,

<img>   to display an image

            Line break

<HR>   header

Attributes

Attributes are added within a tag to extend a tag's action. We can add multiple attributes within a single tag. Attributes appear after the tag name and each attribute should be separated by one or more spaces. Most attributes take values, which follow an equal sign "=" after the attribute's name. The attribute values are limited to 1024 characters in length. An example of attributes defined within a tag is given below,

<body bgcolor="khaki" text="#000000" link="blue" vlink="brown" alink="black"

Information which the browser will ignore are tabs and multiple spaces will appear as a single space.

For example if the text appears as below in the document,

"Hello,

How are you?"

The browser will ignore the blanks and new line and displays

Hello, How are you?

**Line break
**

This tag breaks the line and starts text at a new line. It will not add an empty line like the paragraph tag. Multiple
  tags will display multiple line breaks in the text.

Paragraph Tag <P>

<P> is a Paragraph tag. It creates more space than a <BR> tag. It leaves one empty line after the tag. Multiple <P> tags with no intervening text is interpreted as redundant by all browsers and will display a single <P> tag.

Horizontal Rule <HR>

<HR> tag creates a Horizontal Rule. We can use attributes with <hr> such as

<hr width="70%">

Comments <!-- -->

The text enclosed within the comment tag will not be displayed. It is used to insert *comments* in the source code.

>   <!-- This is a comment   -->

>   <!-- This is another

>   comment -->

We can also use the following tag as a comment,

>   <comment>   This a comment   </comment>

Headings: <h1> .. <h6>

We can create headlines of various sizes on our web page. Headlines normally appear as bold letters. An empty line will also follow the headlines. It is used for representing titles and subtitles of various sizes.

For example, H1 is the largest font heading and H6 is the smallest font heading. These Headings tag need an end tag </H1>.

Character Formatting

Special types of text that can be displayed using HTML are:

- Bold text
- Important text
- Italic text
- Emphasized text
- Marked text
- Small text
- Deleted text
- Inserted text
- Subscripts
- Superscripts

The following HTML tags are used to format the appearance of the text on a web page.

\<B> Bold \</B>

The text enclosed within the <B> tag are displayed as Bold or it appears as dark letters.

\<I> *Italic* \</I>

The text enclosed within the <I> tag are displayed as *Italic* letters.

\<U> Underline \</U>

The text enclosed within the <U> tag are displayed as Underlined text.

\<PRE> Preformatted \</PRE>

The text enclosed by PRE tags is displayed in a mono-spaced font. Spaces and line breaks are supported without additional elements or special characters.

\<EM> *Emphasis* \</EM>

The text enclosed by <EM> tags are usually displayed as italics.

\<STRONG> STRONG \</STRONG> Browsers display this as bold. The HTML **** element defines **strong** text, with added semantic "strong" importance.

Example

<p>This text is normal.</p>

<p>This text is strong.</p>

The above text is displayed as follows.

This text is normal.

This text is strong.

<TT> TELETYPE </TT> The text enclosed by <TT> tag is displayed in a mono-spaced font. It looks like a typewriter text, e.g. fixed-width font.

<CITE> A Beginner's Guide to HTML **</CITE>** The text enclosed by <CITE> tag represents a document citation usually in *italics*. For example, it appears like this,

(*A Beginner's Guide to HTML*)

It is used to display f**or titles of books, films, etc.**

<FONT SIZE="+2"> Two sizes bigger**</FONT>**

The size attribute can be set within the <FONT> tag as an absolute value from 1 to 7 or as a relative value using the "+" or "-" sign. Normal text size is 3 (from -2 to +4).

The color attribute can be set within the <FONT> tag,

Color = "#RRGGBB" The COLOR attribute of the FONT element.

For Example, **<FONT COLOR="#RRGGBB">this text has color</FONT>**

Lists

Lists are used to organize items in the browser window. There are two ways to create a list namely, Unordered list which is a Bulleted list and it is most popular. Here, list items have no particular order. The other list is an Ordered list or Numbered list.

- Ordered Lists (OL): e.g. 1,2,3
- UnOrdered Lists (UL): e.g. bullets.

Basic Syntax:

Unordered Lists

Fruit

```
<UL>
    <LI>Banana
    <LI>Grape
</UL>
```

Fruit
• Banana
•Grape

Exploring Different Attributes

We have the choice of setting the TYPE Attribute to one of five numbering styles.

<OL
type=I or i (for large or small roman numerals)
type=A or a (for capital or small letters)
type=1 (for numbers, which is the default)
>

<UL
type=disc (the default for first level lists)
type=round (the default for second level lists)
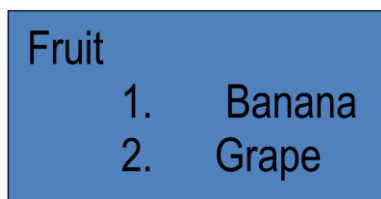type=square (the default for third level lists)
>

Numbered list/ Ordered List:

Fruit

<OL>
    <LI> Banana
    <LI>Grape
</OL>

Fruit
1.   Banana
2.   Grape

We can also specify a starting number for an ordered list.

<OL TYPE="i" START="3">
    <LI> List item …</LI>

</OL>

<OL TYPE ="i">
<LI> List item ...</LI>
<LI> List item ...</LI>
</OL>

<P> text ....</P>
<OL TYPE="i" START="3">
<LI> List item ...</LI>
</OL>

List Elements

DL: Definition List.

This kind of list is different from the others. Each item in a DL consists of one or more **Definition Terms (DT elements),** followed by one or more **Definition Description (DD elements).**

<DL>
<DT> HTML </DT>
    <DD> Hyper Text Markup Language </DD>
<DT> DOG </DT>
    <DD> A human's best friend!</DD>
</DL>

The above HTML code appears as below in the browser.

HTML

        Hyper Text Markup Language

DOG

    A human's best friend!

Summary

In this module the different Markup languages have been introduced. The Web programming language HyperText Markup Language (HTML) has been explored and the basic tags has been explained in HTML.