Web Design & Development I

Topic 2: Markup languages

Objectives

- ⇨ Define is a markup language.
- ⇨ Define what is SGML.
- ⇨ Characteristics of XML and XHTML.
- ⇨ Guidelines for using XHTML
- ⇨ History of HTML.

1.0 Overview of Markup languages

A markup language combines text as well as coded instructions on how to format that text and the term "markup" originates from the traditional practice of 'marking up' the margins of a paper manuscript with printer's instructions. Nowadays, however, if you mention the term 'markup' to any knowledgeable web author, the first thing they are likely to think of is 'HTML'.

1.1 What is Markup language?

It's a computer language that uses tags for the definition of elements within a document. It is human-readable and contains standard words instead of typical programming syntax. These markup languages are used to create web pages. While several markup languages exist, the two most popular are HTML and XML.

1.2 What is SGML?

Standard Generalized Markup Language

The concept behind SGML is that documents contain structural and other semantic elements that can be represented without regard to how they should be displayed. Depending on the performance medium and style preferences, the actual display of such a document can differ.

This is a standard on how to determine a document's markup language or tag set especially electronic documents exchange, document management, and document publishing. It's metadata for specifying the description of a document.

Note: SGML is just a standard not a language

Advantages of using SGML

- ⇨ They can be created by thinking in terms of document structure rather than appearance characteristics (which may change over time).
- ⇨ They will be more portable because an SGML compiler can interpret any document by reference to its document type definition (DTD).
- ⇨ Documents originally intended for the print medium can easily be re-adapted for other media, such as the computer display screen.

1.3 What is XML?

XML (Extensible Markup Language) is a markup language similar to HTML, but without predefined tags to use. Instead, you define your own tags designed specifically for your needs. This is a powerful way to store data in a format that can be stored, searched, and shared. Most importantly, since the fundamental format of XML is standardized, if you share or transmit XML across systems or platforms, either locally or over the internet, the recipient can still parse the data due to the standardized XML syntax.

Characteristics of XML

⇨ **It's extensible**

Allows developers to create their own self-descriptive tags, or language according to their application.

⇨ **Carries the data does not present it**

Allows storage of data irrespective of how it will be presented.

⇨ **Public standard**

It's an open standard developed by World Wide Web Consortium (W3C).

XML is called the "Extensible Markup Language" since custom tags can be used to support a wide range of elements. Each XML file is saved in a standard text format, which makes it easy for software programs to parse or read the data. Therefore, XML is a common choice for exporting structured data and for sharing data between multiple programs.

An example of an XML file that stores information about computer models may include the following section:

```
<computer>
  <manufacturer>Dell</manufacturer>
  <model>XPS 17</model>
  <components>
    <processor>2.00 GHz Intel Core i7</processor>
    <ram>6GB</ram>
    <storage>1TB</storage>
  </components>
</computer>
```

1.4 HTML

HTML is a markup language used for creating webpages. The contents of each webpage are defined by HTML tags. Basic page tags, such as <head>, <body>, and define sections of the page, while tags such as <table>, <form>, <image>, and <a> define elements within the page. Most elements require a beginning and end tag, with the content placed between the tags

HTML refers to a unified system for tagging text files with the aim to achieve font, color, graphics and hyperlink effects on the World Wide Web pages.

1.5 What is xHTML?

XHTML is an acronym for eXtensible HyperText Markup Language. It is the next step in the internet's evolution. The first document form in the XHTML family is XHTML 1.0. With just a few exceptions, XHTML is almost identical to HTML 4.01. HTML 4.01 has been updated to be cleaner and more stringent. If you are already familiar with HTML, you will just need to devote a small amount of time to learning this most recent edition. The World Wide Web Consortium (W3C) created XHTML to assist web developers in transitioning from HTML to XML. Web developers can access the XML environment with all of its advantages today by converting to XHTML, while remaining secure in the content's backward and future compatibility.

Note: XTHML = HTML + XML

Characteristics of xHTML

- ⇨ Elements must be in lowercase
- ⇨ Attribute minimization is forbidden
- ⇨ Elements must be closed
- ⇨ Elements must be properly nested

Advantages of using xHTML.

- ⇨ XHTML documents are XML conforming as they are readily viewed, edited, and validated with standard XML tools.
- ⇨ XHTML documents can be written to operate better than they did before in existing browsers as well as in new browsers.
- ⇨ XHTML documents can utilize applications such as scripts and applets that rely upon either the HTML Document Object Model or the XML Document Object Model.
- ⇨ XHTML gives you a more consistent, well-structured format so that your webpages can be easily parsed and processed by present and future web browsers.
- ⇨ You can easily maintain, edit, convert and format your document in the long run.
- ⇨ Since XHTML is an official standard of the W3C, your website becomes more compatible with many browsers and it is rendered more accurately.
- ⇨ XHTML combines strength of HTML and XML. Also, XHTML pages can be rendered by all XML enabled browsers.
- ⇨ XHTML defines quality standard for your webpages and if you follow that, then your web pages are counted as quality web pages. The W3C certifies those pages with their quality stamp.

1.5.1 Guidelines for writing an xHTML document.

- ⇨ Must include the DOCTYPE declaration at the start of the document.

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
```

- ⇨ XHTML is case sensitive only lower case is allowed.

```
<!-- This is invalid in XHTML -->
<A Href="/xhtml/xhtml_tutorial.html">XHTML Tutorial</A>

<!-- Correct XHTML way of writing this is as follows -->
<a href="/xhtml/xhtml_tutorial.html">XHTML Tutorial</a>
```

- ⇨ Close all XHTML tags properly

```
<!-- This is valid in XHTML -->
<p>This paragraph is not written according to XHTML syntax.</p>

<!-- This is also valid now -->
<img src="/images/xhtml.gif" />
```

- ⇨ All the tags must be well nested

```
<!-- This is valid in XHTML -->
<p>This paragraph is not written according to XHTML syntax.</p>

<!-- This is also valid now -->
<img src="/images/xhtml.gif" />
```

- ⇨ Quote all the attribute values.

```
<!-- This is invalid in XHTML -->
<img src="/images/xhtml.gif" width=250 height=50 />

<!-- Correct XHTML way of writing this is as follows -->
<img src="/images/xhtml.gif" width="250" height="50" />
```

- ⇨ Forbid attribute minimization

```
<!-- This is invalid in XHTML -->
<option selected>

<!-- Correct XHTML way of writing this is as follows -->
<option selected="selected">
```

More attributes

Checked → "checked", readonly → "readonly", disabled → "disabled", nowrap → "nowrap", multiple → "multiple", etc.

- ⇨ Replace the **name** attribute with the **id** attribute.

```
<!-- This is invalid in XHTML -->
<img src="/images/xhtml.gif" name="xhtml_logo" />

<!-- Correct XHTML way of writing this is as follows -->
<img src="/images/xhtml.gif" id="xhtml_logo" />
```

- ⇨ Avoid using the language attribute of the script tag.

```
<!-- This is invalid in XHTML -->

<script language="JavaScript" type="text/JavaScript">
   document.write("Hello XHTML!");
</script>

<!-- Correct XHTML way of writing this is as follows -->

<script type="text/JavaScript">
   document.write("Hello XHTML!");
</script>
```

An example of a xHTML document

```
<?xml version="1.0" encoding="UTF-8"?>

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">

<html xmlns="http://www.w3.org/TR/xhtml1" xml:lang="en" lang="en">
   <head>
      <title>Every document must have a title</title>
   </head>

   <body>
      ...your content goes here...
   </body>
</html>
```

Figure 1. An example of a compact xhtml document

Note: The above guidelines differentiate xhtml from html

1.6 History of HTML

HTML —which is short for **H**yper**T**ext **M**arkup **L**anguage— is the official language of the World Wide Web and was first conceived in 1990. HTML is a product of SGML (Standard Generalized Markup Language) which is a complex, technical specification describing markup languages, especially those used in electronic document exchange, document management, and document publishing. HTML was originally created to allow those who were not specialized in SGML to publish and exchange scientific and other technical documents. HTML especially facilitated this exchange by incorporating the ability to link documents electronically using *hyper*links. Thus, the name *Hyper*text Markup Language. However, it was quickly realized by those outside of the discipline of scientific documentation that HTML was relatively easy to learn, was self-contained and lent itself to a number of other applications. With the evolution of the World Wide Web, HTML began to proliferate and quickly spilled over into the mainstream.

1.6.1 HTML versions

- **HTML 1.0**
  - Released in 1993
  - The main objective was to share the information which is readable and be accessed through web browsers.
- **HTML 2.0**
  - Published in 1995.
  - New features were introduced in addition to the features of HTML 1.0
- **HTML 3.0**
  - Developed by Dave Raggett in January 1997
  - Included improved new features of HTML. This provides web developers with more powerful characteristics although these improvements slowed down the browsers.

- **HTML 4.01**
  - Published in 2012.
  - It's the most and widely used markup language worldwide. It is also the most successful version of HTML before the coming of HTML 5.
- **HTML 5**
  - Recently released and used worldwide.
  - HTML 5 is the extended version of HTML 4.01.

1.7 Other technologies that we can use in web design and development.

As a student and web developer there are other trends in web development apart from HTML you would need to know. Modern websites and systems are designed and developed used more sophisticated, more complex and secure ways to ensure that the security of customers' data is enhanced. Modern websites apart from using html and css (we will look at it later) are developed using new tech trends like frameworks, newer programing languages, APIs, Data formats etc. Therefore, it is important for a web developer to dive in research to find out which is the best tool and approach to develop a system.