

Introduction to (X)HTML

This lecture is quite technical, but don't worry too much if it becomes overwhelming – it will start to make sense when we actually start writing some code. You probably have no intention of becoming a programmer and learning code, however a basic understanding of coding (what it is, how it works) will give you a good foundation in using programs such as Dreamweaver and understanding the fundamentals of website production and construction.

The reading for this topic will also help to set the context.

Let's go over some basics first.

The Internet

The Internet is a network of networks. It is a collection of computers that are all linked together. You link into the Internet via a broadband connection (such as cable or satellite) or a modem. Once you are connected you and your computer become part of the Internet and you are linked to every other computer that is also connected at that time.

The World Wide Web

The World Wide Web is an ever-changing collection of documents that are all stored someplace on the Internet. The Knowledge Base website at the Indiana University explains it this way:

“The World Wide Web is a way of exchanging information between computers on the Internet, tying them together into a vast collection of interactive multimedia resources.”

Knowledge Base, University Information Technology Services

To find out more about the difference between the Internet and the World Wide Web visit http://www.webopedia.com/DidYouKnow/Internet/2002/Web_vs_Internet.asp

These documents, better known as webpages, are usually written in some form of HTML (code). The HTML format consists of tags describing the data that your pages contain. The file extensions are typically htm or html. A website is made up of a number of webpages. These pages have hypertext links to enable navigation from one page or section to another. Webpages can consist of text, photographs, illustrations and are displayed using a web browser.

HTML

HTML, Hypertext Markup Language, is the common language used to communicate on the World Wide Web, and is the basic underlying structure of webpages. It was developed by Tim Berners-Lee at CERN, the European Particle Physics Laboratory in Geneva, Switzerland, in the early 90's.

HTML lets you format text, add graphics, sound and video, and save it all in a Text Only or ASCII file that any computer can read. Therefore the beauty of HTML is its simplicity. Of course, to project or view video and play sound, the computer must have the necessary hardware and devices.

HTML has two essential features:

- hypertext
- universality

Hypertext

Hypertext allows you to create a link in a webpage that leads the visitor to any other webpage or to practically anything else on the Internet. Information can be accessed from many different directions, more like a person's brain rather than something static like a book.

Universality

Because HTML documents are saved as Text Only files, virtually any computer can read a webpage. It doesn't matter if your visitors are using Mac or Windows machines or even hand-held devices like a Palm or mobile phone.

The web is open to all. This was Berners-Lees' vision for the web.

While HTML is available to all, that doesn't mean that everyone experiences it the same way. Any computer can display web pages, but what those pages actually look like depends on:

- type of computer
- monitor
- speed of the internet connection
- software used to view the page: that is, the Browser

The most popular browsers today are Internet Explorer, Firefox, and Chrome with handhelds and PDAs gaining momentum everyday (see http://www.w3schools.com/browsers/browsers_stats.asp). Unfortunately none of these displays a webpage exactly like the next.

As the Web and the technologies that supported it evolved, so did HTML. Designers wanted more control over the appearance of pages and also the code needed to support the many kinds of media formats that were also starting to be developed. Problems soon started to arise with early versions of HTML.

Browsers started to develop different ways of doing the same thing, which meant that sites would look different in different browsers. Netscape started to put up the first 'fences' on the Web in 1994. People liked these extensions so much they flocked to Netscape and by 1996 Netscape had become the most popular computer program in the world. (Castro:2007)

Not wanting to lose valuable customers Microsoft started fencing off their part of the Web. They added non-standard extensions that only Internet Explorer could recognise. Let the browser wars begin!

This made it very difficult for designers to create web pages that both browsers would recognise and read in a similar way. Different users had different sized screens and different sized screen resolutions and therefore pages looked different on different computers. If you wanted to change some aspect of the site design such as the background colour of the pages, or the typeface, every page had to be changed individually, because this information was in the HTML of every page.

As a result, designers found they were creating a number of different versions of the one site for different browsers and different devices.

The push for Standards

The World Wide Web Consortium (W3C) www.w3c.org was set up by Tim Berners-Lee in 1994 to try to quell the war. Their mission was to convince the Web community of the importance of universality while attempting to satisfy the thirst for beautiful looking pages. They wanted to take down the fences.

Some of the organizations that are members of the W3C are:

- Apple
- Microsoft - Explorer browser has taken over no 1 position from Netscape
- Opera – a browser for desktop computers and handhelds
- Mozilla - makers of Firefox browser, which are now giving Explorer a run for its money
- America Online - absorbed Netscape Communications
- Adobe

And search engines such as

- Google
- Yahoo!

The idea is that these companies come together and agree on standards while still trying to differentiate their products.

The first step by The W3C was to standardise the proprietary extensions used by different browsers and also try to encourage browser manufacturers to support these specifications. This was released as HTML 3.2. As web pages began to become visually more interesting and designers strived to incorporate more stylistic formatting in their pages, the documents started to become larger. Old versions of HTML pages had all the content, structure and formatting instructions in each document.

The W3C came up with a simple solution: stop authors from mixing *presentational* (or stylistic) markup in with the markup that dictated the structure of the document. With HTML 4, the W3C marked most of the formatting tags for later removal. These tags became deprecated tags. However HTML pages were still a mess in terms of a standard system for adding new features and most browsers still displayed pages slightly differently. It also was fairly accepting of sloppy code. The W3C decided that a much sturdier platform was needed for more structure.

The idea of separating style from content stripped HTML back to the basics and allowed for the development of a new system for formatting instructions called Cascading Style Sheets (CSS).

At the same time the W3C decided that it was time for a large-scale overhaul in the way websites were created. What was evident was that “different users require different presentations of the same information.” (Duckett:2005)

Enter XML - Extensible Markup Language - a language for creating other languages, for example, a tag (used in html) identifies data, therefore, that data becomes available for other tasks (they can be used for different markup languages and they can speak to each other. Reformulating HTML into XML would prepare the language for the next decade and beyond.

The problem with XML is that it is not nearly as lenient as HTML - XML calls for careful attention to upper and lower case, quotation marks and opening and closing tags. To ensure a smoother transition (as there were billions of web pages already written in HTML and browsers that knew how to read html) the W3C rewrote HTML in XML. Instead of calling the next version HTML 5.0, they called it XHTML.

However, stay tuned for HTML 5.0. See

http://www.w3schools.com/w3c/w3c_html.asp

XHTML

Extensible Hypertext Markup Language

- Stronger
- More flexible
- More likely to be supported in the future

XHTML had all the features of HTML and people who already knew HTML only had to learn a few basic syntax rules. You can create an (X)HTML document with any word processor or text editor like SimpleText or Notepad or by using a WYSIWYG (What You See Is What You Get) software program like Dreamweaver which writes the (X)HTML for you.

The Basics of (X)HTML

URL's

Uniform Resource Locator is basically a fancy name for an address. The URL contains information about where a file is and what the browser should do with it. Each file on the Internet has a unique URL.

Scheme	Server name	Path	File Name
http:// www.site.com/donna/ file.html			

The Scheme tells the browser how to deal with the file it is about to open. The most common scheme we see is:

HTTP - Hypertext Transfer Protocol

Other schemes are:

HTTPS - for secure web pages

FTP - file transfer protocol

Mailto - for sending email

These all generally have a colon and two forward slashes after them except for mailto.

The Server Name is where the file is located, followed by the path that leads to the file and the file's name itself.

When a user jumps to the URL that corresponds to your web page, the browser needs information right away about what kind of document it is and how it should be displayed. This is called the DOCTYPE, and looks typically like this:

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

This sits at the top of the page. This doctype is for XHTML transitional documents. Each page will then have the basic structure using tags. We will go through the basic structure and set up of XHTML pages in the tutorials.

What does HTML look like?

HTML is made up of 'tags'. The tags store the information that the browser needs to be able to interpret the document.

- These tags are enclosed in brackets < >
- There must be a beginning and ending <tag>....</tag>
- They must be nested (more about this in the tutorials)

Most web pages are divided into two sections -

the Head tag <head>....</head>

and the Body tag <body>....</body>

HTML tag <html>....</html> encloses a document

The Head section provides information about the URL of your webpage as well as its relationship with the other pages at your site.

The only element in the Head section that is visible to the user is the title of the webpage

Title tag <title>....</title>

The Title sits between the Head tags is used in search indexes as well as in browsers' history lists and bookmarks

The Body tag follows after the end Head tag, and is where the content of your Web pages are contained. This includes basic formatting of your pages.

Basic Formatting

Headings

One way of formatting headings in your pages is by using the heading tags.

h1, h2, h3, h4, h5, h6

`<h1>HEADING 1</h1>`

`<h2>HEADING 2</h2>`

`<h3>HEADING 3</h3>`

`<h4>HEADING 4</h4>`

`<h5>HEADING 5</h5>`

`<h6>HEADING 6</h6>`

Heading 1 is more important to a browser than Heading 6

Paragraphs and Breaks

(X)HTML does not recognise the returns that you enter in your text editor. To start a new paragraph in your webpage you must use the `<p>` tag

`<p>` text of paragraph goes between the tags `</p>`

To begin a new line without so much space, use the line break tag `
`

Other tags

`<a href>` tag (hypertext reference)

indicates a link

`Apple`

`` tag - image tag

``

References

Castro, E (2007) Visual Quickstart Guide *HTML, XHTML & CSS* Sixth Edition, Peachpit Press

Knowledge Base, University Information Technology Services, Indiana University
<http://kb.iu.edu/data/acqj.html>

Duckett, Jon (2005), *Accessible XHTML and CSS Web Sites: Problem - Design - Solution*, John Wiley & Sons, Inc, Indianapolis IN

The World Wide Web Consortium, <http://www.w3c.org>