2. HTML - Part I

What is a Web page? Can be regarded as a document (page) when you enter a URL in the browser address bar and hit go. You may right click on the page and read the content. It contains many words other than the actual content you see, especially symbols < and >. These words are from a special language used to write (mark up) the page.

HTML (Hyper Text Markup Language) is the language. You already know what hypertext is but what is a mark-up language? It originates from the newspaper and publishing industry when editors write notes (mark up) for manuscripts to determine the final look of the page before publishing. HTML used concepts from earlier mark-up languages like IBM's Generalized Markup Language (GML) and later Standard Generalized Markup Language (SGML).

2.1. What is HTML? | 1.6. the universal resource locator (URL)

2.1. What Is HTML?

It is declarative. Think about other computer languages you may know of, such as Java, C++ or Scratch. They are procedural or imperative. You need to tell the computer step by step to do something, i.e. how?. For mark up language, you don't care about how but what; I want to have a heading ... a 3 by 3 table, etc. The browser will produce them for you.

HTML is the standard publishing language of the World Wide Web now. It gives the writer functionalities for:

- publish online documents with desired structures (headings, text, tables, lists, photos, etc)
- retrieve online information via hypertext links, at the click of a button
- design forms for conducting transactions with remote services, for use in searching for information, making reservations, order products, etc.
- include spread-sheets, video clips, sound clips, and other applications/objects directly in their documents.

2.2. HTML tag (element)

2.2. HTML Tag (Element)

The < > pair together with the enclosed name is called a tag or formally element, e.g. <body>, . Most tags come in pairs open <head> and close </head>, some are empty tags, i.e. not in open/close pair, e.g.
, <hr>

Why use these tags? Because Tim Berners-Lee made a decision that any content in HTML or HTTP should be readily read by human and hence should be in text rather than binary like programs in other language such as Java. The mark up (tag) has to be distinguished from the content and so the special symbols < > are used and have special meaning. How about if we actually want a < or > in our content?

Note what the content is - the text enclosed by the open & close tags.

```
<p>this is the content</p>
```

An empty tag therefore has no content. Tags can also have properties, formally known as attributes used to specify characteristics of a tag, e.g. give it a name or id.

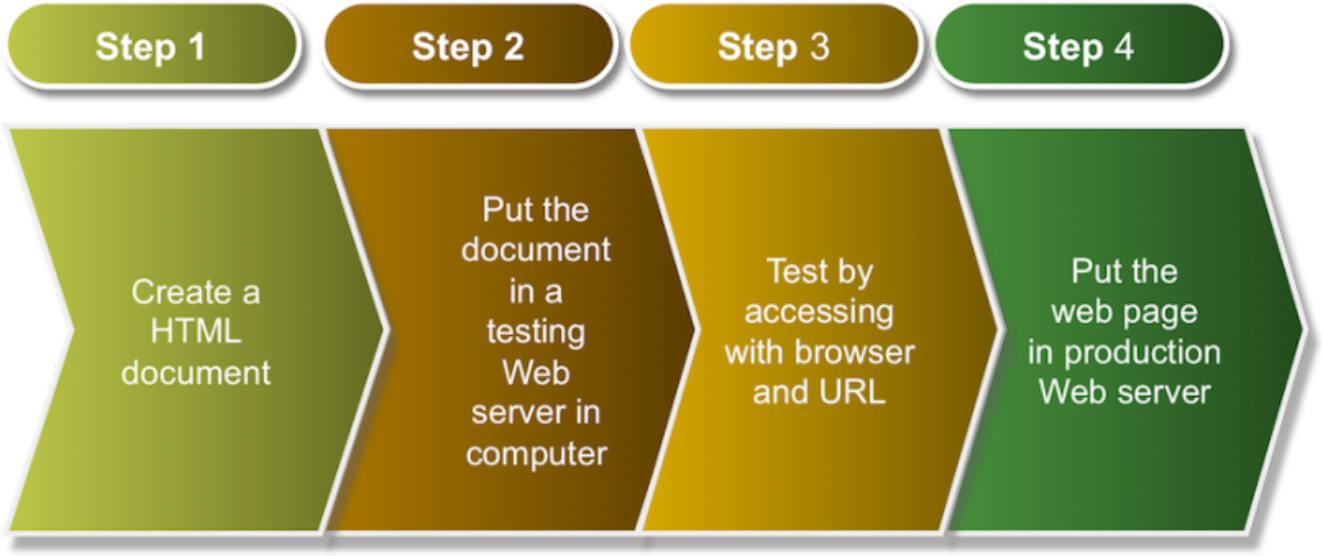
```
<div id="me">div's content</div>
```

2.3. Publish A Web Page

Publish means putting the Web page in a Web server for public access.

The Web page can be written with any editor. The editor introduced in this course is Atom. It is up to you to use any tool you like, or even a text editor (which we do not recommend). The purpose of this course is not to learn a particular tool, but the idea behind so that you can learn to use any tool easily later.

4 steps to publish:

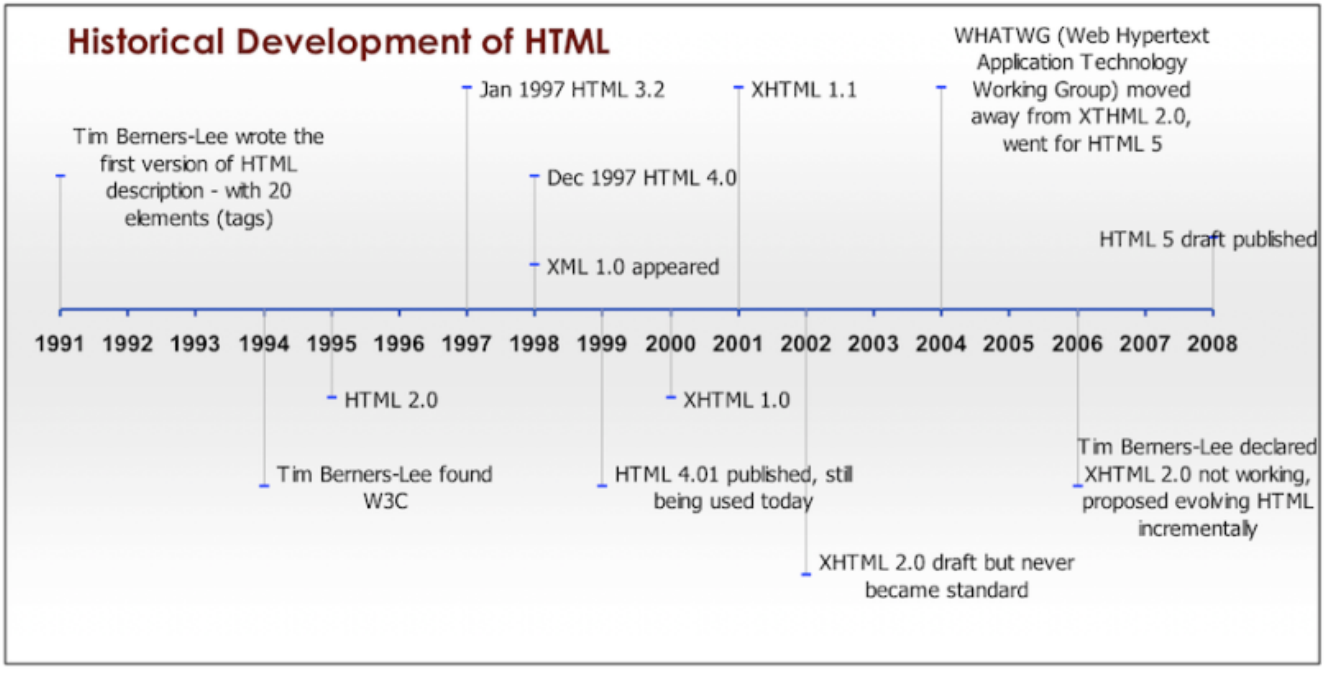


2.4. Versions of HTML | **1.6. the universal resource locator (URL)** | **Komodo - https://www.activestate.com/komodo-ide/downloads/edit** | **Atom - https://atom.io**

2.4. Versions Of HTML

3 main versions: HTML, XHTML & HTML5. Like human languages, markup language develops and changes over time. HTML5 is the prevailing one but old versions can still function in modern browsers

Why do we care? The versions had been standards at a time; new versions have new functionalities; browsers behave differently for different versions and different tools support different versions. HTML5 will be learnt in this course.



2.5. Structure Of A Web Page

Learning HTML is all about structure. 2 main parts: declaration - a line before the head section; followed by one big element <html> inside which divides further into two sub parts <head> <body>. Let's take a look at the example source code.

Head element provides information about the document to the browser by containing other elements. Head section is not supposed to have tags for display (render) by the browser but now most browsers render any tag.

Body element contains all other displayable element for the page content.

2.5.1. Inside Head Section | **Example - http://courses.cs.cityu.edu.hk/cs1303/example/html/01-StructureBasic.html**

2.5.1. Inside Head Section

Common elements:

Comment Element	Description
base	It specifies an explicit URL used to resolve links and references to external sources, such as images, and style sheets.
link	It enables the current document to establish links to external documents, such as style sheets.
meta	It conveys information about the document to the server and the client.
script	It specifies a script for the page that is interpreted by a script engine.
style	It specifies a style sheet for the page.
title	It contains the title of the document.

2.5.2. Meta Tags

The <meta> tags provide information about your page but will not be shown (rendered), that is why it is called meta information. Different use of meta tag determined by the attributes:

- the http-equiv attribute controls the http protocol header, browser would react differently
- the name attribute determines the usage, e.g. author, description or keywords and the content attribute contains the actual value
- description and keywords allow you to influence crawler/spider programs that support the tags so that search engines can index your pages better and hopefully the quality of search results may be improved

Most meta tags are optional but the kewords, descriptions, etc. should be used.

```
<meta http-equiv=“   “ content=“   “ />
<meta name="author" content=“…" />
<meta name="description" content=“…" />
<meta name="keywords" content=“…" />
```

2.5.3. Multi-language handling with Meta tag

2.5.3. Multi-Language Handling With Meta Tag

Multi-language support is a must nowadays in Web page. It is straight forward in HTML5 although troublesome and tedious in XHTML. First make sure the encoding is correct and then set the character set with <meta> tag.

Encoding is the way characters are represented internally in computer as well as storage. Need to consider how many bits are used for a character - 8 bits can only represent 256 characters, how many commonly used characters in Chinese? and Japanese? Korean?

Unicode is a universal system for encoding all of the characters in all of the world's languages. Each character in each language is assigned a unique code. Can be implemented by different methods, e.g. UTF-8 (Unicode Transformation Format) from 1 to 6 bytes (some bits are for control purpose) - most commonly used in Web pages now. Other schemes include UTF-16, UTF-32.

Use the meta tag in the head section. Don't forget to save in utf-8 format with your Web page editor.

```
<meta charset="utf-8">
```

2.6. Body Section

Represent the beginning of the visible part of the document body. Contain the content of the Web page in the form of other elements including text, images, links, etc.

Example common tags are:
section headings
`<h1> to <h6>`
paragraph
`<p>`
comments
`<!-- -->`
line break and horizontal line
`<br> <hr>`

2.6.1. Building Structure In The Body

Structure is important in Web page. It should reflect the structure of your content, also more structures can help later in setting the styles.

Structural tags are used to create structure, they are not seen in the Web page. The most common ones are division - grouping different elements into a single unit; span group some text within a line of text or paragraph.

Newer Structure not found in XHTML:
<header> <nav> <section> <footer>, using these give more meaning to the structure of a page. Will understand more when we learn layout techniques (CSS).

2.6.2. Links | **Example - http://courses.cs.cityu.edu.hk/cs1303/example/html/02-StructureDivSpan.html** | **Example - http://courses.cs.cityu.edu.hk/cs1303/example/html/03-StructureNewer.html**

2.6.2. Links

Links are the distinguishing features of the WWW (World Wide Web) - Hypertext from which HTML/HTTP is called. By setting links, you will be able to jump from one page to another or pointing to any resource.

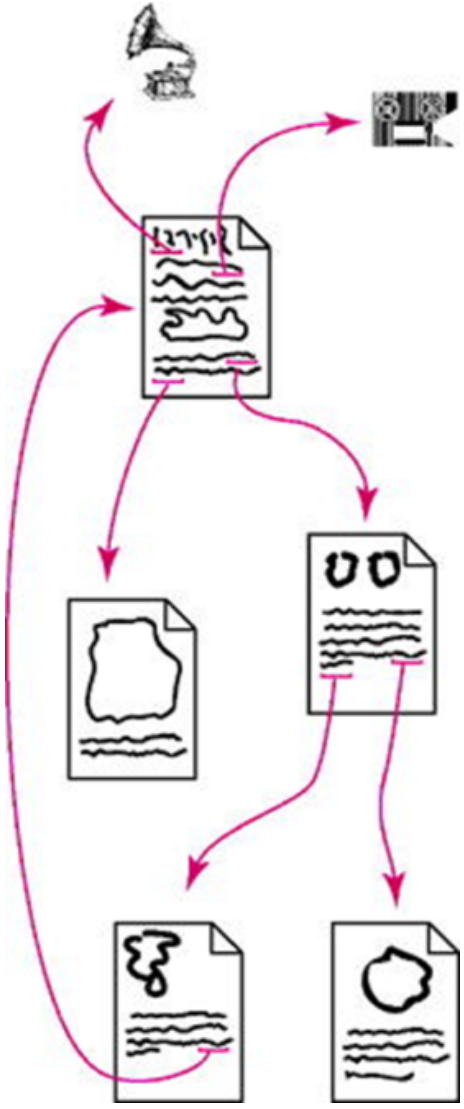
A link has three parts:

- destination: it specifies, in the form of a URL, the pointed at location
- label: it is the words shown on the screen and is clickable; it can be text, an image or both
- target: it determines how the new page will be displayed, a new window (_blank) or a particular named window (window name)

You can create links from one page to the next and further.

```
<a href="destination">label</a>
```

Be careful with the lower/uppercase letter for the URLs, i.e. the destination page name. Don’t make the link’s label too long.

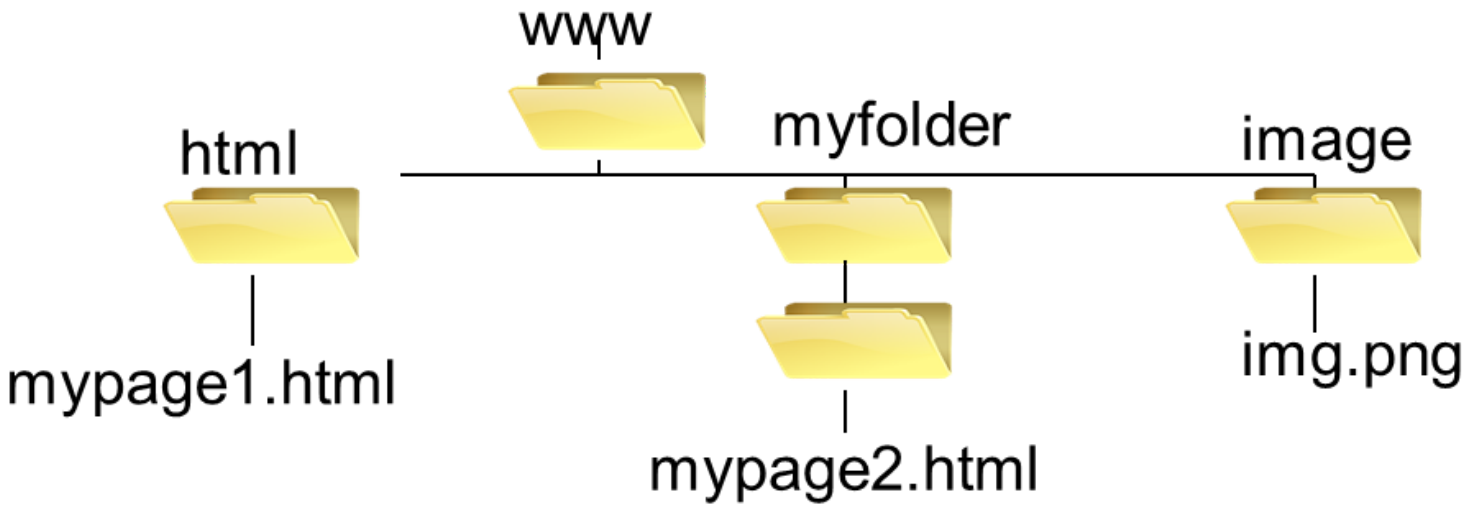


2.6.2.1. Web Site Organization and Links

2.6.2.1. Web Site Organization And Links

Files and resources in a Web site are usually stored in folders arranged in different levels (tree like structure).

In the URL, the parts up to the domain name (with port no. if used) will point to the folder at the highest level - root folder (e.g. www, this is an internal name not visible to outside). the remaining path will then point to different resources level by level (e.g. http://www.site1.com/html/mypage1.html)



2.6.2.2. Absolute And Relative Links

Let's look at how the href can be written. Recalling the URL format:
protocol | domain name | port no. | path

Absolute link - use the full URL and the path is a complete path starting from the Web site root folder (may not be the computer system root folder); usually used to point to other server's resource from your own server.

Relative link - use only the path and skip all other parts; the path is written with reference from the current HTML file; usually used to point to resources in the same server.

You should know the difference of the two and be able to use them appropriately. This is important when you use your local computer for testing and publish to a Web server for production. If you use absolute links all the time (for convenience?), what would happen when you move your pages to the Web server? Be aware some WYSWUG editors generate absolute links for you.

2.6.2.3. Example absolute and relative links

2.6.2.3. Example Absolute And Relative Links

Referring to the Web sites below, pointing to the image file in different ways:

- from mypage1.html - ../image/img.png (relative)
- from mypage2.html - ../../image/img.png (relative)
- from otherpage.html - http://www.site1.com/image/img.png (absolute)

2.6.2.4. Internal Link

Links pointing to specific locations inside the same page. Used in long page to move to the top or bottom, e.g. discussion forum, table of content, etc.
From:

```
<a href="#pagelocation">Go there</a>
```

Point to:

```
<p id="pagelocation"></p>
```

You can set up target locatons by giving any HTML element an id. Note that id must be unique within the page, i.e. no two elements can have the same id.

Can extend this internal link idea to locations in other pages (even in other server) provided that the target id is known. Just put the page link before the # sign.

```
<a href="http://abc.com/xyz.html#pagelocation">Go</a>
```

2.6.2.5. Other Forms Of Link

There are other useful ways of setting up link with the target attribute.

Open new window instead of re-loading the current page; useful when we want to refer to new information together with the existing.

```
<a href=" …." target="_blank">Go</a>
```

The value _blank is defined in the standard, must be used.

Re-loading part of the page using iframe or other window if multiple windows are in use. Set the target attribute to the name of the iframe or window. You will know better when iframe is discussed later.

Image link - use an image instead of text as the label. Useful when we use icon (image) in navigation bar/menu but note that sometimes users have to guess what the icon means.

```
<a href="destination.html"><img src="image.gif" alt="name"></a>
```

2.6.3. Image

The first element that make Web page multimedia. The image itself is not included in the HTML, a source file is needed to store the actual image which could be in different formats: e.g. Bitmap (BMP), Interchange Format (GIF), Photographic Experts Group (JPEG), Network Graphics (PNG), etc.

The image file is pointed with a URL set in src (source) attribute. Basic syntax:

```
<img src="url-for-image-file">
```

The image tag is also good for introducing the concept of good practice. The img tag used in the example is:

```
<img src="../images/loans1.jpg" width="105" height="85">
```

which is correct in syntax and gives the ability to control the dimensions of the image, however, this is not a good practice as HTML is supposed to declare the structure of the page only. Anything related to appearance, layout or style should not be used even though, due to historical development, HTML provides functions to do it.

The good practice is separate the structure from style, do not mix.

2.6.3.1. Alternate Text & Title

The image tag used in the example page has 2 more attributes: alt and title.

- alt contains a text to be displayed when the image cannot be shown due to whatever reason. This is a mandatory attribute defined in the standard missing which your page will fail in valiation
- title contains a text to be displayed whenever the image is pointed by the mouse cursor (mouse over).

Alternate text has another function when the page is processed by non-visual browser (e.g. voice) for visually impaired users. It provides a description about the image which will be read out. This allows different kinds of users to access the Web page.

Title serves similar purpose by showing the image description when mouse-over so that the user does not need to guess the meaning (useful if the image is an icon).

2.6.3.2. Accessibility

Enabling different kinds of users to access your Web page with a good experience is an importance issue nowadays. Have to consider different situations, visually impaired user, hearing impaired, people cannot use mouse (or mouse available in tablet), color blinded or elders who cannot see clearly, etc.

In some countries, a Web site could be sued if it does not follow accessibility guidelines. In Hong Kong, the Office of Government Chief Information Officer (OGCIO) has a unit to promote accesibility working with universities and give out certifications.

Some common considerations:

- must include alt for all images
- allow font size adjustment by user
- choose color carefully, preferrably with high contrast
- provide subtitle if using video or audio

Guidelines are available in W3C. It allows different levels of conformance because lots of effort may be required to fulfill all requirements.

2.7. Conclusions

Extension readings : Text book chapter 1 - 3. http://www.w3cschools.com can be the main reference site containing both tutorials and references. Can refer to it when you are working online.

Critical thinking:

- Why do I need to worry about all these?
- Why not just learn to use a tool?
- What kind of tools should I use?