

Week 8 - Day 2

Ch. 4 HTML (hypertext markup language)

- Inventor: Tim Berners Lee (1989), a British computer scientist while he worked @ CERN, the European Particle Physics Laboratory (Geneva)
check out: <http://www.w3.org/People/Berners-Lee/>
- HTML = an encoding (or a language) describing structured documents & that allows documents to be linked

Structured document: document containing

- title
- chapters, formatting,
- sections & sub-sections
- figures
- tables, etc...

With HTML, we can represent struct. docs with unstructured text (a long sequence of alphabet characters, punctuation marks & other symbols, perhaps encoded using ASCII table)

Why? Networks can transfer text (unstructured). Extra information about the structure of document (font, italic, bold, sections, title, etc) must be encoded in text.

- HTML + HTTP = the web.
HTTP = hypertext transfer protocol
→ follows the client/server communication model

→ it specifies a set of rules obeyed by web servers & browsers when they transfer data

→ it is built on top of internet protocol for transmission of data (TCP/IP)

→ consists of simple text messages exchanged between a browser (client) and web server.

(similar to the protocol used in wall climbing gym, for ex, between climber and person on belay)

Homework

Try this:

- on Windows, click Start / run

- type "telnet [address of a web server] 80"

ex "telnet www.google.ca 80"

- in the window, type "get index.html"

→ you should receive text representing the HTML encoded index web page.

→ explanation: you have just carried out one step of the many steps taken by a web browser when contacting a web server; you used "telnet" (a simple program designed to send text data to computers across the net) to send message "GET" which is part of the messages used by HTTP protocol.

How to encode structure & formatting info? (back to HTML)

→ tags :- special words enclosed in angle brackets.

(ex) `<title>` (title tag)

most tags are paired $\left\{ \begin{array}{l} \text{begin} \quad \text{<title>} \\ \text{end} \quad \text{</title>} \end{array} \right.$
(end tags start with slash '/').

tags assign meaning to the text between begin & end pair

(ex) `<title> CPSC 1000 </title>`

specifies that text "CPSC 1000" is the title of the web page

- Mandatory tags. Any HTML web page has:

```
<html>
  <head>
    ... preliminary material ...
  </head>
  <body>
    ... page content ...
  </body>
</html>
```

→ preliminary material: or the title of the document

Page content, structuring documents

a) headings (section titles)

`<h1>` Heading text `</h1>` → main section

`<h2>` - | - `</h2>`

o
o
o

`<h8>` - | - `</h8>` → least important section

→ the browser interprets these tags & displays the heading text according to heading importance.

b) White space (newlines, tabs, space characters)

in HTML code is ignored by browser.

For new paragraph, use "paragraph tag":

`<p>` paragraph text ... `</p>`

c) Lists of items

→ numbered
↘ un-numbered
↓ definition lists

Numbered:

1. One
2. Two
- ⋮

Un-numbered

- One
- Two
- ⋮

Definition list

- One
A number
- Two
Another number
- ⋮

Numbered list: (ol = ordered list)

 One (li = list item)

 Two

Un-numbered list: (ul = un-numbered list)

- same as above, but replace "ol" tag with "ul" tag

Definition list (dl = definition list)

<dl>

<dt> One </dt> (dt = definition term)

<dd> A number </dd> (dd = definition data)

<dt> Two </dt>

<dd> Another number </dd>

</dl>

Options - tags may be given optional arguments (or attributes)

(ex)

<dl compact>




o

→ argument

</dl>

→ tells the browser to display the definition term on the same line with definition data.

Attributes go between angle brackets, after tag name.

- attributes for paragraphs: align → left 
right 
center 

ex: $\langle p \text{ align} = \text{"center"} \rangle \dots \langle /p \rangle$

↑
paragraph is displayed by browser as centered.

→ align = attribute name
center = its value

Rule is: attribute name equal sign value between double quotes.

Exception: numeric values are not written between double quotes.

(ex) tag $\langle hr \rangle$ → horizontal rule
(tells the browser to draw a horizontal line).

• hr tag with attributes:

$\langle hr \text{ width} = \text{"50\%"} \text{ size} = 1 \rangle$

→ two attributes (width, size)

← how long the line is

↓ how thick the line is

• hr tag has no ending pair because it does not tag text, it just inserts a line

Anchors (anchor tag)

→ links to external information (info stored on other web pages)

` University of Lethbridge `

→ attributes: href (hypertext reference); its value is the URL (uniform/universal resource locator) or location for the external information.

URL: `http://www.uleth.ca/index.html`

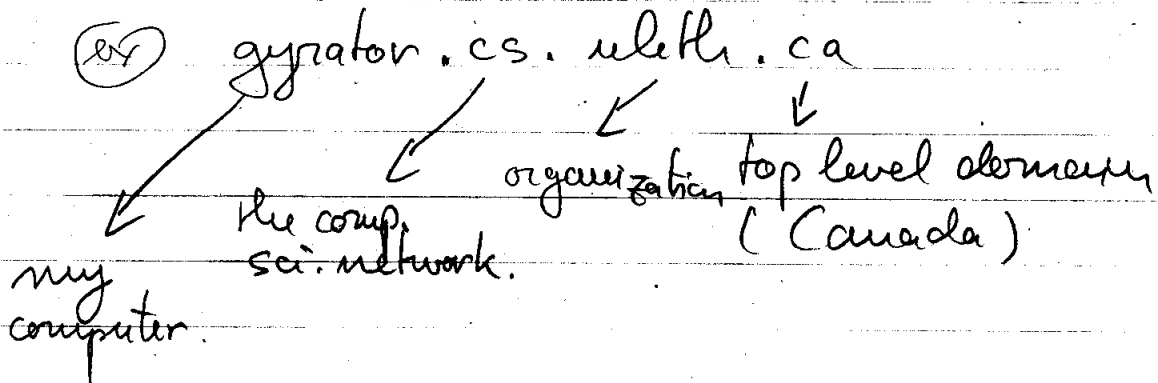
(or web address)

<u>http://</u>	<u>www.uleth.ca</u>	<u>/index.html</u>
protocol name used for accessing the info	address (or name) of the computer containing the info	filename containing the information
↓	↓	↓
tells the browser how to actually retrieve the info (eg: using HTTP)	computer name or internet address	the location for the information on the remote computer.

Computer name: any machine on the internet is identified by the internet address (4 bytes or 16 bytes).

↓
difficult to remember, so any internet address has a name associated (domain name)

- there are servers (computer programs) on the internet that have tables associating addresses with names (domain name servers or DNS)
- names are hierarchical.



- Not all internet names & addresses are stored by one server; there are many DNS each storing a small portion of all names.

- When a browser follows a link:
 - a) browser asks a DNS for the address of web server
 - b) browser connects to the computer whose address it just received & starts dialog according to protocol.

e) Eventually, browser receives content of file & displays it on screen.

Obs about anchors

→ The location of the webpage is ONLY given by the href attribute. The text between `<a...>` and `` tags is just displayed as a link & is NOT used to locate data. The value of href attribute is not displayed by the browser.

Eg. `www.royalbank.com `

→ The connection is established with a computer @ address 192.168.1.54 and not with Royal Bank's web server as one might believe (cheap trick used by phishers to trick you in getting personal info)

Absolute & relative links

Absolute vs relative web page links.

- a value for href that starts with

`http://computer name_or_address/location_for_file`

is absolute

- if the value for href starts with
/directory/filename

then it is assumed that the connection is established with the server providing the page that is currently viewed.

→ this simplifies construction of web sites composed of many different files. We can link to these files without specifying the server. If the address of the web server changes, we don't have to update all the links to the files making up the web site.

Images (img tag)

There is no ending tag for img.

Attributes are used to specify

- position of image relative to nearby text (align)
- size of image (height, width in pixels)

Colors

- given using names → restricted to a few samples
- given using RGB values → more general

→ used : - globally in the <body> tag

ex: <body bgcolor = "red" text = "blue" >

- locally, using tag.

ex: <p> This is default color.
This is red. </p>

RGB values : specify the intensity for Red, Green, and Blue. They are written as hex numbers (base 16)

(ex) # FF 8B 2A

 red green blue

FF = 255 = max intensity
00 = 0 = min. intensity

Colors in HTML (color)

Hex numbers use digits 0...9 and

A=10 B=11 C=12 D=13 E=14 F=15

$$\text{Hex \# } 8E = 8 \cdot 16^1 + 14 = 128 + 14 = \underline{\underline{142}}$$

(This is the intensity of green in the previous example)

Formatting text

Bold text is specified using `...` tag

↓
this text is displayed in bold characters

Italic text

`<i>...</i>`

Tables

A	B	C	D
1	2	3	4
x	y	z	z

`<table>` → table row; defines a row

`<tr>` → table data; defines a cell

`<td> A </td> <td> B </td>`

`<td> C </td> <td> D </td>`

`</tr>` → end of a row

`<tr> <td> 1 </td> ...`

`</tr>`

...

Summary creating a web page using html code:

<html>

<head>

<title> My document </title>

</head>

<body text = "white" bgcolor = "black">

↑ foreground ↑ background (optional)

<h1> My title </h1>

<h3> Abstract </h3>

<p> This book ... </p>

<h2> Chapter 1. Introduction </h2>

<p> insert paragraphs ... </p>

... insert links as desired, pictures etc ...

</body>

</html>