CP 476
Internet Computing

Instructor: Masoomeh Rudafshani

Agenda

- HTML History
- HTML Syntax and Rules

HTML: HyperText Markup Language

- Purpose
  - General layout and structure of documents that could be displayed by a wide variety of computers
- HTML was defined with SGML
  - Standard Generalized Markup Language
  - SGML is a meta-markup language

HTML Versions

- HTML 1.0
  - Initial version of HTML
  - Developed along with the structure of Web and the first browser at CERN by Tim Berners-Lee
  - 20 tags, a proof of concept
  - TBL started W3C in 1994 to standardize HTML
    - marketing competition between Netscape and Microsoft war
- HTML 2.0
  - Released in 1995
  - The first officially released HTML standard
  - 50 tags, widely used

HTML Versions

- HTML 3.2
  - 1997
  - 67 tags
  - since 1998 the evolution of HTML has been dominated by W3C
    - in part because Netscape gradually withdrew from its competition with microsoft.
  - There are now several different organization that produce and distribute browsers, all of which except microsoft followed relatively closely the HTML standard produced by W3C
- HTML 4.0
  - Released in 1997
  - Introduction of style sheets
  - Introduced many new features and deprecated many older features

- HTML 4.1
  - 1999
  - A cleanup of 4.0
  - Problems:
    - Specifies loose syntax rules
    - Its specification does not define how a user agent (most often a browser) is to recover when erroneous code is encountered
  - Solution: XHTML

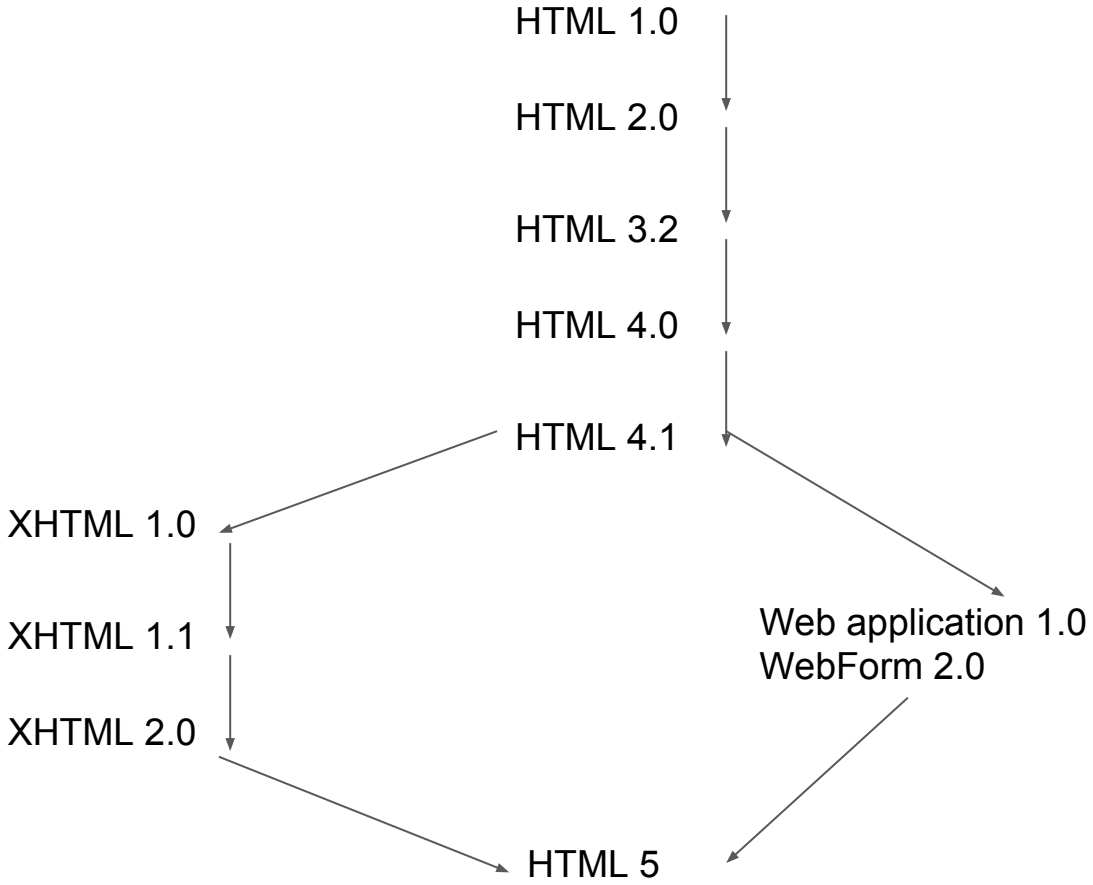
- XHTML 1.0
  - Redefinition of HTML 4.1 using XML
  - XML specification requires that XML processor not accept XML documents with any errors
  - Address one of the problems in HTML 4.01 by providing complete rules stating what is and what is not syntactically acceptable
  - XHTML documents must be served with the application/xhtml+xml MIME type
    - This means user agents were required to halt interpretation of an XHTML document when the first syntactic error was found, as is the case with XML documents.
  - This was a drastic response to the error-ignoring nature of HTML 4.01
    - XHTML 1.0 specification included Appendix C, which allowed XHTML documents to be served as HTML (with text/html MIME Type)
      - Allows the continuation of the practice of user agents ignoring syntax errors

- XHTML 1.1
  - RELAtively minor additions to XHTML 1.0, but eliminated the Appendix C
  - Officilaly specified that XHTML 1.1 documents had to be served as application/xhtml+xml MIME type and that user agents were required to reject all syntactically incorrect documents.
    - Draconian error handling
  - Recommended by W3c in 2001
  - A modu;aroizatopm pf XHTML 1.0
  - Drops some of the features of its predecessor
  - Although the value of the consistent and coherent syntax rules of XHTML were widely recognized and acceptedm the draconia error handling was not
  - It was generally approved by developes and provides of markup documents that browser users should not be given error messages dute to the syntax errors found in documents they were attempting to view
    - XHTML documents were still served with the text/html MIMe type and browser continued to use forgiving HTML parser.

- XHTML 2.0
  - W3C ignored these issues
  - Not requiring it to be backward compatible with HTML 4.01 or XHTML 1.1
- WebForm 2.0 and Web application 1.0
  - In reaction to the XHTML 1.1 and 2.0
  - Developed by WHAT (Web Hypertext Application Technology) working group
    - Based on HTML 4.01 rather than XHTML 1.1
  - Goals:
    - Backward compatibility with HTML 4.01
    - Error handling that is clearly defined in the specification
    - Users would not be exposed to document syntax errors
    - WebForm 2.0: Extended HTML form
    - Web application 1.0: A new version of HTML,An algorithm for user agenet error handling

- HTML 5
  - AFter several years of separate work, W3C on XHTML 2.0 and the WHAT working group on a new version of HTML
    - Head of W3C made the momentous decision in 2006 that W3C would begin working with WHAT working group
    - In 209

History of HTML



HTML History

https://www.w3schools.com/html/html5_intro.asp

Reasons to use XHTML syntax rules:

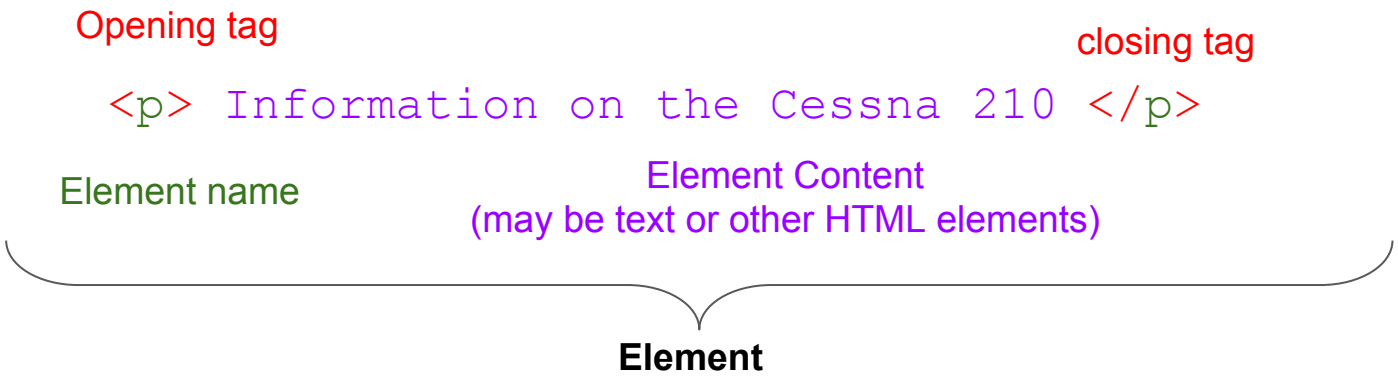
- HTML has lax syntax rules
  - leading to sloppy and sometimes to ambiguous documents
- XHTML syntax is much more strict
  - Leading to clean and clear documents in a standard form
- HTML processors do not even enforce the few syntax rule that do exist in HTML
- The syntactic correctness of XHTML documents can be validated
- In this course, HTML5 is used, but with XHTML syntax rules
  - HTML validator

Demo in Eclipse

Different version of HTML

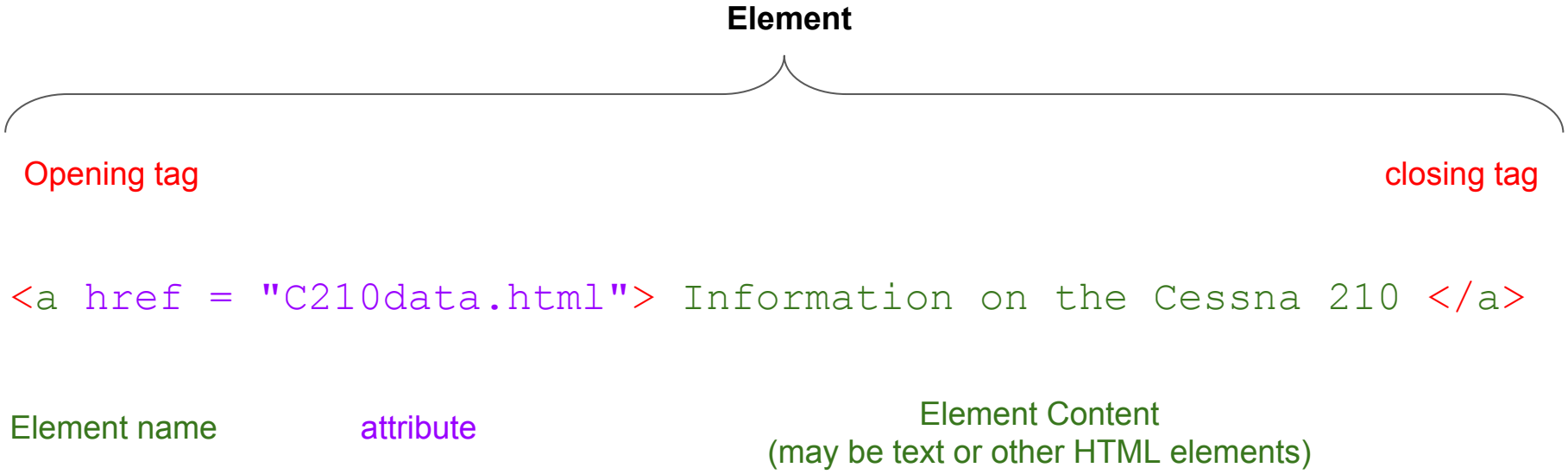
Basic HTML Syntax: HTML Element

- HTML documents is composed of:
  - Textual element
  - **HTML element**
    - **Element name** within angle brackets (**tags**)
    - Opening and closing tags
      - Together they specify a container for the content
    - **The content** within the tag



Basic HTML Syntax

- Each element may contain **attributes**



HTML Syntax

- **Empty element**
  - Does not contain any text content
  - Instructs the browser to do something
  - Trailing slash
    - Had to be there in XHTML
    - Optional in HTML5

Trailing slash

Opening tag

```
<img src = "C210.jpg" alt = "something" />
```

Element name

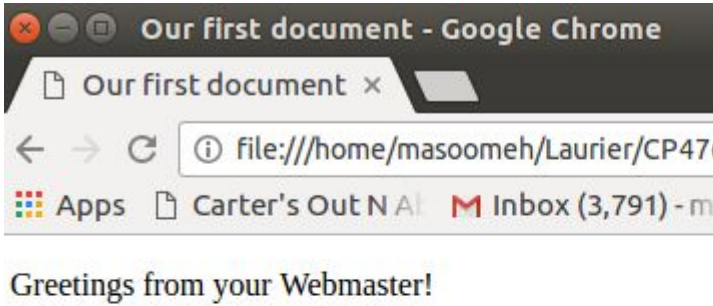
attribute

HTML suntax

- Comment form: <!-- … -->
- Browsers ignore comments, unrecognized tags, line breaks, multiple spaces, and tabs
- Style guide
  - https://www.w3schools.com/html/html5_syntax.asp
  - In XHTML, element and attribute names must be in all lowercase letters
  - In HTML, they can be any combination of uppercase and lowercase

Structure of HTML Documents

```
<!DOCTYPE html>
<!-- greet.html A trivial document -->
<html lang = "en">
<head>
<title> Our first document </title>
<meta charset = "utf-8" />
</head>
<body>
<p>
Greetings from your Webmaster!
</p>
</body>
</html>
```



HTML Structure: DOCTYPE

- Placed at the top of a web page document
- Document Type Definition (DTD)
- Tells the browser what type of document it is about to process
    - Identifies the version of HTML contained in your document.
- Examples

HTML(5): <!DOCTYPE html>

XHTML

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.1//EN"
"http://www.w3.org/TR/xhtml11/DTD/xhtml11.dtd">
```

HTML Structure: <html> </html>

- contains all the other HTML elements in the document
- Root element of the document
- `<html lang = "en">`
  - Specifies the language in which the document is written
    - HTML document will be indexed more effectively by search engines if its language is set
      - allowing it to appear correctly in language-specific results, for example)
    - More accessible web design
      - It is useful to people with visual impairments using screen readers (for example, the word "six" exists in both French and English, but is pronounced differently.)

HTML Structure: head and dody

- <head> contains information that describes the web page document
  - title, style sheets, JavaScript files etc.

```
<head>
…head section info goes here
</head>
```

- <body>
  - Contains the documents'content

```
<body>
 …body section info goes here
 </body>
```

HTML Structure: <head>

- Provide information about the document
  - The <title> tag is used to give the document a title,
    - which is normally displayed in the browser's window title bar (at the top of the display)
    - When bookmarking a page
  - The meta element
    - Used to provide additional information about a document
    - It has not content and no closing tag
    - Necessary element by XHTML, not by HTML
      - <meta charset = ″utf-8″ />
      - declares that the character encoding for the document is UTF-8.
      - is used to provide the character set used to write the document

Basic Text Markup

- Paragraph Elements
  - The <p> tag breaks the current line and inserts a blank line - the new line gets the beginning of the content of the paragraph
  - The closing tag is required in XHTML, not in HTML
  - Line breaks inside paragraph are ignored by the browser
  - The closing tag is required in XHTML, not in HTML
  - Examples
  - https://www.w3schools.com/tags/tryit.asp?filename=tryhtml_paragraphs2
- Line breaks
    - The effect of the <br /> tag is the same as that of <p>
    - except for the blank line (in HTML, it
    - It could be just
      - No closing tag!

Basic Text markup

- Preserving whitespace
  - The text content of a <pre> element is displayed as it is entered
  - Example:
  - <pre> element can contain other tags
    - Except those that cause a paragraph break
    - Special characters must be avoided
      - Character entities
    - https://www.w3schools.com/code/tryit.asp?filename=FZ7IXY7QFK6Y

Basic Text markup: Font styles and sizes

- Font Styles and Sizes (can be nested)
  - <i> and <b>
    - deprecated
  - Emphasis - (often set in italics)
  - Strong - (often set in boldface)
  - Monospace - <code> (often set in Courier)
  - , , and <code> are not affected if they appear in the content of a <blockquote>,
    - unless there is a conflict (e.g., (italics))

Basic Text markup: Font styles and sizes

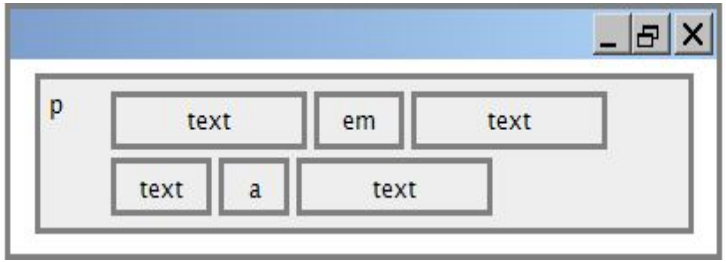
- Superscripts and subscripts
  - Subscripts with <sub>
  - Superscripts with <sup>
  - Example: x₂³
    - https://www.w3schools.com/code/tryit.asp?filename=FZ7J7TOVBD31

Basic Text markup: Headings

- Headings
  - Six sizes, 1 - 6, specified with <h1> to <h6>
    - 1, 2, and 3 use font sizes that are larger than the default font size
    - 4 uses the default size
    - 5 and 6 use smaller font sizes
    - Example
- Block Quotation
  - Content of <blockquote>
  - To set a block of text off from the normal flow and appearance of text
    - In many cases it is a long quotation
  - Browsers often indent, and sometimes italicize

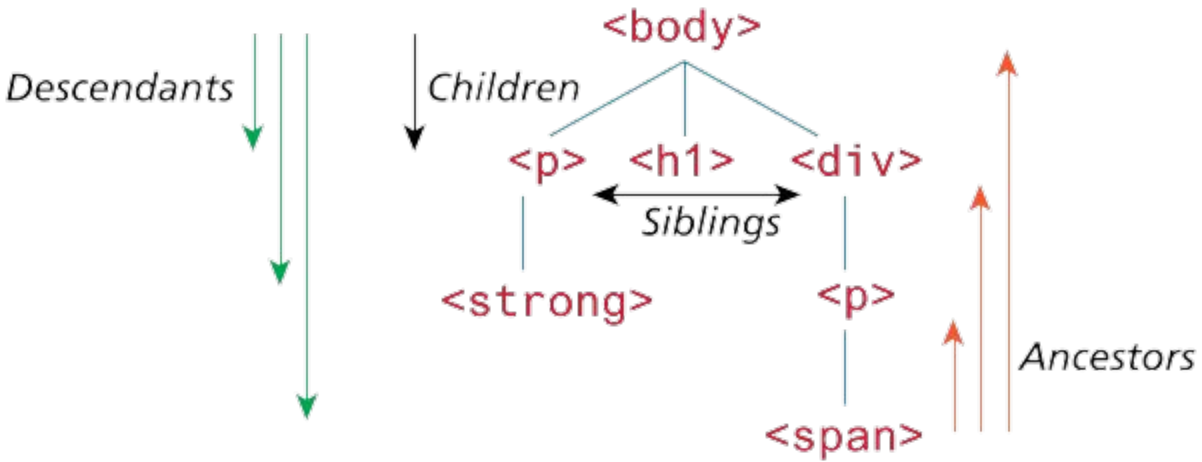
Block and inline elements

- Block
  - Examples: paragraph, Heading and blockquote
  - Contain an entire large region of content
  - Cannot appear in the content of inline element (in XHTML)
  - the browser places a margin of whitespace between block elements for separation
- Inline
  - Examples: em, strong
  - Must be nested inside a block element
    - Greet.html example: text content of its body nested in a paragraph element
  - Inline elements and text cannot be nested in body or form element



Taken from [2]

HTML Element Tree



Block and inline elements

- Distinction was used in HTML specifications up to HTML 4.01
- HTML5
  - Content categories
    - Metadata content
    - Flow content
      - Block-level and inline-level elements combined
    - Section content
    - Heading content
    - Embedded content
    - Phrasing content
      - Inline category roughly correspond to this category

Taken from [2]

Basic Text markup: Character Entities

- These are special characters for symbols for which
  - Full list available at:
    - https://www.w3schools.com/html/html_entities.asp
  - there is either no easy way to type them via a keyboard
  - or which have a reserved meaning in HTML (like“<“)

Char	Entity	Description
space		Non-breaking space
&	&amp;	Ampersand
<	&lt;	Less than
>	&gt;	Greater than
"	&quot;	&quot;
'	&apos;	Single quote

Basic Text markup: Horizontal rules

- <hr /> draws a line across the display, after a line break
  - should be immediately closed with />
  - In HTML can be written as <hr>
  - https://www.w3schools.com/code/tryit.asp?filename=FZ7KF8A2RCEU

Image

- Most browser can render images in these formats:
  - GIF (Graphic Interchange Format)
    - 8-bit color representations for pixels
    - 256 different colors
  - JPEG (Joint Photographic Experts Group)
    - 24-bit color (16 million different colors)
    - Compression is better
    - No support for transparency
  - Image files are compressed
    - Reduce storage
    - Allow faster transfer over the internet
  - Portable Network Graphics (PNG)
    - Relatively new
    - Files are bigger than jpeg – no lost data!

Image Element

- Images are inserted into a document with the <img /> tag
- Specifies an image that is to appear in a document
  - Inline element
  - Has no content (no closing tag)
  - 30 different attributes:
    - Src: specifies the file containing the image
    - Alt: specifies the text to be displayed
      - Required in XHTML
      - Non-graphical browsers
      - Browsers with images turned off
    - Width, hegiht
      - Specify the size of the rectangle of image in pixels

Proper Nesting

https://www.w3schools.com/code/tryit.asp?filename=FZ7KF8A2RCEU

List

- Unordered list
  - The list is the content of the <ul> tag
  - List elements are the content of the <li> tag
- Ordered list
  - The list is the content of the <ol> tag
  - Each item in the display is preceded by a sequence value
- Nested List
  - Any type list can be nested inside any type list
  - The nested list must be in a list item
- Definition List
  - List is the content of the <dl> tag
  - Terms being defined are the content of the <dt> tag
  - The definitions themselves are the content of the <dd> tag

HyperText Link

- Acts as a pointer to some particular place in some web resource
  - An HTML document anywhere on the web
  - The document currently being displayed
- Links are an essential feature of all web pages
  - Links use the <a> element (the “a” stands for anchor).
  - A link is specified with the href (hypertext reference) attribute of <a>

HyperText link

- Source: anchor element that specifies the link
- Target: the document whose address is specified in a link
  - A complete document in the same directory as the HTML document
  - If the document is in another directory, relative addressing should be used
  - Content: text or image (cannot be another anchor element)

 Information on the Cessna 210

HTML Validator

- https://validator.w3.org/
  - Validate against many different versions of HTML
- Download the validation tool from:
- http://totalValidator.com
  - Validate against XHTML 1.0

References

1. Introduction to HTML, https://developer.mozilla.org/en-US/docs/Learn/HTML/Introduction_to_HTML
2. ***Web Programming Step by Step***, Marty Stepp, Jessica Miller, and Victoria Kirst