What is HTML?

- Hypertext Markup Language
- Markup language
  - "Set of words or symbols"
  - Assigns properties to text
  - Not actually part of the text
- HTML specifies the *content* of a page
  - Used to also specify the appearance and behaviour
  - Those are now done by stylesheets and scripts
- World Wide Web Consortium
- Develops and recommends specifications
  - Producing web pages (e.g., HTML, CSS)
  - Rendering and searching web sites
  - Protocols (e.g., HTTP)
  - Internationalization
  - Mobile devices
- Headed by Tim Berners-Lee
- Does not actually set standards
  - The Internet Engineering Task Force (IETF) does
- www.w3c.org

A Brief History of HTML

- WorldWideWeb
- First web browser
- Written in Objective-C
A Brief History of HTML

- 1991: created by Tim Berners-Lee at CERN
  - Developed first browser called WorldWideWeb
- 1994: HTML2 standard published by IETF
- 1996/97: HTML3.2 standardized
  - Features forms and internationalization
- 1997: HTML4 released by W3C
  - Features stylesheets and scripting

- 1999: HTML4.01 approved by W3C
  - Was commonly used for many years
- 2000-01: XHTML1 released by W3C
- 2004: Work on HTML5 begun
  - Goal: replace both HTML4 and XHTML1
  - Initially begun by WHATWG
    - Web Hypertext Application Technology Working Group
    - Includes Apple, Mozilla and Opera
- 2007: W3C begins work with WHATWG on HTML5
2011-12: W3C splits with WHATWG on HTML5
- W3C wanted an “HTML5 Recommendation”
  - A final specification
- WHATWG wanted a “Living Standard for HTML”
  - An evolving specification

Presently: there are two HTML5 specifications
- WHATWG's is evolving
- W3C's is a recommendation (28 October 2014)
  - Contains portions of WHATWG's

Some Notes on HTML

From both the W3C and WHATWG specifications (sections 1.5 and 1.7, respectively):

> It must be admitted that many aspects of HTML appear at first glance to be nonsensical and inconsistent.

> HTML, its supporting DOM APIs, as well as many of its supporting technologies, have been developed over a period of several decades by a wide array of people with different priorities who, in many cases, did not know of each other's existence.
Some Notes on HTML

- The idea of a “living standard” in a good one
  - Browsers have often supported features ahead of specifications
  - Can encourage fast adoption of new features and improvements

- Standards are also good
  - Promote stability and interoperability
    - Same support across platforms and devices
    - Forces updates for elements which browsers may otherwise choose to ignore

- What ultimately matters is what browsers implement
  - What browsers and versions are required for your site?
  - What do they support and how?
  - That's what you'll use and/or have to work around

- Focus on what's widely supported
  - Many HTML5-specific things
    - e.g., backwards-compatible DOCTYPE
  - Many (X)HTML things that are also in HTML5
### HTML Comparison

<table>
<thead>
<tr>
<th>Approach</th>
<th>Older HTML</th>
<th>XHTML</th>
<th>HTML5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syntax Rules</td>
<td>Lax</td>
<td>Strict</td>
<td>Less Strict</td>
</tr>
<tr>
<td>Manages</td>
<td>Content, Style,</td>
<td>Content</td>
<td>Content</td>
</tr>
<tr>
<td></td>
<td>Behaviour</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Browser Support</td>
<td>Waning</td>
<td>Excellent</td>
<td>Excellent</td>
</tr>
<tr>
<td>Validation</td>
<td>Waning</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Closing Tags</td>
<td>Not Always</td>
<td>Required</td>
<td>Not Always</td>
</tr>
<tr>
<td></td>
<td>Required</td>
<td></td>
<td>Required</td>
</tr>
</tbody>
</table>

### HTML Deprecation

- As frameworks evolve, not everything is carried forward
  - Some functionality may no longer be valid
    - Redesigned and replaced
    - Flawed or not really necessary
- Deprecation
  - Disallowed, but not disabled
    - It may work, it may not
  - Attempt to preserve *backwards compatibility*
    - Previous versions may still work
- Many older HTML elements are now deprecated or no longer supported
Extension for HTML files is .html

Main page for a website is named index.html
  • Recognized by most web servers

Ignores most whitespace

Comments about the webpage can be inserted
  • They are not displayed by the browser
  • They may help you manage your content

<!-- comment -->

Markup is done with tags
  • Attributes and values are mostly optional

<tagname attribute ="value" />
  Self-closing tag

<tagname attribute ="value">
  content
</tagname>
  Element
Special Characters

- Reserved by HTML for markup use
- Special entities render these characters as content

<table>
<thead>
<tr>
<th>Character</th>
<th>Entity</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;</td>
<td>&quot;</td>
</tr>
<tr>
<td>&amp;</td>
<td>&amp;</td>
</tr>
<tr>
<td>&lt;</td>
<td>&lt;</td>
</tr>
<tr>
<td>&gt;</td>
<td>&gt;</td>
</tr>
<tr>
<td>~</td>
<td>˜</td>
</tr>
<tr>
<td>space</td>
<td> </td>
</tr>
</tbody>
</table>

and many more …

XHTML Document Structure

```html
<!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/1999/xhtml">
  <head>
    <title>Your title</title>
    <meta http-equiv="content-type" content="text/html;charset=utf-8"/>
    <meta name="author" content="Your name"/>
  </head>
  <body>
    <p>Your content</p>
  </body>
</html>
```
DOCTYPE

- Tells the browser how to render
- HTML5 version is backwards-compatible
  - Will work with older browsers
- XHTML components
  - Version number
  - Standard
    - Strict: must conform
    - Transitional: permits deprecated items
    - Frameset: for use with frames
  - Link to file with document type definition

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

- Identify the beginning and end of the page
- Opening tag must always follow DOCTYPE
- XHTML version assigns a value to the xmlns attribute
  - Determines the XHTML namespace
    - Element and attribute names
    - Not a link to a file

```
<html xmlns="http://www.w3.org/1999/xhtml">
```

---

Header

- Provides information about the page
  - The title and character set should always be specified
    - The example below is HTML5

- Can set metadata
  - “Data about data”
    - `<meta name="item" content="value"/>

- Some predefined meta tags
  - description, ...
- You can specify whatever you like

```
<head>
  <meta charset="UTF-8">
  <title>Your title</title>
  <meta name="description" content="Your description"/>
</head>
```
Container for the page's content

Two types of elements

- **Block elements**
  - Represent significant pieces of content
  - Covers a rectangular block on the page
  - Can be nested within other block elements

- **Inline elements**
  - Represent smaller pieces of content within a block
  - Can be nested within other inline elements
  - Can **not** contain block elements

**Basic Block Elements**

- **Paragraph**
  - `<p>content</p>`
  - Intended for prose paragraphs

- **Headings**
  - `<h#>content</h#>`
  - # ranges from 1-6
    - The lower the #, the larger the text
  - Organize and separate other blocks
Basic Block Elements

- Horizontal rule
  - `<hr/>`
  - Horizontal line that divides block elements

- Block quotation
  - `<blockquote>content</blockquote>`
  - Intended for paragraph-length quotes
  - Must nest another block-level element in this one for XHTML

- Preformatted text
  - `<pre>content</pre>`
  - Preserves whitespace

Lists

- Specify lists of items
  - Displayed in the order given (top to bottom)

- Unordered list
  - `<ul>list items</ul>`
  - Produces a bulleted list

- Ordered list
  - `<ol>list items</ol>`
  - Produces a numbered list

- Both are populated with `list items`
  - `<li>content</li>`

- Lists can be nested in list items
Definition Lists

- Creates a list of terms and definitions
  - `<dl>definition terms and descriptions</dl>`
- Definition term
  - `<dt>content</dt>`
- Definition description
  - `<dd>content</dd>`
- Each term should be followed by a description
  - `<dt>... </dt>`
  - `<dd>... </dd>`

Tables

- Arrange data in a tabular form
  - `<table>caption and table heading and body</table>`
- Caption
  - `<caption>content</caption>`
  - Optional name of the table
- Table header
  - `<thead>table rows with table headings</thead>`
- Table body
  - `<tbody>table rows with table data</tbody>`
Tables

- Table rows
  - `<tr>table headings or data</tr>`
  - Specifies a row in the table

- Table headings
  - `<th>content</th>`
  - Denotes a heading for a column or row

- Table data
  - `<td>content</td>`
  - Specifies a standard cell

- Cells within the table can be merged
  - `<th rowspan="#" colspan="#">content</th>`
  - `<td rowspan="#" colspan="#">content</td>`
  - `rowspan` merges # cells vertically
  - `colspan` merges # cells horizontally