REPRESENTING DATA: HTML, HYPERTEXT & MARKUP

Dr Rachel Salzano
r.salzano@napier.ac.uk

TL/DR

Using HTML to create HyperText Documents...

AIMS

- At the end of this section of the module you will be able to:

  - Understand the major elements of the web

  - Use web browser development tools to support development & investigation of web sites

  - Use HTML to create HyperTextDocuments

OVERVIEW

- At the end of this (sub-section) of the topic you will:

    - understand how HTML has developed & why it works the way it does

    - be aware of the range of tags supported by HTML

    - be able to assemble basic HTML documents

HTML & THE WEB

- HTML is generally processed by Web browsers

- The Web was designed for ease of publication

- To achieve this browsers have traditionally been very accommodating in what they will accept

- So you will see code from lots of versions of the Web & should be in a position to handle it (HTML in general) but we should aim to develop using the latest tools (e.g. HTML5)

HTML

- **H**yper**T**ext **M**arkup **L**anguage

- A **language** for turning **text** into **hypertext** by using **markup**

- The standard markup language for creating web pages

- Not a programming language - no support for programming constructs

- Part of the triad of foundational web technologies (alongside CSS & Javascript)

    - Describes the semantic structure of the data, which CSS presents, and Javascript manipulates

WHAT DOES AN HTML DOCUMENT LOOK LIKE?

```
<!DOCTYPE html>
<html>

  <head>
    <title>My first HTML 5 document</title>
  </head>
  <body>
    <p>Hello World from HTML 5</p>
  </body>
 </html>
```

W3C

Edinburgh Napier
UNIVERSITY

- World Wide Web Consortium (W3C) defines the standards for the web, e.g.

  - HTML 4.01

    - Based on Standard Generalised Markup Language (SGML)

  - HTML 5

    - Backwards compatible but no longer based on SGML

- & various related standards: XHTML 1.0, 1.0, 2 (cancelled)

  - Based on eXtensible Markup Language (XML)

CLASSICAL TO MODERN

- Until 4.01 defined the visual presentation of a web page

  - Mixed structure & presentation

- Modern HTML

  - Describes the content, its structure, and its relation to other content
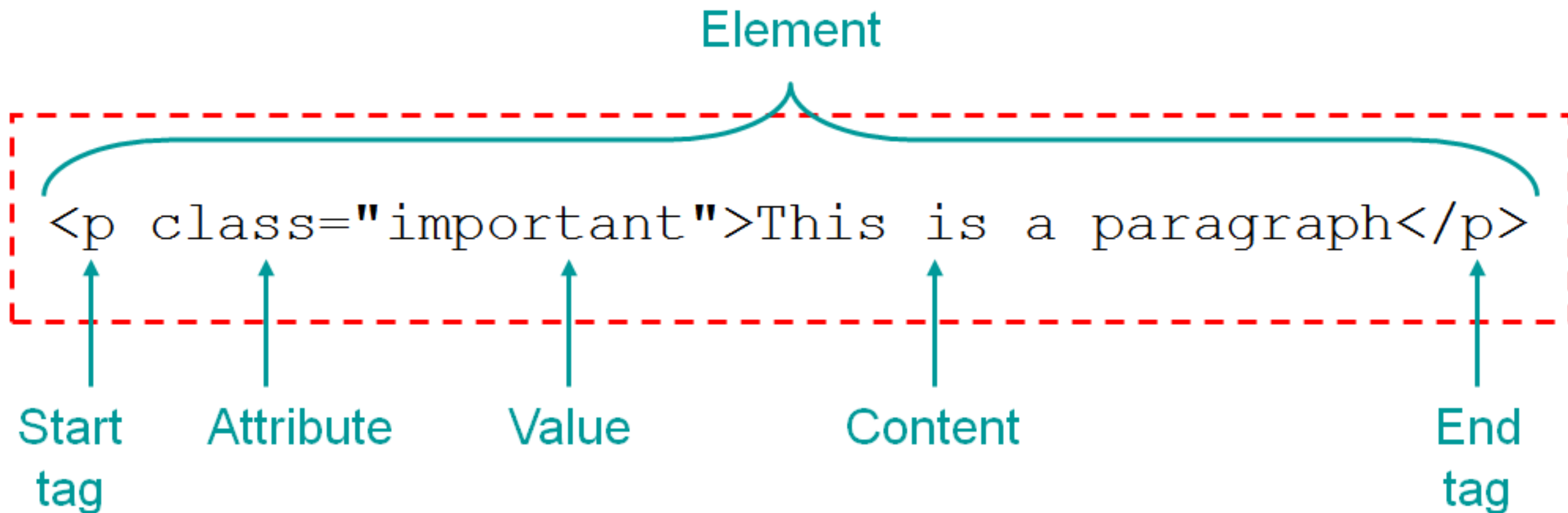
  - Visual presentation delegated to CSS

HTML ELEMENTS

- HTML documents are constructed from HTML Elements

- Elements are **keywords** encapsulated within angle brackets, e.g. <html>

- Elements are represented using opening and closing tags, e.g. <html></html>

  - Most tags delineate the start & end

  - Some stand alone amongst the text, e.g.

- Combined to create structured documents

HTML ELEMENT STRUCTURE

HTML VERSIONS

- As you investigate various web pages you will notice considerable variation amongst versions of HTML

- Let's compare HTML documents from two versions…

HTML 4.01

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN"
    "http://www.w3.org/TR/html4/strict.dtd">
<HTML>
  <HEAD>
    <TITLE>My first HTML 4.01 document</TITLE>
  </HEAD>
  <BODY>
    <P>Hello World from HTML 4.01</P>
  </BODY>
</HTML>
```

HTML 5

```
<!DOCTYPE html>
<html>
  <head>
    <title>My first HTML 5 document</title>
  </head>
  <body>
    <p>Hello World from HTML 5</p>
  </body>
</html>
```

Edinburgh Napier
UNIVERSITY

VALIDITY

- HTML is a language

  - It has a syntactic structure

- There are tools to automatically verify that a given HTML document is correct (or otherwise):

  - https://validator.w3.org/

W3C Markup Validation Service
Check the markup (HTML, XHTML, ...) of Web documents

Validate by URI   **Validate by File Upload**   **Validate by Direct Input**

Validate by URI

Validate a document online:

Address: [   ]

▸ More Options

Check

This validator checks the markup validity of Web documents in HTML, XHTML, SMIL, MathML, etc. If you wish to validate specific content such as RSS/Atom feeds or CSS stylesheets, MobileOK content, or to find broken links, there are other validators and tools available. As an alternative you can also try our non-DTD-based validator.

I ♥ VALIDATOR

The W3C validators rely on community support for hosting and development.
Donate and help us build better tools for a better web.

Flattr us!

Home   About...   News   Docs   Help & FAQ   Feedback   Contribute

This service runs the W3C Markup Validator, v1.3+hg.
COPYRIGHT © 1994-2013 W3C® (MIT, ERCIM, KEIO, BEIHANG), ALL RIGHTS RESERVED. W3C LIABILITY, TRADEMARK, DOCUMENT USE AND SOFTWARE LICENSING RULES APPLY. YOUR INTERACTIONS WITH THIS SITE ARE IN ACCORDANCE WITH OUR PUBLIC AND MEMBER PRIVACY STATEMENTS.

I ♥ VALIDATOR

TAGS

- Document Structure: <html><head><body>

- Within <head> section: <title>, <base>, <meta>, <style>, <link>

- Tags for text blocks: <address>, <blockquote>, <div>, <h1>...<h6>, <p>, <pre>, <xmp>

- Tags that define lists: <dir>, <dl>, <dt>, <dd>, <menu>, <ol>, <ul>,

- Tags that define text format: <b>, <basefont>, <big>, <cite>, <code>, <em>, <font>, <i>, <kbd>, <strike>, <sup>, <tt>, <u>, <var>

- Tags that define anchors and links: <a>

- Tags that define images and image maps: <img>, <area>, <map>

- Tags that define tables: <table>, <caption>, <thead>, <tbody>, <tfoot>, <tr>, <th>, <td>

- Tags that define forms: <form>, <fieldset>, <input>, <select>, <option>, <textarea>, <label>, <legend>, <isindex>

- Tags that define frames: <frame>, <frameset>, <iframe>

- Tags that define scripts: <script>, <noscript>

- Tags that define applets & plug-ins:<applet>, <param>, <object> (<embed> not standard)

- Tags that adjust text: <br>, <center>, <hr>

TEXT FORMATTING

- Headings: <h1>, …, <h6>

- Physical Styles: <b>, <i>

- Logical Styles: <cite>, <code>, <em>, <strong>

- <font face="" size="">

  - You can do this but don't. EVER.

LISTS

- Definition Lists

  - <dl>, <dt>, <dd>

- Ordered Lists

  - ,

- Unordered lists

  - ,

LINKS

- From the Hypertext perspective links are the most important element of HTML

- Hyperlinks turn text into hypertext using two types of link (internal & external):

  - **Internal Links**

    - Link: …

    - Target: …

    - *Target also known as an anchor*

  - **External Links**

    - To another document in same site:

    - To a target within another document: …

    - To another site: …

TABLES

- For data representation

- Uses mix of <table>, <tr>, <th>, <td> tags

- Also: <thead>, <tbody>, <tfoot>, <caption> - more semantic structure

```
<table>
 <tr>
   <th>Heading 1</th>
   <th>Heading 2</th>
 </tr>
 <tr>
   <td>data 1</td>
   <td>data 2</td>
 </tr>
</table>
```

IMAGES

- <img> with mandatory attributes: src, alt

- Optional attributes::

  - width, height, longdesc

- Image types: GIF, JPG, PNG

FORMS

- Everything so far has been about retrieving HTML pages from the server - using the **HTTP GET** method

- Sometimes we want to send data from the client to the server (using the **HTTP POST** method)

  - More commonly using a **for**

  - We'll exploit this more when we start using Javascript

<form name="name"
     action="page.html"
     method="method">
… various controls …
</form>

FORM CONTROLS

- **Buttons:**

    `<input type="submit">`

    `<input type="reset">`

    `<input type="button">`

    `<input type="image">`

- **Check boxes:**

    `<input type="checkbox">`

- **Radio buttons:**

    `<input type="radio">`

- **Text boxes:**

    `<input type="text">`

- **Password textboxes:**

    `<input type="password">`

- **Hidden fields:**

    `<input type="hidden">`

MORE FORM CONTROLS

- **File Upload:**

    <input type="file">

- **Selection Lists:**

    <select> <option> <optgroup>

- **Text Areas:**

<textarea>

- **Label (for a control)**

- **Group of controls:**

<fieldset>, <legend>

RESOURCES

- MDN HTML Reference:
- https://developer.mozilla.org/en-US/docs/Web/HTML/Element

- W3Schools HTML Examples:

- https://www.w3schools.com/tags/

SUMMARY

- You should now:

    - understand how HTML has developed & why it works the way it does

    - be aware of the range of tags supported by HTML

    - be able to assemble basic HTML documents

Questions?

BREAK

TL/DR

Edinburgh Napier
UNIVERSITY

Using HTML5 tags to define structure…

AIMS

- At the end of this (sub-section) of the topic you will:

    - understand the difference between syntax & semantics

    - be aware of the need for semantic representation within markup

    - know about the range of semantic tags within HTML5

OVERVIEW

- Syntax & semantics

- HTML 5 for semantic markup

semantic | sɪˈmantɪk |

adjective

relating to meaning in language or logic.

semantics | sɪˈmantɪks |

plural noun [usually treated as singular]

the branch of linguistics and logic concerned with meaning. The two main areas are logical semantics, concerned with matters such as sense and reference and presupposition and implication, and lexical semantics, concerned with the analysis of word meanings and relations between them.

• the meaning of a word, phrase, or text: such quibbling over semantics may seem petty stuff.

MEANING IN MARKUP

- <p> & </p> are used to mark up a paragraph.

- Has meaning because:

    - People know what paragraphs are

    - Browsers know how to handle paragraphs

- Contrast to tags like <b> or <i>

    - Doesn't convey anything about the content, just how it should look

- Taken to an extreme: *replace body content markup with <div> tags and style using CSS. What do we gain or lose?*

<h1>This is a top level heading</h1>

Is this a top level heading?

MORE TAG MISUSE

- **blockquote -** used to indent text because default presentation is indented (instead of using CSS *margins*)

- **paragraph -** used to add space between page elements instead of defining actual paragraphs (instead of *margin & padding* style properties)

- **Unordered List -** used to indent text but without list elements the HTML is invalid (as well as being semantically incorrect (*margin* or *padding* styles)

- **Heading -** used to make text bigger and bolder. Semantically incorrect if the text isn't actually a heading (*font-weight & font-size* CSS properties)

DIV SOUP

AMBIGUOUS STRUCTURE

(AKA "<DIV> SOUP")

IDENTIFIABLE SECTIONS

(AKA "SEMANTIC MARKUP")

STRUCTURE & MEANING

- Search only see content.

- They understand how to handle content because the content, & its markup, communicate the meaning

- Where we place content on a page, & how we mark it up, can alter how that content is dealt with.

  - e.g. Content in <h1> is more likely to be weighted as important, than content in <p> (which is more likely than content in <b>).

SEMANTIC MARKUP

- Drive toward marking up meaning in addition to typographic elements

- Designers have tried to bring semantic markup to HTML in their use of <div> and <span> to group elements of a page together - but <div> & <span> don't mean anything.

  - Designers defined their own id/class names but this was *ad hoc* & so inconsistent

- Tags that communicate more clearly defined meaning about content are useful:

  Can lead to better structure, increased accessibility, more reliable automated processing, maintainability, & reuse.

HTML 5

- HTML5 introduced the addition of meaningful grouping tags (**sectioning elements** - "divs but with added meaning"):

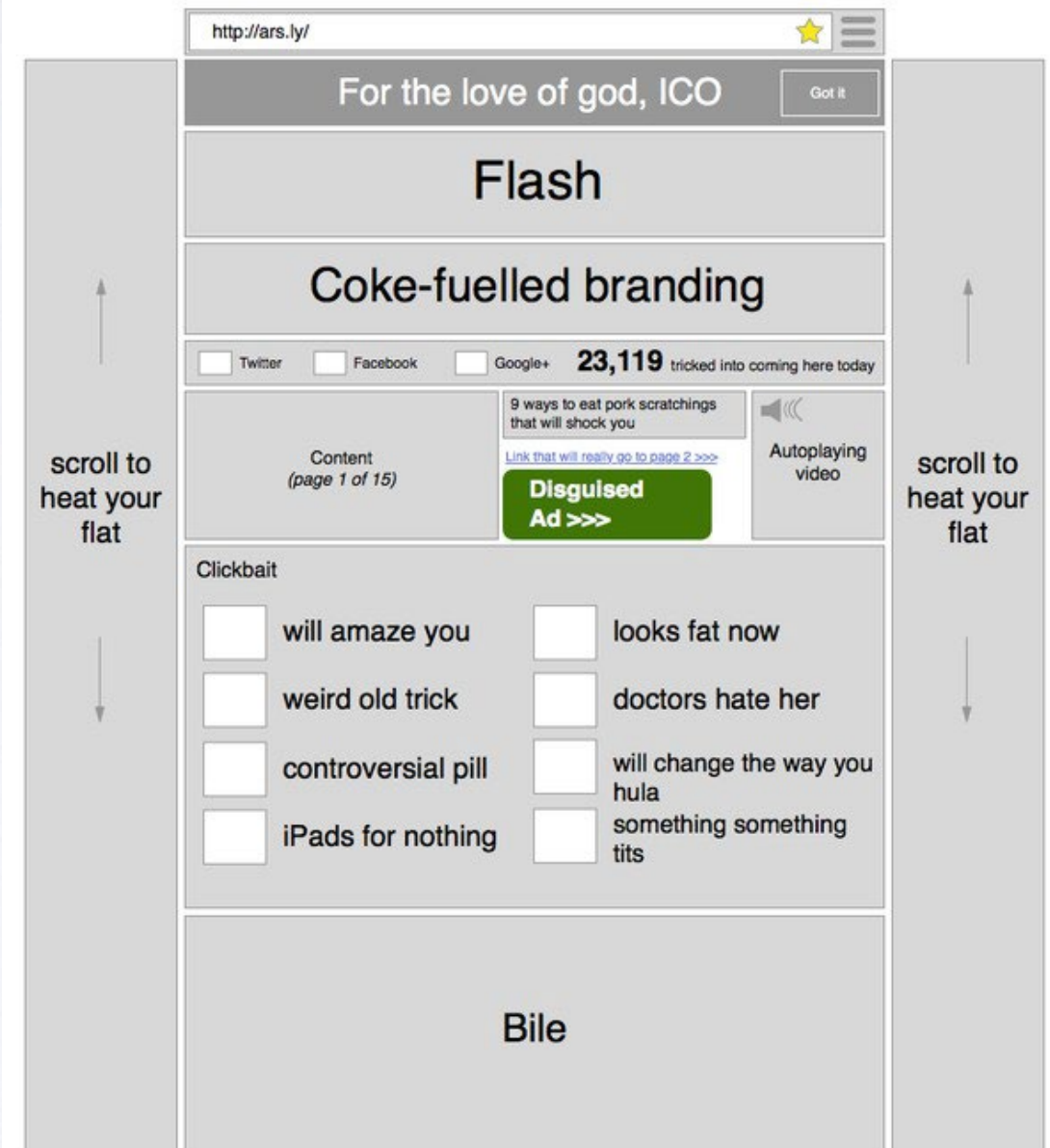 <article>, <aside>, <details>, <figcaption>, <figure>, <footer>, <header>, <main>, <mark>, <nav>, <section>, <summary>, <time>,

WHY?

SEMANTIC MEANING (HTML 4)

Tag	Meaning	Tag	Meaning
<abbr>	Abbreviation	<h1>	First-level headline
<acronym>	Acronym	<h2>	Second-level headline
<blockquote>	Long quotation	<h3>	Third-level headline
<dfn>	Definition	<h4>	Fourth-level headline
<address>	Address for author(s) of the document	<h5>	Fifth-level headline
<cite>	Citation	<h6>	Sixth-level headline
<code>	Code reference	<hr>	Thematic break
<tt>	Teletype text	<kbd>	Text to be entered by the user
<div>	Logical division	<pre>	Pre-formatted text
<span>	Generic inline style container	<q>	Short inline quotation
<del>	Deleted text	<samp>	Sample output
<ins>	Inserted text	<sub>	Subscript
<em>	Emphasis	<sup>	Superscript
<strong>	Strong emphasis	<var>	Variable or user defined text

NEW SEMANTIC TAGS IN HTML5

- <section> - a thematic grouping of content, typically with a heading

- <article> - independent, self-contained content

- <header> - container for introductory content

- <footer> - information about its containing element (author, copyright,&c.)

- <nav> - major blocks of navigational links

- <aside> - related additional content

- <figure> & <figcaption> - visually explain an image

- <main> - the core content of the document

- <mark> - highlighted/emphasised sections

- <details> - additional information that can be hidden/shown

- <summary> - visible heading associated with <details>

- <time> - date/time information

SUMMARY

- You should now:

  - understand the difference between syntax & semantics

  - be aware of the need for semantic representation within markup

  - know about the range of semantic tags within HTML5

Questions?

COMING UP…

- How to make you sites look a bit nicer,

- or

- Cascading Style Sheets (CSS) & their role in styling the information that is captured in an HTML document

RESOURCES

- MDN HTML Reference:

https://developer.mozilla.org/en-US/docs/Web/HTML/Element

- W3Schools HTML Examples:

https://www.w3schools.com/tags/