

# CSE308

## HTML Review



Tim Berners-Lee

1

## Reading

- WebReference tutorials  
[www.webreference.com/html/tutorials/](http://www.webreference.com/html/tutorials/)
- Wiki  
[en.wikipedia.org/wiki/Html](http://en.wikipedia.org/wiki/Html)
- Character sets  
[en.wikipedia.org/wiki/Character\\_encodings\\_in\\_HTML](http://en.wikipedia.org/wiki/Character_encodings_in_HTML)

## References

- WWW Consortium - HTML 4  
[www.w3.org/TR/html4/](http://www.w3.org/TR/html4/)
- W3C HTML Validator - [validator.w3.org/](http://validator.w3.org/)
- Web Design Group Validator  
[www.htmlhelp.com/tools/validator/](http://www.htmlhelp.com/tools/validator/)
- nVu - <http://www.nvu.com/>
- Tidy GUI  
[perso.wanadoo.fr/ablavier/TidyGUI/](http://perso.wanadoo.fr/ablavier/TidyGUI/)
- HTML character entity references  
[www.htmlhelp.com/reference/html40/entities/](http://www.htmlhelp.com/reference/html40/entities/)

© Robert Kelly, 2001-2016

3

## Lecture Objectives

- Become familiar with HTML syntax
- Understand the relationship between an HTML document and the corresponding element tree
- Know the different versions of HTML
- Know the structure of an HTML page
- Become familiar with HTML validation

© Robert Kelly, 2001-2016

4

## HTML Issue

- Approaches to HTML
  1. Primitive abstractions and integrated view
  2. Valid html in which styling information is contained within style sheets
- It's easy to produce bad HTML, but a little more difficult to produce correct HTML
- Html is not a programming language, but think of it as a programming language for which most compilers allow errors

© Robert Kelly, 2001-2016

5

## Evolution of HTML

- Began as a subset of SGML
- Implemented as vendor standards
- Evolved to vendor independent standards that were well implemented by vendors
- Continued evolution to remove styling
- Further evolved to XML structure
- HTML 5 is emerging as the new standard, and reducing burden of SGML legacy (and XML)

© Robert Kelly, 2001-2016

6

## Anatomy of an HTML Page

This IBM Registration page is the start of your project

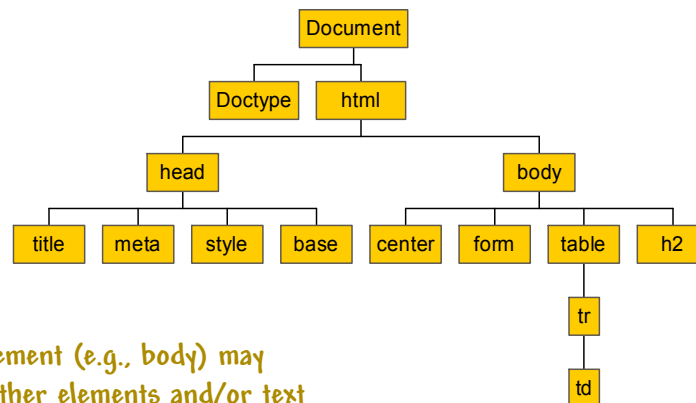
Your project will use a reduced version of a form page - no long drop down lists, reduced number of components.

© Robert Kelly, 2001-2016

7

## Example Document Structure

- An HTML page is a tree of html elements



An element (e.g., body) may contain other elements and/or text

© Robert Kelly, 2001-2016

8

## HTML Element

- An element consists of a begin tag, an end tag, and element content

```
<td><div align="center">  
   </div>
```

Fill out the form below and your information will be sent to a sales representative.

Be sure to specify what price range you are wanting to stay in.

If you prefer, you can call us toll free at 877-456-7223 or </td>

This html produces this browser display

Contact Us

Fill out the form below and your information will be sent to a sales representative. Be sure to specify what price range you are wanting to stay in. If you prefer, you can call us toll free at 877-456-7223 or

© Robert Kelly, 2001-2016

9

## What are the Components of HTML?

tag  
↓  
<html>

element name  
↓  
<body bgcolor="#ffffff" ...

An element name appears in a start tag and (usually) in an end tag

<body> and </BODY>

Element names are case insensitive in html

© Robert Kelly, 2001-2016

10

## Rules for Element Types

- An element type describes a start tag, an end tag, and the content
  - Some element types allow the end tag to be omitted (e.g., p and li), but not in xhtml
  - Some element types have no content and can be denoted as empty (in xhtml):  
(e.g., <br /> )
  - Basic rules:
    - Must be properly nested
    - An end tag closes all intervening tags
- Convention is to leave a space before the /
- HTML5 does not require adherence with XML structure

© Robert Kelly, 2001-2016

11

## What are the Components of HTML?

attribute name                      attribute value

```
<body bgcolor="#FFFFFF" ...
```

Attribute values can be set by authors, scripts, or by default

Attribute name/value pairs are separated by spaces

Attribute name/value pairs may appear in any order

© Robert Kelly, 2001-2016

12

## Attribute Values

- Are usually enclosed in quotes (single or double), but quotes are
  - Not required in html if the value of the attribute does not contain special characters
  - Always required in xhtml
- May be restricted to a specific set of values

© Robert Kelly, 2001-2016

13

## Terms to Know

- Document - a message entity with a content type of text/html (also applies to other text documents)
- HTML user agent - a device that interprets HTML documents (includes browsers)

© Robert Kelly, 2001-2016

14

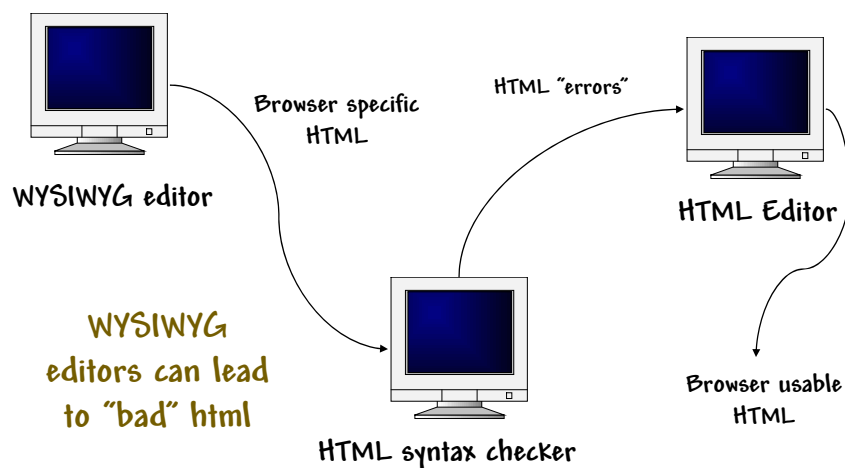
## Well-Formed (Parsable) XML

- Basic Rules (common to all XML documents)
  - No unclosed tags
  - No overlapping tags
  - Attribute values must be enclosed in quotes
  - The text characters >, <, and " must always be represented by character entities
- Extended rules (Specific to each XML application)

© Robert Kelly, 2001-2016

15

## One Way to Generate HTML



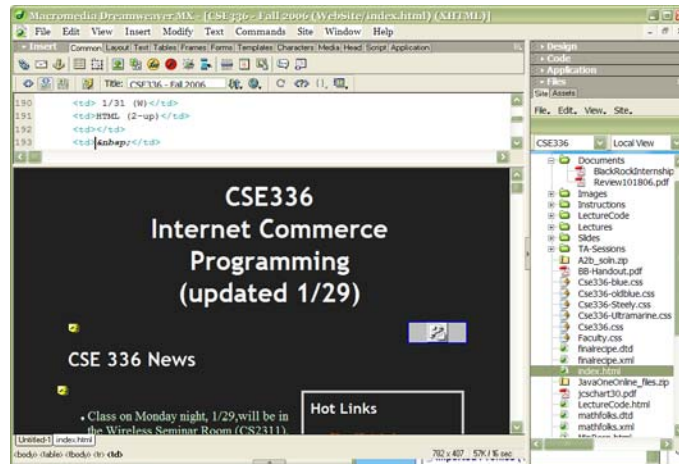
© Robert Kelly, 2001-2016

16



## HTML Editor - DreamWeaver (CS6)

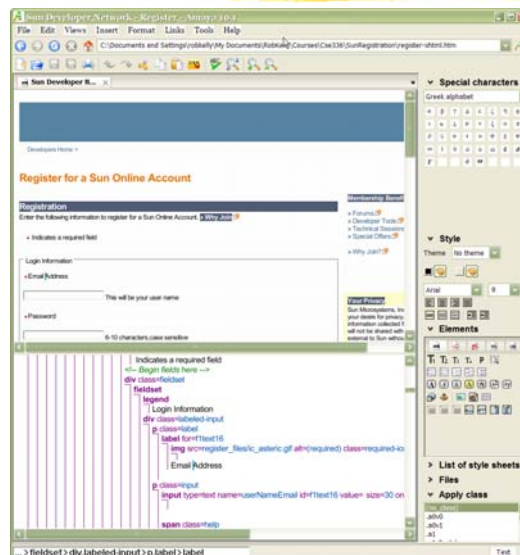
Available to students through DOLT



© Robert Kelly, 2001-2016

17

## HTML Editor - Amaya

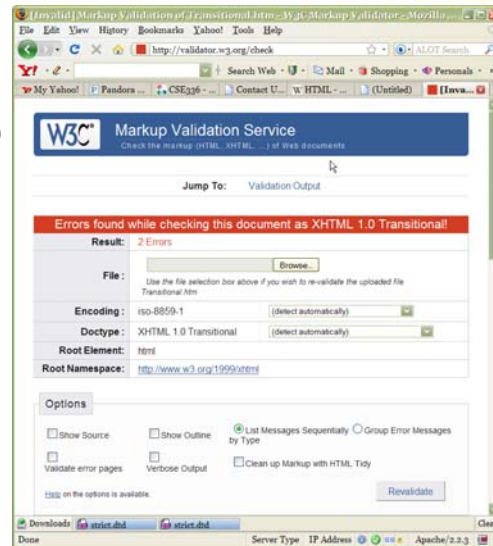


© Robert Kelly, 2001-2016

18

### What is an HTML "Compiler" Error?

- HTML is validated against a version (e.g., XHTML Transitional)
- Web based validation
  - WDG
  - W3C
- Tool-based validation (and conversion)
  - Dreamweaver feature



© Robert Kelly, 2001-2016

19

### Common "Errors"

- Errors give you the line number and a brief explanation

✖ Line 406, Column 87: required attribute "alt" not specified.

```
...http://www.sun.com/" target="_blank" / >
```

The attribute given above is required for an element that you've used, but you have omitted it. For instance, in most HTML and XHTML document types the "type" attribute is required on the "script" element and the "alt" attribute is required for the "img" element.

Typical values for type are type="text/css" for <style> and type="text/javascript" for <script>.

✖ Line 408, Column 21: ID "Map2" already defined.

```
<map name="Map2" id="Map2">
```

An "id" is a unique identifier. Each time this attribute is used in a document it must have a different value. If you are using this attribute as a hook for style sheets it may be more appropriate to use classes (which group elements) than id (which are used to identify exactly one element).

© Robert Kelly, 2001-2016

20

### Doctype

```
<!DOCTYPE html >
```

- First line of your document
- Identifies the version of HTML the document should comply with
- Reference to DTD no longer needed (with HTML5)
- Above example will be validated as HTML5

© Robert Kelly, 2001-2016

21

### Are We on Track

- For the Project page located at:  
[https://www.ibm.com/developerworks/dwwi/jsp/Register.jsp?lang=en\\_US&appname=developerworks&d=http%3A%2F%2Fwww.ibm.com%2Fdeveloperworks%2Ftopics%2](https://www.ibm.com/developerworks/dwwi/jsp/Register.jsp?lang=en_US&appname=developerworks&d=http%3A%2F%2Fwww.ibm.com%2Fdeveloperworks%2Ftopics%2)
- Be sure that you can set up your project
  - Download the Web page
  - Open it in your HTML editor
  - Validate
  - Automatically (?) convert to HTML5

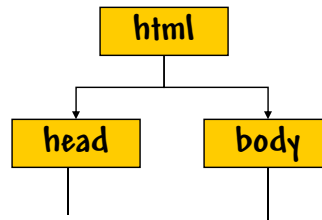
<http://validator.w3.org/>

© Robert Kelly, 2001-2016

22

## Head Element

- The head element contains header information about the document, such as its title, keywords, description, and style sheet.



```
<head>
<title>2008 CSE336 Conference</title>
<meta http-equiv="Content-Type" content="text/html;
charset=utf-8" />
<link href="Registration_files/dt_javaone.css"
rel="stylesheet" type="text/css" />
<style type="text/css">
  .smaller {
    font-size:11px;
    TEXT-DECORATION: none
  }
</style></head>
```

Style sheet

MIME type

© Robert Kelly, 2001-2016

23

## Head Elements

- Title - user agents always make the contents of the title available to users
- Meta - specifies various name,value pairs
  - meaning of a name, value attribute is specified in a profile
  - E.g., `<meta name="Author" content="Dave Raggett">`
  - http-equiv attribute used in place of name attribute to create an HTTP header
  - Used for search engine data, code signing, privacy, defaults, copyrights
  - E.g., `<meta name="keywords" content="Java,image,">`

© Robert Kelly, 2001-2016

24

## Body Elements

- Viewable content in HTML (e.g., text, images, graphics) is painted (top to bottom) onto the visible page for visual browsers
- Contains elements that are either **block level** or **inline**
  - Block level - begin on a new line
  - Inline - text level
  - div and span are used to provide additional structure (block and inline)

© Robert Kelly, 2001-2016

25

## Text Elements

```
<body>
  <p>This is a paragraph tag</p>
  <ul>
    <li>This is an item in an unordered list.</li>
    <li>This is another item in that list.</li>
  </ul>
  <ol>
    <li>This is an item in an ordered list.</li>
    <li>This is another item in that list.</li>
  </ol>
</body>
```



This is a paragraph tag

Definition lists  
are also available  
in html

- This is an item in an unordered list.
  - This is another item in that list.
1. This is an item in an ordered list.
  2. This is another item in that list.

Be sure that you  
understand the html tags  
for ordered and  
unordered lists

© Robert Kelly, 2001-2016

26

## Text

- Inline elements - em, strong, cite, code abbr, acronym, Q, sub, sup, etc.
- Block elements - blockquote, p

The elements that dictate appearance are best replaced by CSS (covered in the next session)

© Robert Kelly, 2001-2016

27

## Characters

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

- The meta element can be used to communicate communications protocol information to the server
- You should place the information early in the document head element

© Robert Kelly, 2001-2016

28

## Document Representations

- Servers send HTML documents to agents as a bytestream; user agents interpret them as a sequence of characters
- HTML allows different computers to interoperate seamlessly, but these computers may use different character encodings
- This process requires a knowledge of:
  - Document character set - characters used in a document
  - Character encodings - the byte representations of characters - referred to as "charset"

© Robert Kelly, 2001-2016

29

## Early Character Codes

- The earliest encoding systems used six bits (BCD), allowing 64 characters
- In 1963
  - 8-bit EBCDIC was introduced by IBM
  - The 7-bit ASCII code was introduced and used by other computer HW manufacturers
- The codes are
  - Clearly inadequate for global commerce
  - Important to understand implementation of current codes (backwards compatibility)

© Robert Kelly, 2001-2016

30

## Characters

- Languages consist of a set of characters, usually defined as the smallest unit of information in the written form of a natural language
- Examples
  - English includes 26 letters (a-z), along with their capital equivalents, digits (0-9), and special symbols (e.g., ",.")
  - Chinese has 4,000 characters for general language coverage and 40,000 characters for more complete coverage
  - Japanese has 2,000 characters for general language coverage
- There are approximately 6,800 living languages in the world today

© Robert Kelly, 2001-2016

31

## Character Code Issues

- Character codes
  - Mapping of characters to strings of binary digits
  - E.g., "S" usually is usually mapped to  $01000011_2$
- Mapping to a 8-bit code usually restricts the language to 256 characters
- Mapping to longer character codes can result in longer strings
  - Length of text strings still a concern, even with much less expensive memory and disk
  - Text is sometimes transmitted over low bandwidth communications links

Each mapping is sometimes referred to as a "code point"

© Robert Kelly, 2001-2016

32



## ASCII Reference Table

Note the ordering of characters

MSD LSD	0	1	2	3	4	5	6	7
0	NUL	DLE	SP	0	@	P		p
1	SOH	DC1	!	1	A	Q	a	q
2	STX	DC2	"	2	B	R	b	r
3	ETX	DC3	#	3	C	S	c	s
4	EOT	DC4	\$	4	D	T	d	t
5	ENQ	NAK	%	5	E	U	e	u
6	ACJ	SYN	&	6	F	V	f	v
7	BEL	ETB	'	7	G	W	g	w
8	BS	CAN	(	8	H	X	h	x
9	HT	EM	)	9	I	Y	i	y
A	LF	SUB	*	:	J	Z	j	z
B	VT	ESC	+	;	K	[	k	{
C	FF	FS	,	<	L	\	l	
D	CR	GS	-	=	M	]	m	}
E	SO	RS	.	>	N	^	n	~
F	SI	US	/	?	O	_	o	DEL

74<sub>16</sub>  
1110100<sub>2</sub>

© Robert Kelly, 2001-2016

33

## Modern Approach to Encoding

- Establish
  - Universal set of characters that can be encoded in a variety of ways
  - Ordering of the characters
- Character repertoire - the full set of abstract characters that a system supports, and might allow
  - No additions - e.g., ASCII
  - Additions
- Examples
  - Unicode
  - ISO/IEC10646

© Robert Kelly, 2001-2016

34

## Unicode

- Can represent the characters of every language in the world
- Contains
  - more than 110,000 characters (Universal Character Set)
  - 100 scripts (e.g., Latin, Arabic) These code points are the HTML numeric references
  - Codepoint for every character
  - A 6-part codespace (e.g., Western alphabet codes)
- Equivalent (almost) to ISO 10646
- Implemented by various encodings
  - UTF-8 - one byte for ASCII characters and up to 4 bytes for other characters
  - UTF-16 - 2-4 bytes for each character

Java uses Unicode as its default character set

© Robert Kelly, 2001-2016

35

## Unicode Codespace Allocation

- The lowest-numbered Unicode characters comprise the ASCII code - preserves backwards compatibility

Character Types	Language	Number of Characters	Hexadecimal Values
Alphabets	Latin, Greek, Cyrillic, etc.	8192	0000 to 1FFF
Symbols	Dingbats, Mathematical, etc.	4096	2000 to 2FFF
CJK	Chinese, Japanese, and Korean phonetic symbols and punctuation.	4096	3000 to 3FFF
Han	Unified Chinese, Japanese, and Korean	40,960	4000 to DFFF
	Han Expansion	4096	E000 to EFFF
User Defined		4095	F000 to FFFE

© Robert Kelly, 2001-2016

36

## Example - HTML

- An HTML document consists of Unicode characters
- When transmitted, the document is encoded according to document / server instructions, as in  

```
<meta charset=UTF-8" />
```
- When the encoding or editor does not support all the Unicode characters used in the document, characters can be escaped using an entity reference

Entity Reference	Category	Displays As
&#x5E7;	Hebrew	פ
&#x645;	Arabic	م
&#x8449;	Chinese	葉
&#xB5AB;	Korean	뎡

© Robert Kelly, 2001-2016

37

## Special Characters

- Characters can be used directly or as a special reference (if it is not in the character set or if there is a "meaning conflict")
- Character references can be numeric or literal

Literal character  
reference



Copyright © 1996-2004 Sun Microsystems, Inc. in the U.S. and other countries. All rights reserved. To send comments about this page, please contact: [Sun Microsystems, Inc.](#)

You should replace  
 Copyright</a> © 1996-2004  
 with

Copyright</a> &copy; 1996-2004

Good practice to  
add the symbol  
when it is missing

© Robert Kelly, 2001-2016

38

## Character References

### ■ Numeric references (decimal or hexadecimal)

■ `&#229;` - å (Norwegian)

Ariel Unicode MS font supports  
Unicode characters

■ `&#x6C34;` - 水 (Chinese character for water)

### ■ Character entity references

■ `&gt;` represents the > sign

Numeric references use either decimal  
notation (`#nnnn`) or hex notation  
(`#xhhhh`), with or without leading zeroes

Numeric references refer  
to Unicode, which is then  
mapped into the specific  
encoding (e.g., UTF-8)

Unicode is like a virtual encoding

© Robert Kelly, 2001-2016

39

## Body Content View Descriptions

### ■ An HTML page can describe some of the styling information in external style sheets

```
<link rel="stylesheet" href="original_files/nav.css"
type="text/css" />
```

```
<link rel="stylesheet" href="original_files/right.css"
type="text/css" />
```

```
<link rel="stylesheet" href="original_files/calander.css"
type="text/css" />
```

### ■ Style information is usually applied to the element (e.g., `td`) or to enclosed elements (e.g., with `font`)

More on this in the  
next class session

© Robert Kelly, 2001-2016

40

## Additional HTML Data Types

### ■ Colors

- attribute value type "color" refers to color sRGB definitions
- A color value may either be a hexadecimal number (prefixed by a hash mark) or one of sixteen color names

### ■ Length - pixels or percentage

### ■ Media descriptors

- Screen, tty, tv, projection, print, handheld, print, Braille, aural, all

© Robert Kelly, 2001-2016

41

## Tables

- The HTML table model allows authors to arrange data (text, preformatted text, images, links, forms, form fields, other tables, etc.) into rows and columns of cells

Most html pages use tables to organize the content on the page (including embedded tables)

- Tables should resize dynamically
- Should allow incremental display
- Allow head, foot, and body groupings
- Cells can span multiple rows and columns

It is usually not a good idea to use exact table (e.g., column) pixel dimensions

© Robert Kelly, 2001-2016

42

## Forms

- A form element usually contains text, along with GUI components and a submit button
- Typical GUI components
  - Text box (input element, with type of text)
  - Dropdown (select element)
  - Check box (input element, with type of checkbox)
  - Radio button (input element, with type of radio)

© Robert Kelly, 2001-2016

43

## Form Example

```
<form method="post" action="Mets/tix" >
  <input name="Team" value="New York Mets" type="hidden" />
  ...
  <div align="right">Opponent:</div>
  <input name="Opponent" size="20" class="nav" type="text" />
  <div align="right">Date:</div>
  <input name="Date" size="10" class="nav" type="text" />
  mm/dd/yy
  <div align="right"> *Number of tickets:</div>
  <select name="Number" class="nav">
    <option selected="selected">Select</option>
    <option>1</option>
    <option>2</option>
    <option>3</option>
    <option>4</option>
    <option>5+</option>
  </select>
  ...
</form>
```

Opponent:

Date:  mm/dd/yy

\*Number of tickets:

Options appear in the drop-down

© Robert Kelly, 2001-2016

44

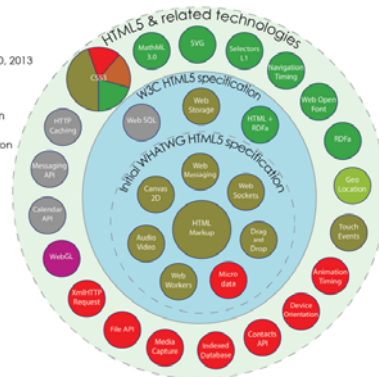
## Where is HTML Headed?

- HTML5 is the future
- Most HTML5 features implemented by current browsers

### HTML5

Taxonomy & Status on January 20, 2013

- W3C Recommendation
- Proposed Recommendation
- Candidate Recommendation
- Last Call
- Working Draft
- Non-W3C Specifications
- Deprecated



by Sergey Mavrody © BY-SA

Issues with Apple  
and Flash

Try <https://html5test.com/>

© Robert Kelly, 2001-2016

45

## HTML5 Features

- Both xml and html syntax included
- New features (e.g., video and audio)
- Enriched semantic structure (e.g., header, section, and article)
- Well-defined handling of incorrect syntax
- New and extended APIs (e.g., DOM)



DOM API is now a part of the  
HTML spec

© Robert Kelly, 2001-2016

46

## New HTML5 Structure Elements

- `<section>` - sections of pages
- `<header>` - header of a page
- `<footer>` - footer of a page
- `<nav>` - navigation on a page
- `<article>` - article or primary content on a page
- `<aside>` - extra content like a sidebar on a page
- `<figure>` - images that annotate an article

Is this an aside? → How are these handled in  
HTML4 / xhtml 1.0?

© Robert Kelly, 2001-2016

47

## HTML5 Features

- New form elements - `datetime`, `datetime-local`, `date`, `month`, `week`, `time`, `number`, `range`, `email`, `url`
- New elements
  - `<canvas>` - gives you a drawing space in JavaScript on your Web pages
  - `<video>` - add video
  - `<audio>` - add sound
- Removes elements - many of them replaced by CSS and already deprecated

© Robert Kelly, 2001-2016

48



## HTML5 APIs

- Improved *APIs can be used with JavaScript*
  - DOM - Document Object Model
- New
  - Dynamic rendering of 2D shapes and bitmap images
  - Cross document messaging
  - Microdata - embeds metadata within page content
- Separate from HTML5
  - Web storage - similar to cookies, but with enhanced capacity
  - Geolocation

© Robert Kelly, 2001-2016

49

## Have You Satisfied the Lecture Objectives

- Become familiar with HTML syntax
- Understand the relationship between an HTML document and the corresponding element tree
- Know the different versions of HTML
- Know the structure of an HTML page
- Become familiar with HTML validation

© Robert Kelly, 2001-2016

50

## Assignment 2

- Begin to modify the HTML of the Project page (link in assignment Web page) so that it adheres to correct HTML5
    - Save the HTML from your browser
    - Reduce the size of the html (eliminate all but 5 countries)
    - Modify the html so that it displays the same as the original page (almost)
    - Validate it using one of the HTML validator services
    - Modify your local copy of the HTML using whatever tool you are comfortable with
    - Check your revised HTML (local file) for reduced errors
    - Repeat
- You will not be able to eliminate all errors until we cover CSS**

© Robert Kelly, 2001-2016

51