

Session 3

HTML



Tim Berners-Lee

1

Reading

- **WebReference tutorials**
www.webreference.com/html/tutorials/
- **Wiki**
en.wikipedia.org/wiki/Html
- **Character sets**
en.wikipedia.org/wiki/Character_encodings_in_HTML

References

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

© Robert Kelly, 2001-2012

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
- Understand the importance of HTML validation

© Robert Kelly, 2001-2012

4

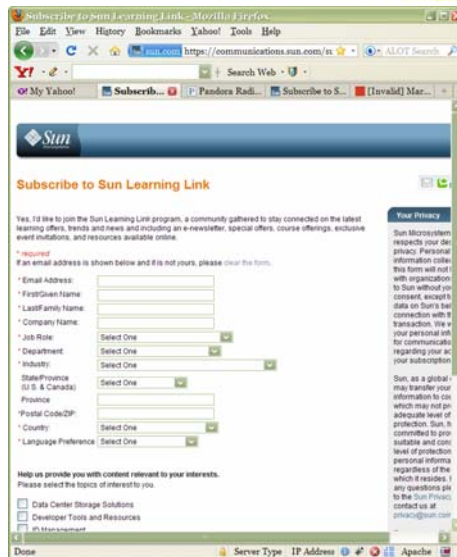
HTML Issue

- Approaches to HTML
 1. "old-style" html - generated by tools, not consistent with standards and not using advanced features
 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
- Advanced Web applications almost always need the second approach
- Think of html as a programming language for which most compilers allow errors

© Robert Kelly, 2001-2012

5

Anatomy of an HTML Page



This Sun Learning Link Registration page is an example your major project (TBD in the next few days)

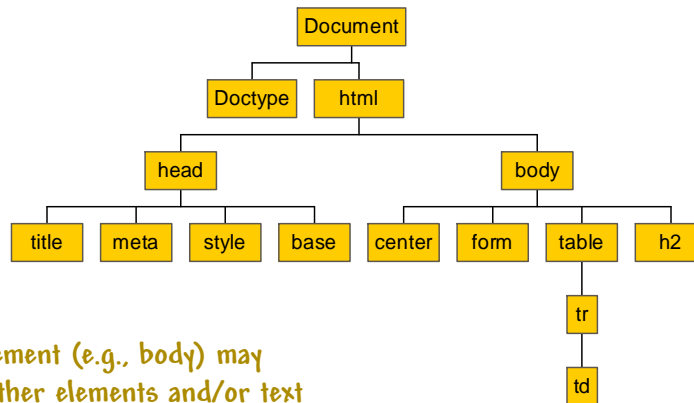
Your project will use a reduced version of a form page - no JavaScript, no long drop down lists, reduced number of components.

© Robert Kelly, 2001-2012

6

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, 2001-2012

7

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

This html produces this browser display

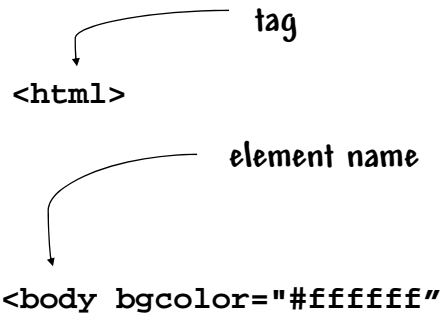
Contact Us

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, 2001-2012

8

What are the Components of HTML?



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

Element names are case insensitive in html

`<body>` and `</BODY>`

© Robert Kelly, 2001-2012

9

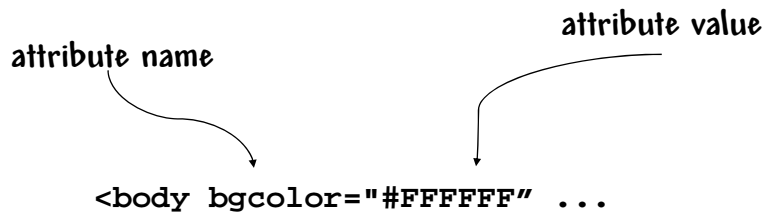
Rules for Element Types

- 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., `
`)
- Basic rules:
 - Must be properly nested
 - An end tag closes all intervening tags

© Robert Kelly, 2001-2012

10

What are the Components of HTML?



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, 2001-2012

11

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, 2001-2012

12

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)
- **Deprecated** - an attribute or element that is outdated
- **Obsolete** - an attribute or element for which there is no guaranteed support

© Robert Kelly, 2001-2012

13

What Exactly is HTML?

- Is it a vendor standard?
 - Internet Explorer
- Or is it a vendor-independent standard?
 - HTML 2.0, 3.2, 4.0, 4.01, 5
 - xhtml 1.0, xhtml 1.2, xhtml 2.0, xhtml 5

Xhtml follows xml
language conventions

© Robert Kelly, 2001-2012

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, 2001-2012

15

XML Document

- Structures textual information
- Does not contain styling information
- Defines a hierarchical structure
- Contains elements and attributes
- Follows basic XML syntax rules
- Usually adheres to a set of domain rules
 - Element names
 - Attribute names
 - Containment rules

© Robert Kelly, 2001-2012

16

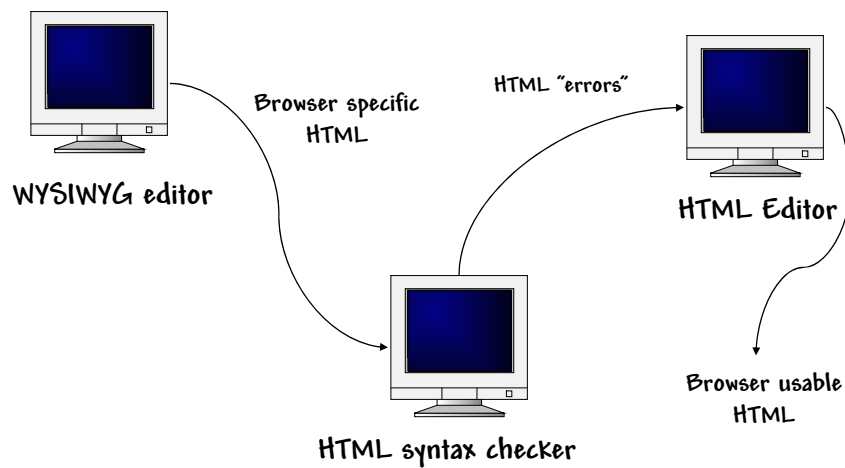
Application-Specific XML Rules

- Rules define each unique XML language (e.g. xhtml Transitional)
- Examples of document rules:
 - Names of the elements and attributes
 - Allowed containment (e.g., tr within table)
- Defined in a schema
 - DTD (Document Type Definition)
 - XML Schema

© Robert Kelly, 2001-2012

17

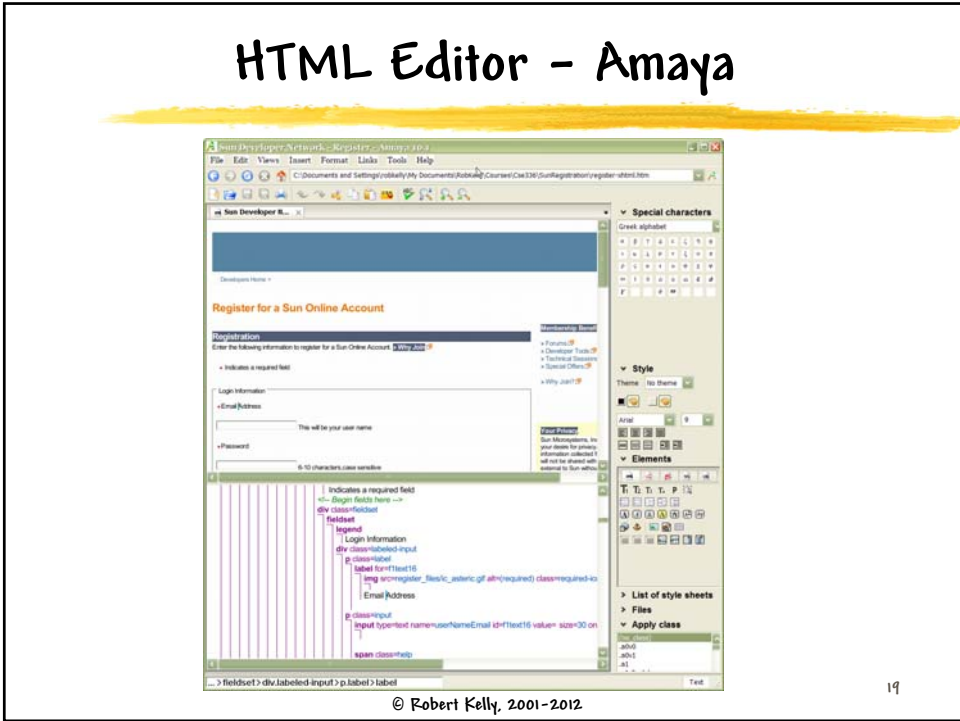
One Way to Generate HTML



© Robert Kelly, 2001-2012

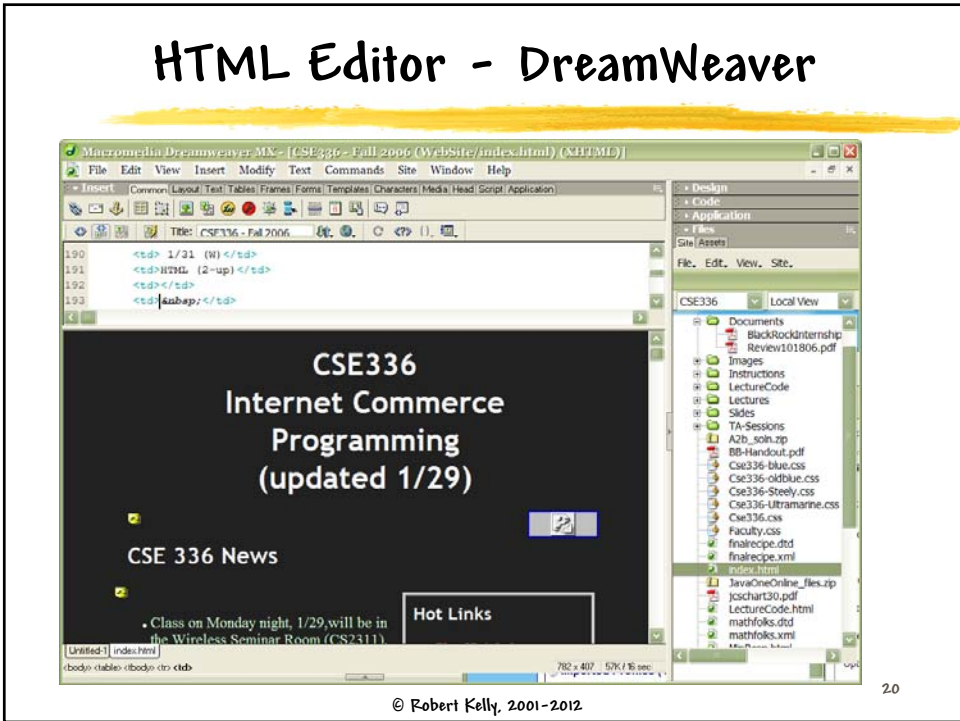
18

HTML Editor - Amaya



19

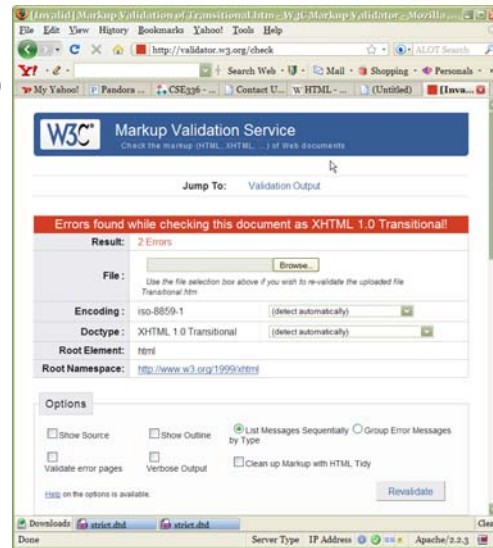
HTML Editor - DreamWeaver



2.0

What is an HTML "Compiler" Error?

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



© Robert Kelly, 2001-2012

21

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, 2001-2012

22

Doctype

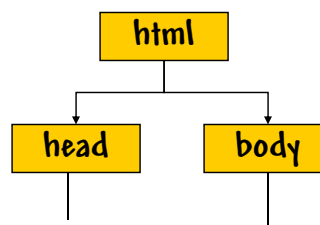
```
<?xml version="1.0" encoding="utf-8"?>  
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0  
Transitional//EN"  
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-  
transitional.dtd">
```

- First line of your document
- Identifies the version of HTML the document (should) comply with
- Refers to the associated Document Type Definition (DTD) which defines the version of HTML used by the document
 - 4.01 Transitional
 - xhtml Transitional
 - Xhtml Strict

Doctype needed when markup language is based on SGML (not true for HTML5)

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

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, 2001-2012

27

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, 2001-2012

28

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

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

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.

29

© Robert Kelly, 2001-2012

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)

30

© Robert Kelly, 2001-2012

Characters

```
<meta http-equiv="Content-Type" content="text/html;  
      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, 2001-2012

31

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, 2001-2012

32

Encodings

- Commonly used encodings
 - ISO-8859-1 (Latin-1)
 - UTF-16
 - ISO-8859-5 (supports Cyrillic)
 - SHIFT_JIS (a Japanese encoding).
- User agents map the encoding to ISO 10646
- As an example, UTF-16 requires transmission in big-endian with a byte order mark (FEFF)
- Encoding is usually identified in HTTP protocol (and META tag)

© Robert Kelly, 2001-2012

33

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



Sponsored by
Copyright © 1996-2004 Sun
Java, J2EE, J2SE, J2ME,
Microsystems, Inc. in the U
MediaLive International is a
To send comments about tl

Literal character
reference

You should replace
Copyright © 1996-2004
with

Copyright © 1996-2004

Good practice to
add the symbol
when it is missing

© Robert Kelly, 2001-2012

34

Character References

- Numeric references (decimal or hexadecimal)
 - `å` - å (Norwegian)
 - `水` - 水 (Chinese character for water)
- Character entity references
 - `>` represents the > sign

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

© Robert Kelly, 2001-2012

35

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, 2001-2012

36

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, 2001-2012

37

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
 - It is usually not a good idea to use exact table (e.g., column) pixel dimensions
- Cells can span multiple rows and columns

© Robert Kelly, 2001-2012

38

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, 2001-2012

39

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, 2001-2012

40

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)



© Robert Kelly, 2001-2012

43

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

How are these handled in
HTML4 / xhtml 1.0?

© Robert Kelly, 2001-2012

44

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, 2001-2012

45

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
- Understand the importance of HTML validation

© Robert Kelly, 2001-2012

46

Semester Project

- Your semester project is to reverse engineer the form processing of the Oracle registration site

The screenshot shows the Oracle MyProfile Create User registration form. It includes sections for 'New User Registration', 'Create Your Account', 'User Information', 'Primary Address', 'Shipping Address', 'My Community Memberships', and 'Subscription Center'. The form contains various input fields, dropdown menus, and checkboxes. A 'Create' button is at the bottom right. The browser window title is 'Oracle - MyProfile Create User' and the URL is 'https://myprofile.oracle.com/EnrollUser/face/jsp/createUser.jsp?nextURL=https%3A%2F%2Fmyprofile.oracle.com'. The page number '47' is visible in the bottom right corner of the browser window.

© Robert Kelly, 2001-2012

Assignment 2

- Modify the HTML of the Project page (link in assignment Web page) so that it adheres to correct XHTML Transitional
 - Save the HTML from your browser
 - Reduce the size of the html
 - Modify the html so that it displays the same as the original page
 - Validate it using one of the HTML validator services using xhtml Strict
 - Modify your local copy of the HTML using whatever tool you are comfortable with (e.g., text editor)
 - Check your revised HTML (local file) with the WDG validator (allows you to upload the file)
 - Repeat

© Robert Kelly, 2001-2012

48

Project HTML Reduction Steps

- You can reduce the size of the html file (without changing what you will learn from the project) by removing the following html:
 - all the JavaScript tags (i.e., `<script>` to `</script>`)
 - any `onxxxx` attributes (e.g., `onsubmit = ...`)
 - any `<option>` tag beyond 5 in a group (e.g., reduce the number of countries)
 - Eliminate the hidden form fields (i.e., input elements with `type="hidden"`)