Web Content

The 3 Components of the Web

Resource Identification

- Using URI (URN/URL)
- Aka Web Addressing
- Interaction
 - Protocol: **HTTP**
- Content/Format
 - How information/data is written
 - HTML: basic web programmers must-know
 - XML based formats
 - Other formats.
 - Specifications on how content is to be rendered.

Content Formats

- Text
 - HTML
 - XML
 - Plain Text
 - CSS*
 - Javascript code*

- Binary
 - Graphics
 - Audio/Video
 - Java Applets
 - Flash
 - Etc...

The Web Browser: Content Renderer

- The Client Application
 - In charge of communication with a web server
 - **Renders** the content (The browser is a document parser).
 - By parsing HTML
 -or XML
 - Apply stylings (default or via CSS)
 - The Browser is the requestor, parser and renderer

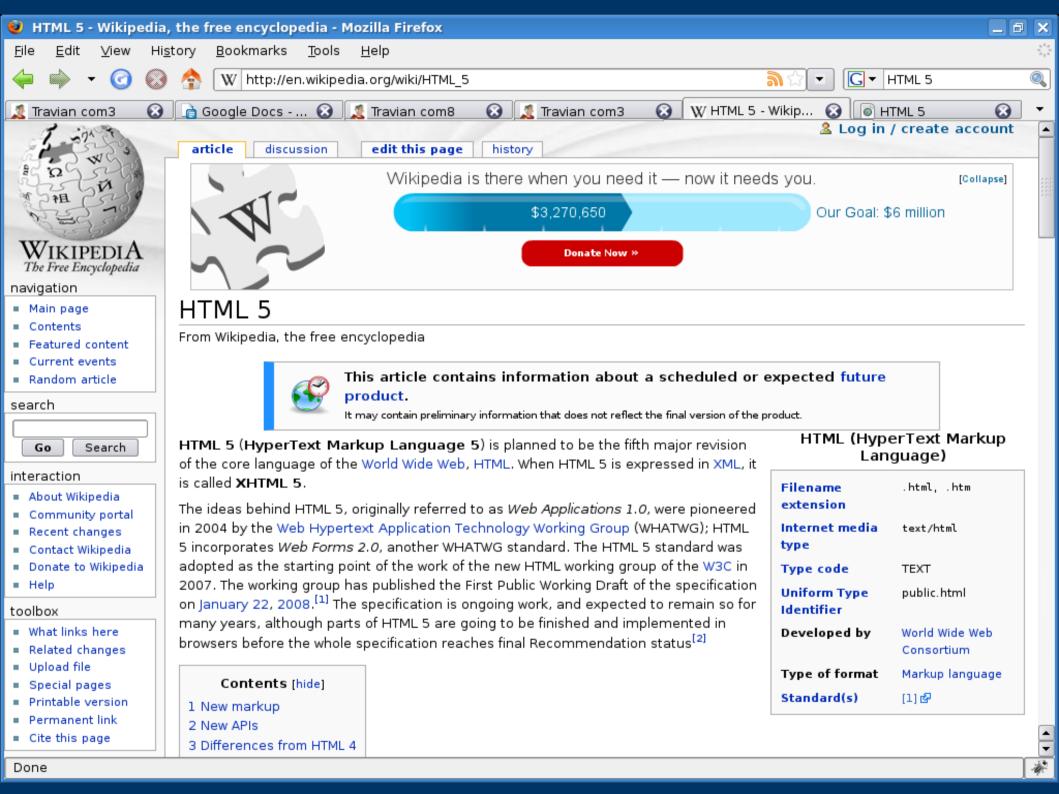
Web Browser

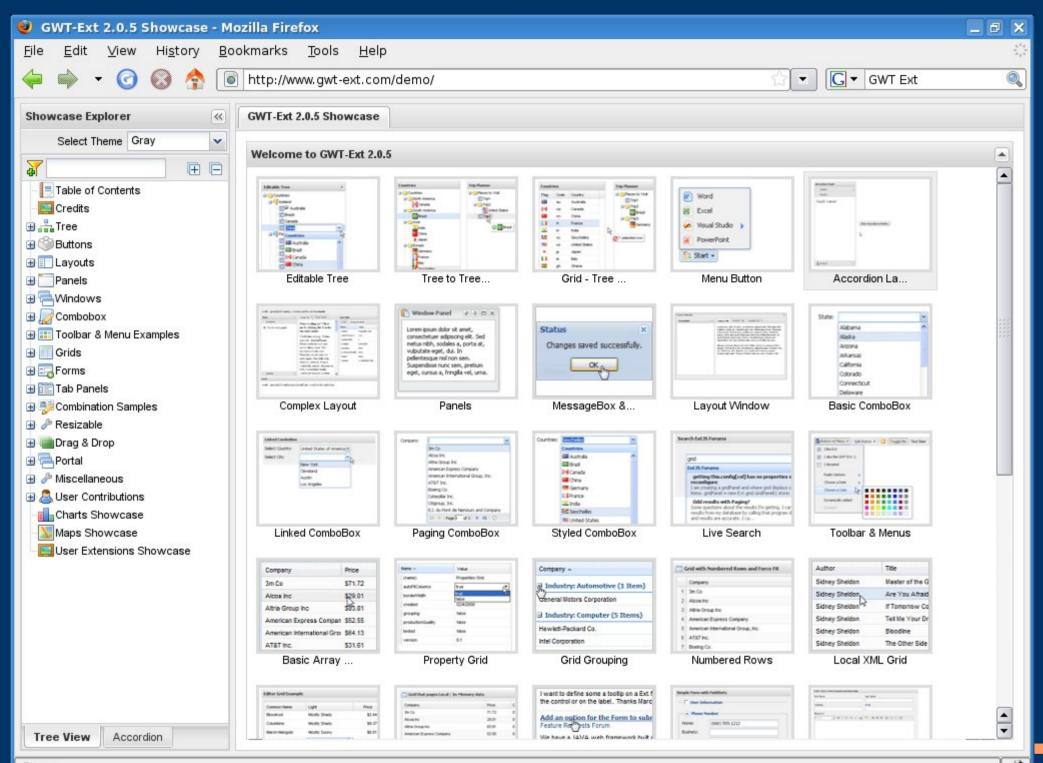
- Thin browser/thin client
 - Renders only basic content: HTML, plain text
- Thick browser/thick client
 - Requires the help of plugins and helper applications to render web page content.
 - E.g. Java applets require a JRE configured witht the browser
 - E.g. Flash animation requires a Runtime environment for flash.
 - E.g. Audio/Video requires media players with codecs

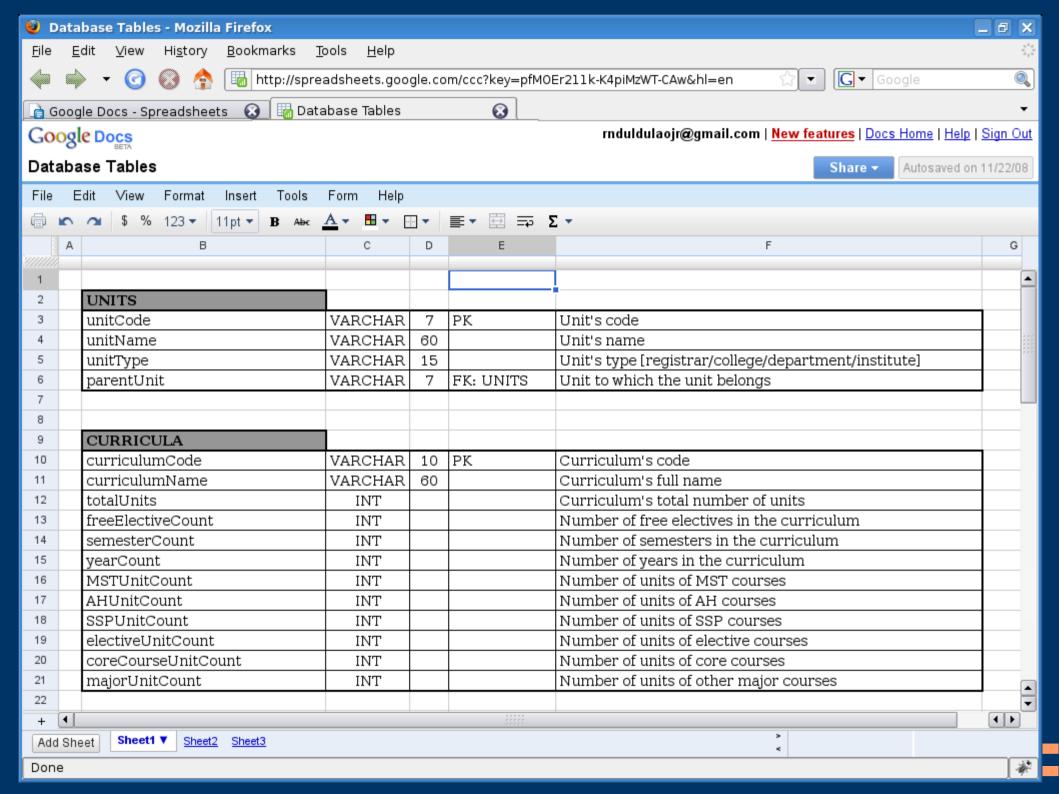
HTML

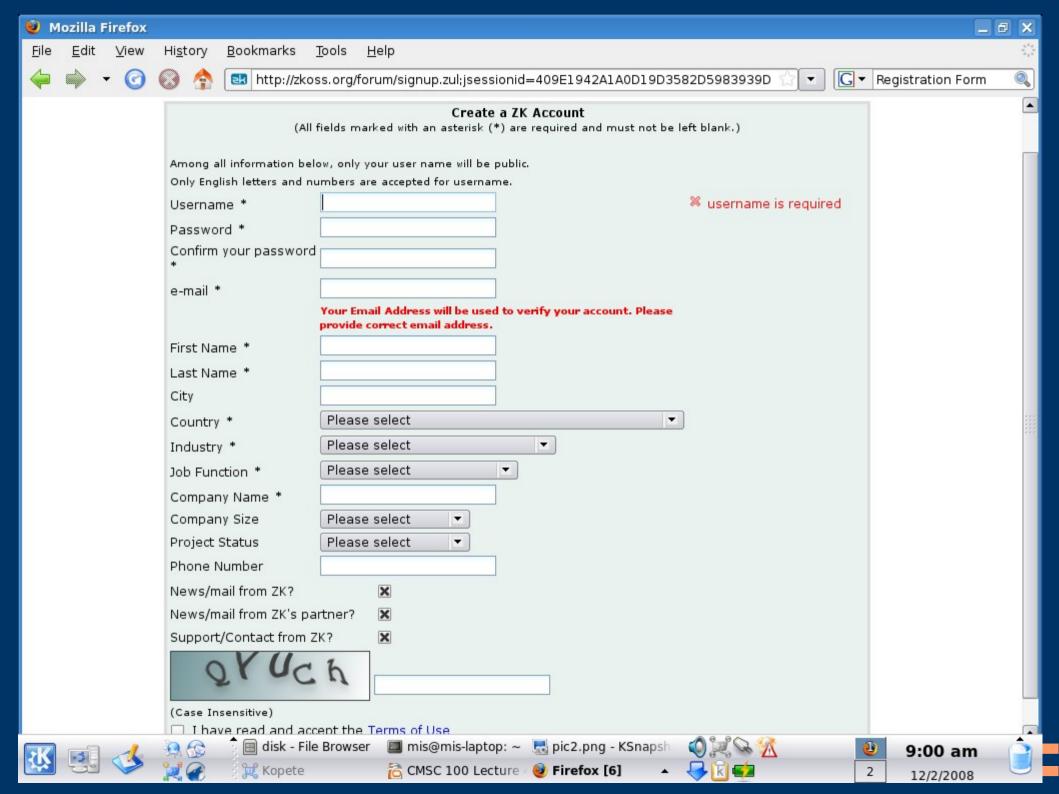
- Hypertext Markup Language
 - A web programmer's basic content markup language.
 - Last Version*: 4.01
 - Update: HTML 5/XHTML 5 is under development.
 - Markup Language
 - Marks up elements of the web page thru the use of tags.
 - An Application of SGML (Standard Generalized Markup Language)
 - A metalanguage

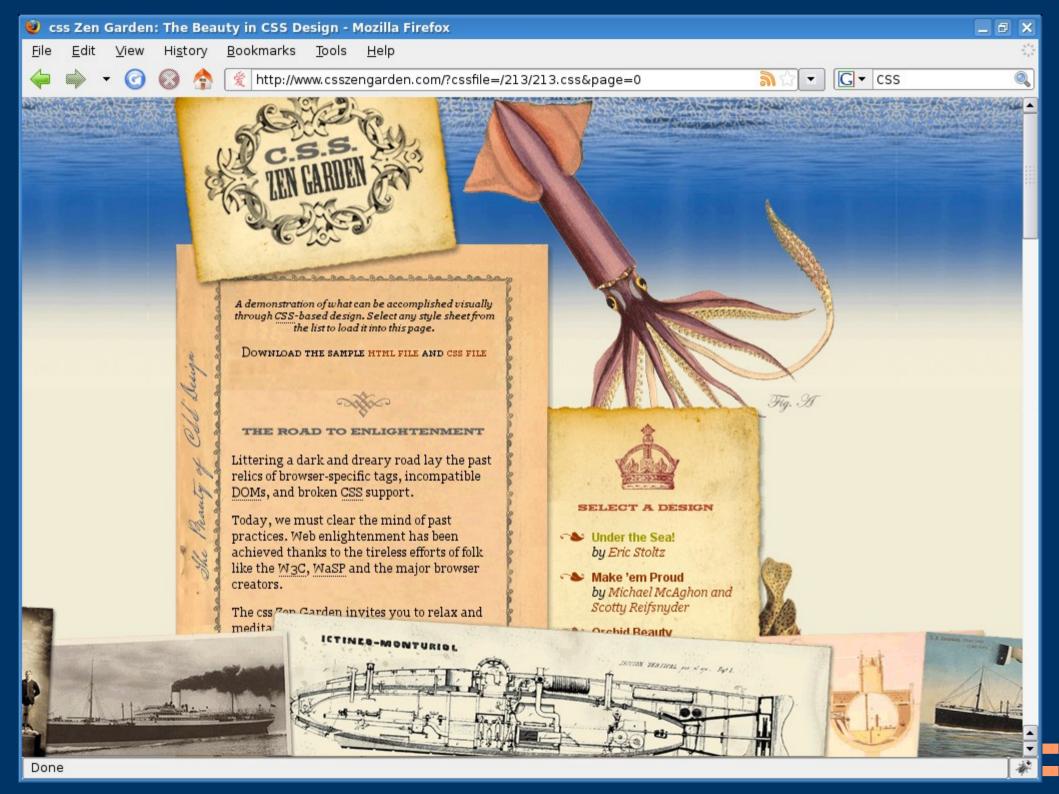












HTML

- Allows Authors to
 - * Publish online documents with headings, text, tables, lists, photos, etc.
 - * Retrieve online information via hypertext links, at the click of a button.
 - * Design forms for conducting transactions with remote services, for use in searching for information, making reservations, ordering products, etc.
 - * Include spread-sheets, video clips, sound clips, and other applications directly in their documents.

HTML Document Design Considerations

- Separate structure and presentation
 - Preferable practice: HTML document contains web page contents only, use Style Sheets to specify rendering.
- Consider universal accessibility to the web
 - Web should be available to everyone.
 - Authors should consider variety of browsers
 - E.g. Speech-based browsers, braille browsers, etc.
- Help User Agents(Browsers) with Incremental Rendering

HTML (SGML) – Language of tags

<singleton/>

Use tags to mark elements of the document.

HTML and Why The Shift to XML-based Formats

- HTML can be written in a bad way.
 - Browsers must therefore be robust in handling "Bad HTML"
 - Implication: Browser code more complicated and larger than it should be.
 - Not standardized and browsers do not impose
- HTML is not relevant/practical in NEW environments where web access is made
 - Consider the different ways to access the Internet esp. on "devices" with limited resources.

XML: Ultimate Web Content Format

- XML
 - Extensible Markup Language
 - A meta language as compared to SGML
 - Simpler and "smaller" than SGML
 - XML is simplified SGML
 - Strict
 - All XML-based documents must follow rules on well-formedness.
 - Standardized Parsers
 - Text Format
 - Tag-based

XML: XHTML --> HTML with an X

- XHTML 1.0
 - Current web content trend
 - HTML with the rules of XML plus more.
 - Imporant Differences with HTML
 - XHTML elements (tags) are properly nested [Universal XML rule]
 - XHTML documents must be well-formed [Universal XML rule]
 - Tag names and attributes **MUST** lowercase.
 - XHTML elements (tags) must be **closed** [still on being well-formed].
 - Some HTML elements not supported, some attributes are required.
 - Doctype required.
 - Can be "mixed" with other XML documents.
 - Authors may create user-defined tags.

Proliferations of XML-based formats

- MathML
 - Mathematical equations
- CML
 - Chemical Markup Language
- SVG
 - Scalable Vector Graphics
- SMIL
 - Synchronized Multimedia Integration Language
- GML
 - G for Geographic
 - G for Geometric
- Etc...

XML – A Must Know

• ...we will study XML in some greater detail in future discussions.

Additional Readings

- HTML 4.01 Specification (W3C)
 - http://www.w3.org/TR/REC-html40/cover.html
- XHTML 1.0 The Extensible Hypertext Markup Language
 - http://www.w3.org/TR/xhtml1/
- XHTML Tutorial -- for your laboratory class
 - http://www.w3schools.com/xhtml/default.asp