HTML

• HyperText Markup Language
• An SGML application
  – (Standardized General Markup Language)
  – HTML is an instance of an SGML language specifically intended to markup general types of documents.
    • a compromise
    • SGML applications intended to be more specific
      – e.g., recipe markup, resume markup
• Now XHTML = an XML Application

Content not Format

• Originally, HTML, as inspired by SGML, was intended to markup content, not specify formatting instructions.
  – resume.html
• The job of the Web browser is to render the marked-up content into human-readable output.
• In addition, SGML always assumed the presence of tools to create the SGML as well as display it.
  – Same with HTML: but, many prefer straight text editing because of tool limitations.
Trouble in Paradise

- Unfortunately, SGML, though complex, was correct in its assumption that different types of documents needed different markup languages
  - e.g., Resumes, cookbooks, Shakespearean plays, …
  - resume.xml

Format not Content

- HTML quickly 'degenerated' into a markup language for format
  - something for which it was not designed
- Graphic designers starting to work on the Web were horrified at the lack of expressibility.
  - Used hacks:
    - e.g., "pouring" of text into tables for precise paragraph layout
    - e.g., using 1-pixel transparent images for precise spacing and alignment
- 2 warring factions ('html purists' and designers)
- Designers won
A Temporary Truce

- CSS (cascading Style Sheets)
- Allows graphic designers much better control over formatting.
- Separate file for detailed formatting instructions from the HTML markup.
  - Designers were happy:
    - more control over formatting
  - Purists were happy:
    - removes the requirement for further formatting tags from HTML proper (a battle purists were headed towards defeat on anyways)

XML: The Solution

- Realizing that they were going to lose anyways, purists abandoned HTML and CSS to the graphics designer camp.
- Invented XML (eXtensible Markup Language) which is a complete return to SGML (a bit simpler)
  - Very useful for inter-application communications as well, as we shall see
- Invented XSL (eXtensible Stylesheet Language) for formatting XML
  - in particular, XHTML (= strictly formatted HTML)
  - Goal of XSL:
    - be able to specify how browsers should display XHTML
    - be able to specify how XHTML should publish
**HTML Standards**

- HTML, first version (Berners-Lee) '92
  - An SGML(ISO8879) application
- HTML 2.0 (Berners-Lee, Dan Connoly)
  - produced by IETF (Internet Engineering Task Force), HTML Working group, RFC 1866 (Request for Comments)
- HTML 3.2 (Dave Raggett) '96
  - W3C
  - tables, applets, text-flow around images, …
- HTML 4.0 (Raggett et. al.) '98
  - frames

- HTML 4.01, Dec 24, '99 (long)
  - cleanup for use with XHTML 1.0
- XHTML 1.0 Jan. 2000 (short)
  - based on XML rather than SGML
    - references HTML 4.01 for semantics of tags
    - references XML 1.0 for language syntax
    - provides XML DTDs for definitions

**Evolving Standards**

- Designed to address client heterogeneity
  - XHTML Modularization, Apr., 2001 (long)
    - individually delimits and defines modules in XHTML 1.0
    - describes how to add new modules
  - XHTML Basic, Dec., 2000 (short)
    - minimal set of modules (e.g., for cellphones)
    - structure, text, hypertext, list, basic forms, basic tables, image, object, meta-info, link, base
  - XHTML 1.1 Mar. 2001 (short)
    - reformulation of XHTML 1.0 Strict based on modularization
  - … more modules to come
    - frames, …
How W3C Works

- World Wide Web Consortium
  - created in October 1994
  - developing common protocols for the Web
- Members:
  - companies, gov't, standards bodies, …
- the Team:
  - full-time employees, paid for by the Members.
  - Working out of MIT, INRIA, and Keio.
  - http://www.w3.org/Consortium

Syntax of XHTML 1.0

- Conforming to XML syntax …
  - W3C XML 1.0, October 2000
- … and 1 of 3 XHTML 1.0 DTD (Document Type Definition)
  - W3C XHTML 1.0 Jan. 2000
  - strict
    - layout-tag and frame-free. Use with CSS for layout
  - transitional
    - compatible with 4.01 no frames
      - e.g., <body bgcolor='red'>
  - frameset
    - transitional + frames
Strict XHTML

prolog  <?xml version="1.0" encoding="iso-8859-1"?>

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Frameset//EN" "DTD/xhtml1-frameset.dtd">

<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en" lang="en">
<head>
<meta http-equiv="Content Type" content="text/html; charset=iso-8859-1"/>
<title>A Document</title>
</head>
<body>
<p>A paragraph</p>
</body>
</html>