lecture 7a xhtml+css
structural markup

Provides structure to the document

Headings (<h1>, <h2>, <h3>, ... ) are structural elements, paragraphs <p> also helps to indicate the structure the content

More structural markups in XHTML: <div> and <span>

stylistic markup

Stylistic markup changes the way the text inside the markup appears, e.g. <center>, <i>, <b>, <u>, <blink>

semantic markup

Semantic (or descriptive) markup gives meaning to the content of the markup, but does nothing to appearance

<contacts>
  <name>Ivo</name>
  <email>ivow@unimelb.edu.au</email>
</contacts>
html: from past to the future

<table>
<thead>
<tr>
<th>HTML 2.0</th>
<th>HTML 3.2</th>
<th>HTML 4.0</th>
<th>XHTML 1.0</th>
<th>HTML5</th>
</tr>
</thead>
</table>

The original HTML, mostly structural markup (<h1>,<h2>,<p>), not intended for formatting a document (no <font> tag).

Stylistic markup like <font> and color attributes were added.

Mix of content and presentation: content chaos, the era of colorful but messy web pages.

Formatting is removed from the document and stored in a separate style sheet (CSS).

Structure is enforced even further.

Stricter and cleaner document.

Presentation is separated from the document structure.

HTML as XML document.

New elements and attributes for modern Web sites.

New semantic markup <nav> (website navigation) and <footer>.

Multimedia <audio> and <video>.

higher level of abstraction
html: the past

Welcome to
THE UNIVERSITY OF MELBOURNE

About the University
Information, News and Events.
Admissions & Scholarships
Prospective students, local and international.
Contacting People
Email, telephone & other details.
The Community
Cultural, alumni & school programs.
Faculties, Departments & Schools
Staff, courses and information.
Careers & Jobs
Job vacancies and career information.
The University Library
References, loans and collections.
Research Information
Projects, grants, centres.
Information Technology Services
Including WWW information.
Student Information
For currently enrolled students.
Procedures & Publications
University documents.
Search the CWIS
A full index of the University Web.

Created: 17 May 1994
Last modified: 1 May 1997
Authorized by: Director, Information Technology Services

Maintainer: CWIS Group, Information Technology Services.
Email: cwis@www.unimelb.edu.au


www.unimelb.edu.au
(May 1997)
HTML 3.2
html: the past

```html
<html>
<head>
    <title>The University of Melbourne</title>
    <link rel="made" href="mailto: cwis@cwis.unimelb.edu.au"/>
</head>
<body bgcolor="#FFFFFF">
    <p align="center"><img src="/images/cwis-banner.gif" alt="The University of Melbourne."/>
    </p>
    <hr>
    <table border="0" width="100%" cellspacing="10">
      <tr align="left" valign="top">
        <td>
          <dl>
            <dt>
              <a href="/about/">
                <img src="/images/black_arrow.gif" alt="*
                    " align="bottom" width="22" height="25"
                    border="0"> <strong>About the University</strong></a>
            </dt>
          </dl>
        </td>
      </tr>
    </table>

...
html: the present day

http://www.unimelb.edu.au
html: the present day

```xml
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
  "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en" lang="en">
  <head>
    ... 
    <title>HomePage of the University of Melbourne, Victoria, Australia</title>
    ...
    <link rel="stylesheet" type="text/css" href="http://www.unimelb.edu.au/homepage/07-04/styles/styles.css" media='screen, projector'/>
    <link rel="stylesheet" type="text/css" href="http://www.unimelb.edu.au/homepage/07-04/styles/handheld.css" media='handheld'/>
    ...
  </head>
  <body id="home">
    <div id="top">
      <a href="#links" class="skiplink">Skip to navigation</a>
      <a href="#content" class="skiplink">Skip to content</a>
      <a href="http://m.unimelb.edu.au" class="skiplink">Mobile version</a>
      <h1><a href="http://www.unimelb.edu.au/">The University of Melbourne</a></h1>
    </div>
    ... 
    <div id="content">
      <div id="dynamic-content">
        <div id="media">
          <h2>Latest news</h2>
          <p class="feeds">
            <a href="http://uninews.unimelb.edu.au/mediareleases.php" class="more" title="link to more latest news">more Latest news</a>
          </p>
        </div>
      </div>
    </div>
</body>
</html>
```

- **xml namespace is used**
- **empty elements are closed**
- **(well formed XML)**
- **language is specified**
- **tell the browser what to expect**
- **style/presentation/formatting is separated from the content**
... XHTML 1.0 is a reformulation of HTML 4.01 in XML, and combines the strength of HTML 4 with the power of XML.

XHTML 1.0 was the first major change to HTML since HTML 4.0 was released in 1997. It brings the rigor of XML to Web pages and is the keystone in W3C's work to create standards that provide richer Web pages on an ever increasing range of browser platforms including cell phones, televisions, cars, wallet sized wireless communicators, kiosks, and desktops.

XHTML 1.0 was the first step: it reformulates HTML as an XML application. This makes it easier to process and easier to maintain. XHTML 1.0 borrows elements and attributes from W3C's earlier work on HTML 4, and can be interpreted by existing browsers, by following a few simple guidelines.

... 

http://www.w3.org/MarkUp/
differences between plain old html (3.2) & xhtml

- an xhtml document is an xml document

- xhtml has stronger structural enforcement *)
  provision of <div> and <span> elements

- separation between content and presentation *)
  e.g. <center>, <i>, <b>, <u>, <blink>
  stylistic markup should be generally avoided within XHTML, use CSS (Cascading Style Sheets)

*) true for HTML 4, too
xhtml & css advantages

making the structure of a web page more sound,
easier processing using XML libraries,

(in conjunction with css)
separation of concerns: content and presentation,
better management of content,
smaller HTML code,
better search engine optimization.
indexing xhtml pages

for every web page:
   load the web page as an xml file
   (thank god for xhtml!)
   get the text content of all elements under <body>
   break down the text into list of words
   for every word:
      normalise each word
      add the word into the index

pickle the index
using the index to search xhtml pages

HTML form

query = "rabbit"

Web Browser

Search results:
- page2.html
- page3.html
xhtml: well formed xml

- elements must be properly closed with an end tag
  - X `<p>This is a para.<p>This is another para.`
  - √ `<p>This is a para.</p><p>This is another para.</p`
- empty elements must also be properly closed
  - X `<br>`
  - √ `<br />`
  - √ `<br /> <hr /> <img src="an_image.jpg" />`
- correct ordering of nested elements
  - X `<em><strong>Something is wrong</em></strong>`
  - √ `<em><strong>Something is good</strong></em>`
- all attribute values must be properly quoted
  - X `<td colspan=3>`
  - √ `<td colspan="3">`
xhtml: well formed xml

using appropriate entities

X  <title>Hansel & Gratel</title>
√ <title>Hansel &amp; Gratel</title>
√ <a href="page.php?id=about&amp;style=5">About Us</a>

element and attribute names are case-sensitive

X  <P ID="sec1">the first paragraph</P>
√ <p id="sec1">The first paragraph</p>

attribute minimization is not allowed

X  <select multiple>
   <option value="1">one</option>
   <option value="2">two</option>
</select>
√  <select multiple="multiple">
   <option value="1">one</option>
   <option value="2">two</option>
</select>
typical xhtml document

```xml
<?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">
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en-au">
    <head>
        <title>My first XHTML page</title>
        <meta name="description" content="Sample XHTML page" />
        <meta name="keywords" content="xhtml, web page" />
        <meta name="author" content="John Doe" />
        <link rel="stylesheet" href="my_styles.css" type="text/css" />
    </head>

    <body>
        <p>Content of the page.</p>
    </body>
</html>
```
deconstructing xhtml

<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
    "http://www.w3.org/TR/xhtml1/DTD/strict.dtd">
<html xmlns="http://www.w3c.org/TR/xhtml1">
  ...
</html>

XML declaration
  optional in XHTML1.0
  often includes encoding attribute to define character set

Document Type Definition (DTD)
  identifies the elements in a vocabulary, specifies which attributes can be used and
dictates how the elements can be used together

Namespace
  every element and attribute in the document is made part of a unique namespace

Meta-Data <meta />
  conveys information about the document
  'name' attribute is interpreted by browser or user agent
    "keyword", "description", "summary", "author" ...
  'content' attribute contains all the meta information
validation of xhtml

Use validation tool e.g. W3C

http://validator.w3.org/ allows validation by direct input

Common validation problems

no ‘alt’ attribute attached to ‘img’
incorrectly nested elements e.g. <a> then <h3>
xhtml 1.0 – 3 flavours

XHTML 1.0 Strict
provides clean structural mark-up, free of any tags associated with layout (free of stylistic markup)
http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd

XHTML 1.0 Transitional
takes advantage of the XHTML features (including CSS) but is more accessible to older browsers.
Allows the use of <body> with bgcolor, text and link attributes.
http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd

XHTML 1.0 Frameset
Use this when you want to use HTML Frames to partition the browser window into two or more frames.
http://www.w3.org/TR/xhtml1/DTD/xhtml1-frameset.dtd
strict and transitional

Presentational elements that are not allowed in Strict DOCTYPEs

  <center>
  <font>
  <u>

Presentational attributes that are not allowed in Strict DOCTYPEs

  align
  background
  bgcolor

http://www.w3schools.com/tags/default.asp
<div> and <span>

Structural markups used to bind a particular (group of) content to a specific logical meaning.

The `<span>` element is used for inline markup:

<p>I am reading <span class="book">Harry Potter</span></p>

The `<div>` element is used for block markup:

<div class="sale_items">
  <p>This week's sale items are:</p>
  <ul>
    <li>Pink Lady apples</li>
    <li>Bananas</li>
  </ul>
</div>

The `<div>` element can enclose any inline or block elements.

Block elements are `<p>`, `<blockquote>`, `<pre>`, `<h1>...<h6>`, `<div>`

The `<span>` can only enclose inline elements.

Inline elements are text and `<img>`
summary

You should be able to

- outline the limitations of HTML
- discuss the benefits offered by XML
- distinguish between the 3 ‘flavours’ of XHTML
- discriminate between the 3 types of markup
- list the key differences between HTML and XHTML
- identify the key elements of an XHTML document