Lënda: Programimi në Internet

XML, XHTML, & HTML5
Dr. Ing. Lule Ahmedi

Slides mostly adopted from Tufts University, Prof. Ming Chow

•Extensible Markup Language (XML) vs. HTML

- HTML 4.0 and its predecessors are defined using SGML
- SGML is very expressive, but too complicated to implement, e.g., in browsers
- HTML comes bound with a set of predefined tags, each with already perceived semantics

XML overcomes these drawbacks:

- (1) it keeps the structure and the presentation separate
- (2) allows to define user-specific tag structures with arbitrary assigned semantics by applications
- (3) Not as general as SGML, i.e., easy to parse (e.g., using PHP)

•Wide Use of XML

- The x everywhere (e.g., doc + xml = docx)
- Configuration files (e.g., for servers, for web application)
- Graphics
- Databases
- Web services
- RSS feeds
- Security (e.g., XML Signature)
- Middleware:
    - acts as a nervous system
- MathML, chemistry, ..

• XML Syntax Requirements

- Every tag closes
- Every attribute is quoted
- Element names are cAsE sEnSiTiVe
  - Example: <pattern></PATTERN> is bad!

• Document Type Definition (DTD)

DTD:

- Contains a set of rules that can be used to validate an XML file
- Create declarations that define:
    - elements, attributes, entities, and notations for any XML files that reference the DTD file
- Establish constraints for how each element, attribute, entity, and notation may be used within any XML files that reference the DTD file

•XHTML

- eXtensible HyperText Markup Language
  - Uses XML (XML + HTML = XHTML)
  - Why? More interoperable across different web browsers
- More rigid and structured
- The DOCTYPE declaration is always the first line in an XHTML document
- Element names are lowercase
- Elements must be properly nested
- Reference: http://www.w3.org/TR/xhtml1/

• Three distinct DTDs

STRICT - Clean markup; no HTML presentation features; best to use with CSS

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML
1.0 Strict//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-
strict.dtd">
```

TRANSITIONAL - Not as tight; you can use HTML presentation features (e.g., center, font tags)

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML
1.0 Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-
transitional.dtd">
```

FRAMESET (we will not use!) - You can use HTML frames (ugh!)

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML
1.0 Frameset//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-
frameset.dtd">
```

Extensible Markup Language (XML)

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">

<html xmlns="http://www.w3.org/1999/xhtml">

<head> <!-- Information about the page (e.g., title, abstract,
stylesheet declarations, keywords) goes here. Does not appear in
the rendered document. -->
</head>

<body> <!-- Content of the page goes here. So many ways to
create content. -->
</body>

</html>
```

- `xmlns` attribute specifies the XML namespace for a document
  - here, http://www.w3.org/1999/xhtml for XHTML documents)
- The importance of a namespace:
  - no ambiguity when objects of different origins but same names are mixed together

More XHTML Rules

- Some tags require certain attributes (e.g., id for <map> and alt for <map>)
- Need to encode special characters especially & (&), > (>), < (<)
-   required for empty table cells

•Validation

- Remember, even if you have a malformed XHTML document, it may still render correctly on your web browser
- Use the official XHTML validation tool to check your XHTML code to ensure that it conforms to the strict XHTML specification http://validator.w3.org/
- If errors are found, fix them!

•Really Simple Syndication (RSS)

- A.k.a. Rich Site Summary, a.k.a. feed, a.k.a. channel
- To syndicate content automatically
- Must conform to XML 1.0 specification
  - Current specification: RSS 2.0
- RSS readers / aggregators:
  - Web browsers (e.g., Safari), Google Reader
- Official W3C validation available:
  - http://validator.w3.org/feed/
- For more information:
  - http://cyber.law.harvard.edu/rss/rss.html

Really Simple Syndication (RSS)

- Required channel elements:
  - title
  - link - to your website
  - description
- Optional channel elements:
  - pubDate - in UTC format (e.g., Sat, 07 Sep 2002 09:42:31 GMT)
  - link - to your website
  - description
- Elements for an item:
  - title
  - link
  - description
  - author - an e-mail address
  - pubDate

•Structure of an RSS Feed

```xml
<?xml version="1.0" encoding="UTF-8"?>  <?xml-stylesheet title="XSL_formatting" type="text/xsl"  href="/../nolsol.xsl"?>
<rss xmlns:media="http://search.yahoo.com/mrss/" xmlns:atom="http://www.w3.org/2005/Atom" version="2.0">

<channel>
<title>BBC News - Europe</title>
<link>http://www.bbc.co.uk/go/rss/int/news/-/news/world/europe/</link>
<description>The latest stories from the Europe section of the BBC News web site.</description>
<language>en-gb</language>
<lastBuildDate>Mon, 07 Mar 2011 19:32:32 GMT</lastBuildDate>
<copyright>Copyright: (C) British Broadcasting Corporation, see http://..  for terms and conditions of reuse. </copyright>
<ttl>15</ttl>
<atom:link href="http://feeds.bbci.co.uk/news/world/europe/rss.xml" rel="self" type="application/rss+xml"/>

<item>
<title>Cyber attack targeted Paris G20</title>
<description>The French finance ministry confirms it came under a cyber attack in December that targeted files on the Paris G20 summit.</description>
<link>http://www.bbc.co.uk/go/rss/int/news/-/news/business-12662596</link>
<guid isPermaLink="false">http://www.bbc.co.uk/news/business-12662596</guid>
<pubDate>Mon, 07 Mar 2011 15:35:55 GMT</pubDate> </item>  ...

<item>
<title>Kosovo</title>
<description>An overview of Kosovo including key facts, leaders and notes on the media</description>
<link>http://news.bbc.co.uk/go/rss/int/news/-/1/hi/world/europe/country_profiles/3524092.stm</link>
<guid isPermaLink="false">http://news.bbc.co.uk/1/hi/world/europe/country_profiles/3524092.stm</guid>
<pubDate>Wed, 02 Mar 2011 15:48:26 GMT</pubDate> </item>  ...
</channel>
</rss>
```

- **The Web Developer Plugin for Firefox**

  - https://addons.mozilla.org/en-US/firefox/addon/60

•HTML5

- In summer 2009, the W3C announced they will not continue with any future development of XHTML: http://www.w3.org/News/2009#item119

- To make a long story short, the future direction is HTML5

- Infographic: http://www.focus.com/images/view/11905/

- The three aspects of HTML5:

    - Content (HTML)

    - Presentation of content (CSS)

    - Interaction with content (JavaScript)

- XML-based but no longer SGML-based

- Originally coined as Web Applications 1.0, now still work in progress

- Backing from Google and of course Apple

- Currently supported in some degree in Chrome, Firefox 3.5+, Opera, and Safari

- Will be flexible with error handling (i.e., incorrect syntax)

- W3C now validates HTML5 pages (experimental): http://validator.w3.org/

•Deprecated Tags in HTML5

- The presentation elements: `font`, `center`
- `frame`, `frameset`
- `applet`
- Old special effects: `marquee`, `bgsound`
- `noscript`
- Reference: http://www.html-5.com/avoid/
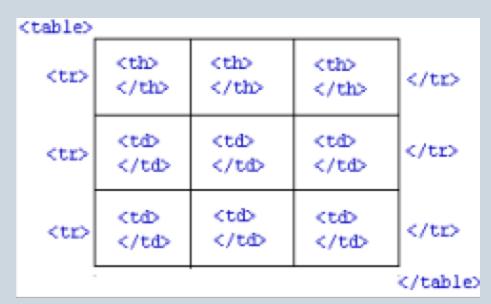
•Deprecated Attributes in HTML5

- `align`
- `border`
- Reference: http://www.html-5.com/avoid/

•HTML5: What is In

- New tags, including `button`, `video`, `audio`, `article`, `footer`, `nav`
- New media events
- New canvas tag for 2D rendering
- New form controls for date and time
- Geolocation
- Local storage
- New selectors
- Example: A Preview of HTML5 by Lachlan Hunt, http://www.alistapart.com/articles/previewofhtml5/

Old Times – HTML Blind Tables

- If text and graphics are to be well-positioned:
  - Have used "blind tables" for the page layout design!

•Structure of an HTML5 Document

```
<!DOCTYPE html>
<html>
          <head><title>An HTML Document</title> </head>
          <body>
                    <header>
                                   <h1>A Preview of HTML5</h1>
                    </header>
                    <nav>
                     <ul>
                       <li><a href="/">Home</a></li>
                       <li><a href="/prodServ">Products and Services</a></li>
                       <li><a href="/about">About</a></li>
                     </ul>
                    </nav>
                    <article>
                                   <section>
                                    <h1>Chapter 1: The Period</h1>
                                    <p>It was the best of times, it was the worst of times,
                                                it was the age of wisdom, it was the age of foolishness,...</p>
                                   </section>
                    </article>
                    <aside>
                     <h1>Archives</h1>
                     <ul>
                       <li><a href="/2007/09/">September 2007</a></li>
                       <li><a href="/2007/08/">August 2007</a></li></ul>
                    </aside>
          </body>
</html>
```