



# CS3220 Web and Internet Programming

## HTML and XML Basics

Chengyu Sun  
California State University, Los Angeles



# HTML

◆ HyperText Markup Language, the language for creating web pages

```
<html>
  <head><title>CS3220</title></head>
  <body>
    <h2>Welcome to CS3220!</h2>
  </body>
</html>
```

# XML

◆ Extensible Markup Language, the *general-purposed* language to describe structured data

```
<user>
  <name>Chengyu Sun</name>
  <office>ET A317</office>
  <phone>(323) 343-6697</phone>
</user>
```

# A Brief History of HTML and XML

Early markup languages (1960s to 1980s)

↓  
SGML (early 1980s)

↙  
HTML (1991)

↘  
XML (1998)

↓  
HTML 4.01 (1999)

↓  
XHTML 1.0 (2000)

↓  
HTML 5 (2014)

↗  
XHTML 1.0 (2000)

# An XML Example

```
<?xml version="1.0" encoding="UTF-8"?>
<user type="faculty">
  <name>Chengyu Sun</name>
  <office>ET A317</office>
  <phone ext="6697" />
</user>
```

# Some Terminology

- ◆ Declaration

- ◆ Tag

- Opening (start) tag
- Closing (end) tag

- ◆ Attribute

- ◆ Element

- Content
- Empty element (i.e. no content)
- Root element

# What About Encoding?

- ◆ An encoding is a way to represent characters in a computer as a sequence of 0's and 1's
- ◆ For more details, please read [this](#)
- ◆ Or, just remember always use UTF-8

# NOT Well-Formed XML

```
<?xml version="1.0" encoding="UTF-8"?>
<users>
  <user type="faculty">
    <name>Chengyu Sun</Name>
    <office>ET A317</office>
    <phone ext="6697">
  </user>
  <user type="faculty">
    <name>Raj Pamula</name>
    <office>E&T A324</office>
  </users>
</user>
```



# Validity of XML

- ◆ Q: How can we decide if some well-formed XML is valid (i.e. *correct*)?
- ◆ A: We need a "specification" of the elements used in the XML.

# Specify XML Grammar – DTD

- ◆ Document Type Definition

- ◆ DTD Example -

[http://www.w3schools.com/xml/xml\\_dtd\\_examples.asp](http://www.w3schools.com/xml/xml_dtd_examples.asp)

# Reference DTD in XML

- ◆ Reference a "private" DTD (DTD intended for use by a single author or a group of authors)

**<!DOCTYPE** root\_element **SYSTEM** "DTD\_location">

- ◆ Reference a "public" DTD

**<!DOCTYPE** root\_element **PUBLIC**  
"DTD\_name" "DTD\_location">

# DOCTYPE Examples

## ◆ Private DTD

```
<?xml version="1.0" encoding="UTF-8"?>  
<!DOCTYPE user SYSTEM "mydtd.dtd">  
<user type="faculty">  
    <name>Chengyu Sun</name>  
    <office>ET A317</office>  
    <phone ext="6697" />  
</user>
```

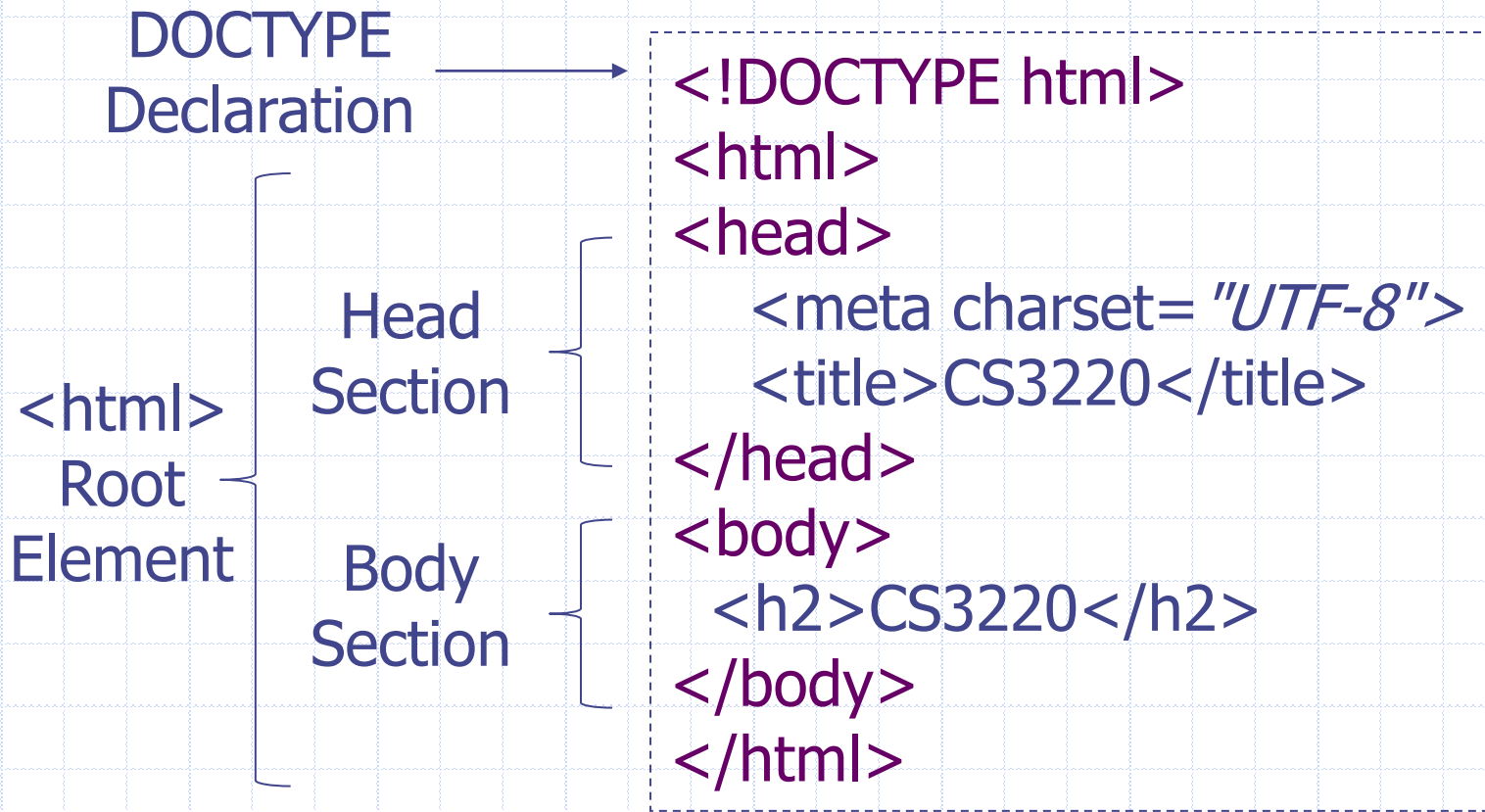
## ◆ Public DTD

See [https://www.w3schools.com/tags/tag\\_doctype.asp](https://www.w3schools.com/tags/tag_doctype.asp)

# XML Recap

- ◆ Terminology
- ◆ Well-formedness
- ◆ Validity
  - DTD and `<!DOCTYPE>`

# Structure of an HTML Page



# DOCTYPE Declaration

- ◆ Specifies which version of HTML/XHTML the document uses
- ◆ Common DOCTYPE declarations
  - HTML 4.01 Strict
  - HTML 4.01 Transitional
  - XHTML 1.0 Strict
  - XHTML 1.0 Transitional
  - HTML 5

# What About Well-formedness of HTML?

◆ HTML (*not* XHTML) is more forgiving

- Tag and attribute names are case-insensitive
- Some tags are self closing, e.g. `<meta>` and `<br>`
- Some closing tags are optional, e.g. `<p>`, and even `<html>`, `<head>`, `<body>`



# <head> Section

- ◆ Contain elements that provide information about the web page rather than the content of the page
  - <title> - title for the browser tab
  - <meta> - metadata about the page
  - <link> - mostly used for style sheets

# <body> Section

- ◆ Contain elements that will be displayed in the browser window

# Tools for HTML Development

## ◆ Text editors for developers

- [Sublime Text](#)
- [Visual Studio Code](#) by Microsoft
- [Atom](#) by GitHub

## ◆ IDEs

- Visual Studio
- Eclipse

## ◆ Professional tools for web designers like Adobe Dreamweaver

# Using VS Code

- ◆ File and folder
- ◆ Auto Save
- ◆ Recommended extensions
  - HTMLHint
- ◆ Emmet - <https://emmet.io/>
  - `html:5`
  - `+`, `>`, and `*`
  - `lorem`

# Headings and Paragraphs

- ◆ `<h1>`, `<h2>`, ..., `<h6>`

- ◆ `<p>`

- ◆ A block element always starts on a new line and can contain other elements



# ... Other Commonly Used Block Elements

◆ `<ul>` - unordered list

◆ `<ol>` - ordered list

◆ `<li>` - list item

# Common Inline Elements

- ◆ `<i>` - italic, `<b>` - bold, `<u>` - underlined
- ◆ `<em>` - emphasized
- ◆ `<strong>` - strongly emphasized
- ◆ `<code>` - for computer code
- ◆ `<span>` - a general-purposed inline element (like an inline version of `<div>`)
- ◆ `<br>` - line break



# Anchor <a>

- ◆ Creates a hyperlink to other web pages, files, locations within the same page, email addresses, or any other URL.

**<a href="http://www.calstatela.edu">CSULA</a>**



URL

# Absolute and Relative URL ...

## ◆ Absolute URL

**http://**www.calstatela.edu/logo.gif

# ... Absolute and Relative URL

## ◆ Relative URL

`/logo.gif`

`/img/logo.gif`

Relative to the  
root path

`logo.gif`

`img/logo.gif`

`../logo.gif`

`../../logo.gif`

`../../img/log.gif`

Relative to the current path

- —→ current folder
- —→ parent folder

# Link to Locations within a Page

```
<a href="#section1">Introduction</a>  
<a href="#section2">HTML Basics</a>  
...
```

```
<h2 id="section1">Introduction</h2>  
<p>...</p>
```

```
<h2 id="section2">HTML Basics</h2>  
<p>...</p>
```

# Some Core Attributes

- ◆ Attributes that are common to all HTML elements
- ◆ `id`: a unique identifier for the element
- ◆ `title`: additional information about the element
- ◆ `class`: CSS class
- ◆ `style`: inline CSS style
- ◆ More at <https://www.w3.org/TR/2012/WD-html-markup-20121011/global-attributes.html>

# Other Uses of <a>

## ◆ Email link

`<a href="mailto:csun@calstatela.edu">Email Me</a>`

## ◆ Phone link

`<a href="tel:+1-323-343-6697">Call me</a>`

# Image <img>

## ◆ Attributes

- `src`: URL to the image
- `alt`: alternate text if the image cannot be displayed
- `height` and `width` in pixels

◆ JPEG, GIF, PNG, SVG images are supported by most browsers

◆ *Clickable image??*

# HTML Special Characters (HTML Entities)

◆ White space: `&nbsp;`;

◆ `<`: `&lt;`;

◆ `>`: `&gt;`;

◆ `&`: `&amp;`;

◆ More at

[http://www.w3schools.com/html/html\\_entities.asp](http://www.w3schools.com/html/html_entities.asp)



# HTML Comment

`<!-- This is a single-line comment. -->`

`<!-- This is a  
multi-line  
comment. -->`

◆ The syntax for comments is the same for both XML and HTML

# HTML Recap

- ◆ Basic page structure
- ◆ Block elements
- ◆ Inline elements
- ◆ `<a>` and `<img>`
  - Absolute and relative URL
- ◆ Special characters
- ◆ Comment