Announcement

• No Plagiarism!
• Attendance is required! If you miss two unexcused lectures, you will automatically fail (excluding the first two weeks.)
• DECO2102 students go to 261 for lab
• DECO1200 students go to 313 for lab
• Assignment #1 due in two weeks
Lecture 3: Markup Language

Dr. Xiangyu WANG

10th August, 2006
A **markup language** combines text and extra information about the text.

The extra information, for example about the text's structure or presentation, is expressed using **markup**, which is intermingled with the primary text.

HyperText Markup Language (HTML)
Examples of Markup Language

- XML
- HTML
- Dynamic HTML
- SAML
- SOAP
- SQL
- UDDI
- VRML

- WSDL
- XAML
- XKMS
- XLANG
- XPath
- XSLT
History of Markup Language

• Historically, markup was (and is) used in the publishing industry in the communication of printed work between authors, editors, and printers.
History of Markup Language

• **GenCode:**
  – First presented by publishing executive William W. Tunnicliffe at a conference in 1967, as "generic coding."
  – Tunnicliffe would later lead the development of a standard called “GenCode” for the publishing industry.

• **TeX:**
  – By Donald Knuth in the 1970s and 80s.
  – Requires considerable skill from the user.
  – LaTeX

• **SGML (Standard Generalized Markup Language)**
  – In the early 1980s, the idea that markup should be focused on the structural aspects of a document and leave the visual presentation of that structure to the interpreter led to the creation of SGML.
  – Developed by a committee chaired by Goldfarb.
  – Allow authors to create and use any markup they wished, selecting tags that made the most sense to them and were named in their own natural languages.
  – Many particular markup languages are derived from it.
  – Cumbersome and difficult to learn
History of Markup Language

• **HTML (Hypertext Markup Language)**
  - Sir Tim Berners-Lee used SGML syntax to create HTML.
  - HTML resembles any other SGML-based tag language
  - Simpler
  - Likely the most used document format in the world today.

• **XML (eXtensible Markup Language).**
  - Developed by the World Wide Web Consortium, in a committee created and chaired by Jon Bosak.
  - Simplify SGML by focusing on a particular problem — documents on the Internet.
  - Allow users to create any tags needed (hence "extensible") and then describe those tags and their permitted uses.
  - Eliminated the complex features of SGML, radically easing learning and implementation.
  - Sweet spot between simplicity and flexibility

• **XHTML (eXtensible Hypertext Markup Language)**
  - The newest incarnation of HTML based on XML
  - More rigorous and robust version that requires documents to be "well-formed" XML documents, but which uses the familiar HTML tags.
Introduction

• A “web page” is a TEXT file containing:
  – text to be displayed (text "elements")
  – style names (e.g., “tags”)
  – “pointers”
    • to other web pages through URL (Uniform Resource Locator)
    • to graphics, for display on the page
    • to sounds, to be played when the page is viewed
Classes of Markup Language

• Presentational: Presentational markup is an attempt to infer document structure from cues in the encoding.

• Procedural: Procedural markup is typically also focused on the presentation of text, but is usually visible to the user editing the text file, and is expected to be interpreted by software in the order in which it appears. For example, in HTML, tag “b” for bold.

• Descriptive: Descriptive Markup or semantic markup applies labels to fragments of text without necessarily mandating any particular display or other processing semantics. For example, the "href=" attribute in HTML.
Classes of Markup Language

• In practice, the classes of markup usually co-occur in any given system.

• **Generic Markup** is another term for "Descriptive Markup". Most modern descriptive markup systems structure documents into trees, while also providing some means for embedding cross-references.
Classes of Markup Language

• Descriptive markup also facilitates the simpler task of reformatting a document as needed, because the format specification is not intertwined with the content.

• “Italic” for emphasis or foreign words
Terms

• **Content**
  – The actual 'meat' of a document -- all of the words, images, and links which a user can read and interact with.
  – "whatever you put in the document."

• **Hyperlink**
  – A link from one document to another, or to any resource, or within a document.

• **Markup Tag/instruction**
  – Tags typically occur in begin-end pairs. These pairs are in the form
    `<tag> ... </tag>`

• **Containers:** These pairs define *containers*.
  – The text within a "boldface container" would be boldfaced.
  – Paragraphs are defined using a "paragraph container."
Example

A common feature of many markup languages is that they intermix the text of a document with markup instructions in the same data stream or file. A small section of text marked up in HTML is shown as follows:

```html
<h1> Anatidae </h1>
<p>
The family <i>Anatidae</i> includes ducks, geese, and swans, but <em>not</em> the closely-related screamers.</p>
```
• Thinking of tag-sets as containers will help you remember that tags should always be balanced.
HTML: Behind the Scenes

• Not every tag in HTML is paired. Some tags, such as the line-break tag </BR>, stand on their own.

• It is often the case that the arrangement of text within a container is irrelevant.

• “whitespace doesn't matter. (Whitespace is all of the blank areas in a text file--empty lines, extra spaces, and so on.)
HTML: Document Tags

- HTML
- HEAD
- TITLE
- BODY

Figure 1
HTML: Basic text structure

• Headings
  <H1>Heading 1</H1>
  <H2>Heading 2</H2>
  <H3>Heading 3</H3>
  <H4>Heading 4</H4>
  <H5>Heading 5</H5>
  <H6>Heading 6</H6>
HTML: Basic text structure
The following HTML text...

Paragraphs can often be quite long, and you may at times want to draw attention to something in the middle of them, or emphasize a certain point. However, `<BR>` this is not the way to do it. `<BG>`

What you need are some of the text-effect tags which are discussed in Chapter 5, and will be a breeze once you’ve mastered things like lists. Be patient.

...will be displayed as follows.

Paragraphs can often be quite long, and you may at times want to draw attention to something in the middle of them, or emphasize a certain point. However, this is not the way to do it.

What you need are some of the text-effect tags which are discussed in Chapter 5, and will be a breeze once you’ve mastered things like lists. Be patient.
HTML: Basic text structure

- Paragraph: `<P>`, and the end by `</P>`.

The actual HTML:

```
<
Paragraphs can often be quite long, wandering on with no clear sense of
direction or purpose, boring the reader and obscuring the point of the
message (if any) contained within the text. This sort of paragraph usually
deserves to be broken up into several paragraphs, since its sheer bulk
dissuades the reader from attempting to plumb its depths.
</P>
<P>
On the other hand, they can be pretty short.
</P>
<P>
Really short.
</P>
<P>
Cool?
</P>

...would appear as follows.
```

Paragraphs can often be quite long, wandering on with no clear sense of direction or purpose, boring the reader and obscuring the point of the message (if any) contained within the text. This sort of paragraph usually deserves to be broken up into several paragraphs, since its sheer bulk dissuades the reader from attempting to plumb its depths.

On the other hand, they can be pretty short.

Really short.

Cool?
HTML: Basic text structure

• Line Break <BR>

This HTML:

<p>
If you want to force a line break for some reason, but it doesn’t make sense to start a new paragraph, then you can use the line break tag just as I will do at the end of this sentence. <br>See the forced display on the next line? <br>How about there? <br>Or there?
</p>

...is displayed as the following.

If you want to force a line break for some reason, but it doesn’t make sense to start a new paragraph, then you can use the line break tag just as I will do at the end of this sentence. See the forced display on the next line? How about there? Or there?
HTML: Unordered Lists

- Monday
- Tuesday
- Wednesday
  - 6am - 9am
  - 9am - 12pm
  - 12pm - 3pm
  - 3pm - 6pm
- Thursday
- Friday
HTML: Ordered Lists

```
<OL>
  <LI>Monday
  <LI>Tuesday
  <LI>Wednesday
  <LI>Thursday
  <LI>Friday
</OL>
```

1. Monday
2. Tuesday
3. Wednesday
4. Thursday
5. Friday
HTML: Special effects tags

• **Boldface**
  – Everything between `<B>` and `</B>` is **boldfaced**.

• **Italics**
  – Everything between `<I>` and `</I>` is *italicized*.

• **Underline**
  – Everything between `<U>` and `</U>` should be **underlined**; however, see the note below.
HTML: Anchors

• The simplest possible anchor starts with <A> and ends with </A>.
• Need to enhance the <A> tag with attributes like HREF: stands for "Hypertext REFerence," which is another way of saying, "The location of the file I want to load." Most anchors are in the form <A HREF="URL">.

• Check out the <A HREF="http://www.cwru.edu/">CWRU Web server</A>-- it's pretty cool!

Check out the CWRU Web server-- it's pretty cool!
The Extensible Markup Language (XML) is a way of describing data. XML provides a text-based means to describe and apply a tree-based structure to information. Its primary purpose is to facilitate the sharing of data across different systems, particularly systems connected via the Internet.
Semantic Web

• The Semantic Web is a project aimed to make web pages understandable by computers, so that they can search websites and perform actions in a standardized way.

• The potential benefits are that computers can harness the enormous network of information and services on the Web. A computer could, for example, automatically book an appointment that fits a person's schedule.
Ajax

- **Ajax**, shorthand for *Asynchronous JavaScript and XML*, is a web development technique for creating interactive web applications.
- The intent is to make web pages feel more responsive by exchanging small amounts of data with the server behind the scene so that the entire web page does not have to be reloaded each time the user makes a change. This is meant to increase the web page's interactivity, speed, and usability.
• Next Lecture: Modern Web Page Design

• DECO2102 students go to 261 for lab
• DECO1200 students go to 313 for lab