HTML, XML & HTML

A web format

COMP6218

HTML is about making documents

This is an example HTML document

1. First paragraph
2. Second paragraph

This is the first paragraph. *It doesn't really say very much.* Now is the time for all good men to come to the aid of their country.

This is the second paragraph. The quick brown fox jumped over the lazy dog.

Simple Code for Simple Layout

```
<html>
<head> <title>My Document</title></head>
<body>
  <h1>This is an example HTML document</h1>
  <ol>
    <li>First paragraph</li>
    <li>Second paragraph</li>
  </ol>
  <p>This is the first paragraph. It doesn't really say very
  much. Now is the time for all good men to come to the aid
  of their country.</p><p>This is the second paragraph. The
  quick brown fox jumped over the lazy dog.</p>
</body>
</html>
```

HTML: The Beginning

- HTML began as a language for encoding simple document semantics

```
<H1>Things to See in London</H1>
<p>The <I>Millenium Dome</I> is great.
```

HTML: Effects through Bloating

- It became more bloated to allow more design precision and visual effects.

```
<H1>Things to See in London</H1>
<p><center>The <I>Millenium Dome</I>
is <blink>great </blink></center>.
```

HTML: Simplicity through Style

■ Until precise style specifications were added.

Large font, bold, red

```
<H1>Things to See in London</H1>
<p>The <I>Millenium Dome</I> is great.
```

Any font denoting emphasis

Normal font, indent 1cm, centered

HTML: Style is All

- But then the names of the tags became irrelevant.

Large font, bold, red

```
<  >Things to See in London</  >
< >The < >Millenium Dome</ > is great.
```

Normal font, indent 1cm, centered

Any font denoting emphasis

HTML: Or are Data and Style Equal?

■ So invent your own for your own applications.

> **Large font, bold, red**

```
<day>Things to See in London</day>
<see>The <place>Millenium Dome</place>
is great.</see>
```

> **Any font denoting emphasis**

> **Normal font, indent 1cm, centered**

HTML: Or is Data All?

- And forget the display semantics.
- Information is for using, not looking at!

```
<day>Things to See in London</day>
<see>The <place>Millenium Dome</place>
is great.</see>
```

- The result of this transformation is XML

XML Example (Viewed in Notepad)

```
<?xml version="1.0"?>
<!DOCTYPE BOOKLIST SYSTEM "books.dtd">
<BOOKLIST>
<BOOKS>
  <ITEM CAT="S">
    <TITLE>The Language of Science</TITLE>
    <AUTHOR>Danzig</AUTHOR>
    <PRICE>5.95</PRICE>
    <QUANTITY>3</QUANTITY>
  </ITEM>
</BOOKS>
</BOOKLIST>
```

XML Example (Viewed in Browser)

```xml
<?xml version="1.0" ?>
<!DOCTYPE BOOKLIST (View Source for full doctype...)>
- <BOOKLIST>
  - <BOOKS>
    - <ITEM CAT="S">
        <TITLE>The Language of Science</TITLE>
        <AUTHOR>Danzig</AUTHOR>
        <PRICE>5.95</PRICE>
        <QUANTITY>3</QUANTITY>
      </ITEM>
    </BOOKS>
  </BOOKLIST>
```
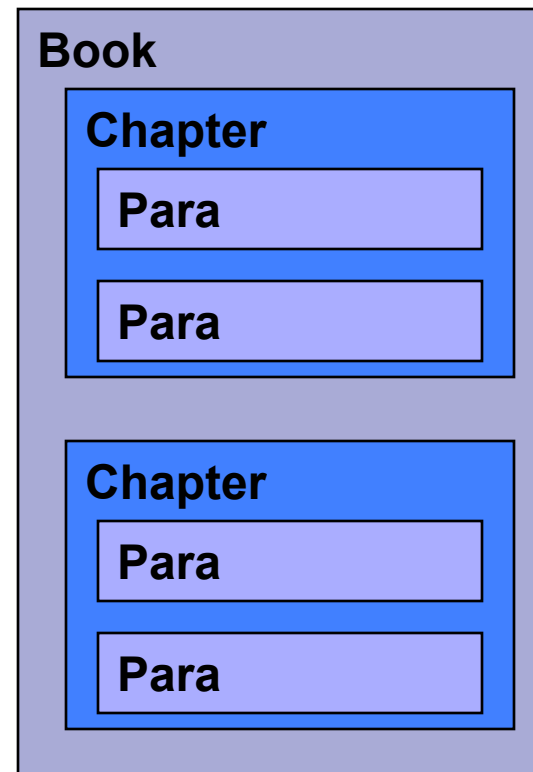
XML Components

- Elements
- Attributes
- Entities
- Declarations
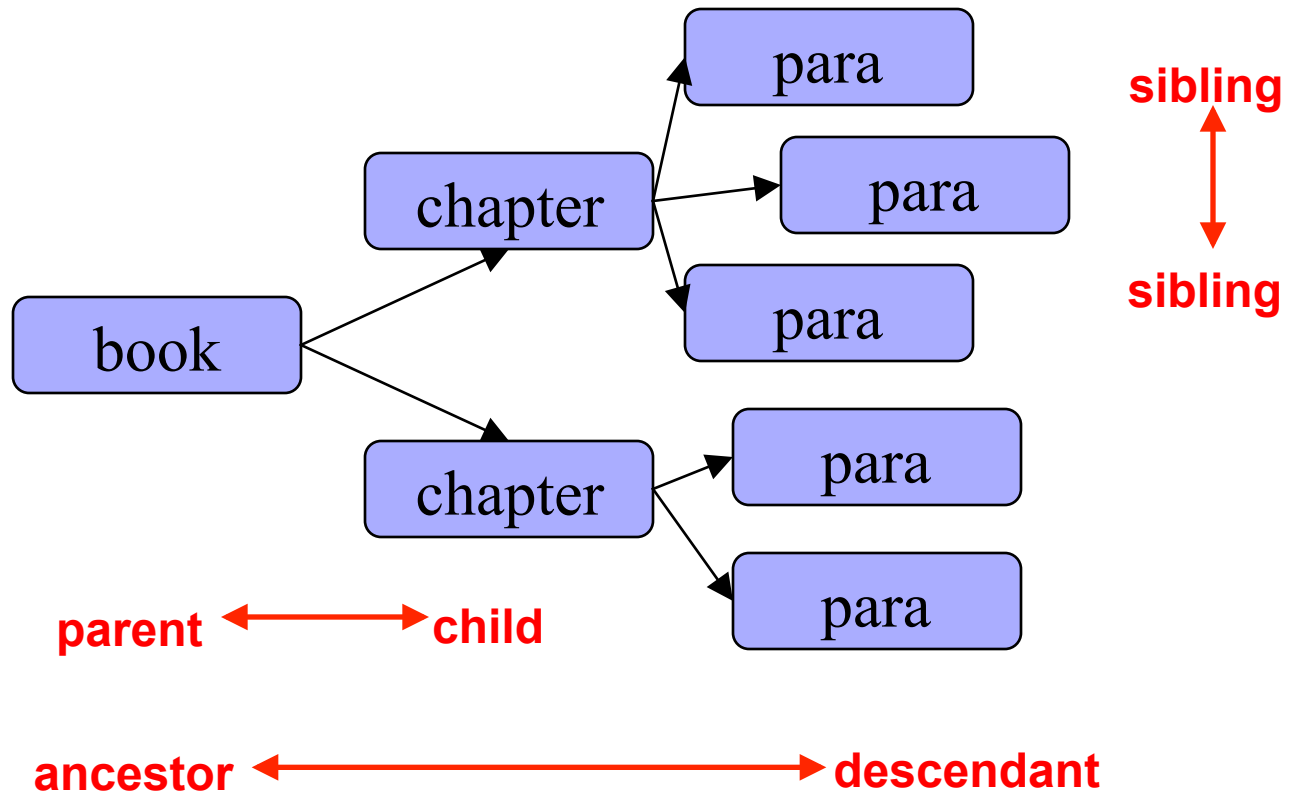- Processing Instructions
- Document Type Definitions (DTDs)

XML Components - Elements
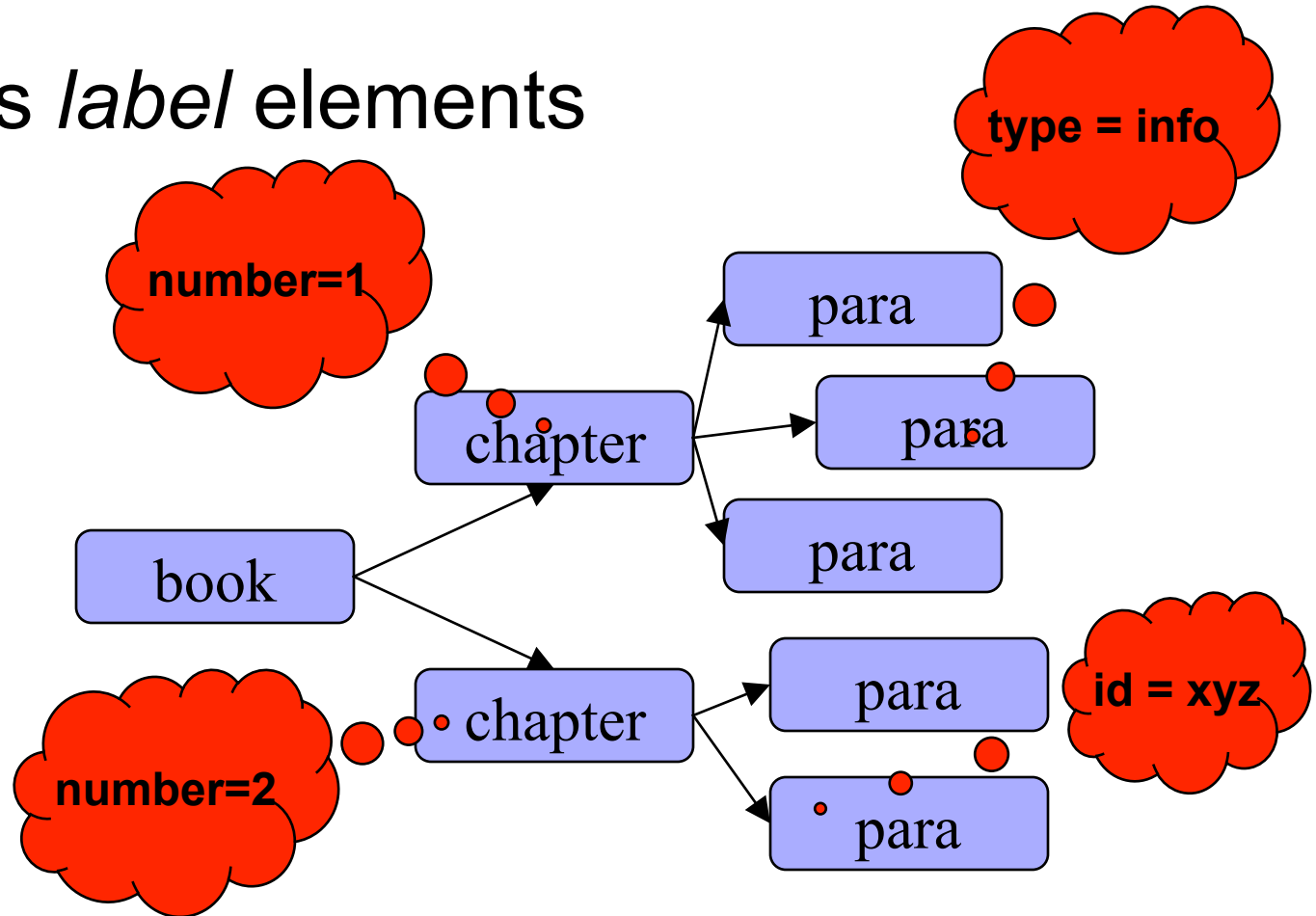
- Elements form a hierarchical decomposition

Book

Chapter

Para

Para

Chapter

Para
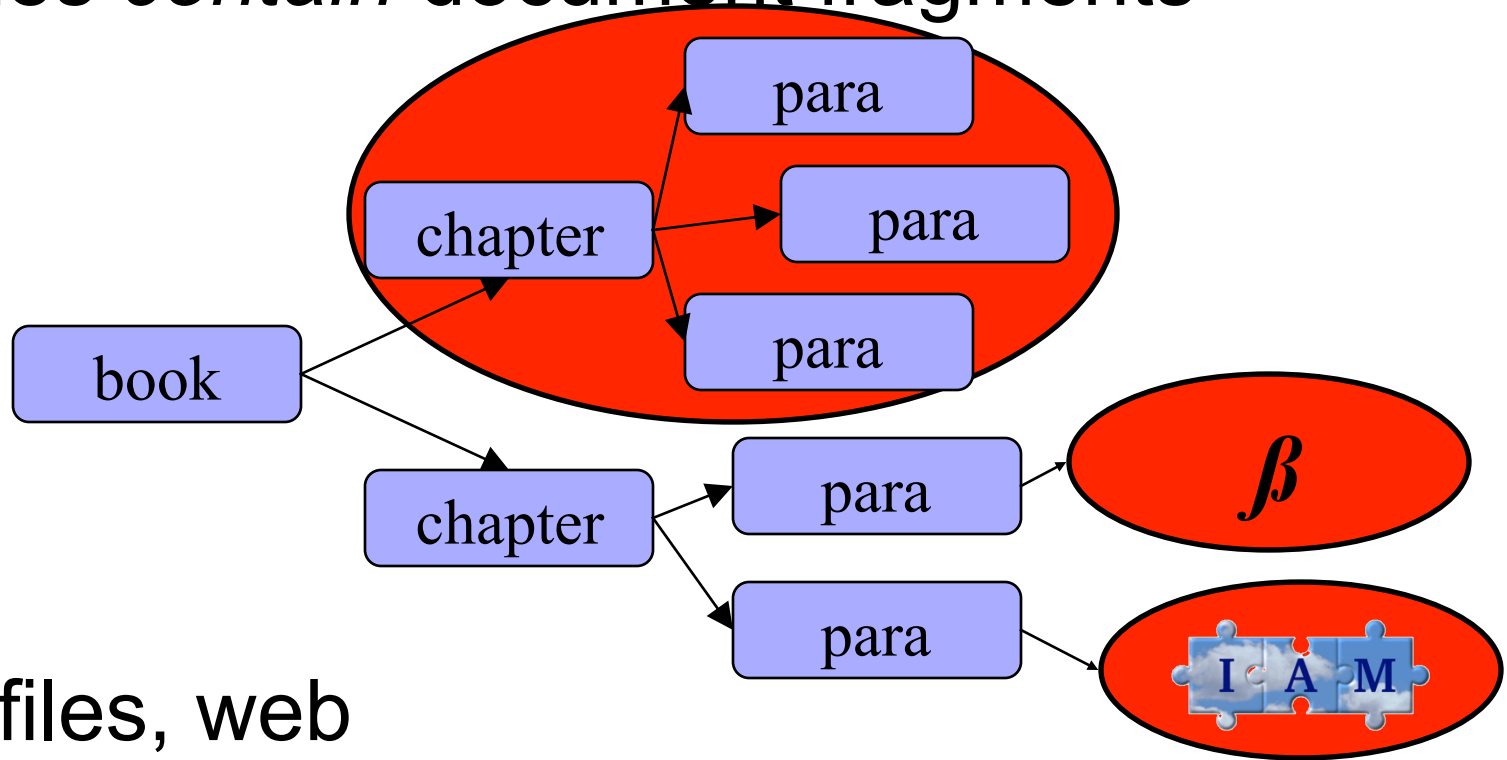
Para

XML Components – Element Structure

■ Elements form a tree

XML Components - Attributes

- Attributes *label* elements

XML Components - Entities

- Entities *contain* document fragments

para

chapter

para

para

book

chapter

para

$\boldsymbol{\beta}$

para

I A M

- Like files, web pages or 'containers'

XML Components - misc

- Markup declaration *defines* entities, elements, attributes, DTDs, comments, marked sections.

- Processing instructions *interpret* elements and content

- DTDs *constrain* elements, attributes and content.

- DTDs *provide* a simple grammar.

XML Syntax (1)

- Elements
  - surrounded by start and end tags

    `<foo>contents of foo element</foo>`

  - empty element tag for no data

    `<foo/>` **equivalent to** `<foo></foo>`

  - names are case sensitive

XML Syntax (2)

- Element content may be
  - ☐ data   `<foo>Just text data</foo>`
  - ☐ mixed   `<foo>Text and <bar>element</bar> data</foo>`

  - ☐ element only

    ```
    <foo>
            <image/>
            <link>Click here</link>
    </foo>
    ```

XML Syntax (3)

- **Element recognition**
  - □ naked angle brackets cannot appear

    `<prog>if(3 < 4)die("Math bug");</prog>`

    must appear as

    `<prog>if(3 < 4)die("Math bug");`…

  - □ applies to
    - **left angle bracket**   **<**   **<**
    - **right angle bracket**   **>**   **>**
    - **ampersand**   **&**   **&**

XML Syntax (4)

- **Attributes**
  - provide extra information

  ```
  <para security="restricted">
    Merge with Microsoft tomorrow!
  </para>
  ```

  - both names and values are case sensitive
  - values MUST be quoted with " or '

XML Syntax (7)

- **Standard entities**
  - ☐ predefined
    - **left angle bracket**   `<`   `<`
    - **right angle bracket**   `>`   `>`
    - **ampersand**   `&`   `&`
    - **apostrophe**   `'`   `'`
    - **double quote**   `"`   `"`
  - ☐ character entities
    - `decimal ASCII`   `A`   `A`
    - `hexadecimal UNICODE` `é`   `é`

HTML5 – adding semantics to webpages

- HTML / XHTML were simple page-oriented structures
  - Headings, paragraphs, lists, tables, images etc
- Gradually generic structures take over as use of the Web explodes
  - <div> & <span>

- HTML5 recognises major new structures that are useful for
  - search engines; Google or Yahoo! can weight content in footer elements lower, or extra weight to content in the header.
  - usability for people with disabilities, screen readers etc

Overall Structure of Web page

Structure of HTML5 Web page

```
<!DOCTYPE html>
<html>
<head>
<title>Title of the document</title>
</head>

<body>
      This is a really simple HTML5 page
</body>

</html>
```

Navigation support

<nav>: Represents a major navigation block. It groups links to other pages or to parts of the current page whose role is simply navigation

```
<nav>
<ul>
<li><a href="/">Home</a></li>
<li><a href="/events">Current Events</a></li>
<li><a href="/contact">Contact us</a></li>
</ul>
</nav>
```

Headers & Footers

<Header>: tag specifies a header for a document or section, the title and datestamp of a blog entry or news article

```
<body>
<header>
<h1>Little Green Guys With Guns</h1>
<nav>
<ul>
<li><a href="/games">Games</a>
<li><a href="/forum">Forum</a>
<li><a href="/download">Download</a>
</ul>
</nav>
</header>
<p>You have three active games:</p>
<!-- this is still part of the subsection entitled "Games" -->
...
```

```
<article>
<header>
<h1>Military Offers Assurances to Egypt and Neighbors</h1>
Published : <time datetime="2011-02-13" pubdate>February 13 2011</time>)
</header>
<p>
CAIRO - As a new era dawned in Egypt on Saturday,
the army leadership sought to reassure Egyptians and
the world that it would shepherd a transition to civilian
rule and honor international commitments like the peace treaty with Israel.
</p>
</article>
```

Header & Footer

<footer>: Material that comes at the base of the page or article, e.g. copyright and contact information.

```
<ARTICLE>
<H1>My Favorite Trains</H1>
<P>I love my trains. My favorite train of all time is a Kof.</P>
<P>It is fun to see them pull some coal cars because they look so
dwarfed in comparison.</P>
<FOOTER>  <!-- footer for article -->
<P>Published <TIME PUBDATE DATETIME="2009-09-15T14:54-07:00"></TIME></P>
</FOOTER>
</ARTICLE>


<FOOTER>  <!-- site wide footer -->
<NAV>
<P><A HREF="/credits.html">Credits</A>-
<A HREF="/tos.html">Terms of Service</A> -
<A HREF="/index.html">Blog Index</A></P>
</NAV>
<P>Copyright © 2009 Gordon Freeman</P>
</FOOTER>


</BODY>
```

Articles

<article>: Articles and blog entries are common, an alternative to <div class="article"> used for distributable content in e.g. RSS feeds

An article may contain a header and footer and a title.

```
<body>
<h1>My blog</h1>
<article>

<header>
<h1>The Very First Rule of Life</h1>
<p><time pubdate datetime="2009-10-09T14:28-08:00"></time></p>
</header>

<p>If there's a microphone anywhere near you, assume it's hot and
sending whatever you're saying to the world. Seriously.</p>
<p>...</p>

<footer>
<a href="?comments=1">Show comments...</a>
</footer>

</article>
<body>
```

Asides (not as important)

<aside>: The "aside" element is a section that somehow related to main content, but it can be

separate from that content

```
<article>
<p>
As of writing, the only web browser completely support date time input is Opera.
In HTML5, it is the job of web browser to ensure user can only enter a valid date time
into the input textbox.
</p>

<aside>
Picking a date from Calendar is not the only way to input a date value even though it's
HTML5 specifications does not mention anything about displaying a calendar for date inp
</aside>
</article>
```

Media - audio & video

- Until now, there has not been a standard for playing media files.

- Today, most audio & video is played by a plug-in (e.g. Adobe Flash).

- HTML5 provides new elements for media
  - But only understands a limited set of formats

```
<video src="movie.webm"></video>

<audio src="music.mp3"></audio>
```

Drawing – Canvas

Provides a surface for programs to draw ion using a standard API

Avoids the need to download pre-generated images from the network.

- Lines
- Arcs
- Text
- Gradients
- Patterns.

Also provides image and pixel manipulation.

<canvas id="myCanvas">
</canvas>

```
<script>
 var myCanvas = document.getElementById("myCanvas");
 var drawingContext = myCanvas.getContext("2d");

 canvasContext.fillRect(100, 100, 100, 100);
</script>
```