The HyperText Mark-up Language (HTML)
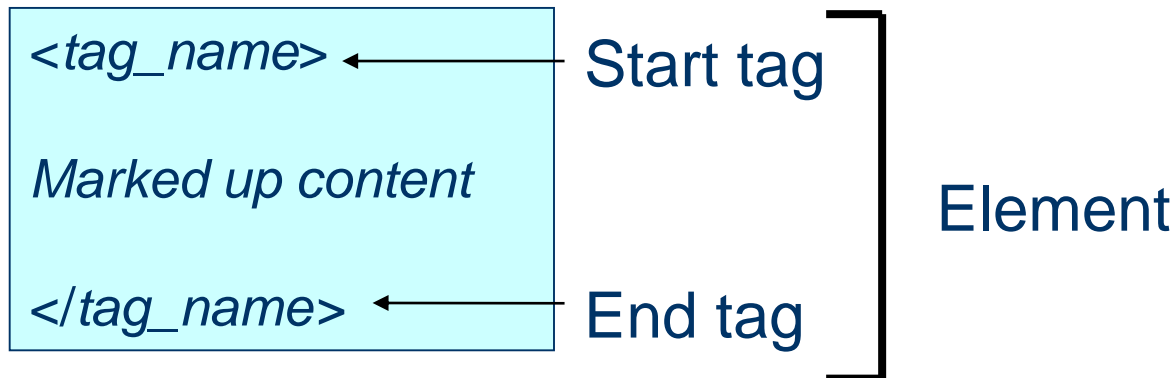
Learning Objectives

- To understand the origins of HTML
- To understand the importance of separating content, structure and presentation in web applications
- To understand some of the most commonly used elements in HTML

Where it all Begins - SGML

- GML – Goldfarb, Mosher and Lorre, 1969
    - Generalized markup that would work across computer systems using *tags*
    - Concept of *validation* – how the tags can be used
- SGML - Standard Generalized Markup Language, 1980
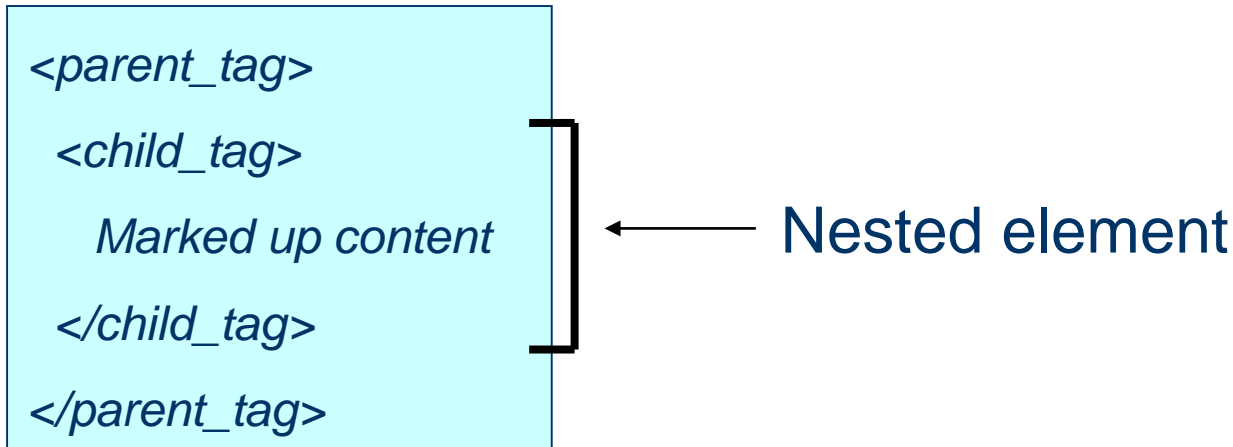   - ISO standard 1986

Markup Concepts

- Tags and elements


```
<tag_name>        ← Start tag

Marked up content                Element

</tag_name>       ← End tag
```

- The characteristics defined by the tag are applied to the content of the element

Nested Elements

- Elements can have other elements nested inside them, to any level of nesting.
    - A 'child' element, begins and ends inside its 'parent'

```
<parent_tag>
  <child_tag>
    Marked up content
  </child_tag>
</parent_tag>
```

Nested element

Attributes

- Attributes configure the tag in some way.
  - Appear inside the opening tag, and consist of one or more name-value pairs

  attribute_name="*attribute_value*"

- e.g.

Empty Elements

- An empty element consists of a single tag, with no closing tag

- There is a forward slash before the closing bracket

```
<tag_name>
```

- Empty elements often include attributes

```
<br>
<hr>
```

Well-Formedness

- Balanced tags

```
<tag>...</tag>
```

- Correct nesting

```
<parent_tag>
 <child_tag>

  …
 </child_tag>
</parent_tag>
```

- Document has a root element

- Quoted attributes

```
<tag name="value">
Or
<tag name='value'>
```

HTML – A Language For Web Pages

- Developed by Tim Berners Lee at CERN in 1991
  - Based on hypertext linking of pages

HTML Documents

- Text files with tags that mark up the content
  - They become web pages when they are made available on a web server and are rendered on the client machine using a web browser
- The browser uses the tags to manage the content
  - e.g. the paragraph tag

```
<p>some text in paragraph one …</p>
<p>some text in paragraph two …</p>
```

Minimal HTML Document

- An HTML document with a minimal set of elements
  - Case sensitive
  - Lower case required for element names
  - Well formed – not necessarily *valid*

```
<html>
 <head>
  <title>Untitled Page</title>
 </head>
 <body>
 </body>
</html>
```

Valid HTML Documents

- All HTML documents must start with a document type declaration.

`<!DOCTYPE html>`

Content Types

- The body of an HTML document is frequently organized into blocks of text

- There may also be other content types
  - images, sound clips, movies etc.

- A content type is some kind of MIME type that defines a particular type of Internet file

- The HTML document type is 'text/html'
  - May contain references to other types of content