The Hypertext Markup Language (HTML)

Pat Morin

COMP 2405

Carleton
UNIVERSITY
Canada's Capital University

Outline

- History of HTML

- Structure of an HTML Document
    - DOCTYPE
    - HEAD
    - BODY

- HTML Tags
    - Paragraphs and Headings
    - Lists and tables
    - Hyperlinks

- Conclusions and Guidelines

Carleton
UNIVERSITY
Canada's Capital University

History of HTML

- Hypertext systems were envisioned as early as 1940 by Vannevar Bush and have a rich history

- Tim Berners-Lee and Robert Caillau at CERN, in 1989-1990 developed HTML as a simplification of SGML

- CERN launched the web in 1991 (HTML+HTTP)

- HTML is now at version 4

Carleton
UNIVERSITY

Canada's Capital University

Versions of HTML

- There are several different version of HTML
    - HTML 1.0, 2.0, 3.2, 4.0, 4.01
    - XHTML 1.0 and 1.1

- HTML 4.01 and XHTML 1.0 come in different flavours:
    - Strict: strictly compliant
    - Transitional: allows some deprecated elements
    - Frameset: a variant of transitional for documents that use frames

- For these reasons, every HTML file must begin with a DOCTYPE definition

Carleton
UNIVERSITY
Canada's Capital University

Structure of an HTML Document

- Every document starts with a DOCTYPE, followed by an HTML section that contains a head and body:

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN"
  "http://www.w3.org/TR/html4/strict.dtd">
<html>
  <head>
    <title>The Hello World Page</title>
  </head>
  <body>
    <p>Hello World!</p>
  </body>
</html>
```

Carleton
UNIVERSITY
Canada's Capital University

5

DOCTYPEs

- A list of DOCTYPEs is available here from the World-Wide-Web Consortium (W3C):
  - http://www.w3.org/QA/2002/04/valid-dtd-list.html

- Help on choosing a DOCTYPE is also available:
  - http://htmlhelp.com/tools/validator/doctype.html

- Pros and Cons of different DOCTYPEs
  - Older DOCTYPEs are supported on older browsers
  - Newer DOCTYPEs are generally better structured, more standards compliant, and more modular

- In this course we will be using *HTML 4.01 Strict*

Carleton
UNIVERSITY
Canada's Capital University

6

The Document HEAD

- The HEAD section of the document contains information about the document as opposed to the contents of the document

- Minimally, it should include a TITLE
  - `<title>My Page Title</title>`

- The title typically appears in a Browser's title bar

- The HEAD may also include metadata such as character encoding and information about how to format the document

Carleton
UNIVERSITY
Canada's Capital University

The Document BODY

- The BODY contains the document itself, i.e., what the user sees

- Any text (also called character data or `cdata`) within the body must be contained within some other tag

Carleton
UNIVERSITY
Canada's Capital University

Paragraphs

- The P tag is used to enclose a paragraph of text
- Paragraphs appear formatted as you would expected

```
<p>In the case of a private wedding
   announcement cards should be mailed the
   following day to all relatives and
   acquaintances of both the contracting
   parties.
</p>

<p>Evening weddings are no longer the custom,
   and the fashionable hour is now high noon,
   ...
</p>
```

Headings

- Headings are marked with H1, H2, ..., H5 tags
- H*n* denotes an *n*th level heading

```
<h1>Etiquette and Courtship</h1>
<p>It is a growing custom ... </p>

<h2>Gifts and Attention</h2>
<p>Just what attention a person is ... </p>

<h2>The Claims of Companionship</h2>
<p>At this period it is a wise person ... </p>

<h1>Morning Customs</h1>
<p>The morning of the wedding, the ... </p>
```

Lists

- HTML has three kinds of lists:

- Unordered information (bulleted lists)

- Ordered information (numbered lists)

- Definitions (like in a dictionary)

Carleton
UNIVERSITY

Canada's Capital University

Unordered Lists

- The UL tag encloses an *unordered list*

- The individual list items are enclosed in LI tags

```
<ul>
  <li>Client/server protocols</li>
  <li>Web site design</li>
  <li>Server-side scripting</li>
  <li>Client-side scripting</li>
  <li>Mixed-mode scripting</li>
</ul>
```

Carleton
UNIVERSITY
Canada's Capital University

Ordered Lists

- The OL tag denotes ordered (numbered) lists

- Again, list items are enclosed in LI tags

```
<ol>
  <li>Mix dry ingredients thoroughly.</li>
  <li>Pour in wet ingredients.</li>
  <li>Mix for 10 minutes.</li>
  <li>Bake for one hour at 300 degrees.</li>
</ol>
```

Carleton
UNIVERSITY
Canada's Capital University

Definition Lists

- Definition lists use the DL tag

- Each item has two parts, the *term* DT and the *definition* DD

```
<dl>
  <dt>Lower cost</dt>
  <dd>The new version of this product costs
      significantly less than the previous one!</dd>

  <dt>Easier to use</dt>
  <dd>We've changed the product so that it's much
      easier to use!</dd>

  <dt>Safe for kids</dt>
  <dd>You can leave your kids alone in a room with
      this product and they won't get hurt (much).</dd>
</dl>
```

Some Other Tags

- To emphasize some text, use EM
  - `The <em>last</em> thing you should do ...`

- To strongly emphasize, use STRONG
  - `The <strong>absolute last</strong> thing ...`

- Use CODE to denote code snippets
  - `Use <code>fseek(fp, 0L, SEEK_SET)</code> to rewind the file pointer`

- There are also short quotes Q, block quotes (BLOCKQUOTE), citations (CITE), addresses (ADDRESS), insertions (INS), deletions (DEL), typed-text (KBD), variable names (VAR), sample output (SAMP), and preformatted text (PRE)

Simple HTML Tables

- HTML tables are created TABLE tag

- Each table contains one or more rows (TR) containing table data (TD) or table headers (TH)

```
<table>
  <tr>
    <th>First Name</th>
    <th>Last Name</th>
    <th>Research Areas</th>
  </tr>
  <tr>
    <td>Pat</td>
    <td>Morin</td>
    <td>Algorithms and data structures</td>
  </tr>
  ...
</table>
```

HTML Tables Cont'd

- HTML4 Tables are actually much more sophisticated than this basic example

- All the gory details can be found at the W3C
  - http://www.w3.org/TR/html4/struct/tables.html

The A Tag

- So far we know enough to create a simple text documents

- What about the *hyper* in hypertext?

- For this we use the A tag

```
<p>Some of the better search sites on the
   Internet include
   <a href="h ttp://www.google.com/"> Google</a>,
   <a href="h ttp://www.a9.com/">A9</a>, and
   <a href="h ttp://www.altavista.com/"> Alta Vista</a>.
   For more information on internet searching, ... </p>
```

Carleton
UNIVERSITY

Canada's Capital University

The A Tag (Continued)

- The attribute `HREF` of the A tag specifies the destination the link will take us to

- This destination can be absolute:
  - `href="http://www.google.com/"`
  - `href="http://www.newgrounds.com/"`

- Or relative:
  - `href="notes/index.html">`
  - `href="/teaching/2405/index.html">`

- Relative locations specify the location relative to the current document and are extremely useful when building large websites

The A Tag (Cont'd)

- The HREF attribute can even specify a target within a document
  - `<a href="http://abc.com/index.html#tuesday">`
  - `<a href="#sectionA">`

- These link to the sections labels "tuesday" and "sectionA" within their respective documents

- To create these targets we again use the A tag in a different way
  - `<a name="tuesday">Tuesday's Schedule</a>`
  - creates a target named tuesday within the current document

Targets and Tags Example

```
...
<body>
<h1><a name=č hap1‴ Chapter 1</a></h1>

<p>Those already familiar with widgets can
   skip to <a href=# chap2‴ Chapter 2</a>.
   ...

<h1><a name=č hap2‴ Chapter 2</a></h1>

<p>If you are reading this and you still don't
   understand widgets then you should go back
   and read <a href=# chap1‴ Chapter 1</a>.
   ...

</body>
```

21

Images

- Images can be added to the document with the `IMG` tag

- The `SRC` attribute specifies the location of the image data

- The `ALT` attribute specifies some text to display if the image can not be displayed

```
<img src="kafka.jpg" alt="Franz Kafka's Portrait">
```

Carleton
UNIVERSITY
Canada's Capital University

Line Breaks

- For a quick and dirty line break you can use the BR tag

- Normally you should avoid this

- Why are you breaking the line?
  - For a list of items (or menu): use
  - To shorten a line: let the browser wrap it
  - For preformatted text: use the `<pre>` tag

- Do as I say, not as I do
  - Some examples may contain `<br>` tag to make them shorter
  - You should avoid them

Carleton
UNIVERSITY
Canada's Capital University

Summary

- There are many versions of HTML
  - You must specify which version using the DOCTYPE tag

- Every document has a HEAD and a BODY tag
  - HEAD contains information about the document
  - BODY contains the contents of the (displayed) document

- Basic document constructs are available
  - Headings
  - Paragraphs
  - Ordered, unordered and definition lists
  - Tables
  - Hyperlinks and hyperlink targets

Carleton
UNIVERSITY
Canada's Capital University

HTML Formatting

- All of the HTML tags discussed have other attributes, examples:
  - `<td align="right" valign="top" bgcolor="red">`
  - `<body background="image1.gif" link="blue" text="red">`

- There are lots of other HTML tags
  - Some of these, like `FONT`, control formatting

- Tags and attributes that control formatting are usually not a good idea

- We will see a better alternative soon