CITS1231 Web Technologies
Understanding URLs and Site Structure
This Lecture

• Understanding URLs
  – Web pages, FTP servers, newsgroups, and e-mail addresses
  – Absolute and relative URLs

• Web site’s structure
  – Storyboarding (structure diagram)
  – linear, hierarchical, mixed
  – Linking between pages
Purposes of URLs

- Link to another document or resource
- Link to an external style sheet or script
- Include an image, object or applets in a page
- Create an image map
- Submit a form
- Create a frame document
- Cite an external reference
- Refer to metadata conventions describing a document
Understanding URLs

- A **Uniform Resource Locator (URL)** specifies the precise location of a resource on the Internet.
- A **protocol** is a set of rules defining how information is exchanged between two resources.
- A Web browser communicates with Web servers using the **Hypertext Transfer Protocol (HTTP)**.
- The **URLs** for all Web pages must start with the scheme “http.”
- Other Internet resources use different **protocols** and have different scheme names.
# Common Communication Protocols

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Used to</th>
</tr>
</thead>
<tbody>
<tr>
<td>file</td>
<td>access documents stored locally on a user’s computer</td>
</tr>
<tr>
<td>ftp</td>
<td>access documents stored on an FTP server</td>
</tr>
<tr>
<td>gopher</td>
<td>access documents stored on a gopher server</td>
</tr>
<tr>
<td>http</td>
<td>access Web pages stored on the World Wide Web</td>
</tr>
<tr>
<td>mailto</td>
<td>open a user’s e-mail client and address a new message</td>
</tr>
<tr>
<td>news</td>
<td>connect to a Usenet newsgroup</td>
</tr>
<tr>
<td>telnet</td>
<td>open a telnet connection to a specific server</td>
</tr>
<tr>
<td>wais</td>
<td>connect to a Wide Area Information Server database</td>
</tr>
</tbody>
</table>
Linking to a Web Page

A sample URL for a Web page

http://www.mwu.edu/course/info.html #majors

protocol server path filename id or anchor name
Linking to a Web Page

- If a **URL** includes no path, then it indicates the topmost folder in the server’s directory tree.

- If a **URL** does not specify a filename, the server searches for a file named “index.html” or “index.htm” or “default.html”.

- For example,
What’s the URL for review.html or plan.html?

http://www.csse.uwa.edu.au/~smith01/CITS1231/review.html

With a URL, a "file system" may no longer be the local network we usually think of. If we open files within our applications via a URL, that file can be on the local hard drive, a local “mounted” network or somewhere that is geographically remote.
Linking to FTP Servers

- FTP servers are one of the main sources for storing files on the Internet
- FTP servers transfer information using a communications protocol called File Transfer Protocol, or FTP for short
- An FTP server requires each user to enter a password and a username to access its files
Linking to Usenet News

- **Usenet** is a collection of discussion forums called **newsgroups** that let users publicly exchange messages with each other on a wide variety of topics.

- When you click a link to a **newsgroup**, your computer opens a program for reading **newsgroups**, known as newsreader, displaying the latest messages from the **newsgroup**.

- **news:alt.surfing**
Linking to a File

file://host/path

- **host** is fully qualified domain name
- **path** is hierarchical directory path
- To access file on your own computer, omit host:
  
  file:///path

- Example:

  file:///D:/proposal.doc
Linking to E-mail

mailto:address

• Where address is the destination email address

• Clicking an email link starts an email program. A new email message is started with “To” field filled in.

• Example:

mailto:admin1231@csse.uwa.edu.au
Spam

- e-mail links can be used for Spam
- Spam is unsolicited junk e-mail
- Spammers use programs called e-mail harvesters that scan HTML code on the Web looking for the e-mail addresses contained in mailto URLs
Protect your Email address from Spams

• Use images of email addresses

• Write a program in a language like JavaScript to scramble e-mail address in the HTML code

• Replace the characters of the e-mail address with character codes

   admin1231&amp;#64;csse.uwa.edu.au

• Replace characters with words in your Web page’s text

   admin1231 at csse dot uwa dot edu dot au
Absolute URL

• In an absolute URL the entire path to the file is specified

• This is analogous to specifying the complete street address, so a letter is unambiguously delivered to the recipient.

• From within your own web page when you include a URL of someone else's server, you must use the absolute form.

• If your friend wants reference your review_csse.html page, the URL would be

  http://student.csse.uwa.edu.au/~your_id/CITS1231/review_csse.html
Relative URL

• To include a link to a page stored in the same folder as the current page, does not require the absolute URL.

• If in the previous example the files `review_csse.html` and `plan.html` are in the same folder, the URL for `review_csse.html` in the page `plan.html` need only contain the relative reference like this:

  `<a href="review_csse.html">Review Document</a>`

• A reference to the `index.html`, one level up (same level as the `CITS1231` folder)

  `<a href="../index.html">Unit Home Page</a>`
Relative URL

- By default, the current page is assumed to be the base URL.
- Relative URLs are resolved to full URLs using the base URL.
- You can change the base URL using the base tag.
- For example in your review_csse.html you could have:

  `<base href=http://student.csse.uwa.edu.au/~your_id/>`
  `<a href="index.html">Unit Home Page</a>`
URLs: Fragment Identifiers

• Some URLs refer to a location within a resource

• This kind of URL ends with a # followed by an anchor or fragment identifier. For example, here is a URL pointing to an anchor named “top”, in the HTML document “lab01.html”:
  – http://web.csse.uwa.edu.au/units/cits1231/labs/lab01.html#top

• This is typically used for long pages that need a table of content at the top and need links within the page.
Storyboard

• A **storyboard** is a diagram of a Web site’s structure, showing all the pages in the site and indicating how they are linked together.

• It is important to **storyboard** your Web site before you start creating your pages in order to determine which structure works best for the type of information the site contains.

• A well-designed structure can ensure that users will be able to navigate the site without getting lost or missing important information.
Linear Structures

• In a **linear structure**, each page is linked with the pages that follow and precede it in an ordered chain

• **Linear structure** works best for Web pages with a clearly defined order

• In an **augmented linear structure**, each page contains an additional link back to an opening page

• For example

  Prev       TOC       Next
Linear Structures

A linear structure

An augmented linear structure
Hierarchical Structures

- In the **hierarchical structure**, the pages are linked going from the most general page down to more specific pages.
- Users easily move from general to specific & back again.
- Implemented using navigations bars - a list of hyperlinks grouped in the same section of the page. Typically, left pane, top (below the title and logo), or on the right pane.
Hierarchical Structures

Pages are arranged in a hierarchy from the general down to the specific. Users can move up and down the tree.
Mixed Structures

• As Web sites become larger and more complex, you often need to use a combination of several different structures

• The overall form can be **hierarchical**, allowing the user to move from general to specific; however, the links also allow users to move through the site in a **linear** fashion
Mixed Structures
Working with Web Site Structures

- A little foresight can go a long way toward making your Web site easier to use
- Each page should contain, at minimum, a link to the site’s home page, or to the relevant main topic page, if applicable
- You may want to supply your users with a site index (site map) which is a page containing an outline of the entire site and its contents
Creating Links Between Documents

• To link to a page, you specify the name of the file using the href attribute of the <a> tag

  <a href="review_csse.html">

• Filenames are case sensitive on UNIX and Macintosh, but not on Windows.

• Recommended practice:
  – use lowercase filenames for all files
  – avoid special characters such as blanks and slashes
  – keep filenames short to avoid typing errors
Link to Location Within Another Document

- When linking to a location within another document, you must use the **anchor** name of the location within the document and the filename

  `<a href="file#id">content</a>`
Linking to Documents in Other Folders

- To create a link to a file you must specify the file’s location, or path.
- Use relative path to ensure the portability of your site.
Open Link in New Window

• The target attribute specifies where to open the linked document:

  <a href="http://www.uwa.edu.au/" target="_blank">Visit U.W.A.</a>

• Clicking the above link will open the main u.w.a. web page in a new browser window.
Popup Title

• If you want to provide additional information to your users, you can provide a **popup title** to your links

• A **popup title** is a descriptive text that appears whenever a user positions the mouse pointer over a link

• Since only some browsers support **popup titles**, you should not place crucial information in them

• Use the title attribute to specify the popup text.

  `<a href="../index.html" title="Return to the Chemistry Home Page">Home Page</a>`
Popup Titles

<a href="../index.html" title="Return to the Chemistry Home Page">Home Page</a>
Summary

• A URL specifies the precise location of an internet resource.
• Use URLs to create links to Web pages, FTP servers, newsgroups, and e-mail addresses.
• You can create links between documents within a Web site.
• You can reference files in different folders using relative and absolute paths.
• Web sites can have linear, hierarchical or mixed structures.
• Storyboarding is an important part of Web page development.
• You can enhance user experience by providing popup titles or force links to open in new window.