

Web Page Design
Mr. Tomlinson

Anatomy of a URL

The Internet is a network of computers and cables that transmit data electronically.

The World Wide Web provides a Graphical User Interface, or GUI, for some of the data being transmitted on the Internet. Browsers are the applications that know how to display the data being transmitted via the World Wide Web. Two examples of browsers are Netscape and Internet Explorer. Programmers tell browsers what to display using a language called HTML. HTML stands for Hypertext Markup Language.

When you want to create a Web page, you first create it on your local computer. Your computer is sometimes connected to the Internet. In order to share your Web page with others or "make it live" you must transfer the files to a server. A server is different than your computer because it has a dedicated connection to the Internet, which makes it easy for other computers to find. When you save a Web page on your computer you need to use the file extension .html to let the computer know that it is a Web page.

If you want your page to be the default page that the browser displays you should name it **index.html or .htm** . This is because Web page locations, or **URLs (uniform resource locator)** always end in a folder/directory name or a file name. The browser will automatically generate a list, or index, of the files in a directory. If the user types in a URL ending in the directory name he will see the list that was generated by the browser. In order to prevent the user from seeing this list, you can provide your own version of an index, named index.html.

A Web page is an html file. A Web site consists of multiple Web pages that are somehow related. The pages in a Web site are usually found in the same folder/directory and are connected by links.

Microsoft FrontPage is a WYSIWYG program, which means what you see is what you get. FrontPage writes the HTML for you. Although FrontPage is a WYSIWYG program, you should always test in different versions of different browsers on different platforms

The Anatomy of a URL

URL = Uniform Resource Locator

HTTP = Hypertext Transfer Protocol

The most common device for storing information for transfer and display on the various computers that are connected to the Internet is the hypertext document. Hypertext documents are prepared using a programming language called Hypertext Markup Language (HTML). Because hypertext documents are created using HTML, they are also known as HTML documents. HTML documents

use a standard set of characters, called a character set, which almost all computers recognize. In addition, each HTML document contains tags that a Web browser interprets to display data in the desired format on a client. A Web page is, in fact, an HTML document that is stored on a Web server.

A URL identifies each Web page and contains the address for the Web server or computer that stores the Web site, and an optional pathname to a specific Web page at that site. The most common method of identifying individual servers or computers on a network is to use an Internet Protocol address. An Internet Protocol (IP) address is a unique number that consists of four sets of numbers, separated by periods (such as 141.209.151.119), and that identifies a specific server or computer. IP addresses are hard to remember, so most users rely on domain names to find Web sites. A domain name is an IP address that uses names instead of numbers, such as www.whitehouse.gov, instead of its IP address equivalent. Whether you use an IP address or a domain name, a Web page's URL identifies its location on the Internet so that client computers can find and retrieve it. You can use the following URL as an example:

<http://www.whitehouse.gov/press/index.html>

This URL identifies the exact address of the server or computer on which the page resides and the type of organization that owns and operates it. The Internet protocol “http” is the communications protocol for accessing a page on the Internet, “www” indicates that the server is a Web server, “whitehouse” indicates the name of the organization that owns the server, and “gov” indicates that the server is owned by a government entity. Other common types of servers in the United States are education (.edu), organization (.org), and commercial (.com).

Finally, all files stored on a Web server must have a unique pathname, just like files stored on a disk. The pathname that follows the domain name www.whitehouse.gov specifies the file named `index.html`, which is stored in the `press` folder. The pathname in a URL includes the folder name(s), filename, and filename extension for locating the Web page. The extension for all Web pages is either `.html` or `.htm`, both of which indicate an HTML document. (The `.html` extension is used with computers that run the UNIX operating system, whereas the `.htm` extension is used most often with personal computers running Windows or Macintosh operating systems.) A Web browser processes files with either of these extensions as HTML documents.

To access a particular Web site or to open a location, type its URL in browser's the Address bar, then press the Enter key. The browser connects to the server specified by the URL, sends a request for information, opens the home page or a specific file identified in the URL, and then displays the information.

The Anatomy of a URL

`http://www.loganelm.net/faculty/teacher_pages/tomlinson.htm`

||
||

`protocol://servername/directory/directoy/filename.fileextension`

protocol (or scheme) = http = hypertext transfer protocol

This lets the browser know what type of protocol to use when reading and transferring the files. Other protocols include, HTTPS (for secure web pages), FTP (File Transfer Protocol, for downloading files from the Net), mailto (for sending mail), and file (for accessing files on a local hard disk).

server name = www.loganelm.net

The server name tells where the file is located (i.e., what server it is on). In this case it tells the browser to look on the World Wide Web, and find a domain named "loganelm.net". Domain names can give you an idea of who owns the site and the nature of the site. The domain name is broken into two parts, the name and the top-level extension, separated by a dot. "net" stands for network. Examples of some top level extensions are:

- .gov = government
- .com = commercial
- .org = non-profit organization

path = `faculty/teacher_pages/tomlinson.htm`

The rest of the URL is called a "path", it tells the browser the path to follow to locate the specific file. "faculty" is the name of a directory (or folder), "teacher_pages" is another directory on the server, and "tomlinson" is the name of the HTML file inside of that folder. "htm" (or "html") is the file extension indicating to the browser that the file can be displayed in that format.

Conventions for Naming Files (and folders):

- Use all lower case letters (technically you can use capitals but URLs ARE case sensitive, so it lessens confusion and mistakes to keep everything lower case.)
- Use **ONLY** alphanumeric characters (no punctuation marks, slashes, etc.)
- You **CAN** use the tilde (~), underscore (_) or a hyphen (-)
- **NEVER** use a space in any filename
- **ALWAYS** use an extension (either .htm or .html for text files and .gif or .jpg for graphics)
- Keep the file name as short as you can

Remember: whatever you name your files and directories, they will eventually become part of a URL, so always follow the above conventions.

Default and Index

Sometimes, you may see a URL that looks like this:

`protocol://servername/directory/`

For example: `http://www.loganelm.net/faculty`

or even this:

`protocol://servername/`

For example: `http://www.loganelm.net/`

In these cases, when there is no file named in the URL, the browser will automatically look for a file in the specified directory (first example) or on the server (second example) named **index.htm** or **default.htm**. Index is the most commonly used file name for the first page of a website which is why we named our HTML file that.

File URLs

The URLs we have been talking about so far are all examples of what an URL looks like when the files have been uploaded (published) on a server. When we look at our files on our local hard drive (before they are uploaded), the URL appears in the browser like this:

`file:///H:/mysite/about_me.htm`

Where the protocol is changed to "file", there are three backslashes instead of two, and the server name is replaced by the local hard drive and a "|". The rest of the URL is the same and consists of the path to the file. In this case, the file "about_me.htm" can be found in a directory (folder) called "mysite".