Addressing web pages

The World Wide Web comprises a series of large-capacity computers, known as "web servers", which are connected to the Internet via telephone lines and satellites. The web servers each use the HyperText Transfer Protocol (HTTP) as a common communication standard to allow any computer connected to any web server to access files across the web.

HTML web pages are merely plain text files that have been saved with a ".htm" or ".html" file extension, such as **index.html**.

In order to access a file across the web its web address must be entered into the address field of the web browser. The web address is formally known as its "Uniform Resource Locator" (URL) and typically has three parts:

● **Protocol** – any URL using the HTTP protocol begins by specifying the protocol as **http://**

● **Domain** – the host name of the computer from which the file can be downloaded, For instance **www.example.com**

● **Path** – the virtual path to the file on the named domain, including any parent directory names where applicable. For instance **/htdocs/index.html**

A URL describing the location of a file by protocol, domain, and path is stating its "absolute address". So the absolute address of the file described by the protocol, domain, and path components above is **http://www.example.com/htdocs/index.html**.

Code contained within an HTML web page can reference other HTML files in any domain by their absolute address. HTML files resident within the same domain can also be referenced more simply by their "relative address", which means that files located within the same directory can be referenced just by their file name. For instance a file named "adjacent.html" located in the same directory can be referenced simply as **adjacent.html**.

Additionally, a relative address can reference a file within the parent directory by prefixing its name with "../". For instance a file named "higher.html" in the parent directory can be referenced from the current directory as **../higher.html**.

How do web servers work?

When you enter a URL into the browser address field the browser first examines the protocol. Where the protocol is specified as HTTP, or assumed to be HTTP if unspecified, the browser recognizes that a file is being sought from a web server. It then contacts a Domain Name Server (DNS) to look up the numerical Internet Protocol (IP) address of the specified domain name. Next a connection is established with the web server at that IP address to request the file at the specified path. When the file is successfully located it is copied back to the browser, otherwise the web server sends an error code, such as "404 – Page Not Found".

Web Browser

User Interface

HTTP Request

HTTP Response

Connection to Domain

Web Server
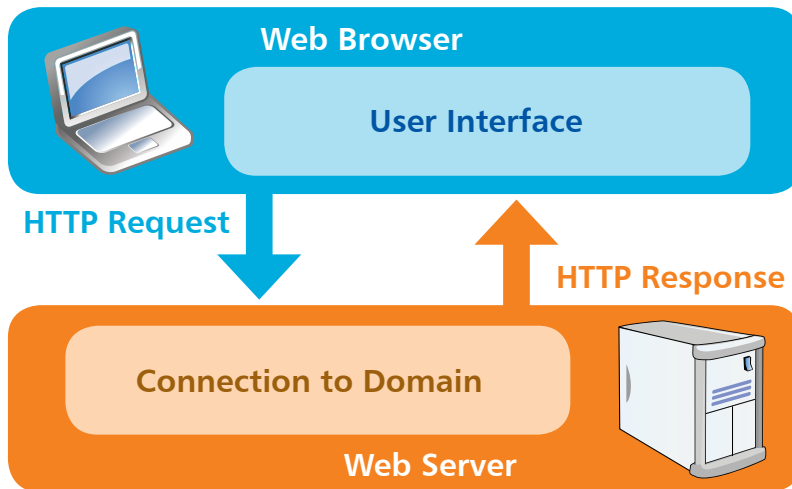
A successful response sends HTTP headers to the web browser, describing the nature of the response, along with a copy of the requested file. The HTTP headers are not normally visible but can be examined using various development tools, such as the Developer Tools feature in the Google Chrome browser:
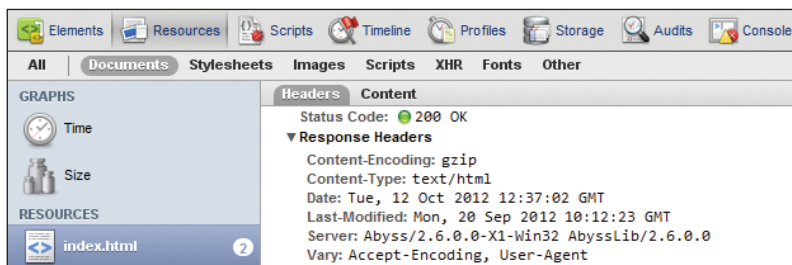
Don't forget

The Domain Name Server is typically run by your Internet Service Provider or by your company.

13

Hot tip

Notice in the headers that the Content-Type is "text/html" – the MIME type used by all web servers to describe plain text HTML files.