

1.3 Web Clients and Web Servers

Web Clients

A web client is software that accesses a web server by sending an HTTP request message and processing the resulting HTTP response.

A web client is an application that communicates with a web server, using Hypertext Transfer Protocol (HTTP). Any web client that is designed to directly support user access to web servers is known as a **user agent**. Browser is the most commonly used web client. The most common interface to the World Wide Web is a browser, such as Mosaic, Netscape Navigator, or Internet Explorer.

Functions of browser:

The primary function of a web browser is to render HTML, the code used to design or mark up webpages. Each time a browser loads a web page, it processes the HTML, which may include text, links, and references to images and other items, such as cascading style sheets and JavaScript functions. The browser processes these items, then renders them in the browser window.

Parts of browser:

Each webpage has an address. This is indicated in the Address Bar. Most webpages will have titles. The actual webpage itself is displayed in the Main Webpage Window. You can reload a page by hitting the "Refresh" button. This will update it to the latest version. If a webpage takes a long time to load you can stop loading it by hitting the "Stop" button.

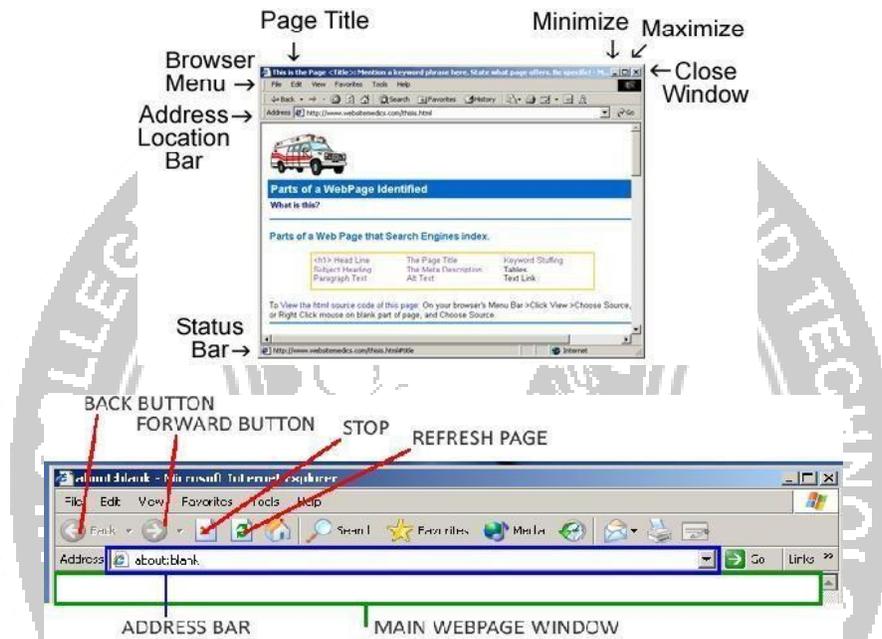


Fig 1.9 Browser Window

Uniform Resource Locators (URL)

The URI (Uniform Resource Identifier) is a string that associates a particular resource on the web. There are two types of URI:

➤ **URN (Uniform Resource Name)**

This identifies the resource using unique names. They do not signify the location of the resource. It consists of three parts: Scheme name, Namespace identifier and Namespace string

| | |
|---|------------------------------------|
| Syntax: Urn: name: resource name | Example: URN: ISBN: 5427877 |
|---|------------------------------------|

➤ **URL (Uniform Resource Locator)**

Uniform Resource Locator (URL) refers to a web address which uniquely identifies a document over the internet.

- This document can be a web page, image, audio, video or anything else present on the web.

URL Types

There are two forms of URL as listed below:

➤ Absolute URL:

Absolute URL is a complete address of a resource on the web. This completed address comprises of protocol used, server name, path name and file name.

Example: `http:// www.abc.com / xyz /index.htm.`

Here `http` is the protocol, `abc.com` is the server name and `index.htm` is the file name.

The protocol part tells the web browser how to handle the file. Other protocols also that can be used to create URL are: FTP, https, Gophe, mailto, news

➤ Relative URL

Relative URL is a partial address of a webpage. Unlike absolute URL, the protocol and server part are omitted from relative URL. Relative URLs are used for internal links i.e. to create links to file that are part of same website as the WebPages on which you are placing the link.

Differences between Absolute and Relative URL

| Absolute URL | Relative URL |
|---|---|
| Used to link web pages on different websites | Used to link web pages within the same website. |
| Difficult to manage. | Easy to Manage. |
| Changes when the server name or directory name changes. | Remains same even if we change the server name or directory name. |
| Take time to access | Comparatively faster to access. |

Web Servers

The computer that supplies files or services to the requesting computer over the internet is called as a web server.

The request to the web pages is sent by the browser to the server. The server will transfer the requested page to the computer over the internet.



Fig 1.10 Web server

Working of web servers

The browser broke the URL into three parts: protocol ("http"), server name ("www.abc.com") and file name ("web-server.htm"). The browser communicated with a name server to translate the server name "www.abc.com" into an IP Address, which it uses to connect to the server machine.

Following the HTTP protocol, the browser sends a GET request to the server, asking for the file <http://www.abc.com/web-server.htm>. The server machine transfers the HTML content to the browser. The browser reads the HTML tags and formats the page onto the screen. In general, all of the machines on the Internet can be categorized as two types: **servers and clients**. The machines that provide services (like Web servers or FTP servers) to other machines are **servers**. The machines that are used to connect to those services are **clients**. It is possible and common for a machine to be both a server and a client.

A server machine may provide one or more services on the internet. For example, a server machine might have software running on it that allows it to act as a Web server, an e-mail server and an FTP server. Clients that connect to a server machine do so with a specific intent, so clients direct their requests to specific software running on the overall server machine. The web server contains **log** records that store information about server activity.

Access log file contains information about every HTTP requests processes by the server. **Message log file** contains a variety of debugging and other information generated by the web applications and the web server. The standard input, output and error streams are also logged.

Differences between web sites and web servers

| Web Sites | Web Servers |
|--|--|
| A website is a set of linked documents associated with a particular person, organization or topic that is held on a computer system and can be accessed as part of the world wide web. | The web server is a computer program, which delivers content, such as websites or web pages. It responds to the request for web pages. |

| | |
|--|---|
| All the web sites reside on the web server. | The web server is a computer with high configuration. |
| Web sites can contain text files, images, videos and audios. | Web servers are hardware or software unit. |

Features of web servers:

1. The server calls on TCP software and waits for connection requests to one or more ports.
2. When a connection request is received, the server dedicates a subtask to handling this connection.
3. The subtask establishes the TCP connection and receives an HTTP request.
4. The subtask examines the Host header field of the request to determine which virtual host should receive this request and invokes software for this host.
5. The virtual host software maps the Request-URI field of the HTTP request start line to a resource on the server.
6. If the resource is a file, the host software determines the MIME type of the file and creates an HTTP response that contains the file in the body of the response message.
7. If the resource is a program, the host software runs the program, providing it with information from the request and returning the output from the program as the body of an HTTP response message.
8. The server normally logs information about the request and response—such as the IP address of the requester and the status code of the response—in a plain-text file.
9. If the TCP connection is kept alive, the server subtask continues to monitor the connection until a certain length of time has elapsed, the client sends another request, or the client initiates a connection close.

Configuring a server

The following are to be given higher importance while configuring a server:

- IP addresses and TCP ports that may be used to connect to this server.
- Number of subtasks that will be created when the server is initialized.
- Maximum number of threads that will be allowed to exist simultaneously
- Maximum number of TCP connection requests that will be queued if the server is already running its maximum number of threads.
- Length of time the server will wait after serving an HTTP request over a TCP connection before closing the connection if another request is not received.

Logging

Web log file is log file automatically created and maintained by a web server. Every hit to the Web site, including each view of a HTML document, image or other object, is logged. The raw web log file format is essentially one line of text for each hit to the web site. This contains information about who was visiting the site, where they came from, and exactly what they were doing on the web site. The following are the key fields:

- Directory- Directory where log file will be written
- Pattern Information- to be written to the log
- Prefix- String that will be used to begin log filename
- Resolve Hosts -Whether IP addresses (False value) or host names (True value) should be written to the log file
- Rotatable -Whether or not date should be added to file name and file should be automatically rotated each day
- Suffix- String that will be used to end log

