Unit II

Introduction to World Wide Web and its work, Web Browsers, Search Engine, Downloading, Hyper Text Transfer Protocol (HTTP), URL, Web Servers, FTP, Web publishing- Domain Name Registration, Space on Host Server for Web Site, Maintain and Updating

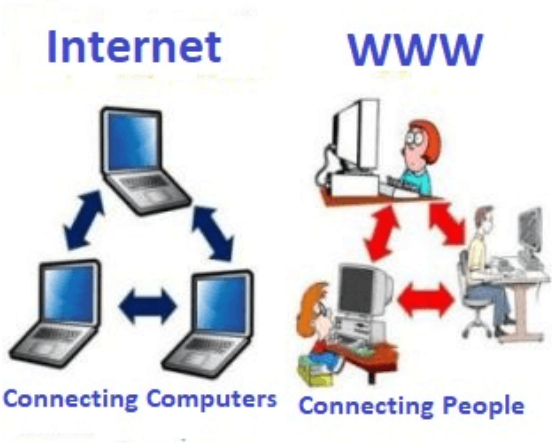
WWW :

World Wide Web, which is also known as a Web, is a collection of websites or web pages stored in web servers and connected to local computers through the internet. A broader definition comes from the organization that Web inventor Tim Berners-Lee helped found, the World Wide Web Consortium (W3C).These websites contain text pages, digital images, audios, videos, etc. Users can access the content of these sites from any part of the world over the internet using their devices such as computers, laptops, cell phones, etc.

The building blocks of the Web are web pages which are formatted in HTML and connected by links called "hypertext" or hyperlinks and accessed by HTTP. These links are electronic connections that link related pieces of information so that users can access the desired information quickly.

Difference between World Wide Web and Internet:

Some people use the terms 'internet' and 'World Wide Web' interchangeably. They think they are the same thing, but it is not so. Internet is entirely different from WWW. It is a worldwide network of devices like computers, laptops, tablets, etc. It enables users to send emails to other users and chat with them online. For example, when you send an email or chatting with someone online, you are using the internet.

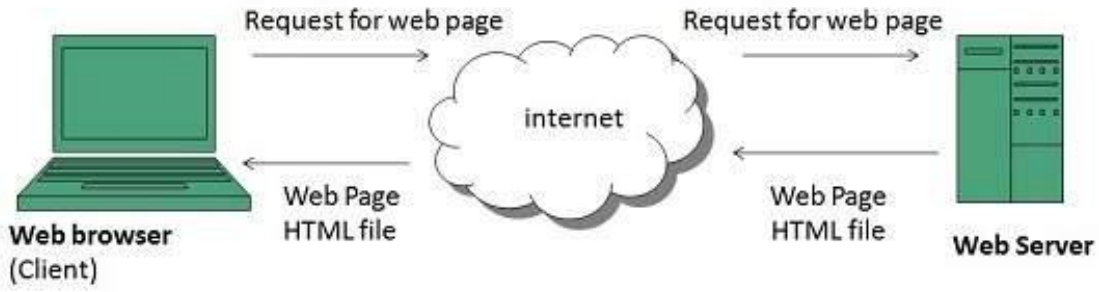


But, when you have opened a website like google.com for information, you are using the World Wide Web; a network of servers over the internet. You request a webpage from your computer using a browser, and the server renders that page to your browser. Your computer is called a client who runs a program (web browser), and asks the other computer (server) for the information it needs.

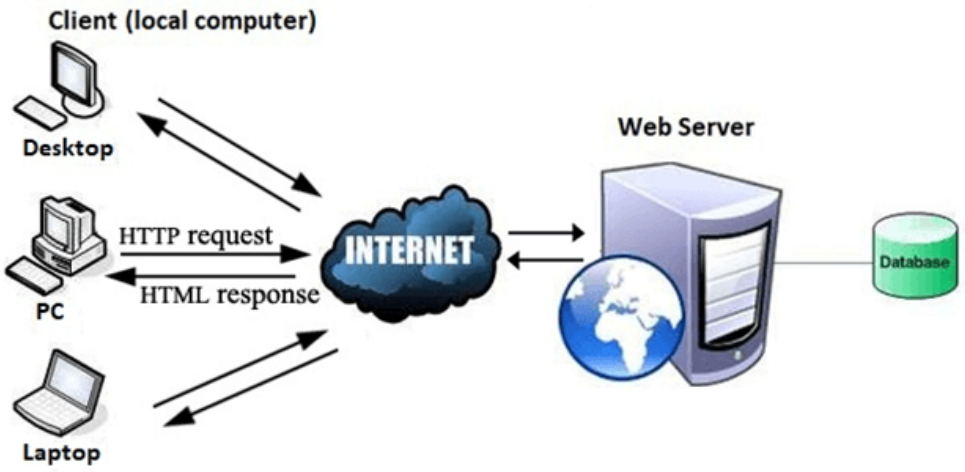
The World Wide Web was invented by a British scientist, Tim Berners-Lee in 1989. He was working at CERN at that time. Originally, it was developed by him to fulfill the need of automated information sharing between scientists across the world, so that they could easily share the data and results of their experiments and studies with each other.

How the World Wide Web Works?

Now, we have understood that WWW is a collection of websites connected to the internet so that people can search and share information. Now, let us understand how it works!

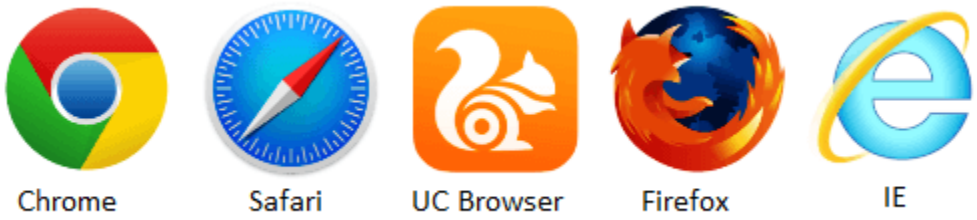


The computer of a user who requests documents from a server is known as a client. Browser, which is installed on the user' computer, allows users to view the retrieved documents.



Web Browser

web Browser is an application software that allows us to view and explore information on the web. User can request for any web page by just entering a URL into address bar.



Web browser can show text, audio, video, animation and more. It is the responsibility of a web browser to interpret text and commands contained in the web page.Earlier the web browsers were text-based while now a days graphical-based or voice-based web browsers are also available. Following are the most common web browser available today:

Browser	**Vendor**
Internet Explorer	Microsoft
Google Chrome	Google
Mozilla Firefox	Mozilla

Netscape Navigator	Netscape Communications Corp.
Opera	Opera Software
Safari	Apple

It provides a software interface that allows you to click hyperlinked resources on the World Wide Web. Today, they are more advanced; along with browsing you can use them for e-mailing, transferring multimedia files, using social media sites, and participating in online discussion groups and more.

Search Engine

The Search engine is a program which is designed to enable the users to browse information or content on World Wide Web. It helps retrieve the desired information in minimum time. It allows you to input specific keywords or phrases and retrieves a list of items matching those keywords and phrases. Thus, it does not provide information straight away; it just retrieves pages which are related to keywords or other search terms. Some of the popular search engines are Google, Bing, and Yahoo! Search.

Search Engine refers to a huge database of internet resources such as web pages, newsgroups, programs, images etc. It helps to locate information on World Wide Web.

User can search for any information by passing query in form of keywords or phrase. It then searches for relevant information in its database and return to the user.

Search Engine Components

Generally there are three basic components of a search engine as listed below:

1. Web Crawler
2. Database
3. Search Interfaces

Web crawler

It is also known as **spider** or **bots.** It is a software component that traverses the web to gather information.

Database

All the information on the web is stored in database. It consists of huge web resources.

Search Interfaces

This component is an interface between user and the database. It helps the user to search through the database.

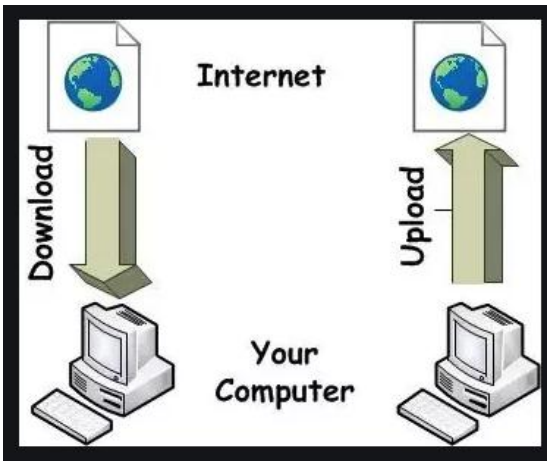
Following are the several search engines available today:

Search Engine	Description
Google	It was originally called BackRub. It is the most popular search engine globally.
Bing	It was launched in 2009 by Microsoft. It is the latest web-based search engine that also delivers Yahoo's results.
Ask	It was launched in 1996 and was originally known as Ask Jeeves. It includes support for match, dictionary, and conversation question.
AltaVista	It was launched by Digital Equipment Corporation in 1995. Since 2003, it is powered by Yahoo technology.

Search Engine Working

Web crawler, database and the search interface are the major component of a search engine that actually makes search engine to work. Search engines make use of Boolean expression AND, OR, NOT to restrict and widen the results of a search. Following are the steps that are performed by the search engine:

- The search engine looks for the keyword in the index for predefined database instead of going directly to the web to search for the keyword.
- It then uses software to search for the information in the database. This software component is known as web crawler.
- Once web crawler finds the pages, the search engine then shows the relevant web pages as a result. These retrieved web pages generally include title of page, size of text portion, first several sentences etc.



Downloading :

'Downloading' is essentially copying data – music, text, images or other information – from a source on the internet and saving it to your computer. It usually involves a process as simple as clicking on a link and following instructions.

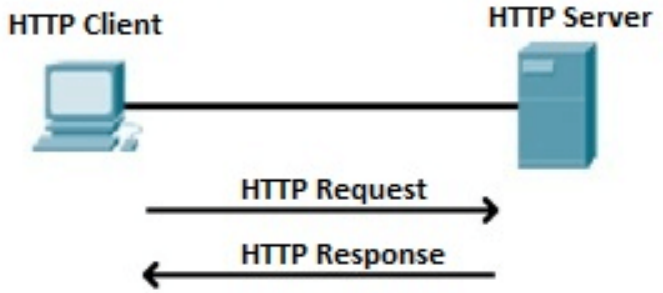
Downloading means your computer is receiving data from the Internet.

Examples of downloading include opening a web page, receiving email, purchasing music files and watching online videos.The File Transfer Protocol (FTP) is the Internet protocol for downloading and uploading files and a number of special applications can furnish FTP services for you.

Hypertext Transfer Protocol (HTTP):

Hyper Text Transfer Protocol (HTTP) is an application layer protocol which enables WWW to work smoothly and effectively. It is based on a client-server model. The client is a web browser which communicates with the web server which hosts the website. This protocol defines how messages are formatted and transmitted and what actions the Web Server and browser should take in response to different commands. When you enter a URL in the browser, an HTTP command is sent to the Web server, and it transmits the requested Web Page.

When we open a website using a browser, a connection to the web server is opened, and the browser communicates with the server through HTTP and sends a request. HTTP is carried over TCP/IP to communicate with the server. The server processes the browser's request and sends a response, and then the connection is closed. Thus, the browser retrieves content from the server for the user.



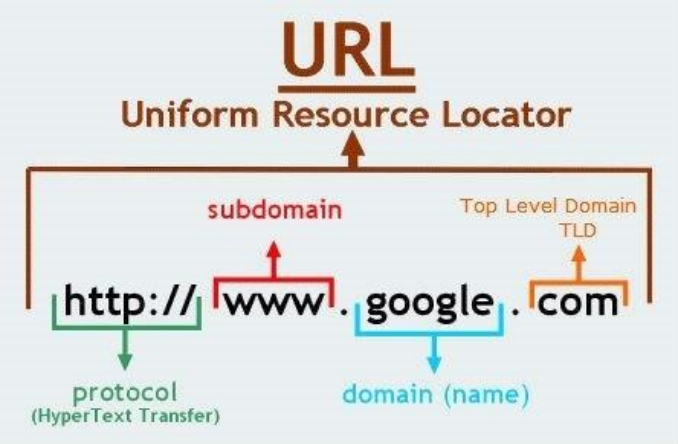
URL

Uniform **R**esource **L**ocator and is defined as the global address of documents and other resources on the World Wide Web. A uniform resource locator (URL) is the address of a resource on the Internet. A URL indicates the location of a resource as well as the protocol used to access it.

A URL contains the following information:

- The protocol used to a access the resource
- The the location of the server (whether by IP address or domain name)
- The port number on the server (optional)
- The location of the resource in the directory structure of the server
- A fragment identifier (optional)

Also known as a Universal Resource Locator (URL) or Web address. A URL is a type of uniform resource identifier (URI).



Web server

Web server is a computer where the web content is stored. Basically web server is used to host the web sites but there exists other web servers also such as gaming, storage, FTP, email etc.

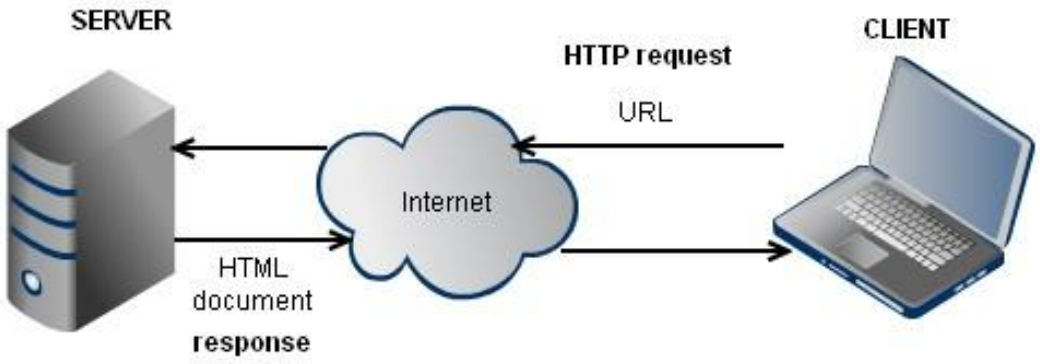
Web site is collection of web pages while web server is a software that respond to the request for web resources.

Web Server Working

Web server respond to the client request in either of the following two ways:

- Sending the file to the client associated with the requested URL.

- Generating response by invoking a script and communicating with database



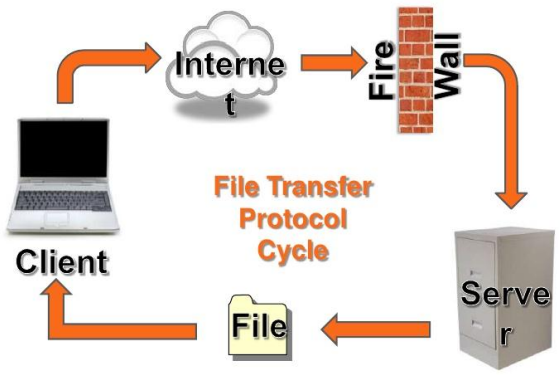
- When client sends request for a web page, the web server search for the requested page if requested page is found then it will send it to client with an HTTP response.
- If the requested web page is not found, web server will the send an **HTTP response:Error 404 Not found.**
- If client has requested for some other resources then the web server will contact to the application server and data store to construct the HTTP response.

FTP

- FTP stands for File transfer protocol.
- FTP is a standard internet protocol provided by TCP/IP used for transmitting the files from one host to another.
- It is mainly used for transferring the web page files from their creator to the computer that acts as a server for other computers on the internet.
- It is also used for downloading the files to computer from other servers.

Objectives of FTP

- It provides the sharing of files.
- It is used to encourage the use of remote computers.
- It transfers the data more reliably and efficiently.



There are two types of connections in FTP:

- **Control Connection:** The control connection uses very simple rules for communication. Through control connection, we can transfer a line of command or line of response at a time. The control connection is made between the control processes. The control connection remains connected during the entire interactive FTP session.
- **Data Connection:** The Data Connection uses very complex rules as data types may vary. The data connection is made between data transfer processes. The data connection opens when a command comes for transferring the files and closes when the file is transferred.

Advantages of FTP:

- **Speed:** One of the biggest advantages of FTP is speed. The FTP is one of the fastest way to transfer the files from one computer to another computer.
- **Efficient:** It is more efficient as we do not need to complete all the operations to get the entire file.
- **Security:** To access the FTP server, we need to login with the username and password. Therefore, we can say that FTP is more secure.
- **Back & forth movement:** FTP allows us to transfer the files back and forth. Suppose you are a manager of the company, you send some information to all the employees, and they all send information back on the same server.

Web publishing

Web publishing is the process of publishing original content on the Internet.

The process includes building and uploading websites, updating the associated webpages, and posting content to these webpages online. Web publishing comprises of personal, business, and community websites in addition to e-books and blogs.

The content meant for web publishing can include text, videos, digital images, artwork, and other forms of media.

Publishers must possess a web server, a web publishing software, and an Internet connection to carry out web publishing.

Web publishing is also known as online publishing.

Website publishing is the process of uploading content on the internet. It includes:

- uploading files
- updating web pages
- posting blogs

1. domain name : A domain name is the part of your Internet address that comes after **"www".** For example, in **www.websol.com** the domain name is **websol.com.**
2. A domain name becomes your Business Address so care should be taken to select a domain name. Your domain name should be easy to remember and easy to type.
3. A domain name is your website name. A domain name is the address where Internet users can access your website. A domain name is used for finding and identifying computers on the Internet. Computers use IP addresses, which are a series of number.
4. A domain name can be any combination of letters and numbers, and it can be used in combination of the various domain name extensions, such as .com, .net and more.

Domain Extensions

The final letter at end of internet address is known as top level domain names. They are called top level because they are read from right to left, and the part after the dot is the highest in a hierarchy.

The following table shows the **Generic** Top-Level Domain names:

Domain	Meaning
.com	Commercial Business
.edu	Education

.gov	U.S. government agency
.int	International Entity
.mil	U.S. military
.net	Networking organization
.org	Non profit organization

Web hosting

Web hosting is a service of providing online space for storage of web pages. These web pages are made available via **World Wide Web.** The companies which offer website hosting are known as **Web hosts.**

The servers on which web site is hosted remain switched on 24 x7. These servers are run by web hosting companies. Each server has its own IP address. Since IP addresses are difficult to remember therefore, webmaster points their domain name to the IP address of the server their website is stored on.

It is not possible to host your website on your local computer, to do so you would have to leave your computer on 24 hours a day.