Chapter 1: Fundamentals

Introduction::

Web technologies related to the interface between web servers and their clients. This information includes markup languages, programming interfaces and languages, and standards for document identification and display. In general web technology incorporates tools and techniques for web development.

Web Development is a broad term for the work involved in developing a web site for World Wide Web. This can include ***web design, web content development, client liaison, client-side/server-side scripting, web server and network security configuration, and e-commerce development.*** However, among web professionals, "web development" usually refers to the main non-design aspects of building web sites: writing markup and coding. Web development can range from developing the simplest static single page of plain text to the most complex web-based internet applications, electronic businesses, or social network services.

Web design is a broad term used to encompass the way that content (usually hypertext or hypermedia) is delivered to an end-user through the World Wide Web, using a web browser or other web-enabled software is displayed. The intent of web design is to create a website—a collection of online content including documents and applications that reside on a web servers. A website may include text, images, sounds and other content, and may be interactive.

For the typical web sites, the basic aspects of design are:

- **The *content***: the substance, and information on the site should be relevant to the site and should target the area of the public that the website is concerned with.

- **The *usability***: the site should be user-friendly, with the interface and navigation simple and reliable.

- **The *appearance***: the graphics and text should include a single style that flows throughout, to show consistency. The style should be professional, appealing and relevant.

- **The *structure***: of the web site as a whole.

Web or Internet?

Internet and the web are not the same thing. The Internet is collection of computers and the other devices connected by equipment that allows them to communicate with each other. The Web is a collection of software and protocols that has been installed on most of the computers on the Internet. Most users of the Internet use it through the Web. In abstract sense, the web is merely a vast collection of documents, some of which are connected by links.

Internet and its Evolution:

Internet is a short form of the technical term ***internetwork***, the result of interconnecting computer networks with special gateways or routers. The Internet is also often referred to as *the Net*. The Internet is a massive network of networks, a networking infrastructure. It connects millions of computers together globally, forming a network in which any computer can communicate with any other computer as long as they are both connected to the Internet. Information that travels over the Internet does so via a variety of languages known as protocols. **The Internet is loosely connected compared with the randomized graph.**

The Internet is a globally distributed network comprising many voluntarily interconnected autonomous networks. It operates without a central governing body. However, to maintain interoperability, all technical and policy aspects of the underlying core infrastructure and the principal name spaces are administered by the **Internet Corporation for Assigned Names and Numbers (ICANN).**

The **history of the Internet** starts in the 1950s and 1960s with the development of computers. This began with point-to-point communication between mainframe computers and terminals, expanded to point-to-point connections between computers and then early research into packet switching.

Since the mid-1990s the Internet has had a drastic impact on culture and commerce, including the rise of near instant communication by electronic mail, instant messaging, Voice over Internet Protocol (VoIP) "phone calls", two-way interactive video calls, and the World Wide Web with its discussion forums, blogs, social networking, and online shopping sites.

World Wide Web:

WWW is a system of interlinked hypertext documents accessed via the Internet. The World Wide Web, or simply Web, is a way of accessing information over the medium of the Internet. It is an information-sharing model that is built on top of the Internet. The Web uses the HTTP protocol, only one of the languages spoken over the Internet, to transmit data. Web services, which use HTTP to allow applications to communicate in order to exchange business logic, use the Web to share information. The Web also utilizes browsers, such as Internet Explorer or Firefox, to access Web documents called Web pages that are linked to each other via hyperlinks. Web documents also contain graphics, sounds, text and video.

The Web is one of the services that runs on the Internet. It is a collection of textual documents and other resources, linked by hyperlinks and URLs, transmitted by web browsers and web servers. The Web is just one of the ways that information can be disseminated over the Internet, so the Web is just a portion of the Internet. In short, the Web can be thought of as an application "running" on the Internet

What is Hypertext?

Hypertext provides the links between different documents and different document types. In a hypertext document, links from one place in the document to another are included with the text. By selecting a link, you are able to jump immediately to another part of the document or even to a different document. In the WWW, links can go not only from one document to another, but from one computer to another.

World Wide Consortium:

The **World Wide Web Consortium** (**W3C**) is the main international standards organization for the World Wide Web. W3C was created to ensure compatibility and agreement among industry members in the adoption of new standards. Prior to its creation, incompatible versions of HTML were offered by different vendors, increasing the potential for inconsistency between web pages. The consortium was created to get all those vendors to agree on a set of core principles and components which would be supported by everyone.

Web Page:

A **web page** is a document or information resource that is suitable for the World Wide Web and can be accessed through a web browser and displayed on a monitor or mobile device. This information is usually in HTML or XHTML format, and may provide navigation to other web pages via hypertext links. Web pages frequently subsume other resources such as style sheets, scripts and images into their final presentation.

Web pages may be retrieved from a local computer or from a remote web server. The web server may restrict access only to a private network, e.g. a corporate intranet, or it may publish pages on the World Wide Web. Web pages are requested and served from web servers using Hypertext Transfer Protocol (HTTP).

Web pages may consist of files of static text and other content stored within the web server's file system (static web pages), or may be constructed by server-side software when they are requested (dynamic web pages). Client-side scripting can make web pages more responsive to user input once on the client browser.

Web Site:

A **website** or simply **site**, is a collection of related web pages containing images, videos or other digital assets. A website is hosted on at least one web server, accessible via a network such as the Internet or a private local area network through an Internet address known as a Uniform Resource Locator. All publicly accessible websites collectively constitute the World Wide Web. *Web sites can be static or dynamic.*

Static Website:

A static website is one that has web pages stored on the server in the format that is sent to a client web browser. It is primarily coded in Hypertext Markup Language, HTML.

Simple forms or marketing examples of websites, such as *classic website*, a *five-page website* or a *brochure website* are often static websites, because they present pre-defined, static information to the user. This may include information about a company and its products and services via text, photos, animations, audio/video and interactive menus and navigation.

This type of website usually displays the same information to all visitors. Similar to handing out a printed brochure to customers or clients, a static website will generally provide consistent, standard information for an extended period of time. Although the website owner may make updates periodically, it is a manual process to edit the text, photos and other content and may require basic website design skills and software.

In summary, visitors are not able to control what information they receive via a static website, and must instead settle for whatever content the website owner has decided to offer at that time.

Dynamic Website:

A dynamic website is one that changes or customizes itself frequently and automatically, based on certain criteria.

Dynamic websites can have two types of dynamic activity: Code and Content. Dynamic code is invisible or behind the scenes and dynamic content is visible or fully displayed.

The first type is a web page with dynamic code. The code is constructed dynamically on the fly using active programming language instead of plain, static HTML.

The second type is a website with dynamic content displayed in plain view. Variable content is displayed dynamically on the fly based on certain criteria, usually by retrieving content stored in a database

Domain Names, DNS, and URLs:

- IP addresses are not convenient for users to remember easily. So an IP address can be represented by a natural language convention called a **domain name**

- **Domain name system (DNS)** translates domain names into IP addresses. DNS is the "phone book" for the Internet, it maps between host names and IP addresses.

- A **uniform resource locator (URL)**, which is the address used by a Web browser to identify the location of content on the Web, also uses a domain name as part of the URL.
- **Syntax: scheme: scheme-depend-part.** Example: In http://www.example.com/, the scheme is http.
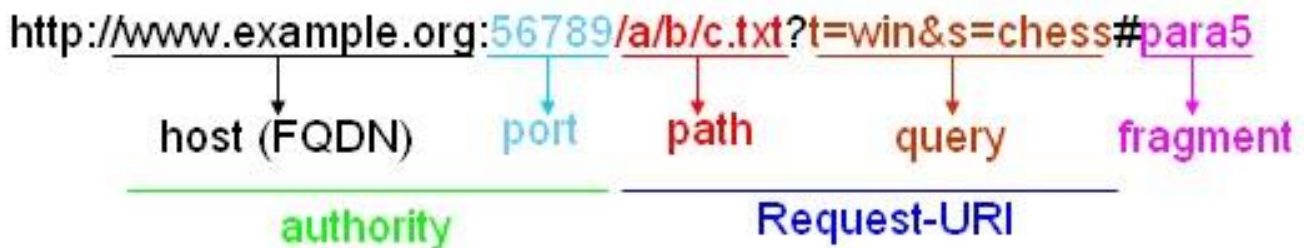
Domain Names

People have difficulty dealing with and remembering numbers, machines on the Internet also have textual names. These names begin with the name of the host machine, followed by progressively larger enclosing collections of machines, called ***domains.*** There may be two, three or more domain names.

- The first domain name, which appears immediately to the right of the hostname, is the domain of which the host is a part
- The second domain name gives the domain of which the first domain is a part
- The last domain name identifies the of organization in which the host resides

For organization in Nepal, edu is the extension for educational institutions, com specifies a company, gov is used for Nepalese Government and org is used for many other kinds of organizations. In most of the countries, the largest domain is often an abbreviation for the country- for example np is used for Nepal, se for Sweden, uk is for United Kingdom etc.

Consider the following example:

> movies.comedy.marxbros.com

Here "movies" is the host name and "comedy" is movies local domain, which is a part of "marxbros's " domain, which is a part of the com domain. The host name and all of the domain names are together called "fully qualified domain name".

Because IP addresses are the addresses used internally by the internet, the fully qualified domain name of the destination for a message must be converted to an IP address before the message can be transmitted on the Internet to the destination. These conversions are done by machines called ***name servers,*** which implement the Domain Name System (DNS).

Name servers serve a collection of machines on the Internet and are operated by organizations that are responsible for the part of the Internet to which those machines are connected. All document requests from browsers are routed to the nearest name server. If the name server can convert the fully qualified domain name to an IP address, it does so. If it cannot, the name server sends the fully qualified domain name to another name server for conversion. Like IP address, fully qualified domain names must be unique. Following figure shows how fully qualified domain names requested by a browser are translated into Ips before they are routed to the appropriate Web Server.
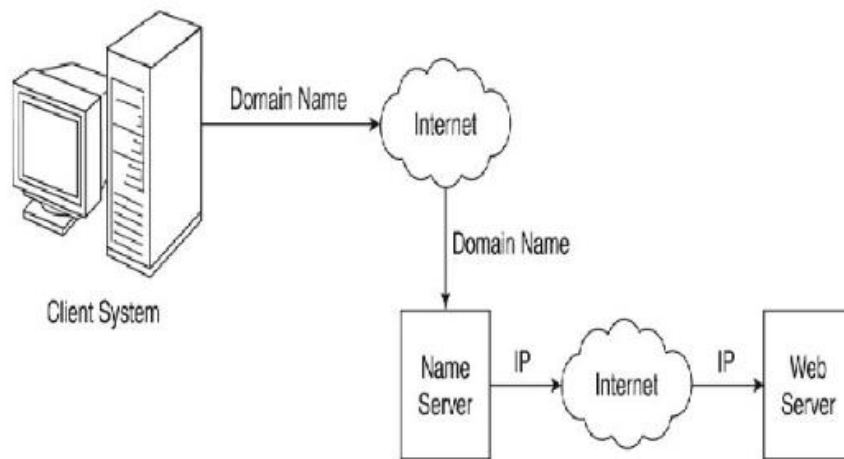
DNS Lookup Process

Step 1: Requesting Website Information

First, you visit a website by typing a domain name into a web browser. Your computer will start resolving the hostname, such as www.liquidweb.com. Your computer will look for the IP address associated with the domain name in its local DNS cache, which stores DNS information that your computer has recently saved. If it is present locally, then the website will be displayed. If your computer does not have the data stored, then it will perform a DNS query to retrieve the correct information.

Step 2: Contact the Recursive DNS Servers

If the information is not in your computer's local DNS cache, then it will query the recursive DNS servers from your (ISP) Internet service provider. Recursive DNS servers have their local DNS cache, much like your computer. Given that many of the ISP's customers are using the same recursive DNS servers, there is a chance that common domain names already in its cache. If the domain is cached, the DNS query will end here and the website displayed to the user.

Step 3: Query the Authoritative DNS Servers

If a recursive DNS server or servers do not have the information stored in its cache memory, the DNS query continues to the authoritative DNS server that has the data for a specific domain. These authoritative name servers are responsible for storing DNS records for their respective domain names.

Step 4: Access the DNS Record

For our example, to find out the IP address for www.liquidweb.com, we will query the authoritative name server for the address record (A record). The Recursive DNS server accesses the A record for www.liquidweb.com from the authoritative name servers and stores the record in its local DNS cache. If other DNS queries request the A record for www.liquidweb.com, the recursive server will have the answer and will not have to repeat the DNS lookup process. All DNS records have a time-to-live value, which shows when a DNS record will expire. After some time has passed, the recursive DNS server will ask for an updated copy of the DNS record.

Step 5: Final DNS Step

The Recursive DNS server has the information and returns the A record to your computer. Your computer will store the DNS record in its local DNS cache, will read the IP address from the DNS record, and pass this information to your browser. The web browser will connect to the web server associated with the A records IP and display the website.

The entire DNS lookup process, from start to finish, takes only milliseconds to complete. For a more profound understanding let's break down the previously mentioned DNS components that are relevant to the DNS lookup process.
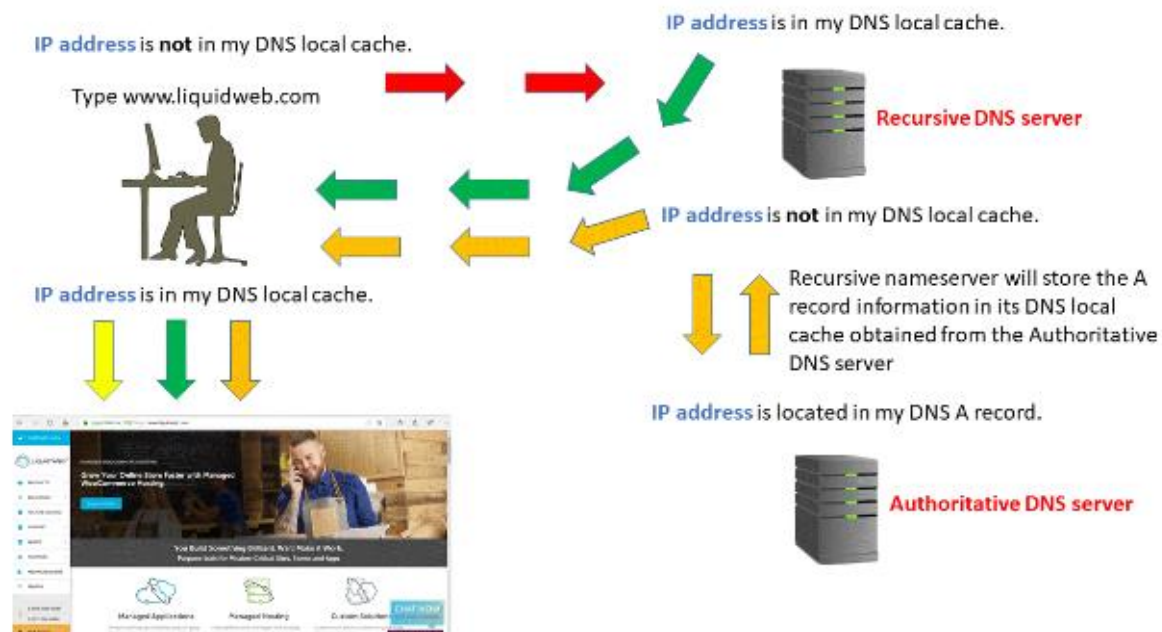
HTTP and HTTPS:

What do they do, and how are they different?

You click to check out at an online merchant. Suddenly your browser address bar says HTTPS instead of HTTP. What's going on? Is your credit card information safe?

Good news. Your information is safe. The website you are working with has made sure that no one can steal your information.

Instead of HyperText Transfer Protocol (HTTP), this website uses HyperText Transfer Protocol Secure (HTTPS).

Using HTTPS, the computers agree on a "code" between them, and then they scramble the messages using that "code" so that no one in between can read them. This keeps your information safe from hackers.

They use the "code" on a Secure Sockets Layer (SSL), sometimes called Transport Layer Security (TLS) to send the information back and forth.

- HTTP is based on the request-response communication model:
    - Client sends a request
    - Server sends a response
    - HTTP is a stateless protocol: where the protocol does not require the server to remember anything about the client between requests.

- Normally implemented over a TCP connection (80 is standard port number for HTTP)

- The following is the typical browser-server interaction using HTTP:
    1. User enters Web address in browser
    2. Browser uses DNS to locate IP address
    3. Browser opens TCP connection to server
    4. Browser sends HTTP request over connection
    5. Server sends HTTP response to browser over connection
    6. Browser displays body of response in the client area of the browser window

Client/Server Computing:

- A model of computing in which powerful personal computers are connected in a network together with one or more servers

- **Client** is a powerful personal computer that is part of a network; service requester

- **Server** is a networked computer dedicated to common functions that the client computers on the network need; service provider

- Web is based on client/server technology. Web servers are included as part of a larger package of internet and intranet related programs for serving e-mail, downloading requests for FTP files and building and publishing web pages. Typically the e-commerce customer is the client and the business is the server. In the client/ server model single machine can be both client and the server The client/ server model utilises a database server in which RDBMS user queries can be answered directly by the server.

- The client/ server architecture reduces network traffic by providing a query response to the user rather than transferring total files. The client/ server model improves multi-user updating through a graphical user interface (GUI) front end to the shared database. In client/ server architectures client and server typically communicate through statements made in structured query language (SQL).
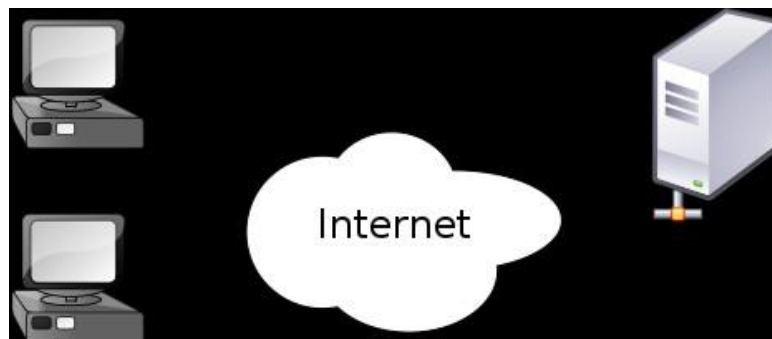
SMTP

is part of the application layer of the TCP/IP protocol. Using a process called "store and forward," SMTP moves your email on and across networks. It works closely with something called the Mail Transfer Agent (MTA) to send your communication to the right computer and email inbox.

SMTP spells out and directs how your email moves from your computer's MTA to an MTA on another computer, and even several computers. Using that "store and forward" feature mentioned before, the message can move in steps from your computer to its destination. At each step, **Simple Mail Transfer Protocol** is doing its job. Lucky for us, this all takes place behind the scenes, and we don't need to understand or operate SMTP.

SMTP provides a set of codes that simplify the communication of email messages between email servers (the network computer that handles email coming to you and going out). It's a kind of shorthand that allows a server to break up different parts of a message into categories the other server can understand. When you send a message out, it's turned into strings of text that are separated by the code words (or numbers) that identify the purpose of each section.

WAP:

Wireless application protocol (WAP) is a communications protocol that is used for wireless data access through most mobile wireless networks. WAP enhances wireless specification interoperability and facilitates instant connectivity between interactive wireless devices (such as mobile phones) and the Internet.

WAP functions in an open application environment and may be created on any type of OS. Mobile users prefer WAP because of its ability to efficiently deliver electronic information.

FTP

FTP is an acronym for File Transfer Protocol. As the name suggests, FTP is used to transfer files between computers on a network. You can use FTP to exchange files between computer accounts, transfer files between an account and a desktop computer, or access online software archives. Keep in mind, however, that many FTP sites are heavily used and require several attempts before connecting.

Graphical FTP clients simplify file transfers by allowing you to drag and drop file icons between windows. When you open the program, enter the name of the FTP host (for example, ftp.microsoft.com) and your username and password. If you are logging into an anonymous FTP server, you may not have to enter anything. Two common FTP programs are Cyberduck (for Mac) and WinSCP (for Windows).

DHCP

DHCP stands for Dynamic Host Configuration Protocol. The built-in DHCP server automatically assigns IP addresses to the computers and other devices on each local area network (LAN). In this section you can view a list of assigned IP addresses and reserve IP addresses for particular devices. It automatically provides and assigns IP addresses default gateways and other network parameters to client devices. Without it, the network admin has to manually setup every client that joins the network.

POP

Post Office Protocol (**POP**) is a type of **computer** networking and Internet standard protocol that extracts and retrieves email from a remote mail server for access by the host machine. **POP** is an application layer protocol in the OSI model that provides end users the ability to fetch and receive email. The Internet Message Access Protocol (IMAP) is a mail protocol used for accessing email on a remote web server from a local client. IMAP and **POP3** are the two most commonly used Internet mail protocols for retrieving emails. Both protocols are supported by all modern email clients and web servers.

PAN

A personal area network (PAN) is the interconnection of information technology devices within the range of an individual person, typically within a range of 10 meters. For example, a person traveling with a laptop, a personal digital assistant (PDA), and a portable printer could

interconnect them without having to plug anything in, using some form of wireless technology. Typically, this kind of personal area network could also be interconnected without wires to the Internet or other networks. Conceptually, the difference between a PAN and a wireless LAN is that the former tends to be centered around one person while the latter is a local area network (LAN) that is connected without wires and serving multiple users.

MIME

MIME (Multi-Purpose Internet Mail Extensions) is an extension of the original Internet e-mail protocol that lets people use the protocol to exchange different kinds of data files on the Internet: audio, video, images, application programs, and other kinds, as well as the ASCII text handled in the original protocol, the Simple Mail Transport Protocol (SMTP). Servers insert the MIME header at the beginning of any Web transmission. Clients use this header to select an appropriate "player" application for the type of data the header indicates. Some of these players are built into the Web client or browser (for example, all browsers come with GIF and JPEG image players as well as the ability to handle HTML files).

Fig: A simple example of protocols's application
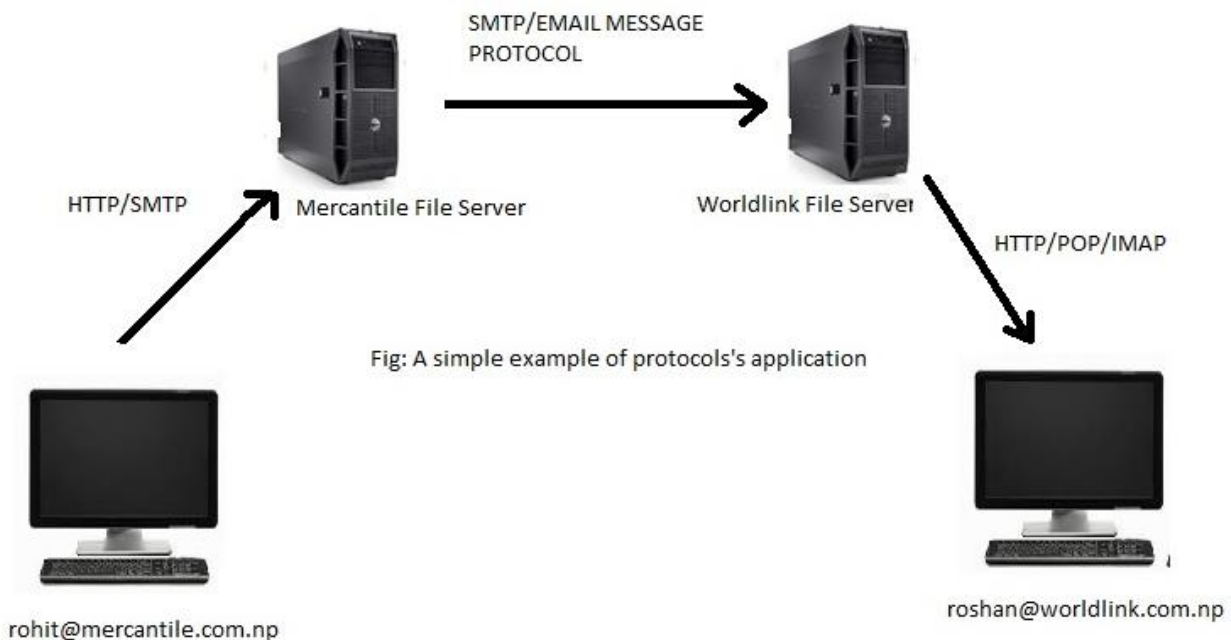
URL & URI

URI can be a name, locator, or both for an online resource where a URL is just the locator. URLs are a subset of URIs. That means all URLs are URIs. It doesn't work the opposite way though. Not all URIs are URLs because a URI could be a name instead of a locator. Here's an illustration of the difference between the two:
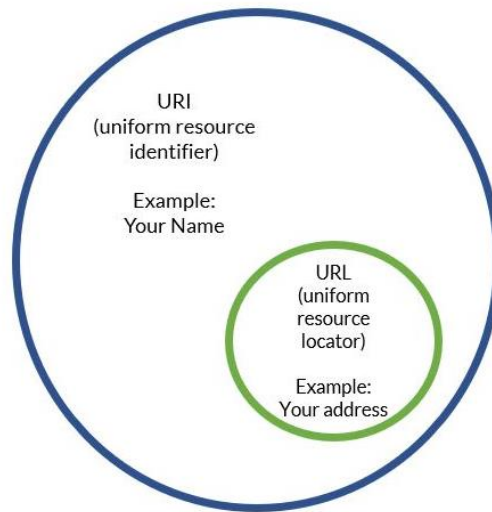
Your name could be a URI because it identifies you, but it couldn't be a URL because it doesn't help anyone find your location. On the other hand, your address is both a URI and a URL because it both identifies you and it provides a location for you.

In the web development world, you might be trying to get different resources by name so you'll have different URIs that have distinct URLs. Remember, a URL is the locator of a resource. The URI is the identifier of the resource and it can just happen to be a locator.

Web Clients:

It typically refers to the Web browser in the user's machine. It is a software application for retrieving, presenting, and traversing information resources on the web server. It is used to create a HTTP request message and for processing the HTTP response message.

User agent: Any web client is designed to directly support user access to web servers is known as user agent. Web browsers can run on desktop or laptop computers. Some of the browsers are: Internet Explorer, Mozilla, FireFox, Chrome, Safari, Opera, Netscape Navigator.

Web Browsers:

Browsers are software programs that allow you to search and view the many different kinds of information that's available on the World Wide Web. The information could be web sites, video or audio information.

Intranet Vs Extranet

An intranet is a network where employees can create content, communicate, collaborate, get stuff done, and develop the company culture. An extranet is like an intranet, but also provides controlled access to authorized customers, vendors, partners, or others outside the company.

Status Bar: You will find the status bar at the very bottom of your browser window. It basically tells you what you are doing at the moment. Mainly, it shows you load speed and the URL address of whatever your mouse is hovering over.

Title Bar: You will find this bar at the absolute top of your browser and in will be the colour blue for the major browsers. The purpose of the Title bar is to display the title of the web page that you are currently viewing.

Menu Bar: The menu bar contains a set of dropdown menus

Navigational Tool: A bar contains standard push button controls that allow the user to return to a previously viewed page, to reverse and refresh the page, to display the home page and to print the page etc.

Toolbar Icons: You will find the Toolbar directly under the Title Bar. The Toolbar is where you will find the back button, home button and the refresh button etc.

Client Area: It is a display window which is the space in which you view the website.

Scroll Bars: The Scroll bars, usually located to the right of the Display Window, allows you to "scroll" (move down or up the web page) so you can view information that is below or above what is currently in the Display Window.
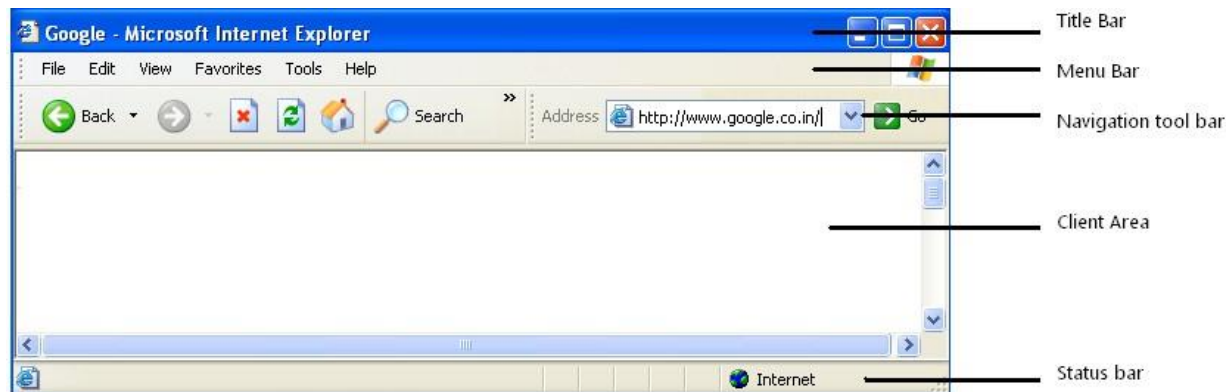
Web Servers:

Basic functionality:
- It receives HTTP request via TCP
- It maps Host header to specific virtual host (one of many host names sharing an IP address)
- It maps Request-URI to specific resource associated with the virtual host
    - File: Return file in HTTP response
    - Program: Run program and return output in HTTP response
- It maps type of resource to appropriate MIME type and use to set Content-Type header in HTTP response
- It Logs information about the request and response
- All e-commerce site require basic Web server software to answer requests from customers like ;
    - **Apache**
        - Leading Web server software (47% of market)
        - Works with UNIX, Linux , Windows OSs
    - **Microsoft's Internet Information Server (IIS)**
        - Second major Web server software (25% of market)
        - Windows-based

Client-Side Scripting:

- Client-side scripting generally refers to writing the class of computer programs (scripts) on the web that are executed at *client-side*, by the user's web browser, instead of *server-side* (on the web server). Usually scripts are embedded in the HTML page itself.

- JavaScript , VBScript, Jscript, Java Applets etc. are the examples of client side scripting technologies. JavaScript is probably the most widely used client-side scripting language.

- Client-side scripts have greater access to the information and functions available on the user's browser, whereas server-side scripts have greater access to the information and functions available on the server. Upon request, the necessary files are sent to the user's computer by the web server (or servers) on which they reside. The user's web browser executes the script, then displays the document, including any visible output from the script.

- Client-side scripts may also contain instructions for the browser to follow in response to certain user actions, (e.g., clicking a button). Often, these instructions can be followed without further communication with the server.

Server-Side Scripting:

- Includes writing the applications executed by the server at run-time to process

client input or generate document in response to client request. So server side script consists the directives embedded in Web page for *server* to process before passing page to requestor.

· It is usually used to provide interactive web sites that interface to databases or other data stores.

· This is different from client-side scripting where scripts are run by the viewing web browser, usually in JavaScript. The primary advantage to server-side scripting is the ability to highly customize the response based on the user's requirements, access rights, or queries into data stores.

· PHP, JSP, ASP…. etc, are the server side scripting technologies.

Web 2.0:

The term **Web 2.0** is associated with web applications that facilitate participatory *information sharing, interoperability, user-centered design, and collaboration* on the World Wide Web. A Web 2.0 site allows users to interact and collaborate with each other in a social media dialogue as creators   of user-generated content in a virtual community, in contrast to websites where users   are limited to the passive viewing of content that was created for them. Examples of Web 2.0 include *social networking sites, blogs, wikis, video sharing sites, hosted services, web applications.*