Digicoders technologies (P) Ltd.

LECTURE NOTES

ON

INTERNET AND WEB TECHNOLOGY

Unit -2

Compiled by

Team Digicoders
B-36, Sector O, Near Ram Ram Bank Chauraha, Aliganj,
Lucknow Uttar Pradesh 226021

CONTENTS

UNIT-2
Internet Connectivity & WWW

Different types of internet connectivity:

1) **Dial-up Connections:**
   - A dial-up connection allows you to connect to the internet via a local server using a standard 56k modem.
   - Your PC literally dials (hence the name) a phone number (provided by your ISP) and connects to the server and therefore the internet.
2) **Leased Line Connection:**
   - Leased connection is also known as direct Internet access.
   - It is the secure, dedicated and most expensive, level of Internet connection.
   - With leased connection, our computer is dedicatedly and directly connected to the Internet using high speed transmission lines.
   - It is on-line twenty-four hours a day, seven days a week.
3) **ISDN**
   - ISDN is acronym of Integrated Services Digital Network. It establishes the connection using the phone lines which carry digital signals instead of analog signals.
   - There are two techniques to deliver ISDN services:
   - Basic Rate Interface (BRI)
   - Primary Rate Interface (PRI)
4) **DSL**

DSL is acronym of Digital Subscriber Line. It is a form of broadband connection as itprovides connection over ordinary telephone lines.

Following are the several versions of DSL technique available today:

- Asymmetric DSL (ADSL)
- Symmetric DSL (SDSL)
- High bit-rate DSL (HDSL)
- Rate adaptive DSL (RDSL)
- Very high bit-rate DSL (VDSL)
- ISDN DSL (IDSL)

All of the above mentioned technologies differ in their upload and download speed, bit transfer rate and level of service.

ADSL (Asymmetric Digital Subscribers Line) Connections:

The connections work by splitting your phone line into two separate channels, one for data (internet) and one for voice (phone calls), which means you can talk on the phone and be connected to the internet at the same time.

5) **Cable connections:**
   - Through the use of a cable modem you can have a broadband Internet connection that is designed to operate over cable TV lines.
   - Cable Internet works by using TV channel space for data transmission, with certain channels used for downstream transmission, and other channels for upstream transmission.

6) **VSAT(very small aperture terminal):**
   - It is an earthbound station used in satellite communications of data, voice and video signals, excluding broadcast television.
   - A VSAT consists of two parts, a transceiver that is placed outdoors in direct line of sight to the satellite and a device that is placed indoors to interface the transceiver with the end user's communications device, such as a PC.

7) **Wireless Internet Connection**

   Wireless Internet Connection makes use of radio frequency bands to connect to the internet and offers a very high speed. The wireless internet connection can be obtained by either WiFi or Bluetooth.

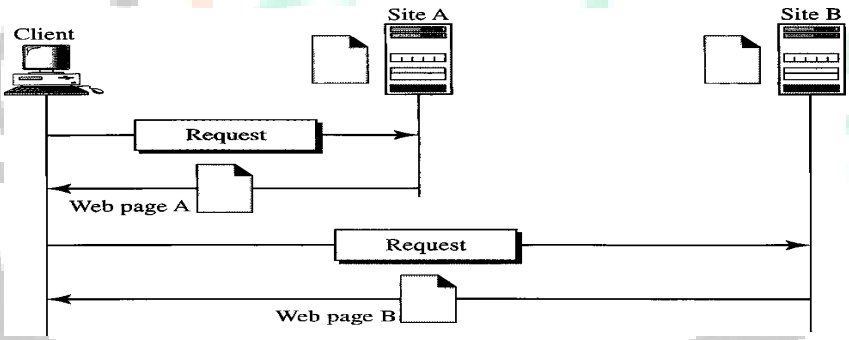
WORLD WIDE WEB (WWW)

- World Wide Web (WWW) is global, seamless environment in which all information (text, image, audio, video, computational service) are accessible in a consistent and simple way by using a standard set of naming and access convention from internet.
- WWW was initially developed by Tim Berners-Lee and others at CERN, Switzerland.
- It is also called WWW or W3 or Web.

ARCHITECTURE:

- The WWW today is a distributed client/server service, in which a client using a browser

  can access a service using a server. However, the service provided is distributed over many locations called *sites*.



- Each site holds one or more documents, referred to as *Web pages*. Each Web page can contain a link to other pages in the same site or at other sites. The pages can be retrieved

  and viewed by using browsers.

Principles of WWW:

- There would be no central control. The web works because people work within the agreed-to guidelines. As part of this the web ethic is that any one publish, and anyone(authorized user) can read information.
- All web servers would use the same protocols/mechanism etc.
- *http* is a fast, stateless, extensible transport mechanism would be used to communicate within the Web.
- httpd or http daemons, would be the base server which receive messages and providing data as requested.
- URLs would be used for network-wide addressing.
- All web browsers would use the same basic language- HTML.

Features of WWW

The Web has unique features:

1. It is grand globally distributed internet service.
2. It is platform independent.
3. It has GUI features with easy navigability through hypertext and multimedia links.
4. It is interactive, because of search engine has push technology features like Web casting etc.
5. It is continuously ex ble and each updated almost by a minute.
6. By processing the Web contents using CGI, Servlets, JavaScript, ASP, and JSP. The services for e-mail, e-com, and m-com are also vi.

APPLICATION LAYER PROTOCOLS

There are several protocols which work for users in Application Layer.Application layer protocols can be broadly divided into two categories:

• Protocols which are used by users. For email for example, E-Mail

• Protocols which help and support protocols used by users. For example DNSFew of Application layer protocols are described below:

1. Hyper Text Transfer Protocol (HTTP) :

The HyperText Transfer Protocol (HTTP) is used to define how the client serverprograms can be written to retrieve web pages from the Web.

• An HTTP client sends a request; an HTTP server returns a response. Theserver uses the port number 80

• HTTP uses the services of TCP. This means that, before any transactionbetween the client and the server can take place, a connection needs to be established between them. After the transaction, the connection should beterminated.

2. Domain Name System:

• DNS Stands for Domain Name System (or Service or Server).

• It is an Internet service that translates domain names into IP addresses.

• Because domain names are alphabetic, they're easier to remember. The Internet however, is really based on IP addresses. Every time you use a domainname, therefore, a DNS service must translate the name into the correspondingIP address. For example, the domain name www.example.com might translate to 198.105.232.4.

3. File Transfer Protocol:

- File Transfer Protocol (FTP) is the standard mechanism provided by TCP/IP for copying a file from one host to another or transferring filesfrom one system to another.
- It is not only a protocol but also a service as well as application.
- Some problems in transferring files are two systems may use different filename conventions, two systems may have different ways to represent textand data, two systems may have different directory structures. All these problems have been solved by FTP in a very simple and elegant approach.

4. Simple Mail Transfer Protocol :

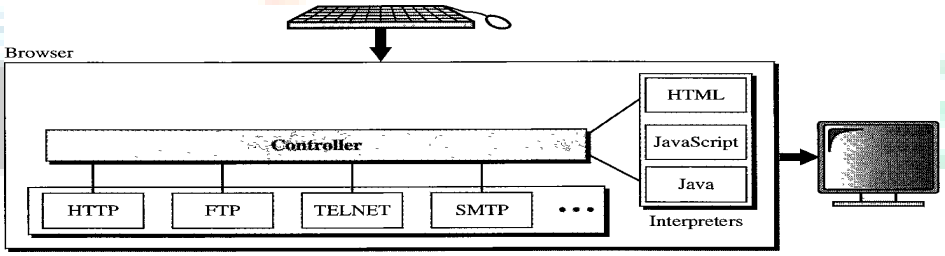
- This is the method protocol computer use to send messages by messagetransfer agents (MTA) on the internet.
- MTA are client & server programs that perform e-mail services such assending or receiving mail for a host computer.

5. TELNET:

- Telnet is a protocol or set of rules that enables one computer to connect another computer. This process is also known as remote login.
- The user computer which initiates connection is referred as localcomputer and the machine being connected to which accepts the connection is referred as the remote or host computer.
- To start telnet session you must login to the server by entering a validuser name and password.
- Telnet is both TCP/IP application and a protocol for connecting a localcomputer to a remote computer.
- Telnet operates on client server principle. The local computer uses a telnet client program to establish the connection and display data on localcomputer monitor.

Web Browser

- Browsers are world Wide Web client side software that enables the user to access resources on the web.
- Each browser usually consists of three parts: a controller, client protocol, and interpreters.
- The controller receives input from the keyboard or the mouse and uses the client programs to access the document.



- After the document has been accessed, the controller uses one of the interpreters to display the document on the screen.
- The interpreter can be HTML, Java, or JavaScript, depending on the type of document.

Features of Web Browser

- The most popular browsers have a number of features in common. They allow users to set bookmarks and browse in a private mode. They also can be customized with extensions, and some of them provide a sync service.
- Most browsers have these user interface features:
- Allow the user to open multiple pages at the same time, either in different browser windows or in different tabs of the same window
- Back and forward buttons to go back to the previous page visited or forward to the next one.
- A refresh or reload button to reload the current page.
- As top button to cancel loading the page. (In some browsers, the stop button is merged with the reload button.)
- A home button to return to the user's home page.
- An address bar to input the URL of a page and display it.
- A search bar to input terms into a search engine. (In some browsers, the search baris merged with the address bar.)

- One of the main feature of a browser is to search the information on the current page as well as search the WWW itself.

  Browser give you the facility to save a Web page in a file on your com pull print a Web page on your computer, and send the contents of a Web page e-Mail to others on the Internet.
- Few Web browser are complete Internet package, means they come with components like e-Mail client, newsgroup client an HTML composer, telnet client, ftpclient, etc.
- Web browser should be able to handle text, images of the World Wide Web, as well as the hyperlinks to digital video, or other types of information.

How Browsers work?

World Wide Web works on the client server model. A user computer works as a client which can receive and send data to the server. When a web page is requested by a user, the browser contacts the requested server (where the website is stored) and by fetching and interpreting the requested files, it displays the web page on the computer screen.

The whole process takes place in these three steps:

1. **Contact to DNS Server**: When a user enters a URL into the address bar and hits enter ”, at first browser contacts the DNS server. A DNS server stores the IP addresses of the server associated with the corresponding domain names. The DNS server takes the domain name from the browser and returns the corresponding IP address to the browser.
2. **Contact to Server**: After getting the IP address of the server for the requested webpage, browser sends a request to that server for the desired files.

3 Rendering :The entire process followed by a browser from fetching the webpage to displaying it on the screen is called Rendering.

***Uniform Resource Locator (*URL)**

- ***U**niform **R**esource **L**ocator is* the global address of documents and other resources on the World Wide Web.
- The URL defines four things: protocol, host computer, port, and path.

- The *protocol* is the client/server program used to retrieve the document. Many different protocols can retrieve a document; among them are FTP or HTTP.
- The host is the computer on which the information is located, although the name of the computer can be an alias. Web pages are usually stored in computers, and computers are given alias names that usually begin with the characters "www".
- The URL can optionally contain the port number of the server. If the *port* is included,it is inserted between the host and the path, and it is separated from the host by a colon.
- Path is the pathname of the file where the information is located. Note that the path can itself contain slashes.

Example:
http://en.wikipedia.org/wiki/File:Raster_to_Vector_Mechanical_Example.jpg
Absolute vs. relative URLs:

An absolute URL is one that completely specifies the desired resource starting from the root of the resource name space. It is unique, meaning that if two absolute URLs are identical, they point to the same resource

An example is:
http://en.wikipedia.org/wiki/File:Raster_to_Vector_Mechanical_Example.jpg

A relative URL points to the location of a resource relative to a base URL. It is preceded by two dots *(../directory_path/file.txt)* for the directory above, one dot *(./directory_path/file.txt)* for the current directory or without the beginning slash (directory_path/file.txt) which is also the current directory. No dots *(/directory_path/file.txt)* for the root directory or domain.

Which results to

http://www.webreference.com/directory_path/file.txt

URI

- Uniform Resource Identifier (URI) is a string of characters used to identify a resource
- Such identification enables interaction with representations of the resource over a network, typically the World Wide Web, using specific protocols.
- Schemes specifying a concrete syntax and associated protocols define each URI
- To guarantee uniformity, all URIs follow a predefined set of syntax rules, but also maintain extensibility through a separately defined hierarchical naming scheme (e.g.http://).

URN

URN stands for Uniform Resource Name. URN is also the subset of URI. One of the best examples of URN is ISBN number which is used to uniquely identify a book. URN is completely different than URL as it doesn't include any protocol.

Difference between URL and URI:

URL	URI
URL is used to describe the identity of an item.	URI provides a technique for defining the identity of an item.
URL links a web page, a component of a web page or a program on a web page with the help of accessing methods like protocols.	URI is used to distinguish one resource from other regardless of the method used.
URL provides the details about what type of protocol is to be used.	URI doesn't contains the protocol specification.
URL is a type of URI.	URI is the superset of URL.

Hypertext

- Hypertext is text which contains links to other texts.
- By clicking on a link in a hypertext document, a user can quickly jump to different content.
- The term was coined by Ted Nelson around 1965.

Characteristics of good hypertext

- **Lots of documents:** Much of the hypertext's power comes from its ability to make large quantities of information accessible. If all the text in your system can be printed on ten pages, it would be just as simple to read through it from beginning to end and forget all this hypertext silliness.

  • **Lots of links:**If each document has just one link, then it is little more than normal, sequential text. A hypertext document should present the reader with several links, offering a choice about where to go next. Ideally, a document should present as many relevant links as the reader can easily comprehend and select among.

- **Range of detail:**The great advantage of hypertext is that it permits readers to explore to a breadth and depth that is simply not feasible in print. To make this accessible, available hypertext documents should range from the broadest possible overview of a subject, down to its gritty details**.**
- **Correct links:**This may seem trivial, but it's amazing how many Web links point nowhere. In general, be careful linking to any hypertext document not under your direct control.

Hyperlink

- A hyperlink is a word, phrase, or image that you can click on to jump to a new document or a new section within the current document.
- Hyperlinks are found in nearly all Web pages, allowing users to click their way from page to page.
- When you move the cursor over a hyperlink, whether it is text or an image, the arrow should change to a small hand pointing at the link. When you click it, a new page or place in the current page will open.

Hypermedia

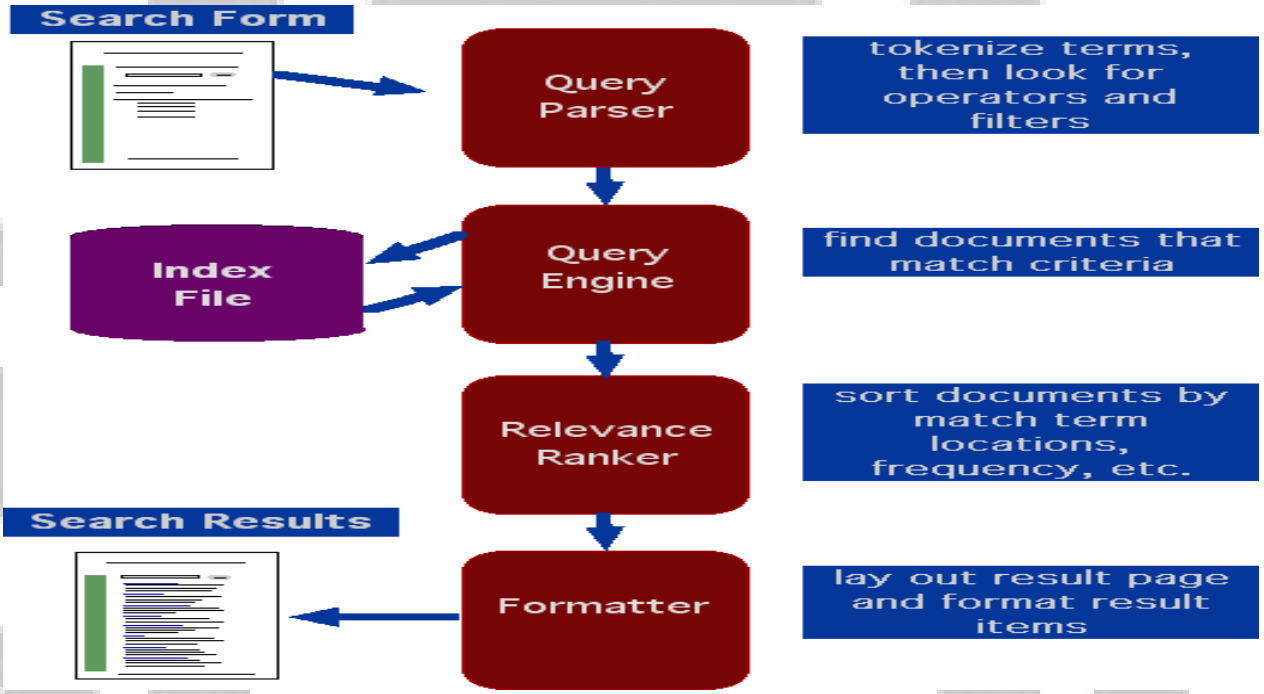
Hypermedia , an extension of the term hypertext, is a nonlinear medium of information that includes graphics, audio, video, plain text and hyperlinks. This designation contrasts with the broader term multimedia , which may include non interactive linear presentations as well as hypermedia

Search Engine

Search engines are programs that search documents for specified keywordsand returns a list of the documents where the keywords were found.

Search Engine Functions

- Accept query
- Look in index for words that match
- Extract documents that match all criteria.
- Sort by relevance (matches, location, link analysis)
- Create results page.



How does a Search Engine work.

Search engine works by following steps:

- Crawling
- Indexing
- Processing Queries
- Ranking

1. **Crawling**
   - Search engines run automated programs, called "robots" or "spiders" that usethe hyperlink structure of the web to "crawl" the pages and documents that make up the World Wide Web.
   - once a new page is found the spider reads the Content & also checks forImages.
2. **Indexing**
   - Once a page has been crawled by robots, its contents can be "indexed" - stored in a giant database of search engines called an "index". This index allows the Search Engine to sort all the documents in fractions of a second.
   - This indexed data is stored in encoded format to save space.
   -

Contents of index

- Basic information for document or record like
  - File name / URL / record ID
  - Title or equivalent
  - Size, date, MIME type
- Full text of item
- More metadata like
  - Product name, picture ID
  - Category, topic, or subject
  - Other attributes, for relevance ranking and display

3. Processing Queries:

- When a request for information comes into the search engine, the searchengine retrieves all the documents from the index and tries to match themwith the query.
- A match is determined if the term or phrase is found on the page in themanner specified in the query.
- The search engine found millions of matching information, so it uses an algorithm to decide in which order to display the results.

4. Ranking:

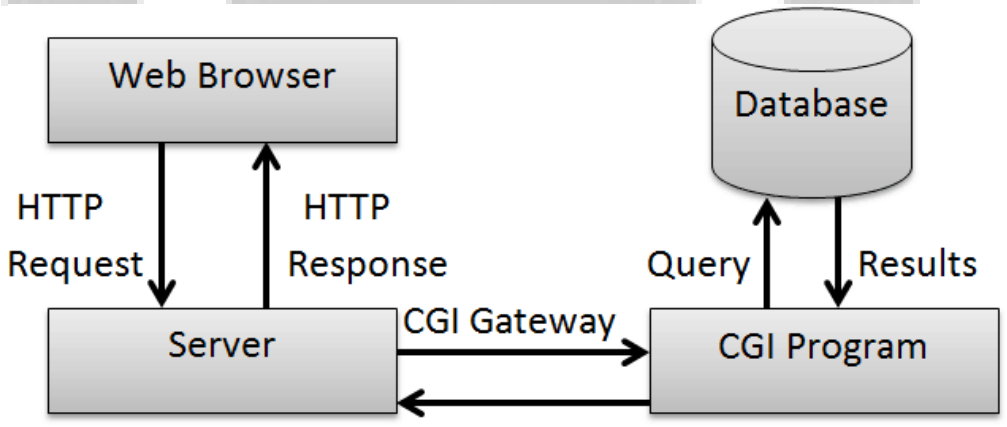
- Once the search engine has determined which results are a match for thequery, the engine's algorithm runs calculations on each of the results to determine which is most relevant to the given query.
- They sort these on the results pages in order from most to least relevant sothat users can make a choice about which to select.

Common Gateway Interface(CGI)

The Common Gateway Interface CGI) is a specification defining how a program interacts with a Hyper Text Transfer Protocol (HTTP) server.

- The Common Gateway Interface (CGI) provides the middleware between WWW servers and external databases and information sources.

Working of CGI (Common Gateway Interface)



The sequence of events for creating a dynamic HTML document on the fly throughCGI scripting is as follows:

1. A client makes an HTTP request by means of a URL. This URL could be typed intothe 'Location' window of a browser, be a hyperlink or be specified in the 'Action' attribute of an HTML <form> tag.

2. From the URL, the Web server determines that it should activate the gateway program listed in the URL and send any parameters passed via the URL to that program.

3. The gateway program processes the information and returns HTML text to the Webserver. The server, in turn, adds a MIME header and returns the HTML text to the Web browser.

4. The Web browser displays the document received from the Web server.

PROXY SERVER

A proxy server provides a gateway between users and the internet. It is a server,referred to as an "intermediary" because it goes between end-users and the webpages they visit online.

When a computer connects to the internet, it uses an IP address. This is similar to your home's street address, telling incoming data where to go and marking outgoingdata with a return address for other devices to authenticate. A proxy server is essentially a computer on the internet that has an IP address of its own.

Proxy Services

The main purpose of a proxy service is to filter requests to ensure that no dangeroustraffic creeps in by applying strict routing rules and to boost the performance of the system. A proxy service works simply – when a proxy service receives a request, forexample, to open a Web page, it looks for the already cached pages. If it finds the requested page in the already cached page, it returns it to the user. If the page is notyet cached, proxy service uses its own IP address to fetch the page from the server for the client.

Proxy services are mainly of two types – forward proxy and reverse proxy. Forward proxy is an Internet-facing proxy that is used to retrieve a range of sources. A reverse proxy is particularly used for protection and security of the server. It includes tasks like caching, authentication and decryption. Other types of proxies includetransparent proxies, anonymous proxies, DNS proxies and highly anonymous proxies.

Dreamweaver

Dreamweaver is a web development tool.

Developed by Adobe Systems, Dreamweaver is a proprietary web development tool first created by Macromedia in 1997. Dreamweaver helps developers build web sitesfor the Internet or an intranet, letting developers design, code and manage websites as well as mobile content. It is currently available for OS X and Windows.
Dreamweaver combines a visual design surface, known as Live View, and a codeeditor with features such as code collapsing, code completion and syntax highlighting

Digi{Coders}